An Empirical Study of Campus-Level Instagram Traffic

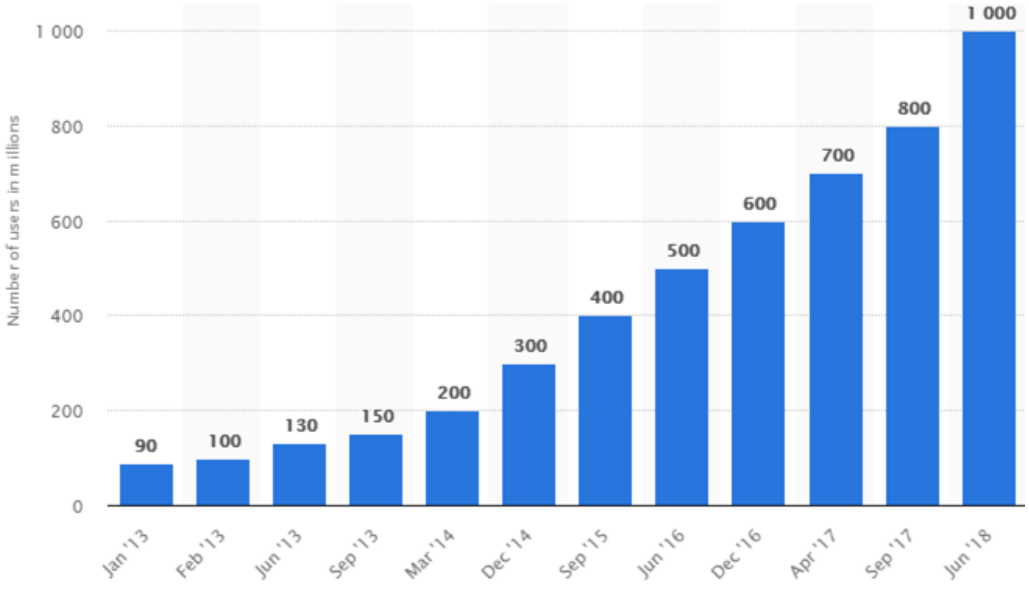
Carey Williamson
Visiting Professor, DAIS, Università Ca' Foscari Venezia
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada

November 27, 2019

[Joint work with S. Klenow, S. Keshvadi, and M. Arlitt]

Instagram

A photo and video sharing service



Monthly Active Users from 2013 to 2018 (in millions)

Features:

- Instagram Explorer
- Instagram Direct
- Instagram Stories
- Instagram TV (IGTV)
- Live Streaming

Research Questions

- What are the key characteristics of Instagram traffic?
- What are its network performance implications?

- How can this better inform our network simulation models?

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

Methodology 1 of 2: Active Measurement

UNIVERSITY OF CALGARY

Objective:

- Study Instagram traffic for 1 user from a single device under test (Android smartphone+Charles)

Rationale:

- Detect IP destinations
- Identify effects of user interactions
- Identify app features

Challenges:

- Different domain names (e.g., Instagram vs Facebook)
- Encrypted data (HTTPS)
- Certificate pinning required (rooted device)
- App security policies

[This part of the work was done by PhD student Sina Keshvadi]

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

Methodology 2 of 2: Passive Measurement

UNIVERSITY OF CALGARY

Objective:

- Study network traffic at scale

Rationale:

- Campus-level traffic characterization
- Identify traffic behaviour
- Identify any network anomalies
- Network performance implications

Challenges:

- Complexity of campus network (DHCP, NAT, VPN, BYOD)
- Traffic capturing process
- Storage/processing of traces
- Analysis of large datasets
- Encrypted traffic (HTTPS)
- Ethical considerations

[This part of the work was done by visiting MSc student Steffen Berg Klenow]

Example of Raw Data Format (Bro Conn Logs)

UNIVERSITY OF CALGARY

Epoch_Timestamp	UID	Src_IP	SPort	Dest_IP	DPort	Prot	Svc	Duration	TCPout	TCPin	State	IPout	B_out	IPin	B_in
1551596628.248886	u29N3fRWgQZb	1.2.3.4	50468	157.240.3.63	443	tcp	ssl	165.901378	9053	86515	S3	100	14297	98	90892
1551596628.250997	CThhn41IttYm27	1.2.3.4	50470	157.240.3.63	443	tcp	ssl	3.334059	489	447	RSTO	11	1093	7	1133
1551596628.301082	z1RCa3W19Ralf	1.2.3.5	50040	157.240.3.63	443	tcp	ssl	329.763400	425964	45413	SF	737	468313	538	73893
1551596628.307782	GcdG3L8hgX7e	1.2.3.6	62558	157.240.3.63	443	tcp	ssl	0.004667	39	39	SF	4	247	5	263
1551596628.316061	CLkeo71KeeiHx1	1.2.3.7	57396	157.240.3.63	443	tcp	ssl	209.412519	5239	91968	SF	67	9983	83	94918
1551596628.348089	OhKpI2eBASJOg	1.2.3.7	57397	157.240.3.63	443	tcp	ssl	209.388824	7489	914542	S3	454	31736	682	933553
1551596628.502214	Vh9Ev3gHpUSga	1.2.4.1	52990	157.240.3.63	443	tcp	ssl	8.459407	1703	126326	SF	83	6031	99	131482
1551596628.504240	pDJJC43UXfYE8	1.2.4.1	52991	157.240.3.63	443	tcp	ssl	8.457381	2222	962556	SF	431	24646	719	998714

Hardware: Endace DAG packet capture card (10 Gbps); headers only (not payloads)

Software: Bro Intrusion Detection System (IDS) (now called Zeek)

We analyze the logs using Vertica (a big data analytics framework available from HP)

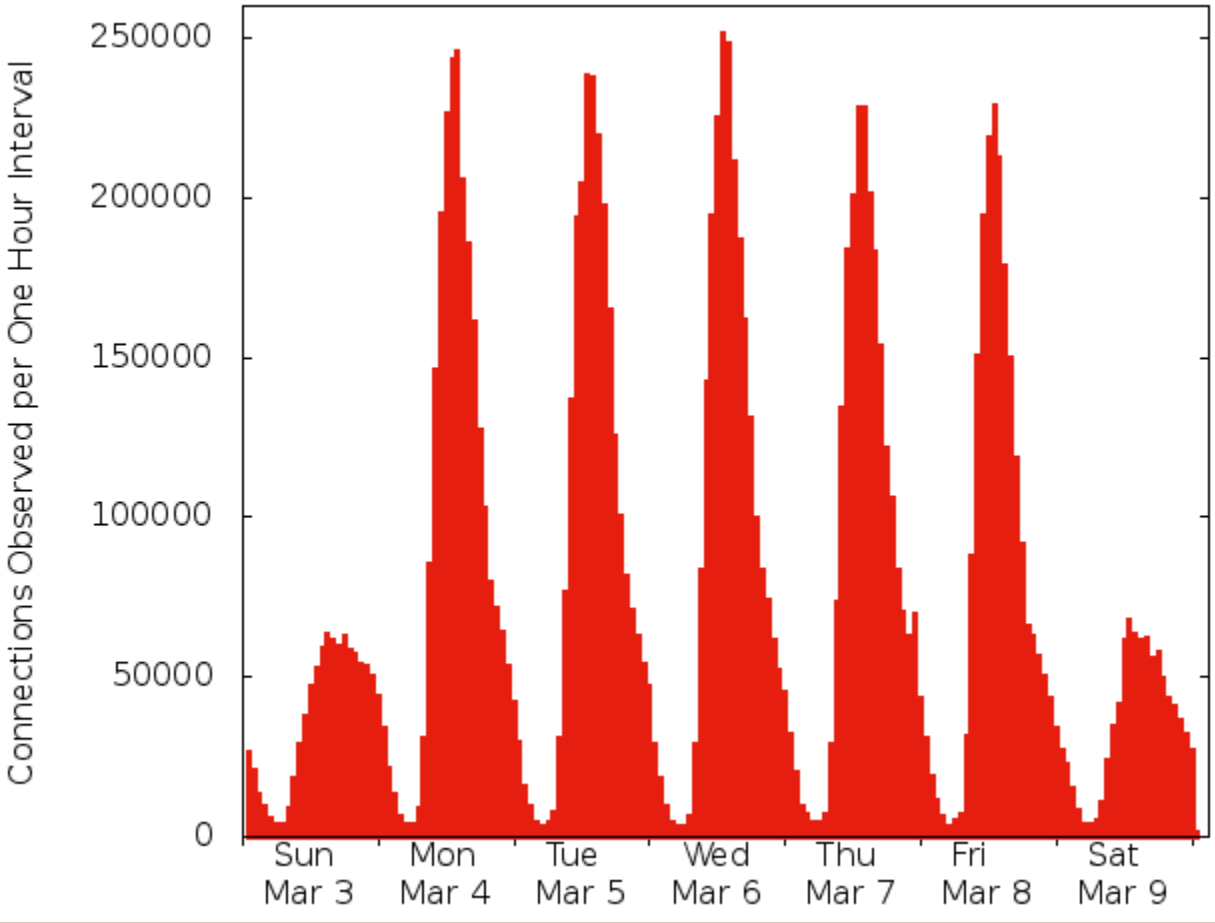
[This data collection/analysis infrastructure was designed by Martin Arlitt]

Overview of UCalgary Instagram Traffic (1 Week)

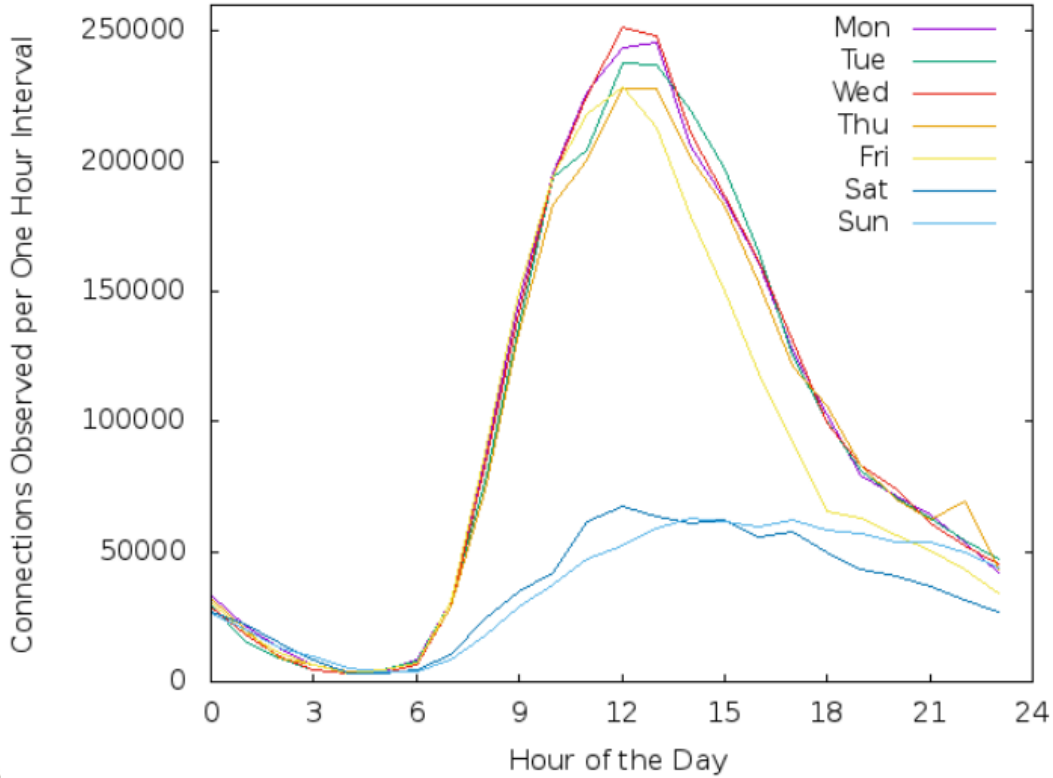
UNIVERSITY OF CALGARY

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

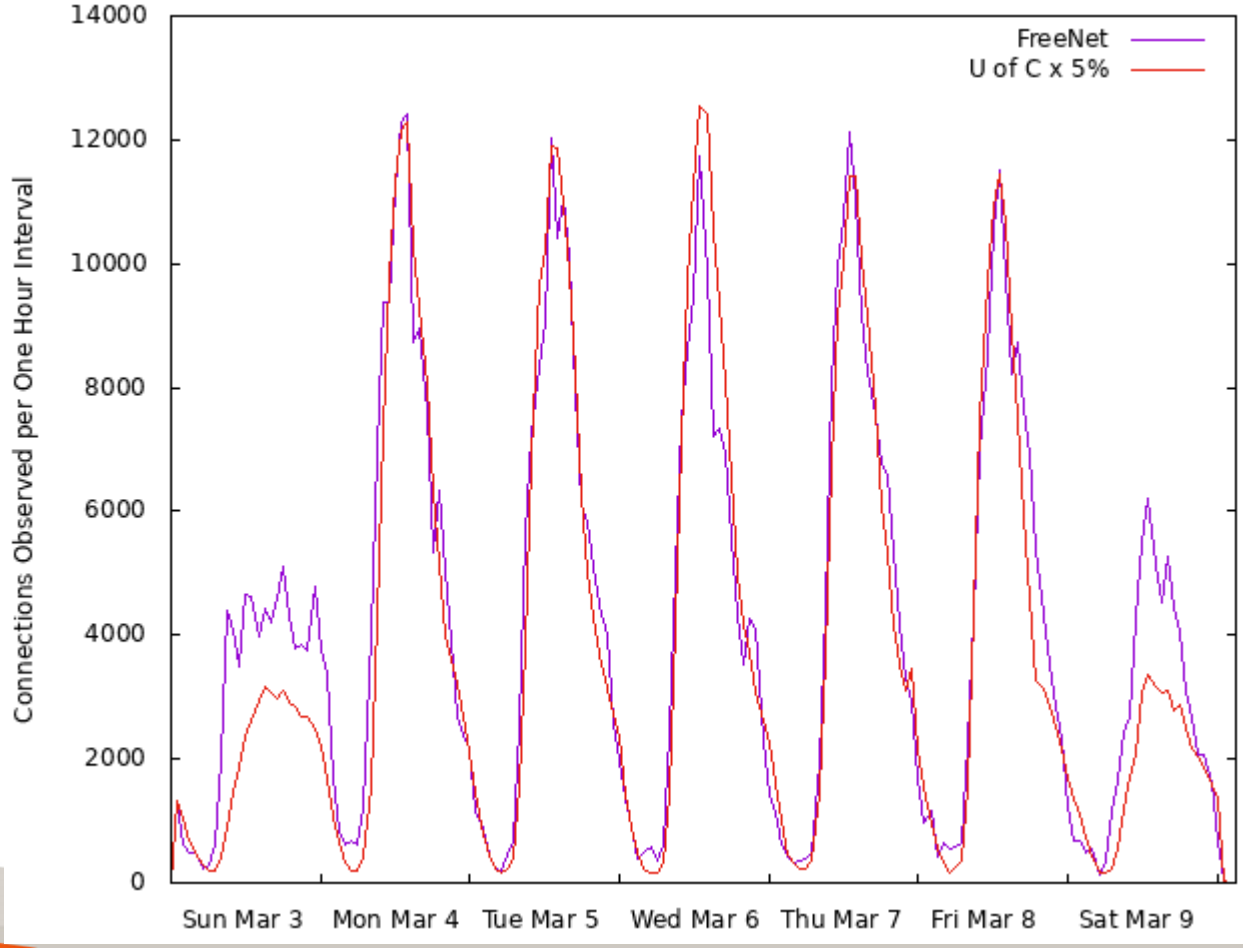
U of C Instagram Traffic Profile



Daily Instagram Traffic

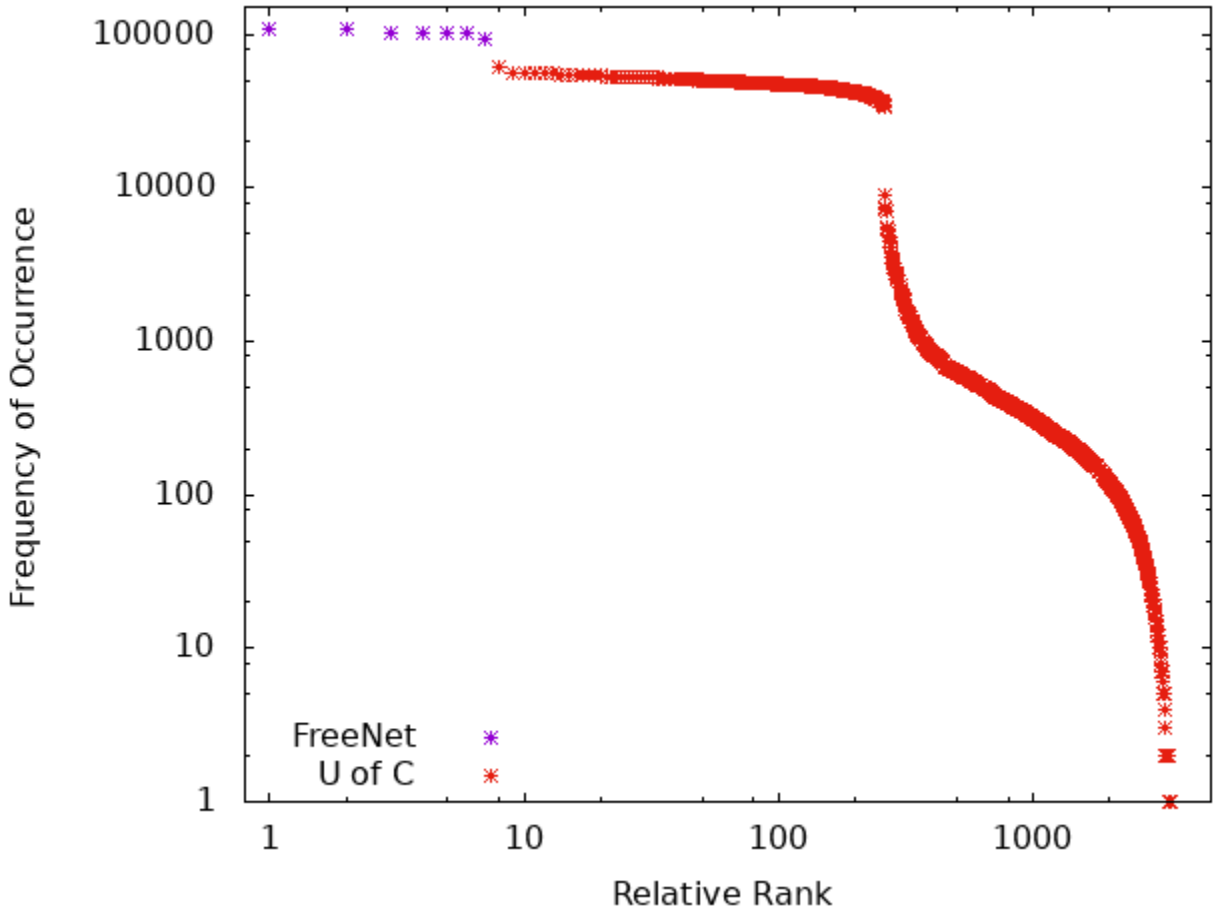


Freenet Instagram Traffic Profile



UNIVERSITY OF CALGARY

IP Address Frequency-Rank Profile

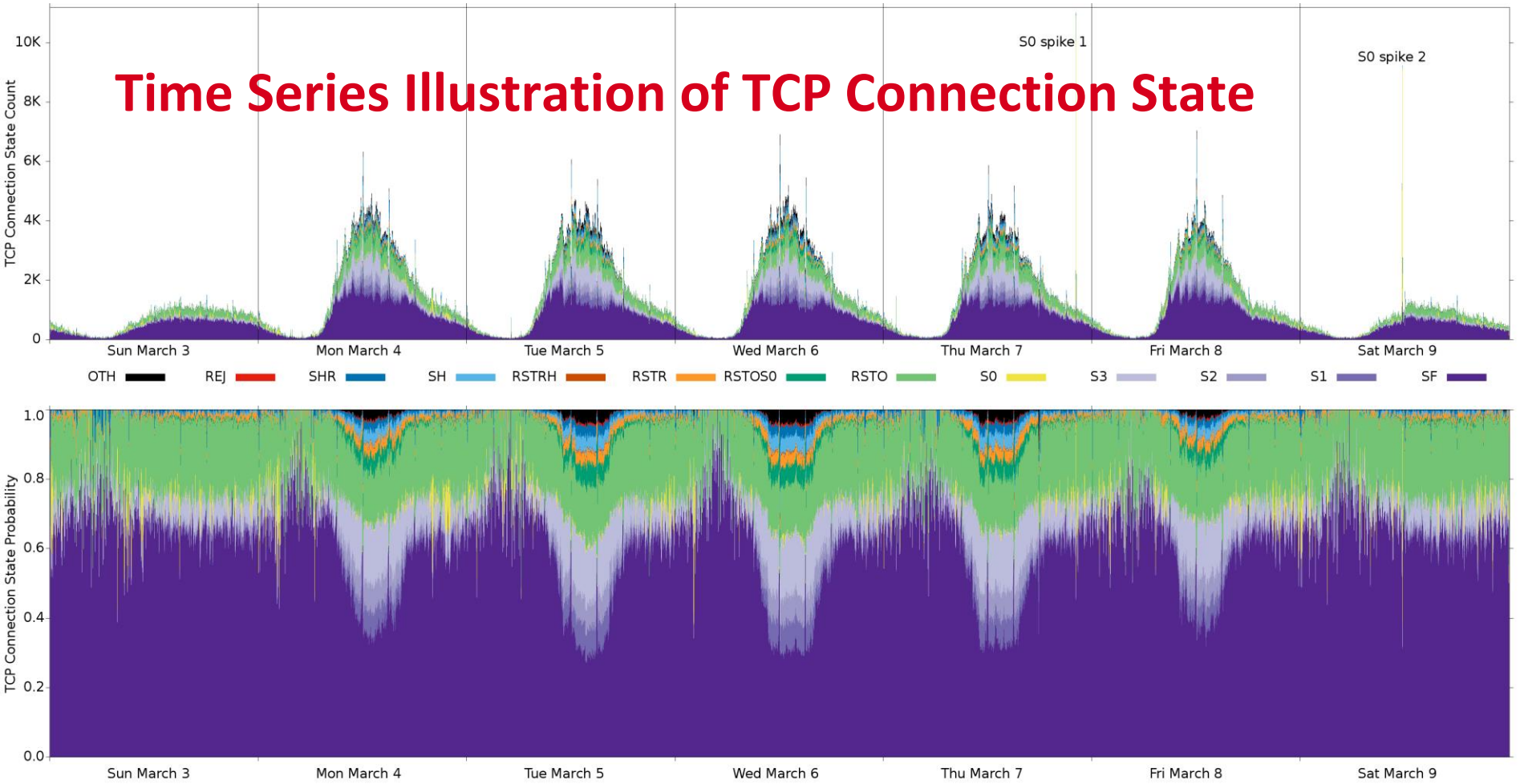


Observed TCP Connection States

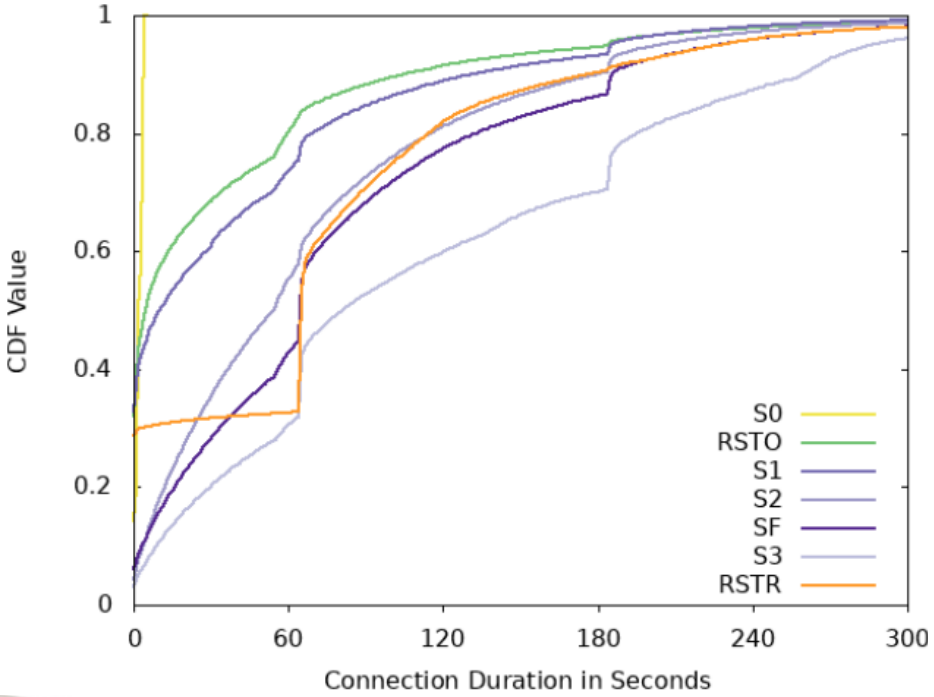
UNIVERSITY OF CALGARY

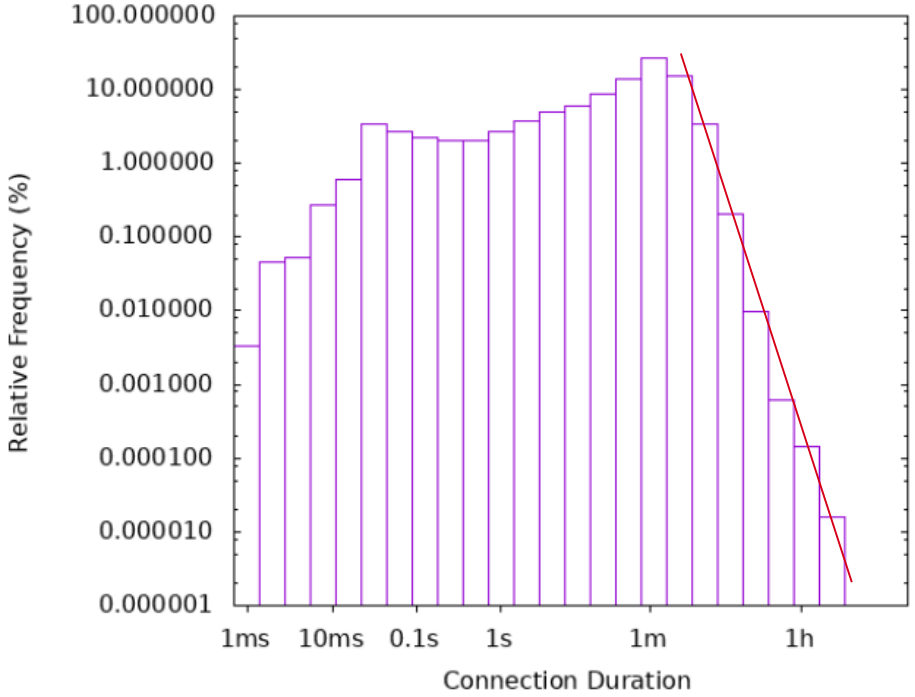
State Description	Conns	%Conns	Bytes	%Bytes
SF: SYN-FIN	6,265,336	47.88%	3.78 TB	52.55%
RSTO: origin reset	2,487,505	19.01%	1.74 TB	22.91%
S3: no FIN seen	1,554,591	11.88%	879.9 GB	11.21%
S2: client FIN only	595,772	4.55%	340.1 GB	4.38%
S1: server FIN only	498,635	3.81%	189.7 GB	2.33%
RSTOS0: fail/RSTO	354,775	2.71%	222.9 GB	2.87%
RSTR: rcvr reset	335,304	2.56%	49.2 GB	0.63%
SH: no SYN-ACK	294,300	2.25%	107.1 GB	1.37%
SHR: no SYN seen	273,951	2.09%	57.3 GB	0.74%
OTH: other state	201,788	1.54%	71.3 GB	0.92%
S0: failed setup	166,822	1.27%	0.03 GB	$< 0.01\%$
REJ: rejected	37,455	0.29%	4.5 GB	0.06%
RSTRH: rcvr reset	20,329	0.16%	2.0 GB	0.03%
Total	13,086,563	100.0%	7.5 TB	100.0%

Time Series Illustration of TCP Connection State

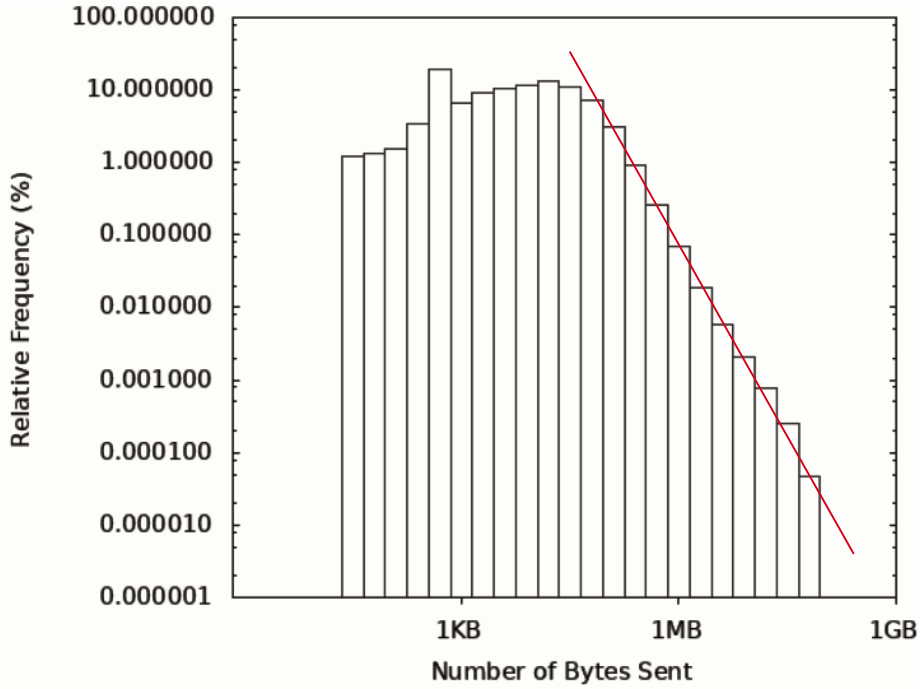


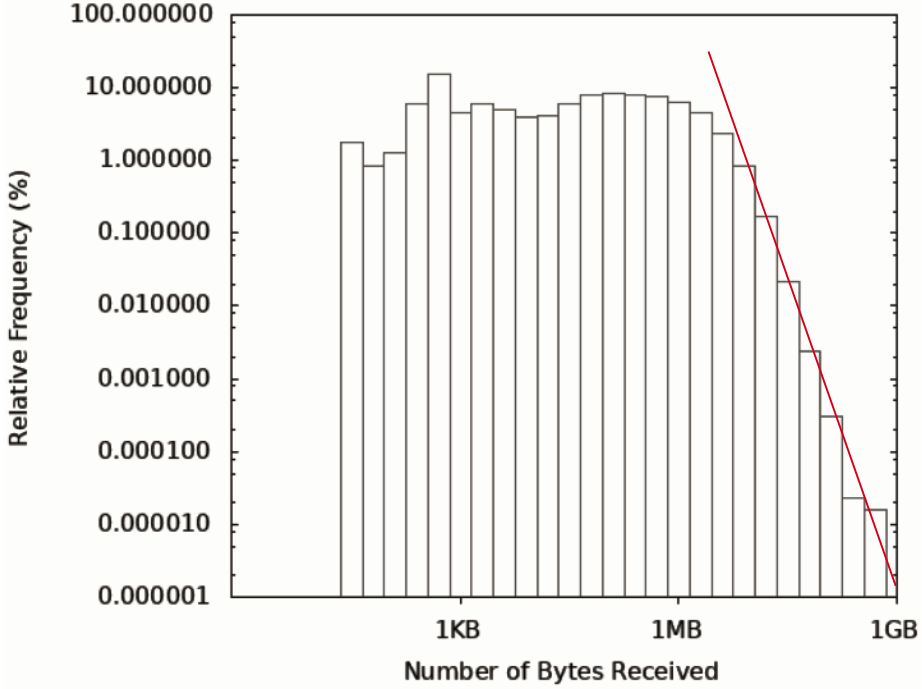
Distribution of TCP Connection Duration



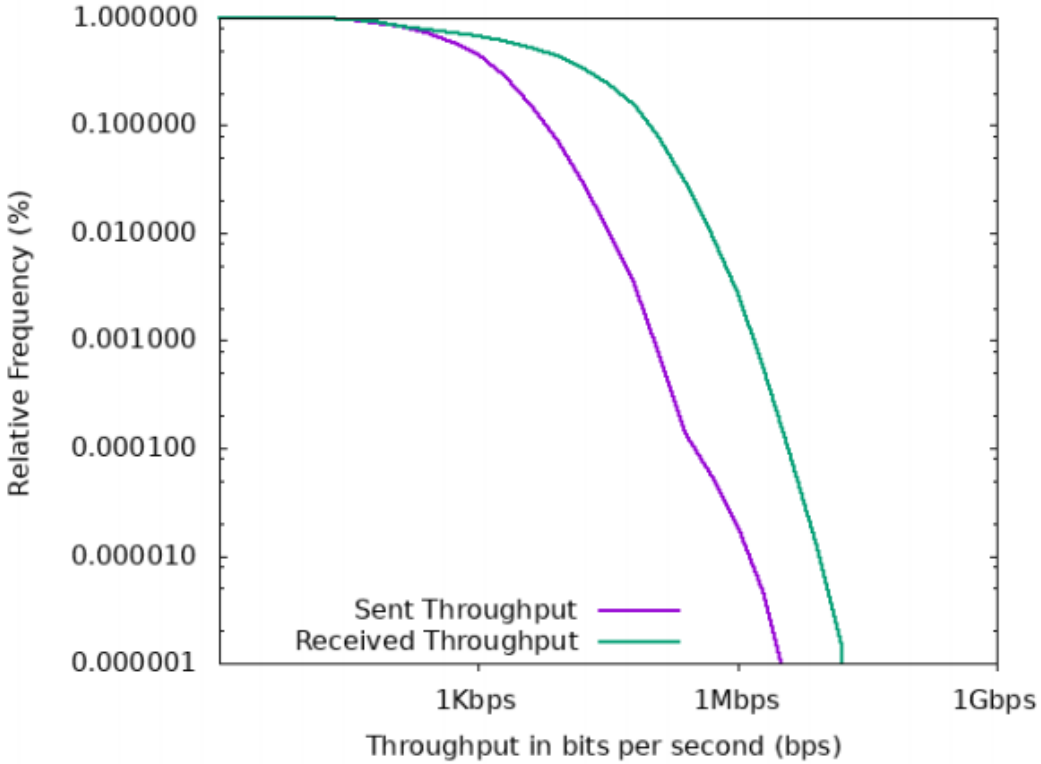


Transfer Sizes in Bytes





TCP Connection Throughput



Conclusions

- 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!

Acknowledgements

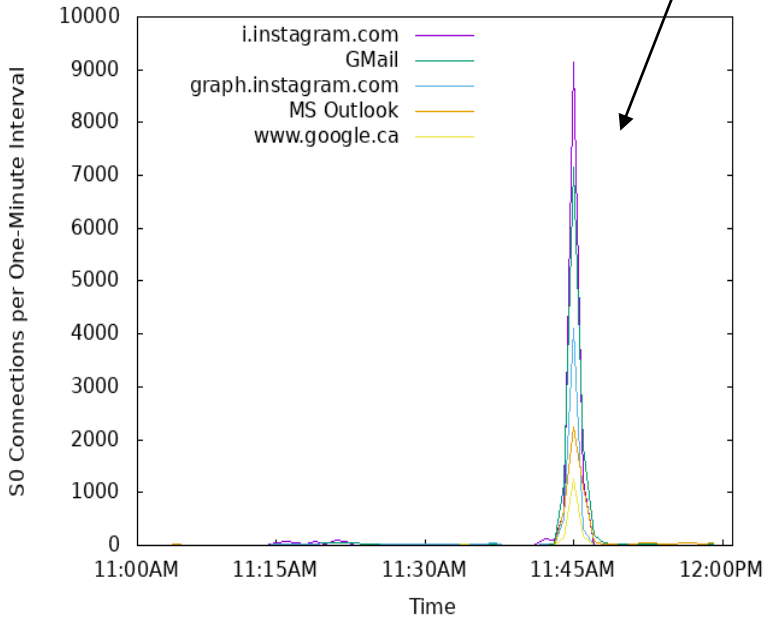
Financial support for this work was provided by Canada's Natural Sciences and Engineering Research Council (NSERC).

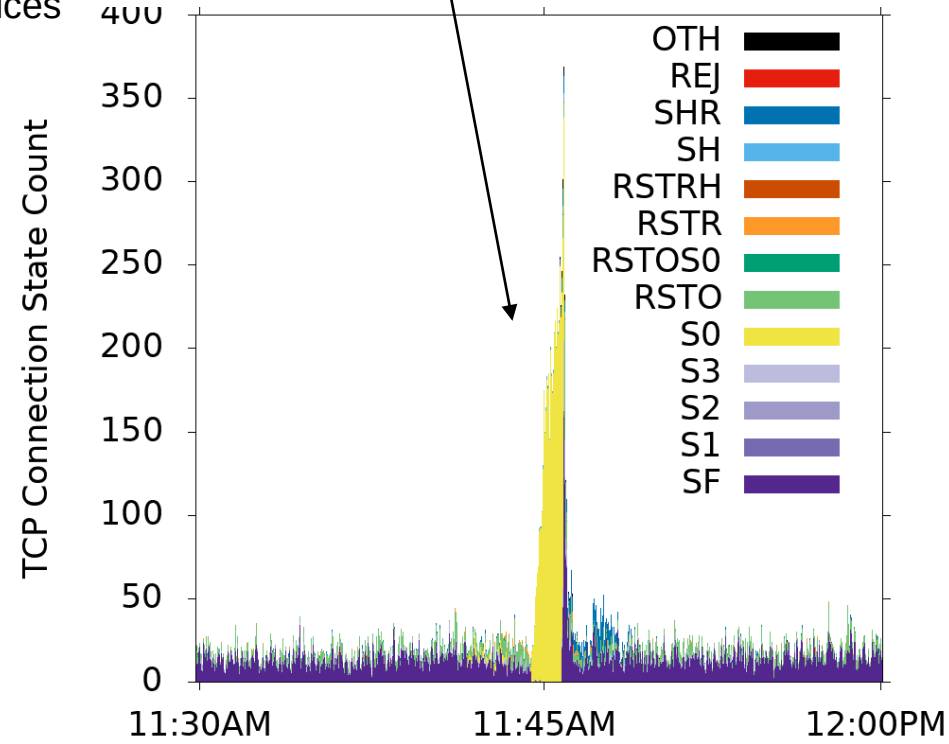
A big thanks to University of Calgary Information Technologies (UCIT) for enabling the collection of our network traffic measurement data.


Thank you!

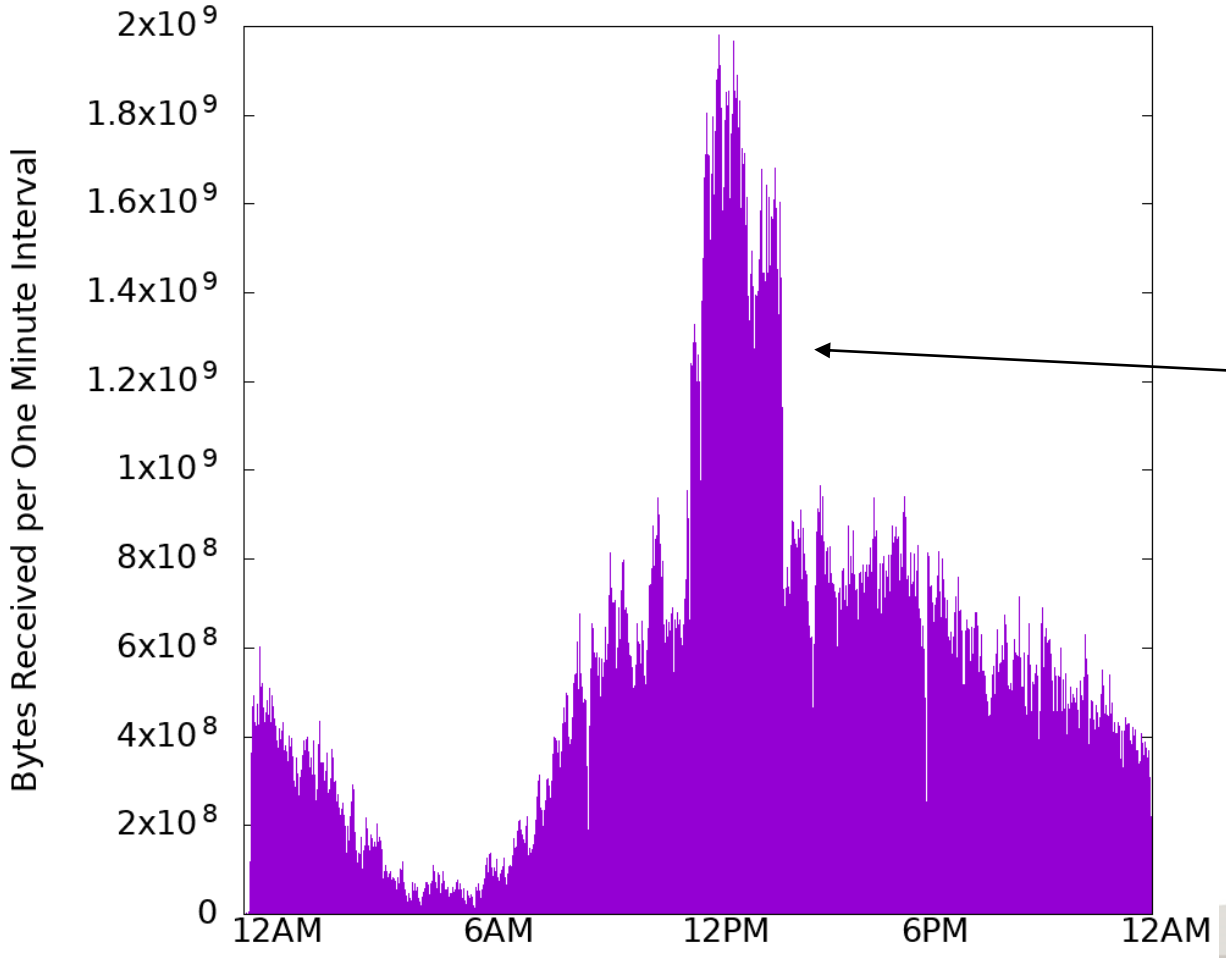
Instagram Traffic Anomaly (1 of 3)

A brief network router outage for about 80 seconds on Saturday March 9, which affected Instagram traffic, and other network services



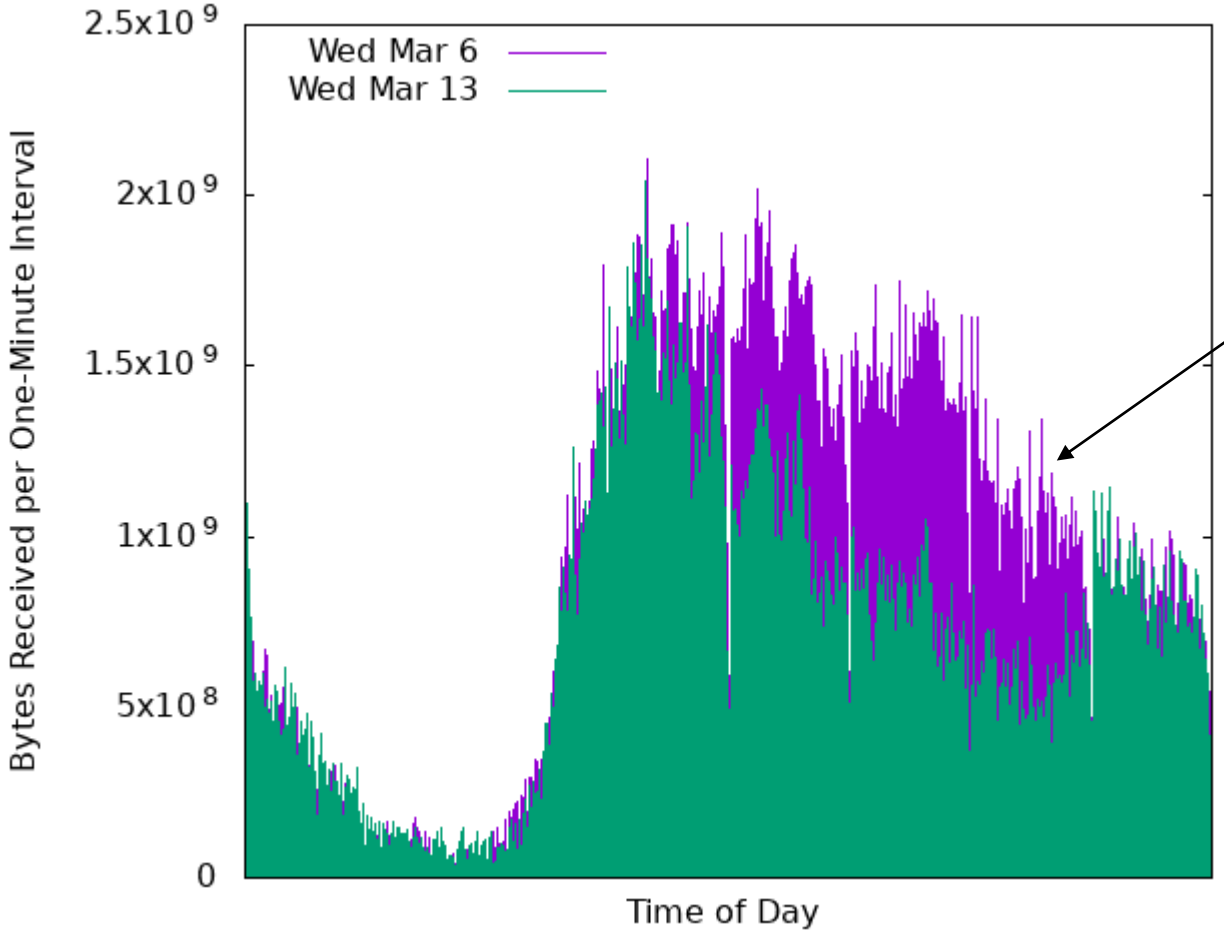


Instagram Traffic Anomaly (2 of 3)



Multiple UCalgary IP addresses partaking in some Instagram video streaming event for about 2 hours on Sat March 9

Instagram Traffic Anomaly (3 of 3)



Partial outage for Facebook, WhatsApp, and Instagram for several hours on Wed March 13/19