Characterization of IMAPS Email Traffic

Mehdi Karamollahi
Sapienza University of Rome
University of Calgary
mehdi.karamollahi@ucalgary.ca

Carey Williamson
University of Calgary
Calgary, AB, Canada
carey@cpsc.ucalgary.ca

***Abstract*—The email delivery ecosystem in modern enterprise networks is large and complex, featuring several email access protocols (e.g., POP, IMAP, HTTP, and the secure versions of each of these). In this paper, we provide a detailed characterization of IMAPS email traffic on a large campus edge network. The main highlights from our study are: (1) commercial cloud-based email services account for most of the campus email traffic; (2) all email access protocols exhibit heavy-tailed transfer size distributions; and (3) the throughputs achieved by large transfers vary widely based on protocol, transfer size, and time of day.**

***Index Terms*—Network traffic measurement; email; IMAPS**

I. Introduction

Electronic mail (email) has evolved a lot over the years, and many different protocols have been implemented to support this classic network application. Simple Mail Transfer Protocol (SMTP) was one of the earliest such protocols [21] for peer-to-peer exchange of messages between mail servers, and is still in use on some networks today for sending emails from user agents to mail servers. On the client side, several email access protocols have been used, including Post Office Protocol (POP), Internet Message Access Protocol (IMAP) [10], and Hyper-Text Transfer Protocol (HTTP).

In recent years, the trends have been toward Web-based email, cloud-hosted email providers, and improved email security. Web-based email services are the most popular, and many people commonly use Web interfaces based on HTTPS (HTTP over SSL/TLS protocols). Nonetheless, a wide variety of email clients are used every day. Some use POP or IMAP to retrieve emails from mail servers, and several use SMTP to transfer outgoing emails from a client to a mail server.

IMAP is a well-known Internet protocol [10]. It is used by email clients to retrieve email messages from mail servers over a Transmission Control Protocol (TCP) connection. IMAP supports remote management of email messages in a mailbox at the server, with a consistent view of email even when accessed from multiple different client devices. Early deployments of IMAP used TCP port 143 for unencrypted data exchanges, though modern implementations use TLS/SSL to transfer emails securely on TCP port 993 (IMAPS).

In this paper, we focus on IMAPS (secure IMAP) traffic, for several reasons. First, many studies have been done on different email services, but IMAPS traffic characterization is missing from the literature, as far as we can tell. Second, the volume of traffic carried over IMAPS is much higher than we expected, given that our campus uses Microsoft Outlook (Office 365) as its official cloud-based email provider. Finally, in our investigations of IMAPS traffic, we found evidence of extremely large transfers, which suggests that this protocol is perhaps being used in different ways (e.g., file transfer and backup services) than a traditional email access protocol.

The research questions underlying our work are as follows:

- What are the characteristics of modern email traffic?
- What email protocols are in use on enterprise networks?
- What are the potential performance implications of these email protocols on a large-scale enterprise network?

In our study, we investigate the IMAPS data traffic viewed at an edge router of the University of Calgary campus network. We have several years of data logs available, but for the purpose of this paper, we focus primarily on one representative week of data. This observational period from a recent busy semester provides sufficient data to study hourly, daily, and weekly patterns, as well as heavy-tailed transfer sizes.

TABLE I
Overview of Email Access Protocols (April 14-20, 2019)

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

As motivational context for our study, Table I shows a statistical summary of the email traffic observed on our campus network for a one-week period from April 14-20, 2019. The table shows the number of TCP connections and the total byte traffic volume exchanged on the ports used by the major email protocols. In this table, Web-based email to Microsoft Outlook using HTTPS is the most dominant, with several Terabytes (TB) of data exchanged during the week. Given that Web-based email has already been quite well studied [12], [23], [25], we focus instead on IMAPS traffic, which is about an order of magnitude smaller than the HTTPS traffic. The volume of traffic (530 GB) exchanged via IMAPS port 993 is about five times larger than the sum (103 GB) of the email traffic on all other remaining protocols/ports (ignoring HTTPS), which is why it is of interest to us. Furthermore, a small subset of this

TABLE II
COMPARISON OF IMAPS EMAIL TRAFFIC AND WEB-BASED EMAIL TRAFFIC (HTTPS TO MICROSOFT OUTLOOK)

Characteristic	Similarities	Differences	Section
Traffic Profile	Both services show strong diurnal traffic patterns.	IMAPS has more noticable spikes in off-peak hours.	IV-A
Traffic Volume	Large number of connections, and high data volume.	HTTPS email traffic volume is 10x larger than IMAPS traffic.	IV-B
Connection State	Both use TCP connections for data exchanges.	Outlook has more TCP resets (and rejects) than Gmail.	IV-C
Origin-Destination	Both have highly non-uniform sources and destinations.	Outlook (HTTPS) uses more servers than Gmail (IMAPS).	IV-D
Asymmetry	Email protocols involve bi-directional data exchange.	IMAPS has far greater asymmetry than HTTPS.	IV-E
Transfer Sizes	Both have heavy-tailed transfer sizes.	HTTPS email transfer sizes have a heavier tail than IMAPS.	IV-F
Throughput	Throughput varies with transfer size and time of day.	IMAPS often has higher throughput than HTTPS.	IV-G

traffic (7.3 GB) actually goes to Microsoft Outlook, facilitating comparisons between HTTPS and IMAPS.

Table II highlights the main insights that emerge from our study. In general, the usage of HTTPS and IMAPS for email services share several similarities (e.g., diurnal patterns, bi-directional traffic, and heavy-tailed transfer sizes). However, there are also noticable differences in how these protocols are used (e.g., connection state, origin-destination pairs, asymmetry, throughput, and heavy hitters). The rightmost column of Table II indicates the section of the paper in which each characteristic is discussed and analyzed in more detail.

The remainder of this paper is organized as follows. Section II provides some background on email access protocols and discusses relevant prior work. Section III presents our research methodology and our empirical datasets. Section IV presents our results, while Section V concludes the paper.

II. RELATED WORK

Internet traffic measurement is a widely used technique to study network-based applications [11]. Two main approaches are *passive* and *active* network measurement. Passive approaches observe a network system without perturbing it, while active approaches use probe packets or test sessions to see how a network or server responds to certain requests. These basic techniques can be used to collect and analyze data to investigate different aspects of network systems. A general tutorial on network traffic measurement is provided in [24].

Network traffic measurements can be used for workload characterization. Classic studies have focused on Web traffic characterization [2], [7], peer-to-peer applications [4], and video streaming services [6], [9], [14]. Other recent works have focused on Netflix [1], [16], online social networks [5], [18], [20], and campus-level email traffic characterization [25].

The same measurement techniques can also be used to assess network security. For example, Durumeric *et al.* [12] conducted a detailed study regarding email security. Their goal was to assess how quickly major email service providers adopt new security settings. Their results showed that top mail providers are more proactive in adopting new security configurations. As another example, Ramachandran *et al.* [22] studied spam email, and developed an approach to detect spam using network-level footprints.

The privacy/security of email is also a recurring theme in the literature. Englehardt *et al.* [13] studied privacy issues in email transfers, and indicated third-party tracking as a potential issue with commercial email services. Schatzmann *et al.* [23] developed a flow-based classification technique for encrypted Web-mail traffic. Gupta *et al.* [15] addressed the tradeoffs between email privacy and essential functions such as spam filtering. Their work showed that spam filtering is still possible, even in the presence of end-to-end encryption.

Some of our own prior work studied Outlook email traffic [25]. To the best of our knowledge, however, IMAPS traffic has not been studied before, which is why we focus on it.

III. METHODOLOGY

Our work uses a combination of passive and active approaches to network traffic measurement.

For passive measurement, we used an Endace DAG packet capture card. This network monitor is installed at the edge router of our campus network, and receives a live feed of inbound and outbound packet traffic via port-mirroring at the router. Conceptually, this monitor is similar to Wireshark, but it captures packets at multi-Gigabit rates using specialized hardware. Our monitor also includes a compute engine for the processing and storage of trace data. It has two Intel Xeon E5-2690 CPUs, 64 GB RAM, and 5.5 TB of hard disk.

To facilitate long-term traffic studies, our monitoring system records data at the flow level, rather than the packet level. Furthermore, only the TCP/IP packet headers are captured, not the payloads. All observed packet headers are fed to the Bro IDS [19], which performs flow-based grouping of all the packets belonging to a certain connection. Bro then creates one entry in the log for that connection with the information including packets sent/received, bytes sent/received, and the duration of the entire connection.

In our implementation, Bro creates hourly connection-level logs based on certain transport-layer protocols and ports. Since in this work we are only interested in IMAPS traffic, we extract from the connection logs all the entries for which the transport-layer protocol is TCP and the destination port is 993. When relevant, we also extract SMTP and HTTPS email traffic as a basis for comparison.

For active measurement, we used common tools to assess email server infrastructure, DNS name resolution, IP address geolocation, routing path, network RTT latency, and so on. For a better understanding of the behavior of IMAPS email sessions, we also conducted some short test sessions from our own laptop and desktop computers. For these test sessions, we used Wireshark to capture the resulting packets, and then relate them to the connections reported in the Bro logs.

IV. Empirical Measurement Results

In this section, we focus on the characteristics of the IMAPS email traffic observed on our campus network.

A. *Traffic Profile*

Figure 1 provides a graphical overview of the IMAPS traffic. In these time series plots, the top graph shows the number of IMAPS TCP connections initiated in each one-hour period during our one-week of observation, while the bottom graph shows the corresponding plot for data traffic volume.

The most obvious observation from Figure 1 is the strong diurnal pattern observed in the traffic, as seen in many other network traffic studies. Email traffic activity is largely human-driven, with strong peaks during the normal working hours, a small secondary peak at bedtime, and then lighter traffic in late evening or early morning hours. There is also a noticable decline in traffic on weekends and holidays (note that April 20 was the Good Friday statutory holiday), since fewer people are on campus. The extraneous spikes in the graph are attributable to machine-generated email traffic, such as scanning attacks.

Another observation from Figure 1 is the positive correlation between the number of connections and the traffic volume exchanged. This is not unexpected, because of the temporal dependence in the traffic. However, the data volume graph shows more spikes than the connections graph, suggesting high variability in the email transfer sizes for each connection.

A final observation from Figure 1 is the asymmetry in the data traffic volumes. That is, the vast majority of the IMAPS traffic is inbound, rather than outbound. This is a design feature of the protocol, since it provides remote access to email stored on the server. Messages and attachments are only retrieved and transferred when requested by the user. Furthermore, outbound email messages typically use other protocol, such as HTTPS, or SMTP on either port 25 or 587.

B. *Data Volume*

Table I earlier in the paper already illustrated the order of magnitude difference in data traffic volume between HTTPS Web-based email and IMAPS email traffic. We next explore this characteristic on a per-connection basis.

Table III shows the median and mean durations, as well as the median and mean transfer sizes, for TCP connections from each email protocol that we studied. This table shows several differences amongst the email access protocols. In particular, the average duration for HTTPS is greater than the average duration for IMAPS, even though the mean transfer size is smaller for HTTPS. The main reason for this is the use of many concurrent persistent TCP connections for Web-based email in Microsoft Outlook [25]. In contrast, IMAP-based user agents tend to have one long persistent connection to manage the email session, and a separate TCP connection for each message or attachment selected for retrieval. The anomalously low median duration and transfer size for HTTPS is attributable to an excessive number of TCP rejects, as mentioned in the next section.

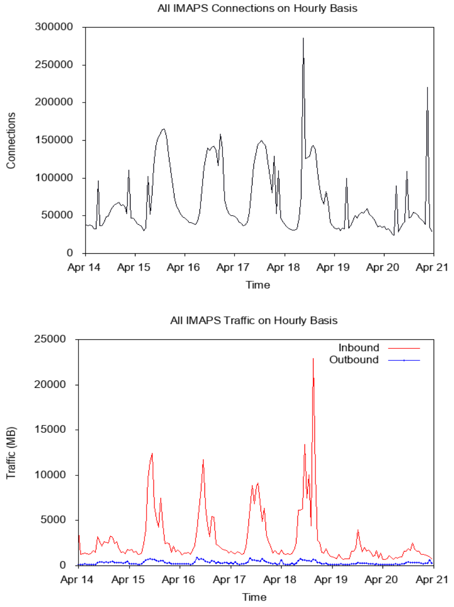


Fig. 1. IMAPS Email Traffic Profile (April 14–20, 2019)

Another observation from Table III is that the mean tends to be much larger than the median, both for connection durations and transfer sizes. This characteristic typically indicates a distribution with many small values (mice) and some extremely large values (elephants). We investigate the distributions in more detail later in Section IV-F.

TABLE III
Comparison of Connection Durations and Transfer Sizes

Email Protocol/Port	Conn Duration (s)		Transfer Size (bytes)	
	Median	Mean	Median	Mean
HTTPS/443	0.019	69.7	208	74,492
IMAP/143	0.00021	117	9,797	135,134
IMAPS/993	2.24	44.4	13,020	99,910
POP/110	1.93	2.44	12,001	38,319
POPS/995	2.06	9.35	14,045	82,834
SMTP/25	2.95	4.46	11,736	28,529
SMTPS/587	61.4	50.0	16,134	408,439

C. *TCP Connection State*

Table IV shows the results from our analysis of the TCP connection states reported in the Bro logs. (Note that the rows in the table are sorted into descending order based on the rightmost column, for IMAPS Gmail traffic volume.) A normal TCP connection starts with a SYN/SYN-ACK handshake, and ends with a FIN/FIN-ACK handshake, and is represented by the SF state. Surprisingly, less than half of the observed TCP connections have the SF state.

The most unusual example in Table IV is the HTTPS email traffic to Microsoft Outlook, which has only 15.5% SF. Nonetheless, close to half (47.4%) of the data bytes are exchanged over SF connections. The biggest surprise is that

TABLE IV
SUMMARY OF TCP CONNECTION STATES OBSERVED FOR HTTPS AND IMAPS EMAIL TRAFFIC TO MICROSOFT (MS) AND GOOGLE (G)

TCP	Description	Connections			Bytes		
State	of State	HTTPS MS	IMAPS MS	IMAPS G	HTTPS MS	IMAPS MS	IMAPS G
SF	Normal SYN-FIN connection	15.50%	46.36%	48.66%	47.40%	34.59%	52.81%
S1	Good conn, but server FIN only	3.33%	8.23%	6.96%	17.88%	25.73%	17.76%
OTH	Mid-stream traffic (no SYN or FIN)	0.77%	14.12%	7.63%	3.54%	24.51%	9.87%
RSTO	Conn reset by originator	3.84%	2.24%	7.77%	9.44%	1.24%	5.78%
S3	Good conn, but no FIN seen at all	2.62%	6.11%	4.51%	7.84%	3.56%	5.64%
SH	Half-open (SYN and FIN, but no SYN ACK)	0.55%	0.78%	0.84%	1.16%	0.57%	3.43%
S2	Good conn, but client FIN only	1.70%	3.02%	1.73%	4.84%	2.17%	2.39%
RSTR	Conn reset by receiver	0.67%	5.32%	0.91%	3.21%	4.46%	1.19%
SHR	Half-open (SYN ACK and FIN, but no SYN)	0.57%	8.37%	19.82%	0.70%	0.85%	0.78%
RSTOS0	Failed conn reset by originator	0.87%	0.50%	0.91%	2.40%	0.10%	0.29%
RSTRH	Conn reset by receiver	7.15%	4.50%	0.16%	1.16%	1.98%	0.03%
REJ	Connection ended with a Reject	57.50%	0.36%	0.07%	0.42%	0.24%	0.02%
S0	Saw SYN, but no SYN-ACK at all	4.93%	0.09%	0.02%	0.01%	<0.01%	<0.01%
Total	All TCP connections	100.0%	100.00%	100.0%	100.0%	100.00%	100.0%

57% of the connections[1] involve TCP rejects. The proportion of TCP resets is also significant for Outlook (7.15%), but this is a known issue for many Microsoft servers [3], especially for persistent connections [25]. The percentage of unsuccessful connection attempts (S0) is 4.93%, which is also anomalous. Overall, about two-thirds of the Outlook HTTPS connections are unsuccessful.

The IMAPS traffic for Gmail ("IMAPS G" in Table IV) has about half (48.66%) of its connections in the SF state, and just over half (52.81%) of its data exchanged on SF connections. Some other states observed here are S1, S2, S3, and OTH, which often represent long-duration connections for which the monitor did not see the SYN and the FIN from both endpoints, perhaps because of network load, or because these events straddled across multiple one-hour logs. For Gmail, resets by the orginator (RST0) are much more prevalent (7.77% conns, 5.78% bytes) than resets by the responder (RSTR, about 1% of connections and bytes).

A subset of the observed IMAPS traffic on our campus network goes to Microsoft Outlook, and is shown as "IMAPS MS" in Table IV. The connection states for this traffic resemble those of the IMAPS G traffic, indicating that the REJ anomaly identified earlier applies only for Web-based Outlook email traffic on HTTPS, and not more generally. For IMAPS MS, resets by the receiver (RSTR 5.32%, RSTRH 4.50%) are more prevalent than those by the originator (RSTO 2.24%). The proportion of bytes exchanged on SF connections (34.59%) is also lower than for IMAPS G, perhaps suggesting that there are even more instances of large and long-duration transfers for IMAPS MS that lead to S1/S2/S3/OTH states. About an equal volume of bytes (25% each) are exchanged over S1 and OTH connections.

D. Origins and Destinations

We next look at the IP addresses of the sources (origins) and target destinations of the observed IMAPS traffic.

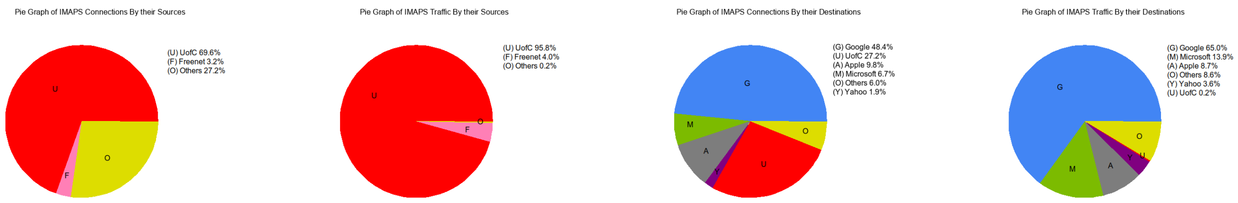
Figure 2 provides a high-level overview of the originators for IMAPS traffic. Figure 2(a) provides a connection-level view. In this pie chart, the red portion (about 75%) represents University of Calgary IP addresses using IMAPS to connect to external email servers, while the yellow portion (about 20%) represents external IP addresses attempting to reach University of Calgary servers using IMAPS, and the pink portion (about 5%) represents the Calgary Community Network Association (FreeNet) using our campus network as a transit ISP to reach email servers on the Internet. Most of the external (yellow) connections represent scanning traffic looking for vulnerable email servers, but these are unsuccessful in doing so. This becomes evident from Figure 2(b), which provides a data volume view. The University of Calgary (95.1%) and FreeNet (4.6%) account for almost all (99.7%) of the IMAPS data traffic volume exchanged.

Figure 2(c) and (d) provide the corresponding pie charts for the destinations of IMAPS traffic. Most of the traffic goes to popular cloud-based email providers (e.g., Gmail, Apple, Microsoft Outlook, Yahoo), but some is inbound scanning traffic to the U of Calgary as well.

Figure 3 shows frequency-rank profile plots for the IP source (top) and destination (bottom) addresses for the IMAPS traffic. In these graphs, the vertical axis shows the frequency with which different IP addresses are seen, while the horizontal axis shows the relative (ordinal) rank of each IP address, based on its activity. Both axes use a log scale. Such graphs are often used to look for evidence of Zipf-like power-law structures in empirical data [7].

The IP source graph (top) in Figure 3 is piece-wise linear, with several distinct parts. The red points on this graph represent about 3800 U of Calgary IP addresses that sent IMAPS traffic. The prominent almost-horizontal red band across the middle of the plot shows about 300 U of Calgary IP addresses that are used roughly equally. These correspond to popular subnets on campus that are used for wireless, DHCP, NAT, and VPN functionality. Many users share this IP space, and the (machine-generated) NAT load balancing leads to the flatness of the graph. The rest of the graph exhibits Zipf-like

[1] We suspect that this anomaly is due to a misconfiguration in our campus NAT, and have reported it to our university-level IT team. We have also pruned the rejected connections from our subsequent analyses.



(a) Source Connections (b) Source Data Volume (c) Dest Connections (d) Dest Data Volume

Fig. 2. Pie Charts Illustrating Sources and Destinations for IMAPS Traffic in Connections and Bytes

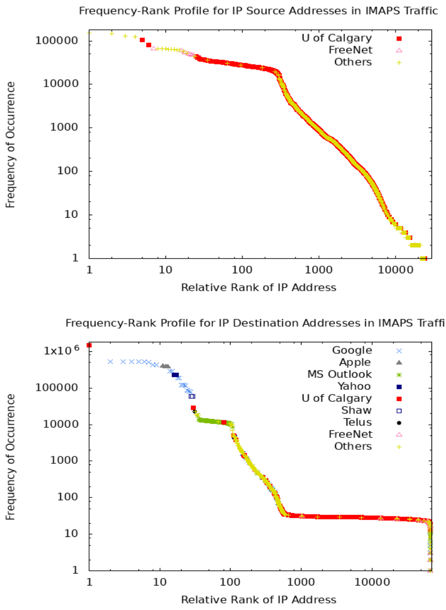


Fig. 3. IP Frequency-Rank Profile for IMAPS Traffic (April 14-20, 2019)

TABLE V
POPULAR DESTINATIONS FOR IMAPS TRAFFIC (APRIL 14-20, 2019)

Destination	Conns	% Conns	Bytes	% Bytes
Google	27,813,551	51.6%	902 GB	53.8%
Apple	6,015,885	11.1%	279 GB	16.7%
Microsoft	3,660,158	6.8%	227 GB	13.5%
Yahoo	1,087,930	2.0%	69.0 GB	4.1%
UCalgary	12,062,030	22.4%	5.54 GB	0.3%
Others	3,309,493	6.1%	194 GB	11.6%
Total	53,949,047	100.0%	1.64 TB	100.0%

structure (a straight line with slope of -2). These IP addresses correspond to a large number of users with widely varying activity levels for their IMAPS traffic. The pink triangles represent FreeNet traffic, which also uses NAT. Finally, the '+' symbols indicate other source IPs. The high-frequency ones in the upper left are mostly machine-generated scanning activities, while the Zipf-like part towards the lower right represents human-initiated email traffic.

The bottom graph in Figure 3 for IP destinations has a much more complex structure, with many piece-wise linear components. Several of these pieces correspond to particular cloud-based email providers, such as Google, Microsoft, and Yahoo. Table V provides some additional information to help interpret this graph. In particular, the table shows that over half of the IMAPS traffic goes to Google's Gmail service, with Apple's iCloud next, and Microsoft Outlook as the third biggest component. These providers are color-coded in Figure 3 for ease of reference. The very first data point is for a campus IMAPS server that acts as a honeypot and attracted over 1.5 million connection attempts during the week. The rest of the top ten IPs correspond to Gmail servers, the next few to Apple, and the middle plateau to about 80 load-balanced email servers used by Microsoft Outlook. The remaining parts of the plot represent a wide set of email servers, as well as the external scanning activity across the entire /16 IP address space for the University of Calgary network.

E. Asymmetry

Email traffic is often asymmetric. On our campus network, the Web-based email traffic (i.e., Outlook HTTPS) typically uses GET and POST, respectively, for retrieving and sending emails [25]. As a result, this traffic is almost balanced between its inbound and outbound data volumes. On the other hand, IMAPS traffic with Outlook is highly asymmetric, with inbound traffic volume exceeding outbound traffic by more than an order of magnitude (recall Figure 1(b)). SMTP traffic (to all destinations) is also highly asymmetric, but in the opposite direction, since it is used by some hosts to send emails to mail servers or to spam filtering services.

F. Transfer Sizes

In this subsection, we analyze the transfer sizes of TCP connections used for email traffic, seeking evidence of heavy-tailed distributions [17].

Figure 4 shows the CDF of IMAPS transfer sizes for the six different destinations identified earlier in Table V. The University of Calgary curve (red) is quite distinct from the rest, since these connections are mostly incoming scan traffic, with no actual data transfers. For the rest of the destinations, the graphs show evidence of heavy-tailed distributions, with the CDF climbing slowly as the transfer size increases. Note the logarithmic scale on the horizontal axis.

We next look at the overall transfer size distribution for all IMAPS traffic. Since the transfer sizes have high variability, we apply a log transform (base 2) to the data.

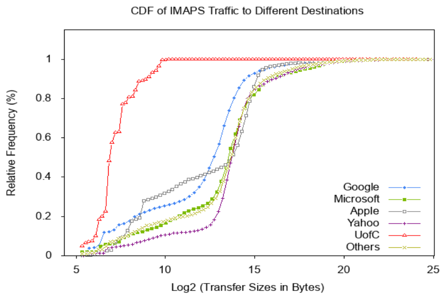


Fig. 4. Distribution of IMAPS Transfer Sizes for Popular Destinations

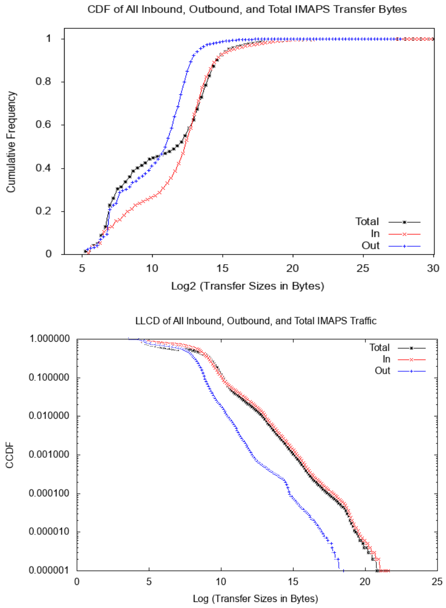


Fig. 5. Transfer Sizes for IMAPS Email Traffic (April 14-20, 2019)

The top graph in Figure 5 shows the CDF of transfer sizes, in which a heavy tail begins to appears for data sizes exceeding 16 KB. The bottom graph shows a log-log complementary distribution (LLCD) plot. A power-law is visible in this graph, indicating a slowly diminishing probability of seeing larger transfer sizes. The straight line in the graph spans across several orders of magnitude of sizes. The largest transfer size observed in our dataset was 10 GB.

G. *Throughput*

It is straightforward to calculate the average TCP throughput for each connection, based on the transfer size and the connection duration. Figure 6 shows the throughput results, with separate graphs for outbound throughput (i.e., bytes sent) and inbound throughput (i.e., bytes received).

The results in Figure 6 show that throughput is highly variable. This makes sense since throughput depends on transfer size, round-trip time, TCP version, network load, packet loss, and many other factors. Nonetheless, there is a pronounced tail to the throughput distribution, with some transfers achieving over 50 Mbps. There is no noticable difference between inbound and outbound throughputs, but there are some small differences between the throughputs for different email protocols. For example, HTTPS for Outlook tends to have lower throughputs near the middle of the graph, but slightly higher throughputs near the tail of the distribution.

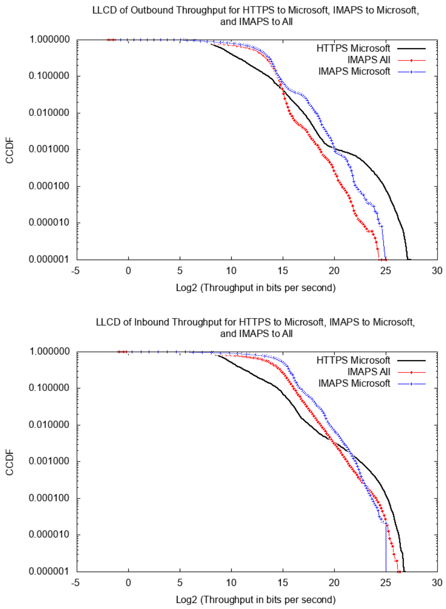


Fig. 6. Throughput Analysis for Email Traffic (April 14-20, 2019)

V. Conclusion

In this paper, we studied the IMAPS traffic on the University of Calgary's campus network, as an example of a large edge network. Our study focused on one week of empirical connection log data from April 14-20, 2019.

Our characterization results show strong diurnal and weekly patterns in the email traffic generated by human users, but with noticable spikes and background traffic that were machine-initiated. Email traffic has high variability in transfer sizes, connection duration, and throughput. Furthermore, email protocols such as IMAPS and SMTP are highly asymmetric in their data transfers, while HTTPS is closer to symmetric. We found evidence of heavy-tailed distributions in transfer sizes, both inbound and outbound.

The key takeaway message from our work is that email traffic is highly complex, and that non-trivial workload models are required to capture these traffic characteristics in network simulations. Our ongoing work is focusing on the implementation and evaluation of synthetic workload models for cloud-based email traffic in network simulations.

Acknowledgements

Financial support for this work was provided by Italy's Torno Subito fellowship program, and by Canada's Natural Sciences and Engineering Research Council (NSERC). The authors thank University of Calgary Information Technologies (UCIT) for facilitating the collection of our network traffic measurement data.

References

[1] V. Adhikari, Y. Guo, F. Hao, V. Hilt, Z-L. Zhang, M. Varvello, and M. Steiner, "Measurement Study of Netflix, Hulu, and a Tale of Three CDNs", *IEEE/ACM Transactions on Networking*, Vol. 23, No. 6, pp. 1984-1997, December 2015.

[2] M. Arlitt and C. Williamson, "Internet Web Servers: Workload Characterization and Performance Implications", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 5, pp. 631–645, 1997.

[3] M. Arlitt and C. Williamson, "An Analysis of TCP Reset Behaviour on the Internet", *ACM Computer Communication Review*, Vol. 35, No. 1, pp. 37-44, January 2005.

[4] N. Basher, A. Mahanti, A. Mahanti, C. Williamson, and M. Arlitt, "A Comparative Analysis of Web and Peer-fo-Peer Traffic", *Proceedings of WWW*, pp. 287-296, Beijing, China, April 2008.

[5] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing User Behavior in Online Social Networks", *Proceedings of ACM IMC*, pp. 49-62, Chicago, IL, November 2009.

[6] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing and Modeling Popularity of User-generated Videos", *Proceedings of IFIP Performance*, Amsterdam, Netherlands, October 2011.

[7] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web Caching and Zipf-like Distributions: Evidence and Implications", *Proceedings of IEEE INFOCOM*, New York, NY, pp. 126–134, March 1999.

[8] R. Caceres, P. Danzig, S. Jamin, and D. Mitzel, "Characteristics of Wide-Area TCP Conversations" *Proceedings of ACM SIGCOMM*, Zurich, Switzerland, pp. 101–112, August 1991.

[9] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User-Generated Content Video System", *Proceedings of ACM IMC*, pp. 1-14, San Diego, CA, November 2007.

[10] M. Crispin, "RFC 3501: Internet Message Access Protocol (IMAP)", IETF RFC, March 2003.

[11] M. Crovella and B. Krishnamurthy, *Internet Measurement: Infrastructure, Traffic and Applications*, John Wiley & Sons, 2006.

[12] Z. Durumeric, D. Adrian, A. Mirian, J. Kasten, E. Bursztein, N. Lidzborski, K. Thomas, V. Eranti, M. Bailey, and J. Halderman, "Neither Snow nor Rain nor MITM...: An Empirical Analysis of Email Delivery Sec urity", *Proceedings of ACM IMC*, pp. 27–39, Tokyo, Japan, October 2015.

[13] S. Englehardt, J. Han, and A. Narayanan, "I Never Signed Up for This! Privacy Implications of Email Tracking", *Proceedings of Privacy Enhancing Technologies Symposium*, pp. 109–126, Barcelona, Spain, July 2018.

[14] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube Traffic: A View from the Edge", *Proceedings of ACM IMC*, pp. 15-28, San Diego, CA, November 2007.

[15] T. Gupta, H. Fingler, L. Alvisi, and M. Walfish, "Pretzel: Email Encryption and Provider-supplied Functions are Compatible", *Proceedings of ACM SIGCOMM Conference*, pp. 169–182, Los Angeles, CA, August 2017.

[16] M. Laterman, M. Arlitt, and C. Williamson, "A Campus-Level View of Netflix and Twitch: Characterization and Performance Implications", *Proceedings of SCS SPECTS*, pp. 15-28, Seattle, WA, July 2017.

[17] A. Mahanti, N. Carlsson, A. Mahanti, M. Arlitt, and C. Williamson, "A Tale of the Tails: Power-Laws in Internet Measurements", *IEEE Network*, Vol. 27, No. 1, pp. 59-64, January 2013.

[18] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks", *Proceedings of ACM IMC*, pp. 29-42, San Diego, CA, October 2007.

[19] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-time", *Computer Networks*, Vol. 31, No. 23, pp. 2435-2463, December 1999.

[20] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding Online Social Network Usage from a Network Perspective", *Proceedings of ACM IMC*, pp. 35-48, Chicago, IL, November 2009.

[21] J. Postel, "RFC 821: Simple Mail Transfer Protocol (SMTP)", IETF RFC, August 1982.

[22] A. Ramachandran and N. Feamster, "Understanding the Network-level Behavior of Spammers", *Proceedings of ACM SIGCOMM Conference*, pp. 291–302, Pisa, Italy, September 2006.

[23] D. Schatzmann, W. Mühlbauer, T. Spyropoulos, and X. Dimitropoulos, "Digging into HTTPS: Flow-based Classification of Webmail Traffic", *Proceedings of ACM IMC*, pp. 322–327, Melbourne, Australia, November 2010.

[24] C. Williamson, "Internet Traffic Measurement", *IEEE Internet Computing*, Vol. 5, No. 6, pp. 70–74, November/December 2001.

[25] Z. Zhang and C. Williamson, "A Campus-level View of Outlook Email Traffic", *Proceedings of the International Conference on Network, Communication, and Computing*, Taipei, Taiwan, pp. 299–306, December 2018.