UNIVERSITY OF CALGARY

Network Traffic Measurement Research at the U of Calgary

Carey Williamson
Department of Computer Science

November 15, 2016

Networks Research Group

- Faculty Members (4):
  - Majid Ghaderi, Zongpeng Li, Mea Wang, Carey Williamson
- Adjunct Faculty (1):
  - Martin Arlitt (HP Labs)
- PhD Students (9):
  - Ali Abbasi, Maryam Elahi, Cyriac James, Mehrnaz Mireslami, Seyed Md. Pakdaman, Ali Sehati, Reza Zakerinesab, Linquan Zhang, Ruiting Zhou
- MSc Students (15):
  - Mohamad Darianian, Wei Fang, Danny Fisher, Sijua Gu, Mackenzie Haffey, Yuhui Lin, Md. Seyed Naghibi, Mahshid Navabi, Keynan Pratt, Sourish Roy, Abolfazl Samani, Maryam Soleimani, Akshita Tyagi, Shunyi Xu, Yao Zhao

Research Overview

- Research area?
  - Computer networks, wireless networks, Internet protocols, computer systems performance evaluation
- Mission: “Make the Internet go faster”
- Approach?
  - Experimental, simulation, analytical
- Key challenges?
  - Citius, Altius, Fortius!
  - Performance, scalability, robustness

My Current Students

- Maryam Elahi (PhD, Dec 2016 – expected)
  - Fairness and efficiency in speed scaling designs

- Mohamad Darianian (MSc, in progress)
  - Experimental evaluation of SAVI OpenFlow controllers
- Mackenzie Haffey (MSc, in progress)
  - Network security analysis tools for enterprise scale
- Keynan Pratt (MSc, Dec 2016 – expected)
  - Distributed caching for Friend-to-Friend (F2F) networks
- Sourish Roy (MSc, in progress)
  - Characterization of Desire-to-Learn (D2L) LMS traffic

Recent Measurement Projects

- Martin Arlitt (adjunct faculty)
  - Monthly network security traffic analysis for UCIT
- Michel Laterman (MSc, Sept 2015)
  - Workload characterization of Netflix and Twitch
- Yang Liu (MSc, Aug 2015)
  - Characterizing scientific Web sites (ASTRO + Aurora)
- Feifei Shi (BSc, June 2016)
  - Redundant traffic elimination (RTE) on email traffic
- Arsham Skrenes (MSc, Aug 2016)
  - Fine-grain energy measurements of Intel i7 processor
- Zhengping Zhang (BSc, June 2016)
  - Characterization of Office 365 email traffic

Summary and Conclusions

- The U of C has a large and active Networks Research Group, some of whom (me!) do very applied network performance research
- Internet traffic continues to grow and evolve in many varied and interesting ways with each new generation of applications (and users!)
- Video streaming is the current bandwidth hog
- Network security issues are quite pervasive
- HTTPS will limit visibility in future studies

The End

- Thank you!

- Questions?

- For more info: carey@cpsc.ucalgary.ca

NETFLIX
TRAFFIC CHARACTERIZATION

Michel Laterman
Department of Computer Science
University of Calgary

Supervisors: Carey Williamson and Martin Arlitt

Introduction

- Video streaming traffic constitutes a large (and growing!) proportion of modern Internet traffic
- Popular video streaming services include:
  - YouTube – user-generated content, short-clips (well-studied)
  - NetFlix – on-demand video, TV shows, movies (some studies)
  - Twitch – live streaming of video game play (few studies)
  - Vimeo – video-sharing site with High-Definition videos
  - Hulu – on-demand video, not in Canada
  - Yahoo Screen – professionally produced content, limited availability in Canada
- On the University of Calgary network, the top video streaming sites observed are YouTube, NetFlix, Twitch

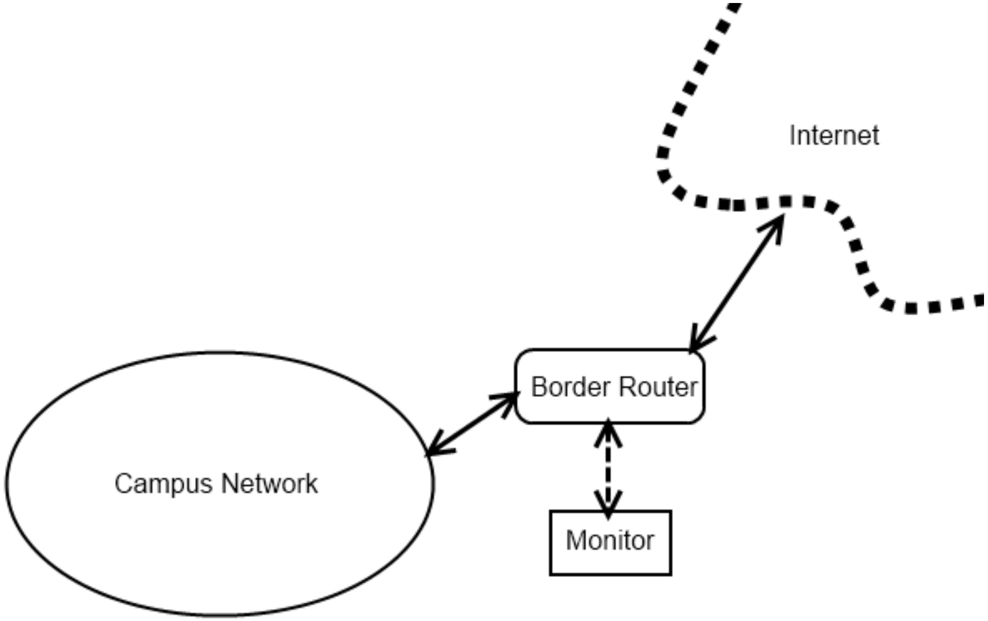
Research Objectives

- General
  - Improve understanding of U of C network traffic
  - Identify network performance problems and anomalies

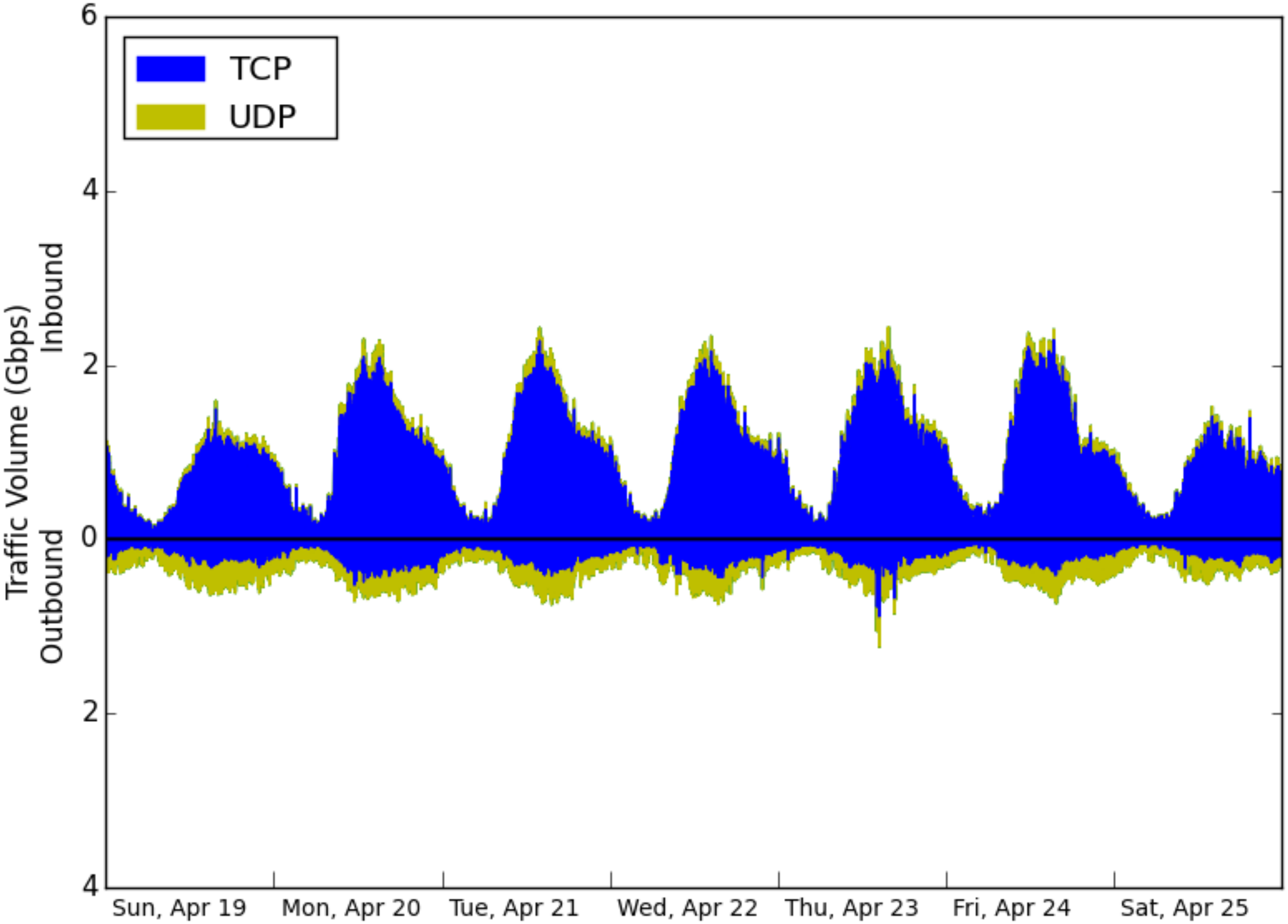
- Specific
  - Characterize video streaming services on U of C network
  - Understand similarities/differences between NetFlix and Twitch

Methodology

- Passive network traffic measurement
- Hardware: Endace DAG packet capture card
- Software: Bro network security monitor
- 5 months of data (December 1, 2014 to April 29, 2015)
- Analysis of TCP connection and HTTP transaction logs



Example: Traffic Overview (April 2015)



HTTP Traffic Overview

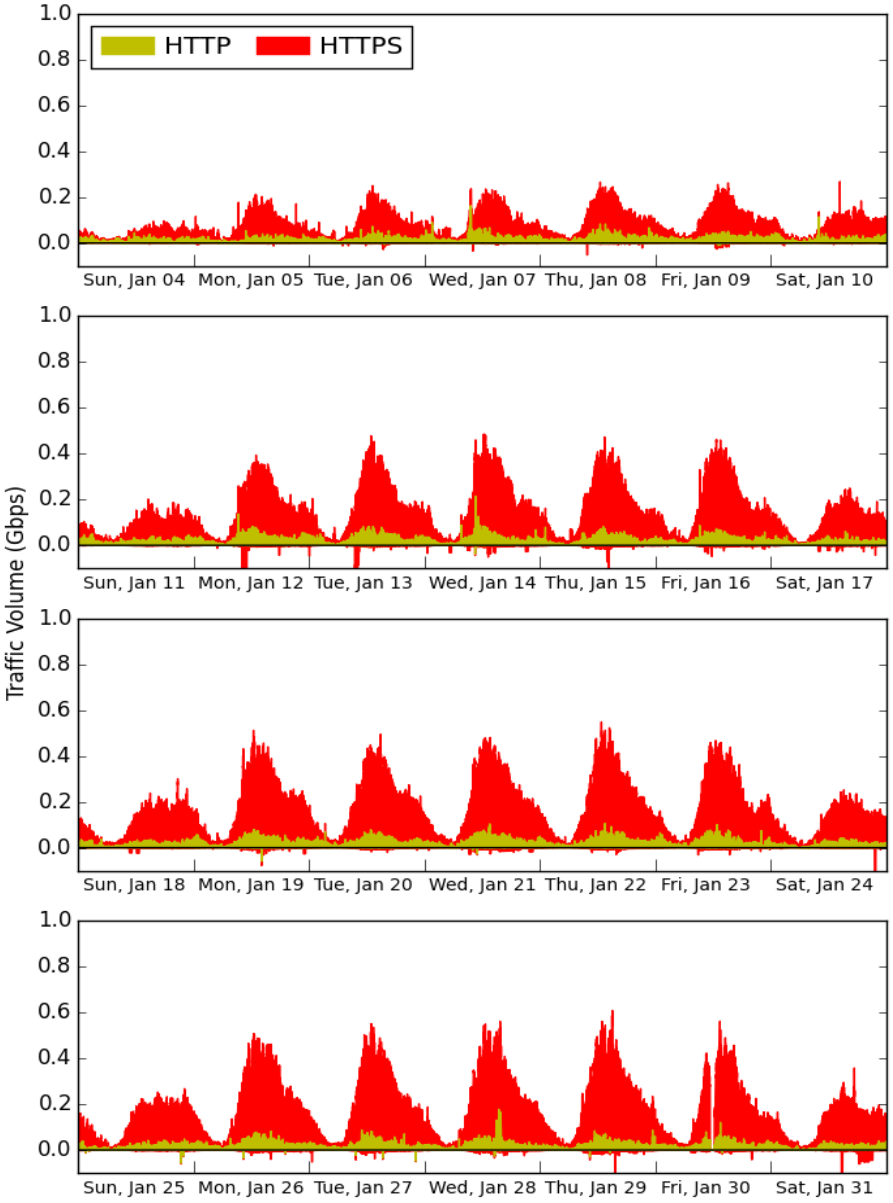
Host	Req. Percent	Volume
netflix.com	33.81%	217.1 TB
apple.com	8.37%	53.75 TB
googlevideo.com	2.43%	15.59 TB
steampowered.com	2.14%	13.79 TB
twitch.tv	2.04%	13.12 TB

HTTPS Traffic Overview

Host	Connections	Percent	Volume
google.com	314 million	7.91%	27.3 TB
apple.com	179 million	4.51%	2.8 TB
majuwe.com	168 million	4.23%	106.7 GB
akamaihd.com	151 million	3.80%	32.7 TB
googlevideo.com	131 million	3.30%	230.1 TB

YouTube Traffic

- January 2015
- Uses HTTPS by default
- HTTP for some embedded clips
- Outbound traffic is for video uploads

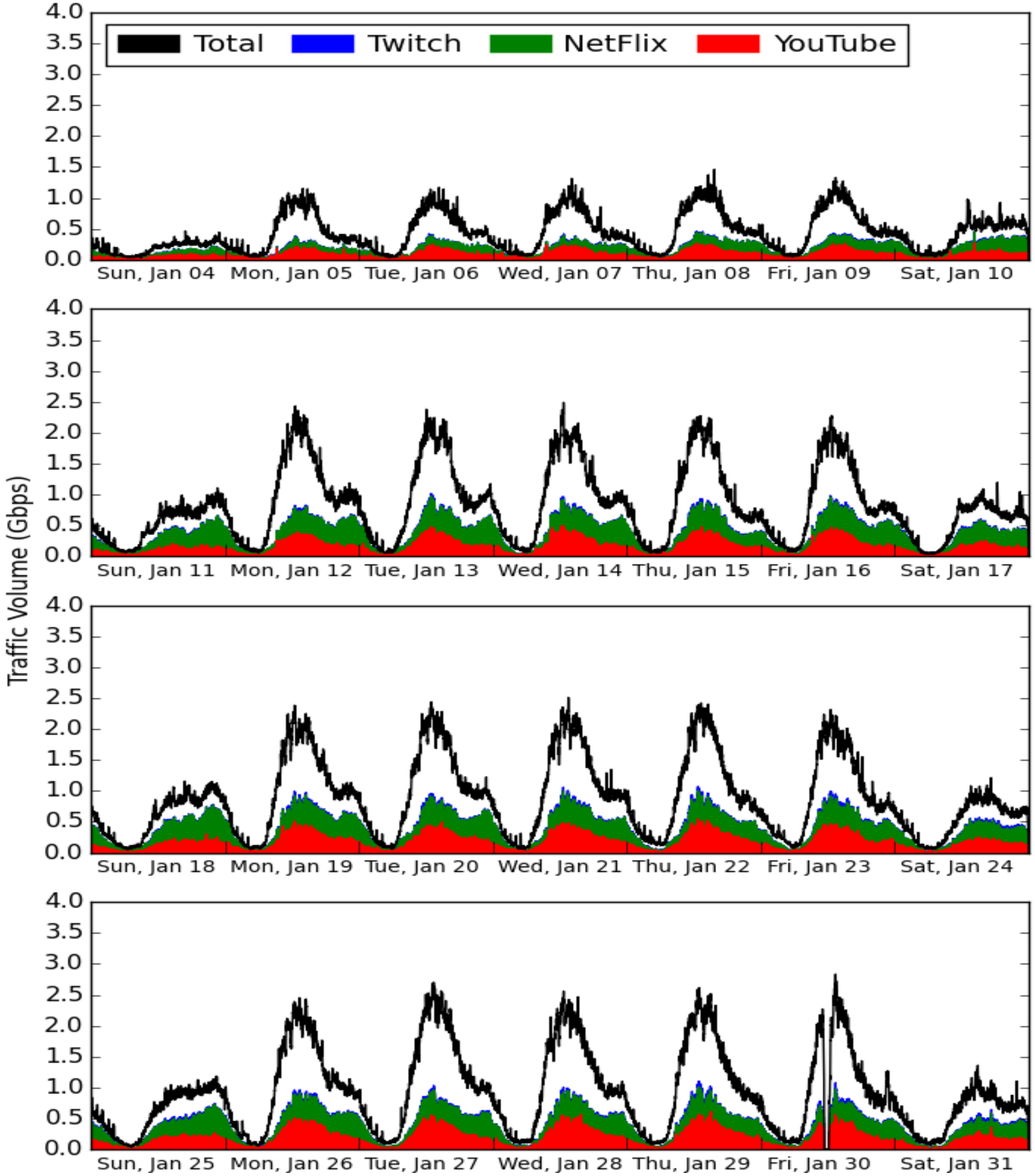


Video Traffic Volume

- Outbound traffic to NetFlix and Twitch is negligible.

	YouTube - HTTP		YouTube - HTTPS		NetFlix	Twitch
	Inbound	Outbound	Inbound	Outbound	Inbound	Inbound
December	1.93 TB	0.14 TB	36.22 TB	0.89 TB	30.77 TB	2.82 TB
January	1.89 TB	0.12 TB	36.31 TB	1.06 TB	44.41 TB	3.14 TB
February	1.79 TB	0.05 TB	45.47 TB	1.14 TB	43.83 TB	3.74 TB
March	2.08 TB	0.05 TB	59.63 TB	1.36 TB	54.29 TB	4.79 TB
April	1.51 TB	0.05 TB	52.43 TB	1.08 TB	43.85 TB	3.74 TB

Video Traffic



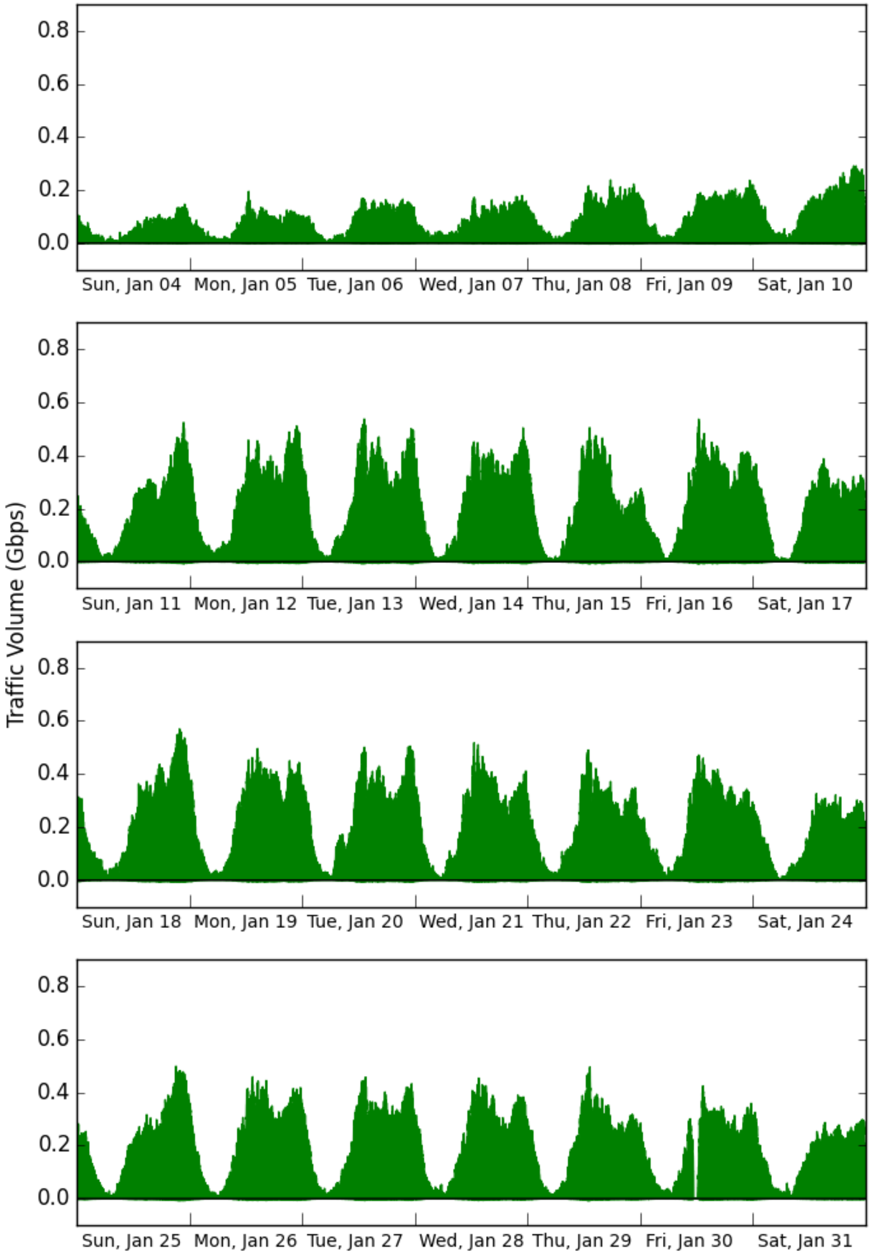
- January 2015
- Top line (Total) is HTTP+HTTPS
- Red is (HTTPS) YouTube
- Green is NetFlix
- Blue is Twitch

NetFlix

- 305 million request-response pairs on 14.3 million connections generating 217.1 TB of volume
- 62.9% of requests had code 200 (OK), 29.9% had 206 (Partial content), 6.09% had no code.
- 35 different content-type headers
  - Application/octet-stream 216.7 TB
  - Text/html 328.8 GB

NetFlix Traffic

- Video content is served from several unnamed servers with NetFlix IP addresses
- 217.1 TB total traffic
- Connections average 26 MB in, 370 KB out
- Average duration 150 seconds



NetFlix – Video Delivery

- HTML5 Player (transitioned away from Silverlight)

- Requests to the Web interface player include a parameter called movieID
- Desktop and Mobile devices use different request paths
  - Can’t see movieid from mobile requests
- 162.6 TB of traffic was responses to content requests from desktop devices, 54.01 TB mobile
- Multiple connections are used to transport video (7-9 for a 22 min episode, 14-16 for 42 min)

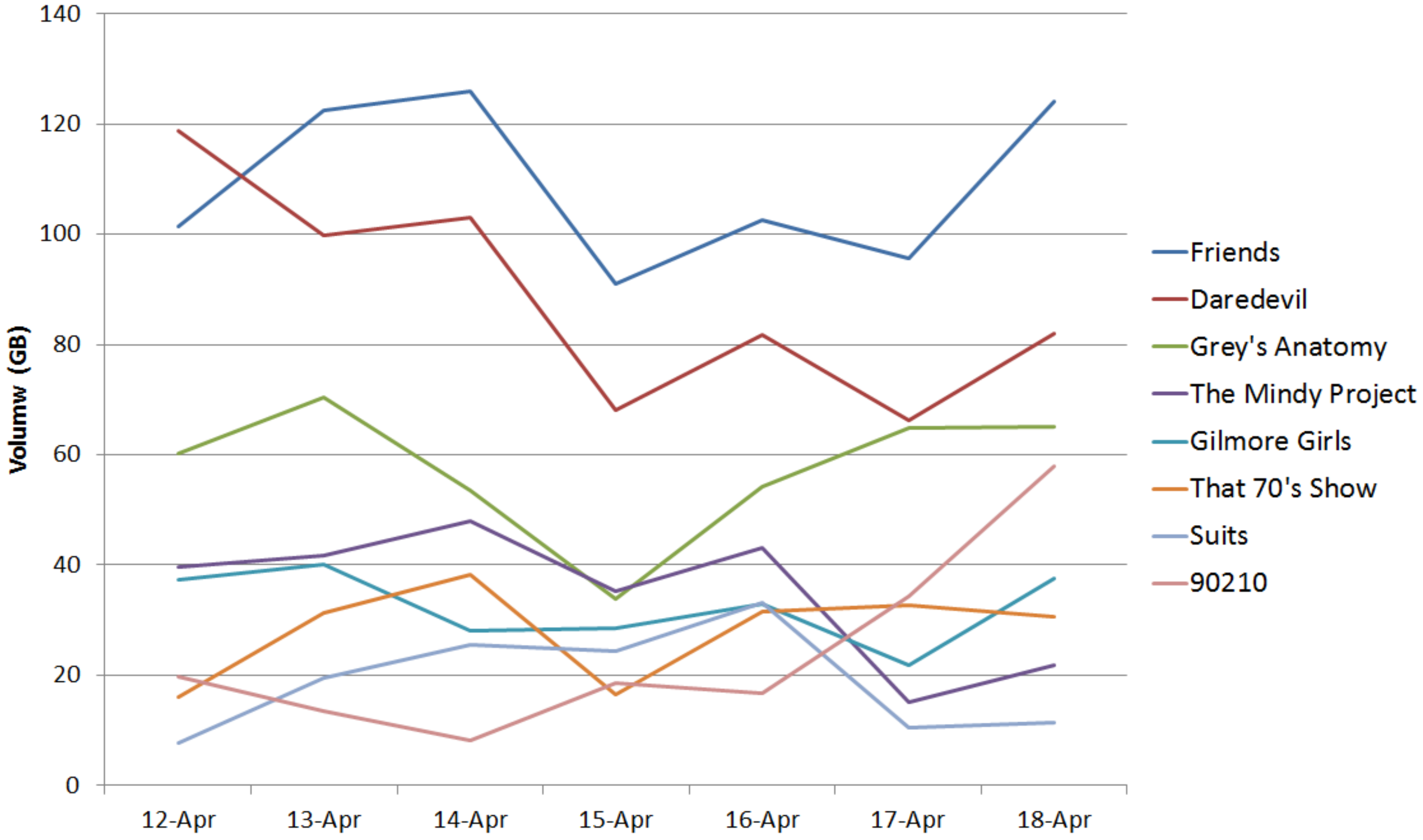
NetFlix – What are people Watching?

Title	Dec	Jan	Feb	Mar	Apr
1. Friends	-	1	1	1	1
2. Grey’s Anatomy	1	2	2	3	2
3. House of Cards	20	16	3	2	9
4. Gilmore Girls	2	4	9	10	5
5. Gossip Girl	3	3	7	7	7
6. That 70’s Show	42	49	4	4	6
…					
18. Daredevil	-	-	-	-	3

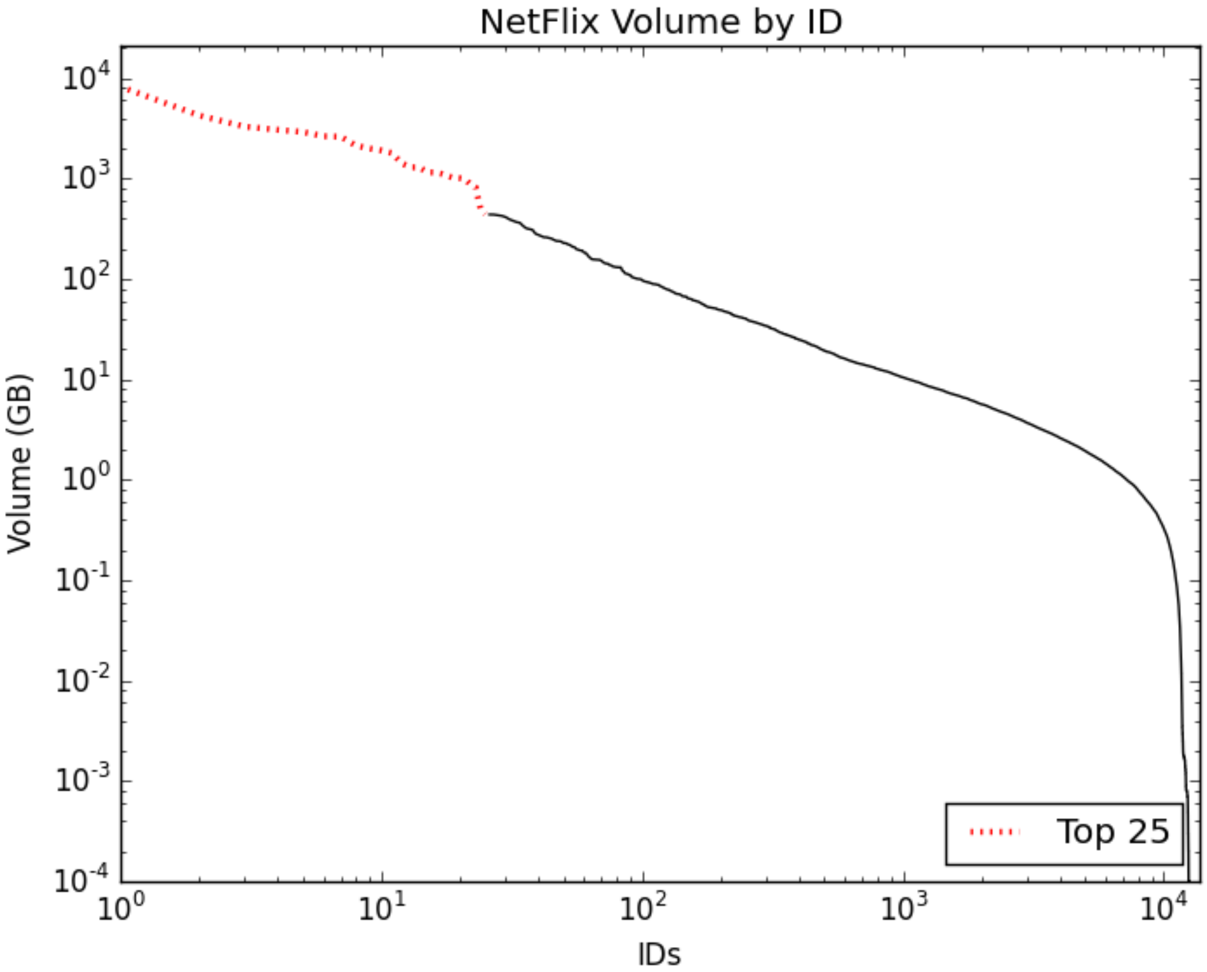
Long-term popularity

Short-term popularity

A Week of NetFlix Traffic – Top Content



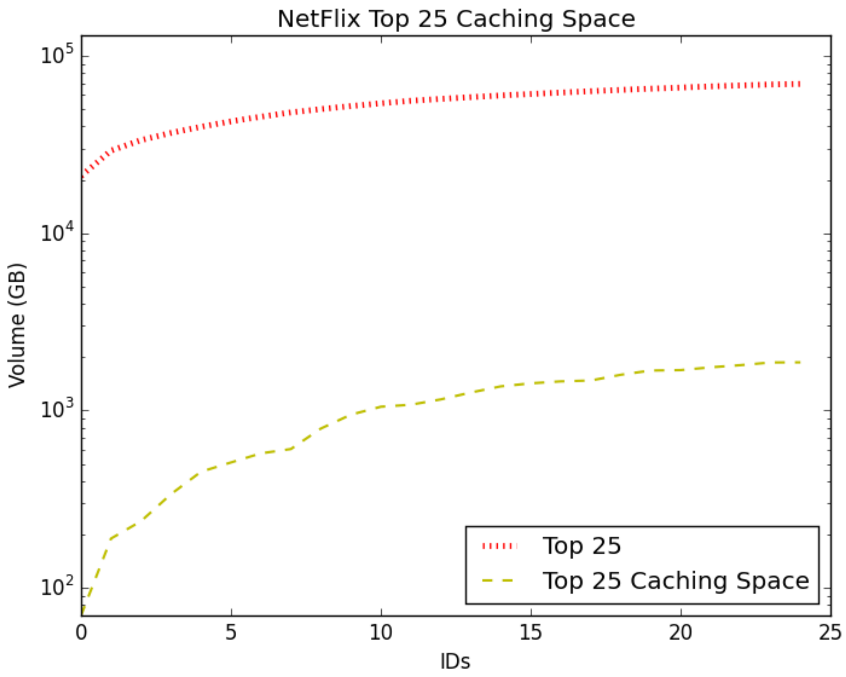
NetFlix movieID Traffic Volumes

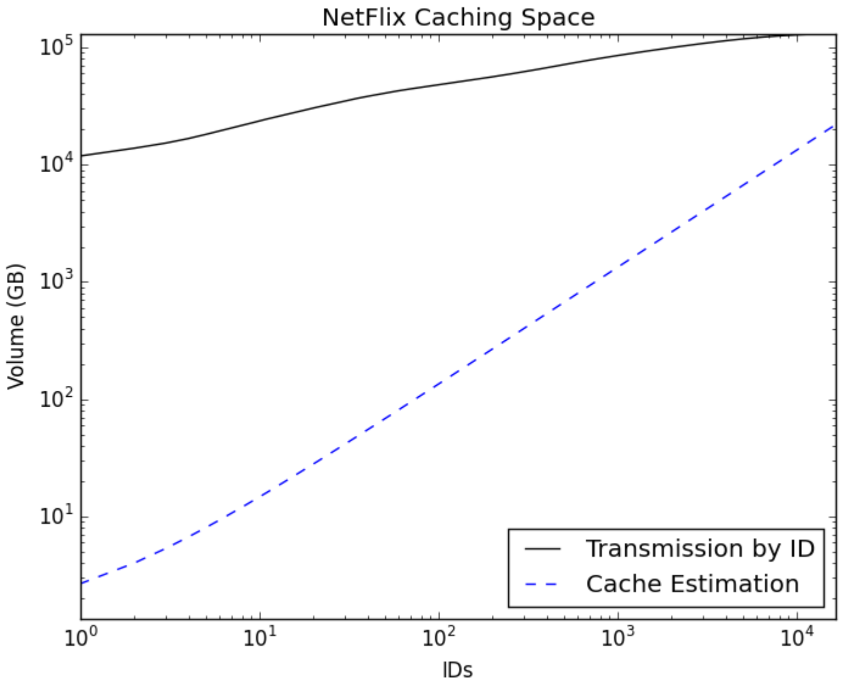


- Top 25 shows (2,801 IDs)
  - 50% of traffic volume
- Friends: 21 TB
- Grey's Anatomy: 8 TB
- House of Cards: 4 TB

Caching NetFlix

- File sizes: 13.23 MB/minute (SD) or 22.58 MB/min (HD)
- 70 GB to cache Friends (21 TB transmission)
- 120 GB to cache Grey’s Anatomy (8.2 TB)
- 40 GB to cache House of Cards (4.25 TB)





Conclusions (Netflix)

- Video streaming services constitute a large proportion of inbound traffic on the U of C network
- YouTube and NetFlix are the most popular currently
- Caching NetFlix could greatly reduce network traffic
    - Caching “Friends” (70 GB) would reduce traffic by 20 TB
- Studies like this will be much more difficult once Netflix moves to HTTPS for all content delivery (mid-2015)

TWITCH
TRAFFIC CHARACTERIZATION

Michel Laterman
Department of Computer Science
University of Calgary

Supervisors: Carey Williamson and Martin Arlitt

Introduction

- Video streaming traffic constitutes a large (and growing!) proportion of modern Internet traffic
- Popular video streaming services include:
  - YouTube – user-generated content, short-clips (well-studied)
  - NetFlix – on-demand video, TV shows, movies (some studies)
  - Twitch – live streaming of video game play (few studies)
  - Vimeo – video-sharing site with High-Definition videos
  - Hulu – on-demand video, not in Canada
  - Yahoo Screen – professionally produced content, limited availability in Canada
- On the University of Calgary network, the top video streaming sites observed are YouTube, NetFlix, Twitch

Twitch

- 19.49 TB total traffic on 1.6 million connections through 54 million request-response transactions
- 25 different content type headers seen
  - Video/mp2t 39.1% of requests 18.68 TB of traffic
    - Greater than Live-stream traffic due to VOD
  - Video/x-flv 0.02% of reqs and 719.0 GB of traffic
  - (6th) Application/vnd.apple.mpegurl 37.8% of reqs, 8.95 GB

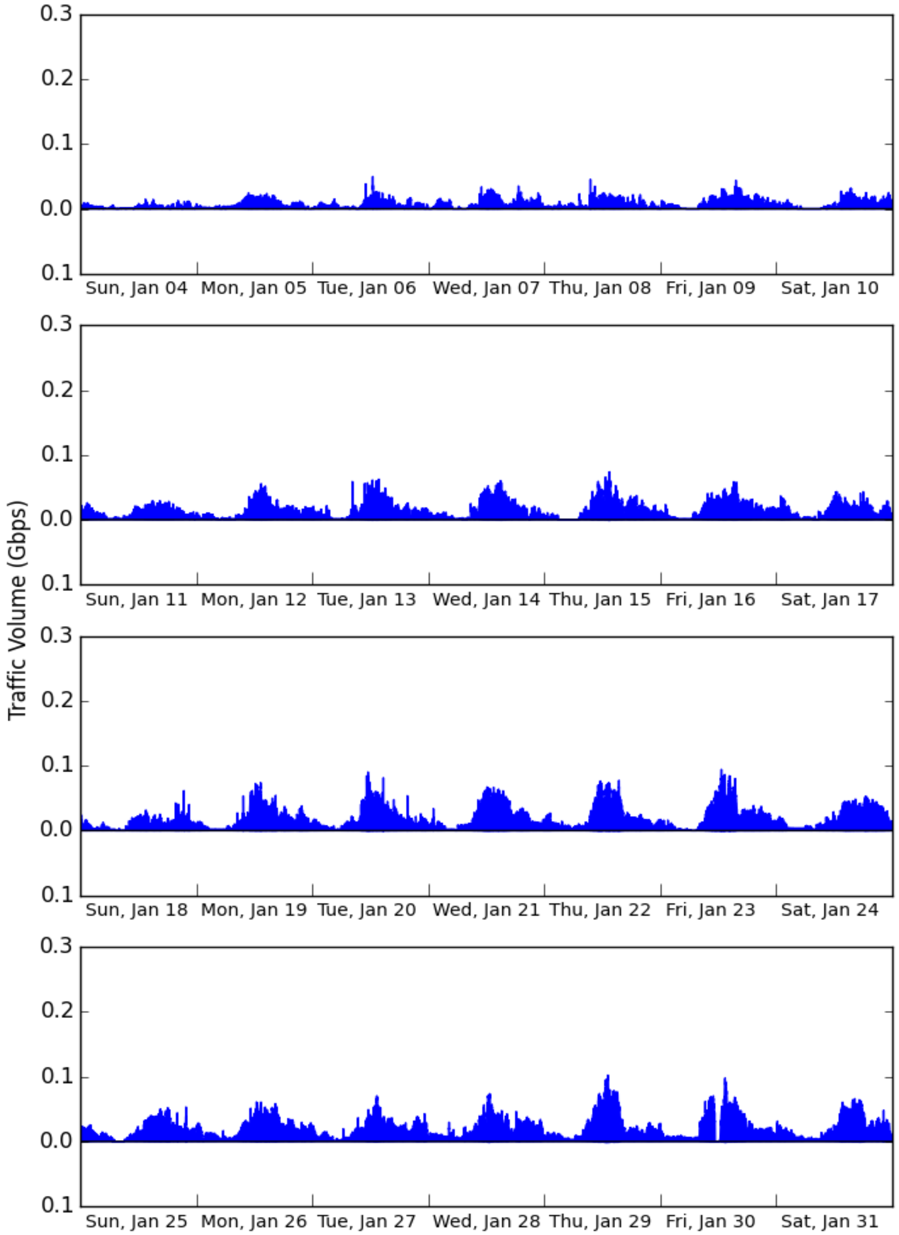
Twitch – Video Delivery

- Uses Apple’s HTTP Live-Streaming (HLS) as a base.
- 18.23 TB live-stream traffic from 40.8 million requests

- Used Flash-based video playback.
- Video qualities: source 1920x1080 (43% of reqs), high 1280x720 (33.7%), medium 852x480(19.9%), low 640x380 (2.63%), mobile 400x226(0.57%), audio only (0.18%)
- Response durations tended to be under 1 second.
- Multiple connections used when viewing a single stream.

Twitch Traffic

- Video content comes from named Twitch servers
  - *.hls.twitch.tv
  - *.hls.ttvnw.net
- 19.49 TB total traffic
- Average connections transmits: 20 MB/300 KB (In/Out) over two minutes



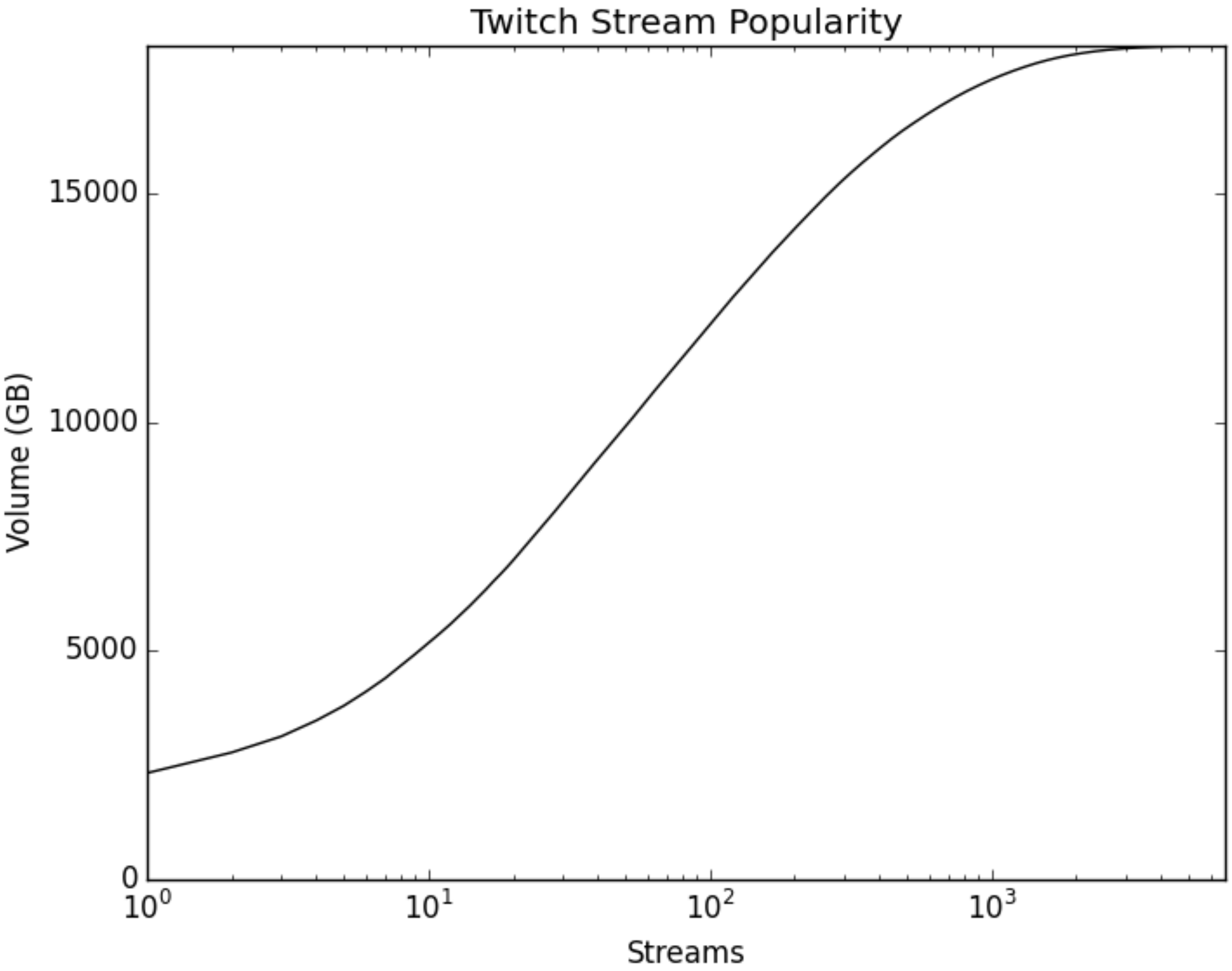
Twitch – What are people watching?

Stream	Dec	Jan	Feb	Mar	Apr
1. Riotgames	338	1	1	1	1
2. beyondthesummit	2	2	2	14	5
3. imaqtpie	13	5	3	4	4
4. lirik	7	3	13	13	8
5. nl_kripp	5	8	5	22	2
6. esltv_lol	1	27	-	-	-
…					
19. esl_csgo	-	-	-	3	61

Long-term popularity

Short-term popularity

Twitch Stream Popularity



- Cumulative GB/stream
- Top 41 streams transmit 50% of volume
- 229 transmit 80%

Conclusions (Netflix and Twitch)

- Video streaming services constitute a large proportion of inbound traffic on the U of C network
- While NetFlix and Twitch are very different services, there are inherent similarities (connection asymmetry, skewed access patterns, short-term and long-term popularity)
- Caching NetFlix could greatly reduce network traffic
  - Caching “Friends” (70 GB) would reduce traffic by 20 TB
- Rebroadcasting Twitch streams locally could lead to lower network traffic and better user viewing experience

University of Calgary – CPSC 329
Guest Lecture: Carey Williamson

Network Security Issues

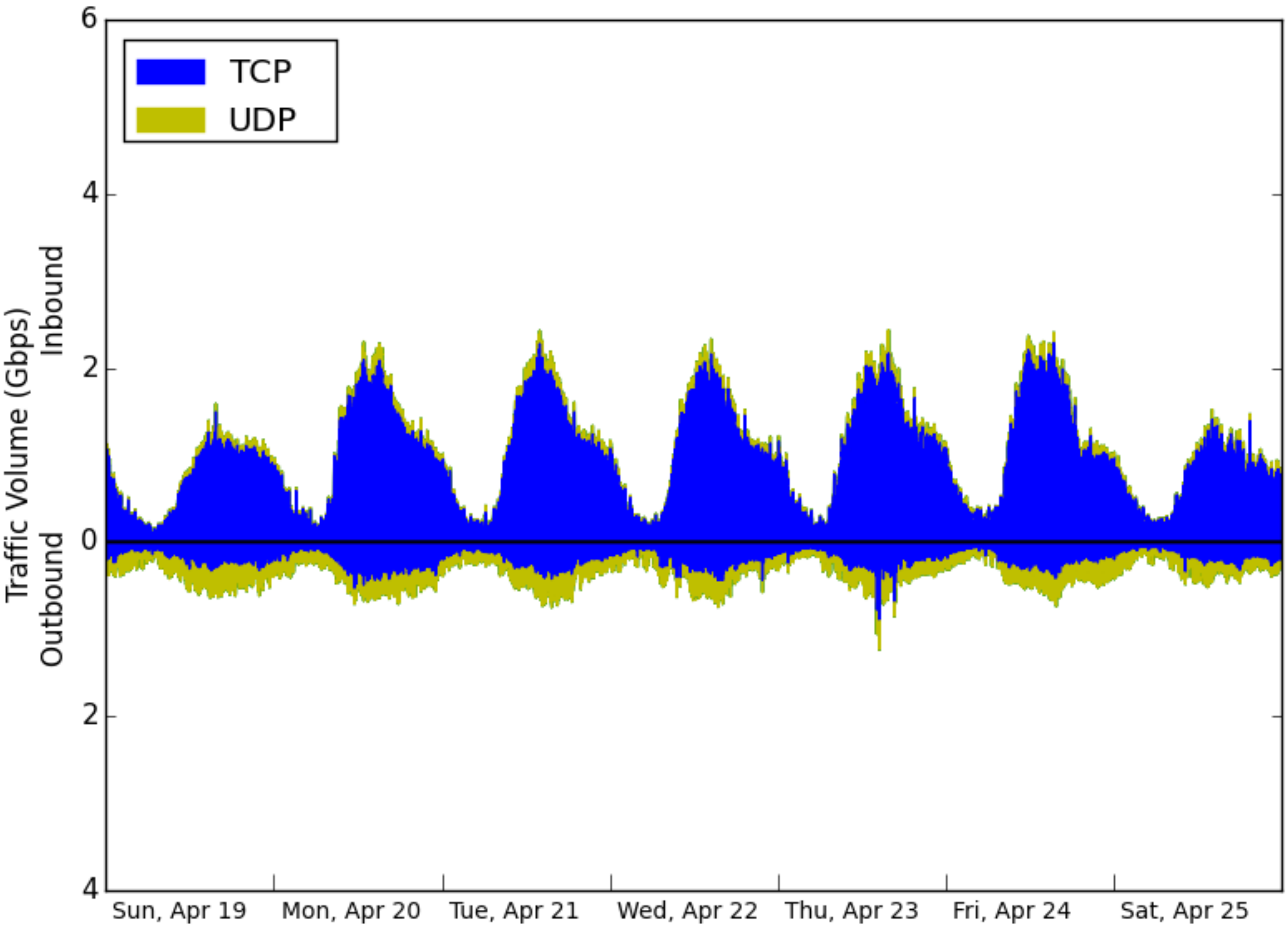
Common Types of Attacks

- Packet sniffing (to steal confidential personal information)
- Spoofing (to forge identity, location, or other credentials)
- Playback (to record and replay valid credentials later)
- Scanning (to actively probe for vulnerable hosts or ports)
- Malware (malicious software, to exploit vulnerabilities)
- DoS: Denial of Service (to make a service inaccessibly slow)
- DDoS: Distributed DoS (like DoS on steroids, using botnets)
- Inference attacks (to learn implicit structural information)

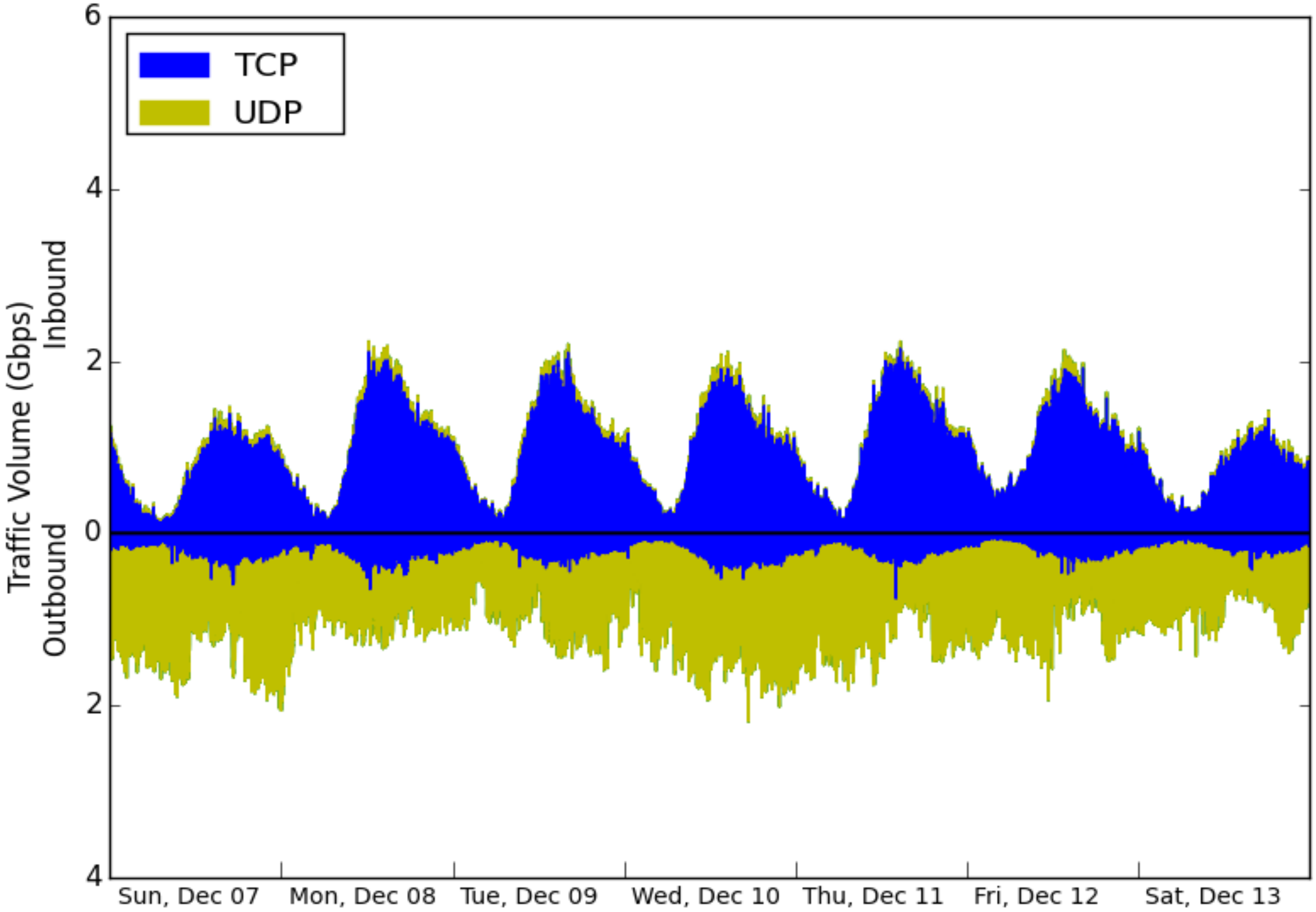
U of C Traffic Examples

- As a networking researcher, I have seen many strange and mysterious things on the U of C network, including these:
- Port scanning
- NTP amplification attacks
- RIP attacks
- Viruses/malware
- SSH attacks
- DoS attacks
- Spam bots

Normal U of C Traffic (Apr 2015)



NTP Amplification Attack (Dec 2014)



Heavy Hitters (outbound)

Outbound Traffic Totals for February 2016

#	IP	Name	Protocol	Port	Service	Volume	Issue?
1	118.90		UDP	123	NTP	9.8 TB	Yes
2	34.148	rb1-s	UDP	53	DNS	6.5 TB	
3	34.130	rb1	UDP	53	DNS	2.9 TB	
4	49.196	gvpn	TCP	10433	VPN	2.9 TB	
5	51.98	aurora	TCP	80	HTTP	2.8 TB	
6	142.7	ns4-a	UDP	53	DNS	2.3 TB	
7	142.5	ns2-a	UDP	53	DNS	2.1 TB	
8	96.25	www	TCP	80	HTTP	1.7 TB	
9	19.141		TCP	443	HTTPS	1.5 TB	Maybe
10	142.6	ns3-a	UDP	53	DNS	1.5 TB	

Heavy Hitters (inbound)

Inbound Traffic Totals for February 2016

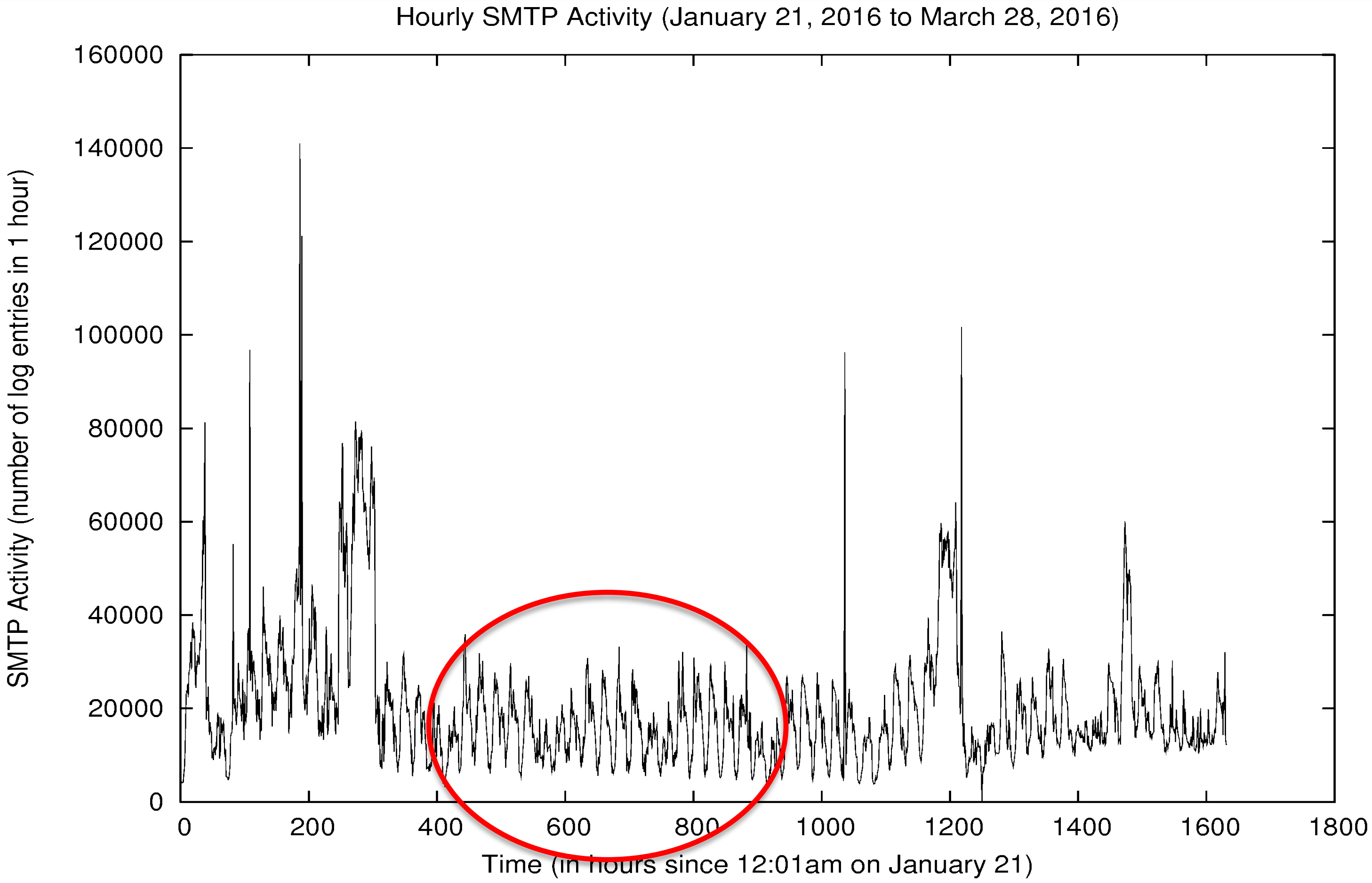
#	IP	Name	Protocol	Port	Service	Volume	Issue?
1	191.61	gop-bio	TCP	22	SSH	2.2 TB	Maybe
2	19.141		TCP	443	HTTPS	1.5 TB	Maybe
3	169.53	ebg	TCP	22	SSH	0.9 TB	Maybe
4	191.45	pc45	TCP	22	SSH	0.5 TB	Maybe
5	49.196	gvpn	TCP	10433	VPN	0.5 TB	Maybe
6	19.143		TCP	25	SMTP	0.4 TB	Maybe
7	191.19	cougar	TCP	22	SSH	0.4 TB	Yes
8	37.45	imap	TCP	993	IMAPS	0.2 TB	
9	129.230	pc230	UDP	137	NetBios	0.2 TB	Yes
10	49.212	itv2	TCP	10433	VPN	0.2 TB	

Strange Connection Activity

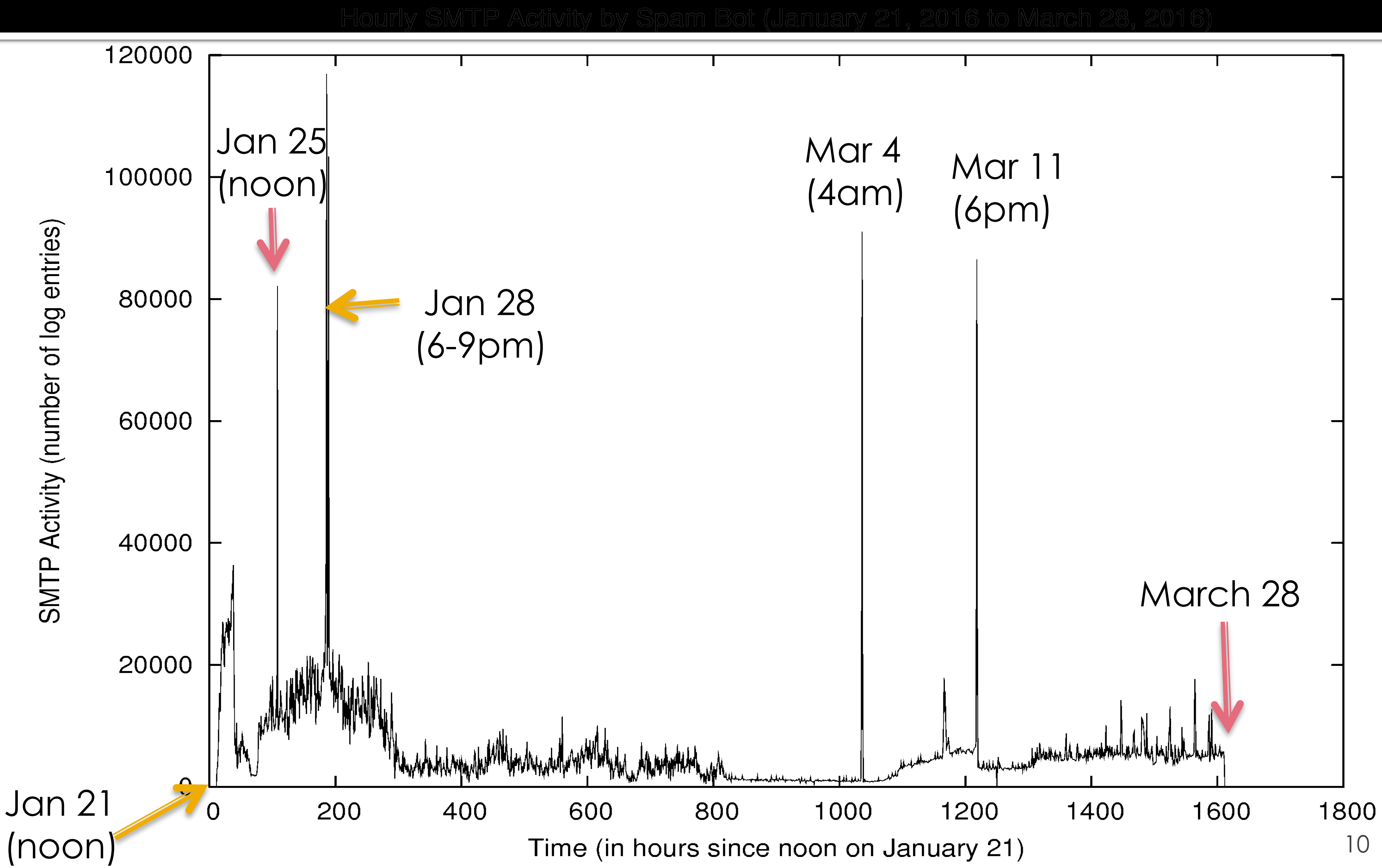
Connection Counts for January 2016

#	IP	Name	Protocol	Port	Service	Conns	Issue?
1	193.8	pc8	UDP	665		908 M	Yes
2	193.9	pc9	UDP	665		778 M	Yes
3	193.7	pc7	UDP	665		702 M	Yes
4	193.8	pc8	UDP	655		538 M	Yes
5	193.9	pc9	UDP	655		502 M	Yes
6	129.230	pc230	UDP	137	NetBios	476 M	Yes
7	193.7	pc7	UDP	655		469 M	Yes
8	118.90		UDP	123	NTP	324 M	Yes
9	34.148	rb1-s	UDP	53	DNS	261 M	Maybe
10	34.51	nassrv3	UDP	520	RIP	240 M	Maybe

SMTP (email) Traffic Activity



Spam Bot Activity



Jan 25 (noon)

Jan 28 (6-9pm)

Mar 4 (4am)

Mar 11 (6pm)

March 28

Jan 21 (noon)

Curious for more?

- Take CPSC 441: **Computer Networks**
  - Learn about the Internet and its protocol stack

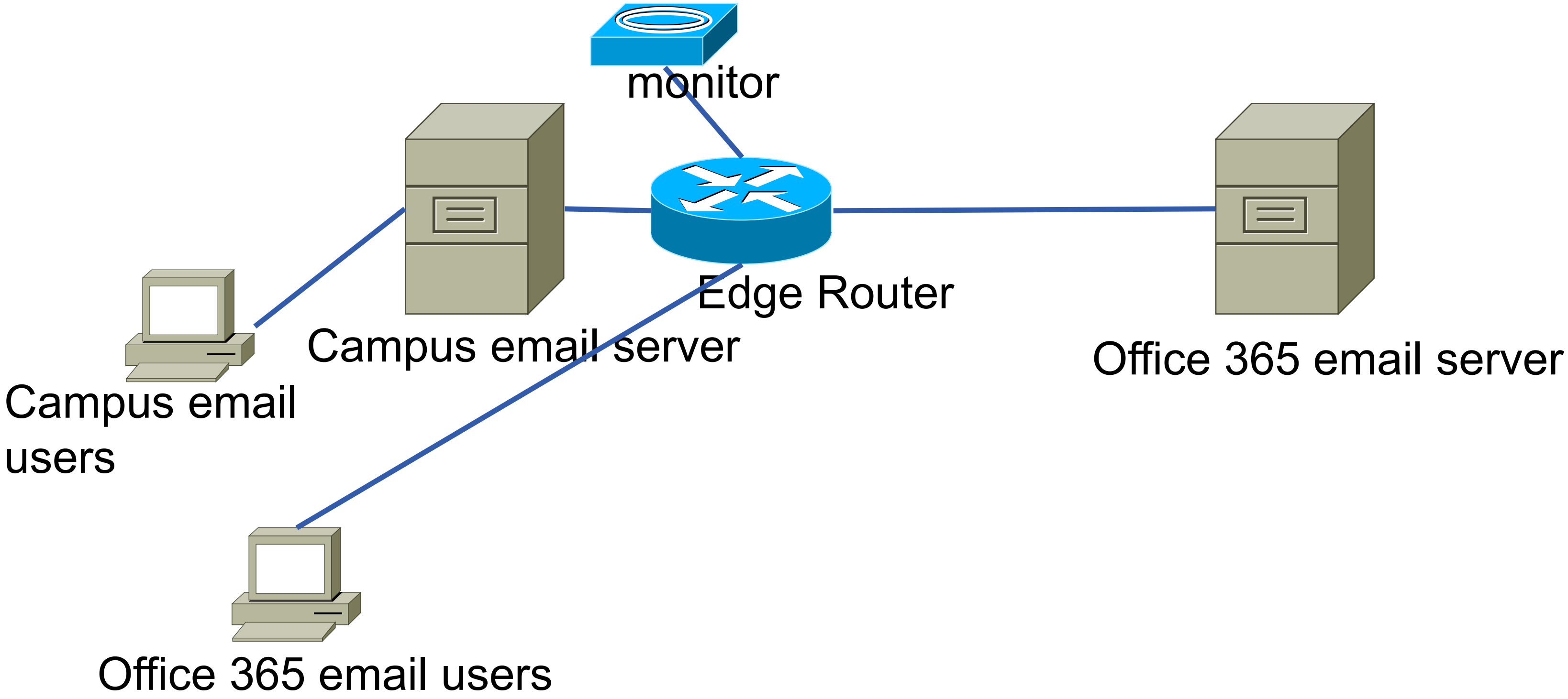
- Take CPSC 526: **Network Systems Security**
  - Course Description: “Attacks on networked systems, tools and techniques for detection and protection against attacks including firewalls and intrusion detection and protection systems, authentication and identification in distributed systems, cryptographic protocols for IP networks, security protocols for emerging networks and technologies, privacy enhancing communication. Legal and ethical issues will be introduced.”

WORKLOAD CHARACTERIZATION OF A CLOUD-BASED EMAIL SERVICE: OFFICE 365

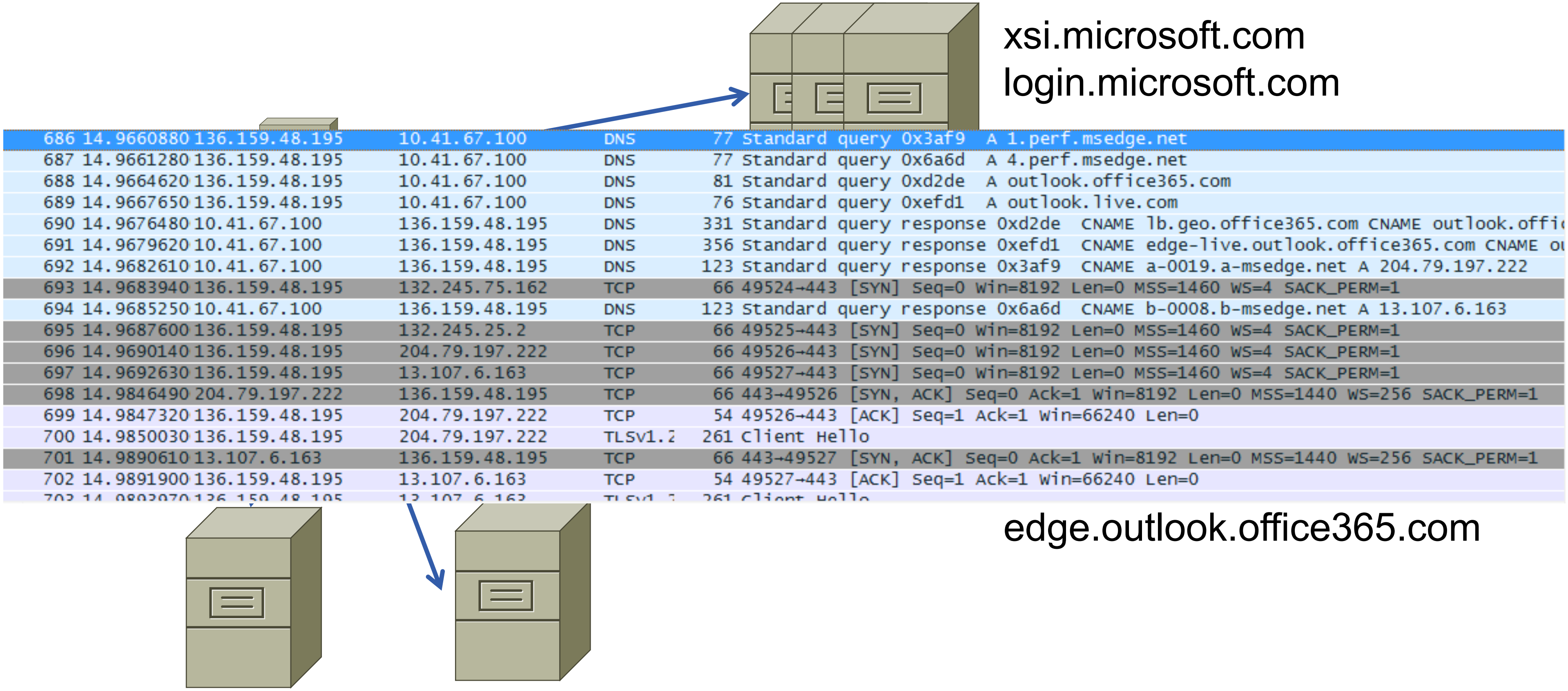
Zhengping Zhang
Department of Computer Science
University of Calgary

Supervisor: Carey Williamson

Background


monitor
Edge Router
Campus email server
Office 365 email server
Campus email users
Office 365 email users

Login Process



xsi.microsoft.com
login.microsoft.com

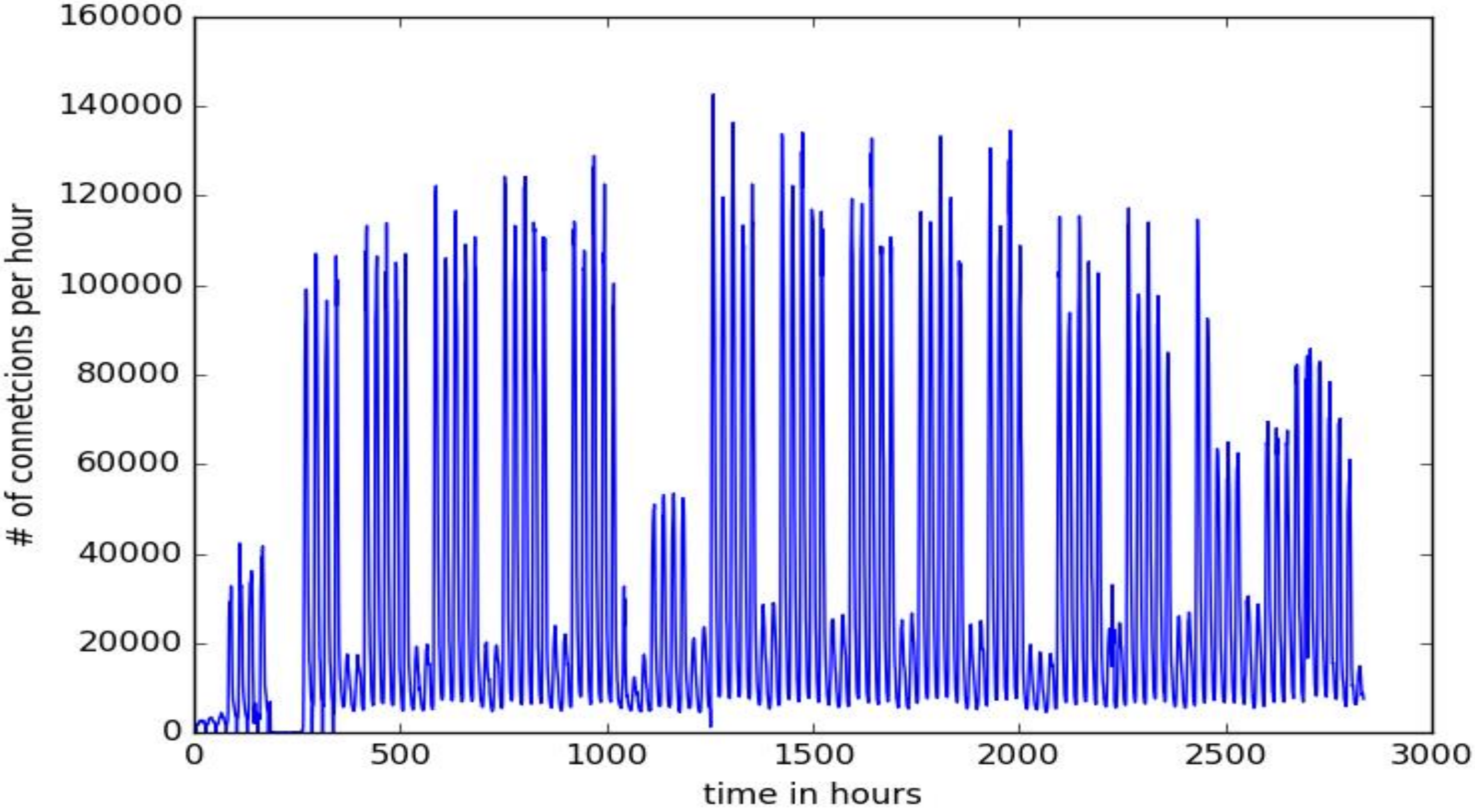
```
686 14.9660880 136.159.48.195   10.41.67.100     DNS       77 Standard query 0x3af9  A 1.perf.msedge.net
687 14.9661280 136.159.48.195   10.41.67.100     DNS       77 Standard query 0x6a6d  A 4.perf.msedge.net
688 14.9664620 136.159.48.195   10.41.67.100     DNS       81 Standard query 0xd2de  A outlook.office365.com
689 14.9667650 136.159.48.195   10.41.67.100     DNS       76 Standard query 0xefd1  A outlook.live.com
690 14.9676480 10.41.67.100     136.159.48.195   DNS      331 Standard query response 0xd2de  CNAME lb.geo.office365.com CNAME outlook.offi
691 14.9679620 10.41.67.100     136.159.48.195   DNS      356 Standard query response 0xefd1  CNAME edge-live.outlook.office365.com CNAME ou
692 14.9682610 10.41.67.100     136.159.48.195   DNS      123 Standard query response 0x3af9  CNAME a-0019.a-msedge.net A 204.79.197.222
693 14.9683940 136.159.48.195   132.245.75.162   TCP       66 49524→443 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
694 14.9685250 10.41.67.100     136.159.48.195   DNS      123 Standard query response 0x6a6d  CNAME b-0008.b-msedge.net A 13.107.6.163
695 14.9687600 136.159.48.195   132.245.25.2     TCP       66 49525→443 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
696 14.9690140 136.159.48.195   204.79.197.222   TCP       66 49526→443 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
697 14.9692630 136.159.48.195   13.107.6.163     TCP       66 49527→443 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=4 SACK_PERM=1
698 14.9846490 204.79.197.222   136.159.48.195   TCP       66 443→49526 [SYN, ACK] Seq=0 Ack=1 Win=8192 Len=0 MSS=1440 WS=256 SACK_PERM=1
699 14.9847320 136.159.48.195   204.79.197.222   TCP       54 49526→443 [ACK] Seq=1 Ack=1 Win=66240 Len=0
700 14.9850030 136.159.48.195   204.79.197.222   TLSv1.2  261 Client Hello
701 14.9890610 13.107.6.163     136.159.48.195   TCP       66 443→49527 [SYN, ACK] Seq=0 Ack=1 Win=8192 Len=0 MSS=1440 WS=256 SACK_PERM=1
702 14.9891900 136.159.48.195   13.107.6.163     TCP       54 49527→443 [ACK] Seq=1 Ack=1 Win=66240 Len=0
703 14.9893970 136.159.48.195   13.107.6.163     TLSv1.2  261 Client Hello
```

edge.outlook.office365.com

outlook.office.com

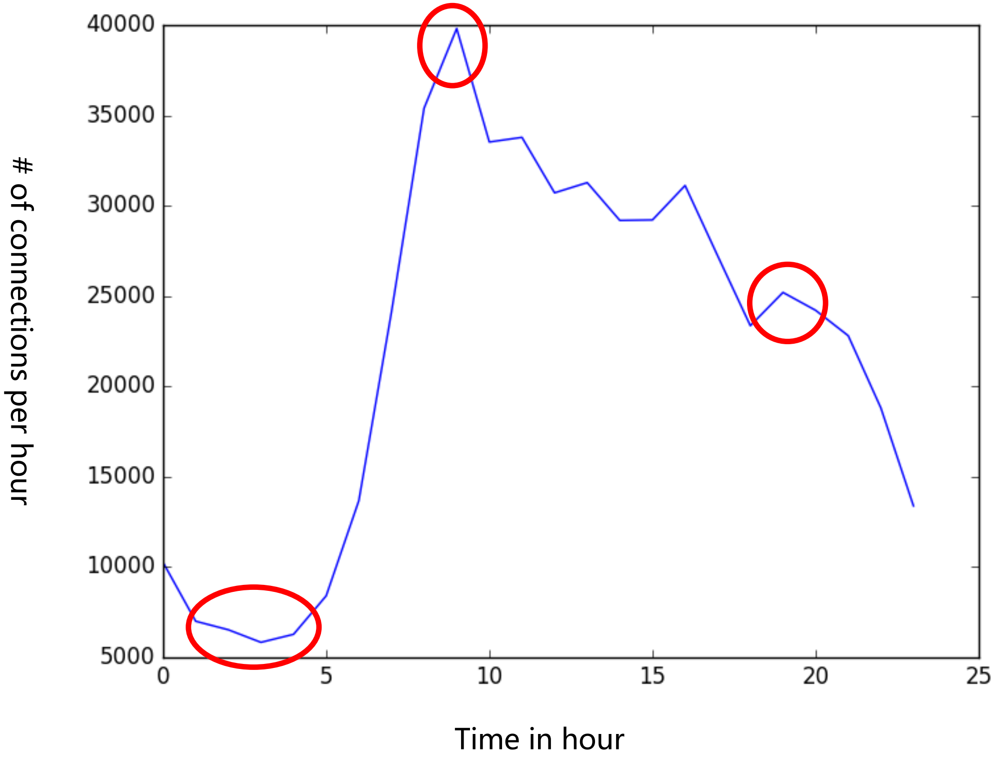
appsforoffice.microsoft.com

Traffic Overview



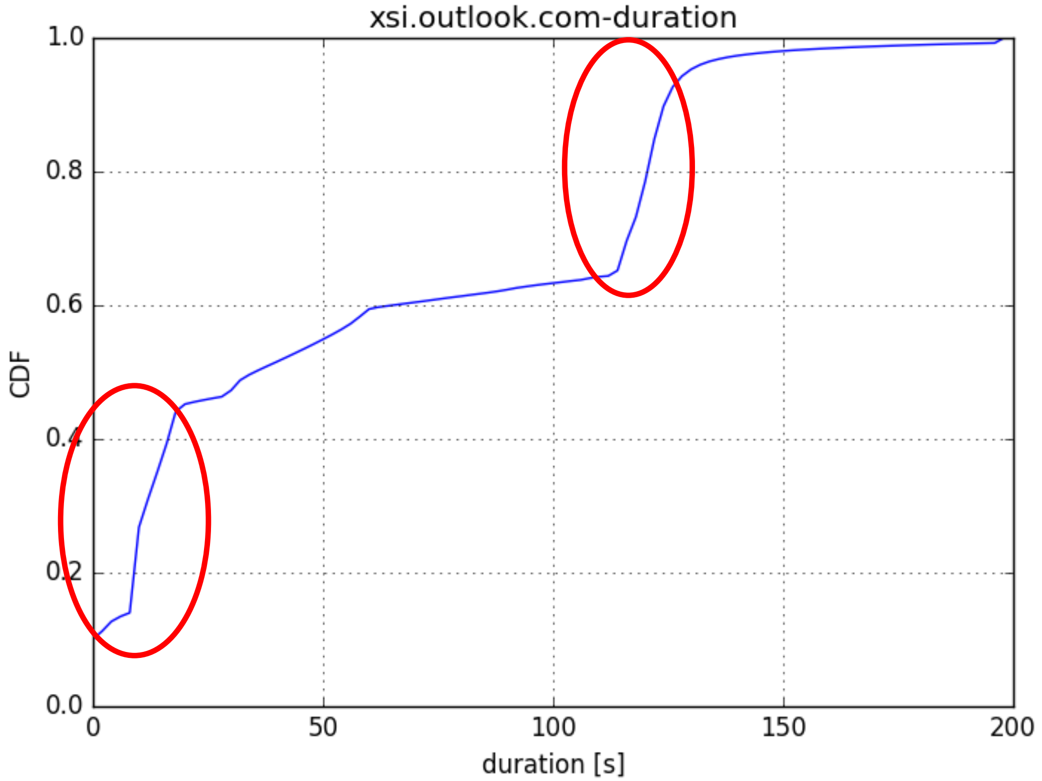
Total traffic of outlook.office.com and outlook.office365.com

Diurnal Pattern



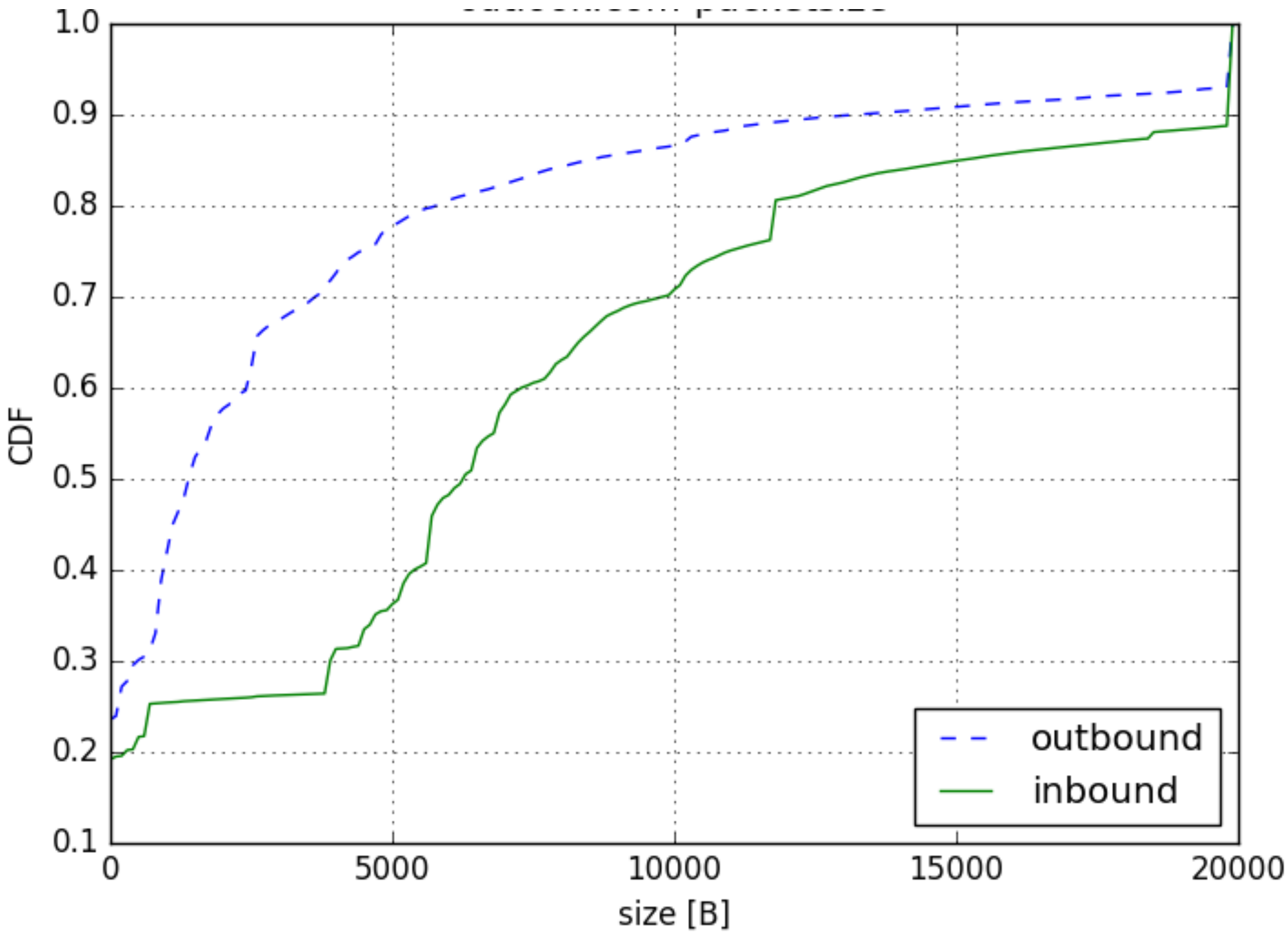
Traffic of campus email server

Connection Duration CDF



Based on xsi.outlook.com

Message Size CDF


CDF
size [B]
outbound
inbound