On the Evolution of Internet Flow Characteristics

Simon Bauer
Technical University of Munich
bauer@net.in.tum.de

Benedikt Jaeger
Technical University of Munich
jaeger@net.in.tum.de

Fabian Helfert
Technical University of Munich
helfert@net.in.tum.de

Philippe Barias
Technical University of Munich
barias@net.in.tum.de

Georg Carle
Technical University of Munich
carle@net.in.tum.de

ABSTRACT

The ongoing evolution of technologies and network services on the Internet indicates ongoing changes in traffic and flow characteristics. Since the analysis of flow characteristics, like duration, size, and rate, has been a frequently studied topic before, results on the evolution of flow characteristics are rare. This paper surveys how flow characteristics have changed over time and whether there are significant trends in such characteristics.

We present a long-term study of TCP flow characteristics based on traffic captures taken between 2008 and 2019. We apply different methods to analyze the distribution of characteristics, the relevance of heavy hitters, and correlations between characteristics. Our analysis shows significant trends in the 99th percentiles of flow characteristics, persistent dominance by heavy hitters regarding transmitted data, and increasing relevance of so-called big-fast flows.

CCS CONCEPTS

• **Networks** → **Transport protocols**; **Network measurement**; *Network monitoring*;

KEYWORDS

Traffic characterization, Flow analysis, Heavy hitters

1 INTRODUCTION

The emergence and evolution of technologies and services on the Internet, such as better network expansion, the Internet of Things, or audio and video streaming, suggest that characteristics of Internet flows, like duration, size, and rate, are also changing. While patterns, distributions, and correlations of TCP flow characteristics have been studied in previous research [1–3], there is little insight into the evolution of characteristics over time.

This paper pursues how flow characteristics have changed during the last years. We contribute a large-scale analysis of flow characteristics of two datasets composed of Internet traffic captures taken between 2008 and 2016, respectively, in 2018 and 2019, published by CAIDA [4]. In total, we analyze over 2.5 billion TCP flows that transmitted over 65 TB of payload data. We apply different taxonomies to assess the relevance of heavy hitters and so-called big-fast flows, as introduced in previous studies [2, 3]. Further, we survey the distribution and correlation of considered characteristics. To encourage further studies on flow characteristics, we present a highly scalable traffic analyzer implemented in Go [5].

Our study shows persistent relevance of bytes transmitted by heavy hitters, while we find significant changes regarding duration and rate of such flows. Applying a two-two taxonomy on flow size and rate results in increasing shares of bytes transmitted by big-fast flows over time. Further, we confirm previously observed correlations between TCP flow characteristics.

The remainder of this paper is structured as follows: First, we describe the extraction and definition of flow characteristics, as well as applied methods for analysis in Section 2. In Section 3, we present the datasets used for our analysis. We then study measured flow characteristics in Section 4. Section 5 reviews related work before Section 6 concludes with a summary of our findings.

2 FLOW ANALYSIS

This section introduces the architecture and implementation of a multi-threaded flow analysis tool, describes the calculation of flow characteristics, and presents methods and taxonomies used for our analysis.

2.1 Scalable Flow Analyzer

To efficiently analyze large packet captures of several 100 GB in size, we implement a multi-threaded flow analysis tool in Go. Cost-intensive packet parsing and composition of flow

ANRW '21, July 24–30, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8618-0/21/07…$15.00
https://doi.org/10.1145/3472305.3472321

characteristics are completely parallelized. The analyzer consists of five major components, as shown in Figure 1, which are described in more detail in the following.

Reader We use the Go library gopacket [6] to read raw packets from PCAPs. As analyzed protocols are known beforehand, DecodingLayerParsers (DLP) [7] provide efficient traffic parsing by only decoding specified protocol headers. So far, the analyzer supports parsing Ethernet, IPv4, IPv6, TCP, and UDP headers. Additional headers can be included by extending the reader's DLP with further layers and protocols from the gopacket library.

Parsers After reading packets and decoding headers, packet features are extracted by a parser component. Multiple parsers can run entirely independently from each other to provide scalability. To save storage capacity and preserve privacy, only required packet information is stored, while payload is ignored. IP addresses are stored as 64-bit integer hashes using xxHash [8]. Using hashed IP addresses provides anonymity regarding input data and benefits from efficient integer comparisons in Go.

Ringbuffer As multiple parsers may analyze packets of the same flow parallel without synchronization, it is required to re-order packets. To bring the packets back into the order they were received, the parsers store the extracted packet information to a ringbuffer. To avoid locking, the reader uses a counter to provide each packet with a packet number, which is used as index for the ringbuffer. Sorting is then done automatically by such packet numbers that already specify the correct packet order. A routine regularly checks the ringbuffer and writes sorted packet information to a pool component.

Pools To collect and store extracted packet information of all packets of a flows until the flow is terminated, we use so-called pools. Pool threads work without any synchronization to provide high scalability to the analysis capabilities of our tool. We calculate a flow key based on hashes of a flow's server and client IP addresses, port numbers, and the used transport layer protocol to assign packets to the correct pool. We sum up hashes of IP addresses to generate a bidirectional flow key for each flow. Bidirectional flow keys are required as packets of both directions of a flow have to be assigned to the same pool. As there may are several flows resulting in the same flow key, the analyzer takes care of detecting such cases by observing header flags and applying timeouts to inactive flows. If a flow terminates, collected flow information is written to the metric component and flushed from the pool afterwards.

Metric Output Finally, flow characteristics are calculated by an own component that receives collected information of terminated flows and corresponding packets in the correct order. Before writing results to files, the analyzer aggregates flow characteristics from collected packet data.

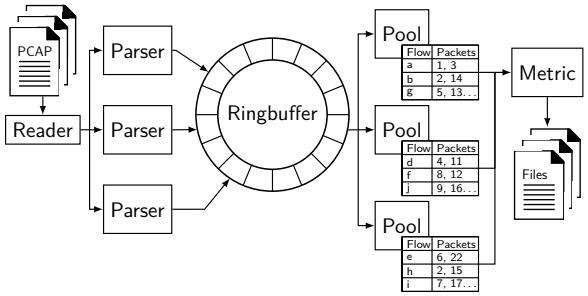

Figure 1: Architecture for scalable flow analysis.

2.2 Flow Characteristics

We refer to a flow as a sequence of packets bidirectionally sent and received by the same IP 5-tuple. To identify flows, we rely on different methods for TCP and UDP traffic. Regarding TCP, we require a flow to start with the TCP 3-way-handshake. As UDP is a stateless protocol, we define a UDP flow to start with the first packet we observe for an unseen IP 5-tuple. Observing the TCP handshake allows detecting the initiator of a connection that we refer to as the client while servers respond to the initially sent packet. For UDP, we assume that the server is often using a system port. If this heuristic is not feasible, we assume that the client sends the first packet. TCP flows end, if we observe the TCP connection tear-down, observe a TCP handshake of an IP 5-tuple already tracked, or in case we cannot observe a packet with the corresponding IP 5-tuple for a specific time interval. For UDP, we only rely on time intervals without observed packets to terminate flows. By default, the analyzer uses a timeout interval of 5 minutes.

We define the duration of a flow as the time interval between the first seen packet and the moment of termination of the corresponding flow by the analyzer. We refer to flow size as the sum of the Layer-4 payload sizes of all packets of a flow. To assess flow rate, we calculate the average rate of a flow based on Layer-4 payload sizes and flow duration.

2.3 Comparison of Flow Characteristics

We apply different methods and taxonomies to describe properties of analyzed flows and differences between single traces to analyze the evolution of flow characteristics.

To describe changes regarding the distribution of measured characteristics, we calculate the cumulative distribution function (CDF) for each characteristic and trace. The cumulative distribution of characteristics is purposed to compare the share of flows for a particular interval in the spectrum of measured characteristics.

To analyze the relevance of so-called heavy hitter flows, we apply the threshold-based taxonomy introduced by Lan et al. [3]. Lan et al. define Elephant, Tortoise, and Cheetah flows

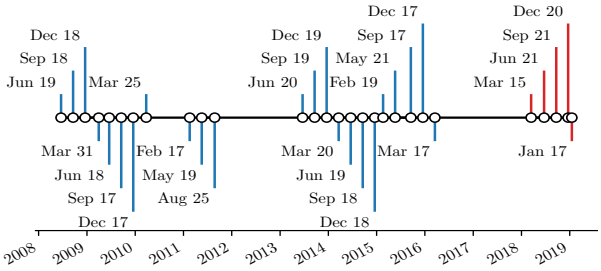

Figure 2: Used datasets: traces taken in Chicago [9] in blue, traces taken in New York [10] in red.

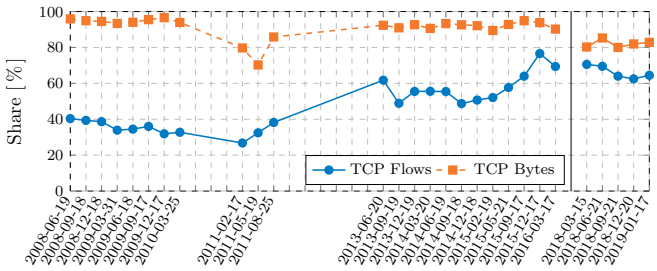

Figure 3: Share of TCP flows and transmitted bytes.

as heavy hitters regarding size, duration, and rate. Thresholds are determined by taking the average of measured values and adding the standard deviation three times, respectively, by calculating the 99th percentile of measured values. This taxonomy allows comparing the share of bytes transmitted by each kind of heavy hitters for different traces.

Further, we apply a two-two taxonomy introduced by Zhang et al. [2] to classify flows regarding their size and rate. Based on threshold values for flow size and rate, the taxonomy classifies flows according to four groups: small-slow, small-fast, big-slow, and big-fast flows. We are particularly interested in the share of big-fast flows and the corresponding share of bytes, as Zhang et al. found that most bytes of Internet traffic are transmitted by a tiny share of big-fast flows.

Next to assessing the relevance of heavy hitters, we are interested in analyzing the correlation between flow characteristics as done by earlier studies [2, 3]. Therefore, we calculate correlation coefficients according to Pearson. Such coefficients describe the linear relation between two characteristics. A correlation coefficient near 1 indicates a strong positive correlation, while -1 implies a strong negative correlation. A coefficient near zero implies no correlation at all. We logarithmically transform input data before calculating correlation due to the extensive spectrum and uneven distribution of observed values.

3 DATASET

This section introduces the traces selected for our analysis, describes the relevance of TCP traffic in the datasets, and assesses the impact of applied pre-filtering of flows for further analysis.

Traces. For our long-term analysis of Internet traffic characteristics, we select traces from two different capturing points provided by the Center of Applied Internet Data Analysis (CAIDA) [4]. All traces provide one hour of network traffic captured on Internet backbone links with a bandwidth of 10 Gbit/s. To respect entities' privacy, CAIDA anonymizes IP addresses of captured traffic in a prefix preserving manner. To survey changes over time, we select 23 traces captured between June 2008 and March 2016 on a Tier-1 ISP backbone link in Chicago [9]. In consideration of distributing traces equally over time, we choose 23 traces mainly at three-month intervals. The selected traces captured in Chicago consist of 1.16 billion TCP flows (2.63 billion flows in total). The long-term dataset captured in Chicago includes two significant periods without traffic, as no captures are available in such periods: First, 11 months without traces between March 2010 and February 2011 and second, 18 months without traces between September 2011 and March 2013. To compare our findings of the introduced long-term dataset to more recent traffic, we select five traces captured on a Tier-1 ISP backbone link in New York between March 2018 and January 2019 [10]. We find significantly larger amounts of traffic in the New York traces. I.e., the five selected traces carry 1.48 billion TCP flows (2.25 billion flows in total). All selected traces are listed in Figure 2.

TCP Traffic. As our analysis focuses on characteristics of TCP flows, we survey how significant TCP is in the sense of observed flows and transmitted data. Throughout our datasets, we find significant dominance of bytes transmitted by TCP. TCP carries over 90 % of bytes in all traces captured in Chicago, except the three traces taken in 2011. The share of TCP flows increases from around 40 % up to shares larger than 70 % in December 2015 and March 2016. Correspondingly, the share of UDP flows shows a significant decrease, while UDP transmits only a small share of bytes. Regarding traces captured in New York, we find similar shares of TCP flows compared to traces captured in Chicago. However, we find smaller shares, i.e., around 80 %, of bytes transmitted by TCP. Figure 3 shows the share of TCP flows and corresponding bytes over time.

Pre-filtering. Previous studies of flow rates [2, 11] filter flows for a minimum duration of 100 ms. Such filtering aims to remove single packet flows, whose duration is zero and, therefore, make rate calculation unfeasible. Further, all packets of very short flows might be sent back-to-back, which

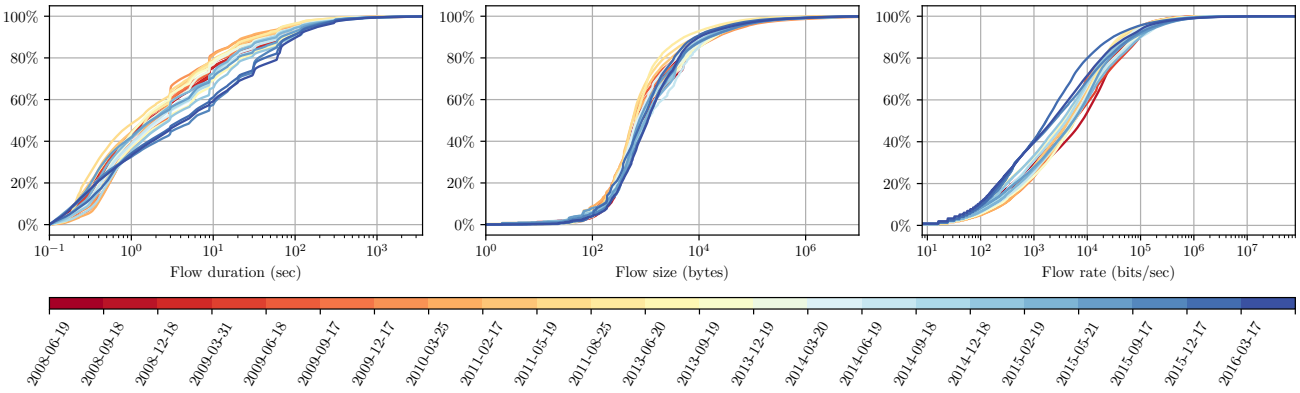

Figure 4: Distributions of flow characteristics for the Chicago dataset.

falsifies the measured rate, too. Therefore, we also filter flows that are shorter than 100 ms. We observe that such filtering excludes nearly 20 % of TCP flows, while the remaining flows still carry 99.94 % of total bytes transmitted by TCP in the long-term Chicago dataset. For traces collected in New York, nearly 35 % of TCP flows get filtered out, while the remaining flows still carry 99.96 % of bytes. We conclude that such filtering of short flows still allows representative analysis of flow characteristics since almost all transferred bytes are considered for further analysis.

4 RESULTS

This section presents analysis results for the datasets presented in Section 3 following the methods and taxonomies described in Section 2.3. We first present results for the long-term dataset captured in Chicago in Section 4.1 and compare our findings to results for the more recent dataset captured in New York in Section 4.2.

4.1 Long-term Analysis

Distributions. To survey the distribution of measured flow characteristics, we analyze the corresponding cumulative distribution function, as shown in Figure 4. Regarding flow rates and sizes, we observe distributions with a very long tail and an interval of a strong positive gradient that covers the values for the majority of analyzed flows. However, the CDFs of rates and sizes do not reveal a significant trend across the traces of the Chicago dataset. We find that around 85 % of TCP flows carry between 100 B and 10 kB, with maximum flow sizes near 10 GB. Regarding flow rates, over 70 % of flows transmit payload with an average rate between 1 kbit/s and 100 kbit/s for most traces, while we observe flow rates up to several Gbit/s. The distributions of TCP flow durations reveal a slight trend towards smaller shares of short flows, especially between the 40th and 90th percentile.

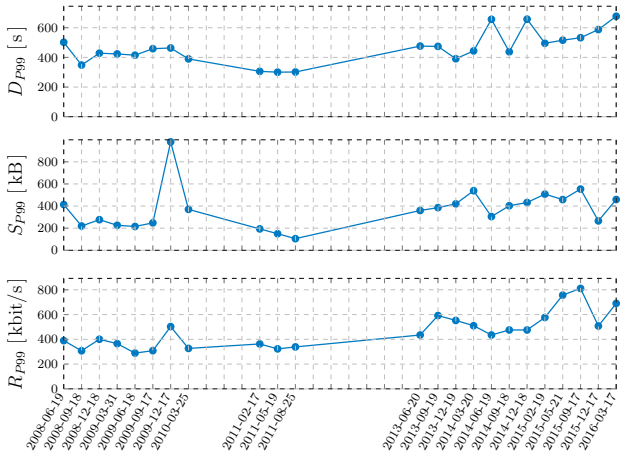

Figure 5: 99th percentiles of flow characteristics.

Heavy hitters. Lan et al. [3] define heavy hitters based on the average and standard deviation of measured flow characteristics and based on the 99th percentile of a characteristic. Just as Lan et al., we find similar results for both definitions. Therefore, we rely on the 99th percentile for further analysis.

Before assessing the relevance of heavy hitter flows, we are interested in trends regarding the longest, largest, and fastest flows. Therefore, we calculate the 99th percentiles of duration, size, and rate, hereafter notated as D_{P99}, S_{P99}, R_{P99}, for each trace and compare such values over time. Figure 5 shows measured 99th percentiles over time. D_{P99} shows little change between 2008 and 2010. Between June 2013 and March 2016, D_{P99} increases for a factor near 1.5, i.e., from around 400 s for early traces to around 600 s for more recent traces. Flow size shows larger variances between measured percentiles of different traces. We observe that S_{P99} for traces between 2008 and 2009 tendentiously cluster around 250 kB, while S_{P99} measured for traces between 2013 and

2016 cluster around 400 kB. However, we do not observe a particular trend between data points. Regarding flow rates, we find increasing R_{P99} over time. For traces before 2013, R_{P99} lies between 300 kbit/s and 400 kbit/s, except an outlier in Dec 2009. After 2013, R_{P99} varies between 400 kbit/s and 800 kbit/s. For traces taken in 2015 and 2016, R_{P99} regularly exceeds 600 kbit/s.

To assess how relevant flows within the 99th percentiles are, we analyze the corresponding share of bytes. Further, we are interested in the share of flows and bytes by intersections of such flow sets to survey relations between different percentiles. As we do not find trend-like evolutions within the share of bytes for the 99th percentiles over time, we calculate the average of measured byte shares, as shown in Table 1. We find that flows within the S_{P99}, on average, carry 89 % of all TCP bytes with a fairly small standard deviation of 2 %. On average, nearly 20 % of bytes are transmitted by the intersection of all three 99th percentiles $D_{P99} \cap S_{P99} \cap R_{P99}$. Note that these 20 % of TCP bytes are transmitted by only 0.009 % of TCP flows. For the intersection of S_{P99} and R_{P99}, we measure the same values as for the intersection of all three percentiles. This indicates that the biggest and fastest flows are also part of the 1 % of longest flows. The difference between bytes transmitted by R_{P99} and $S_{P99} \cap R_{P99}$ is relatively small. This observation implies that flows in R_{P99} are mostly also part of S_{P99}. At the same time, the share of flows in $S_{P99} \cap R_{P99}$ is only a third of the share by R_{P99} while both sets nearly transmit the same amount of data. Same applies for D_{P99} and $D_{P99} \cap S_{P99}$, indicating that the majority of very long flows also carries much data. Considering such observations, our analysis clearly points out the relevance of such small subsets of analyzed flows.

Relevance of Big-Fast Flows. Next, we apply a two-two taxonomy according to Zhang et al. [2] to assess the relevance of so-called big-fast flows and corresponding bytes. The taxonomy is based on two thresholds regarding flow size and flow rate and classifies flows into four groups, considering small or big flow sizes and slow or fast flow rates. We select

Table 1: Relevance of intersections of 99th percentiles.

Flow set	Chicago		New York	
	Share	Bytes	Share	Bytes
D_{P99}	1.000 %	40.5 %	1.000 %	43.1 %
S_{P99}	1.000 %	89.2 %	1.000 %	88.4 %
R_{P99}	1.000 %	55.9 %	1.000 %	68.0 %
$D_{P99} \cap S_{P99}$	0.185 %	39.9 %	0.142 %	42.6 %
$D_{P99} \cap R_{P99}$	0.009 %	19.9 %	0.005 %	31.4 %
$S_{P99} \cap R_{P99}$	0.337 %	54.8 %	0.332 %	67.2 %
$D_{P99} \cap S_{P99} \cap R_{99}$	0.009 %	19.9 %	0.005 %	31.4 %

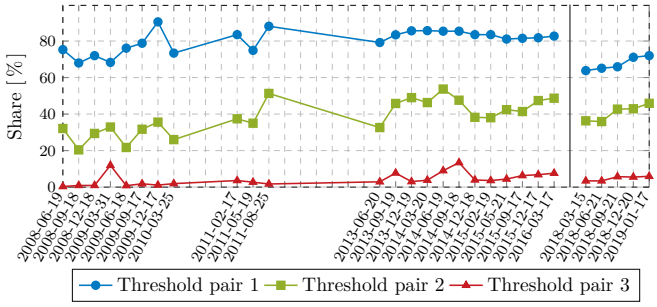

Figure 6: Share of bytes transmitted by big-fast flows.

three pairs of thresholds. First, we rely on the thresholds chosen by Zhang et al. [2], who use 100 kB as a threshold between small and big flows and 10 kB/s as a threshold between slow and fast flows. We find that the share of big-fast flows is below 2 % for all traces. The most common flow type is small-slow which represent 90 % to 95 % of flows. As shown in Figure 6, big-fast flows carry between 70 % up to over 80 % of bytes transmitted by TCP. Analysis over time shows a slight increase of such byte shares between 2008 and 2013, as shown in Figure 6. The increase of bytes by big-fast flows correlates to a decrease of bytes by big-slow flows. For early traces, around 20 % of TCP bytes are transmitted by big-slow flows, decreasing to around 10 % for more recent traces. This indicates that the rates of big flows increase with time, which we also observe for the 99th percentile of TCP flow rates. As the second pair of thresholds, we select significantly larger values to survey the relevance of very big and very fast flows. We set the threshold regarding size to 1 MB and the threshold for flow rates to 100 kB/s. The shares of bytes transferred by very big and very fast TCP flows are significantly smaller than bytes by big-fast flows. However, such shares show a significant increase across all traces of the dataset from around 30 % up to over 50 %. Last, we apply even larger thresholds, i.e., 10 MB for flow size and 1 MB/s for flow rates. We only find very small shares of bytes transmitted by these extreme big and fast flows while we also observe a slight increase of bytes transmitted by such flows. Note that the shares of big-fast flows for the second and third threshold set are by far smaller than 0.5 %.

Correlations. Zhang et al. [2] restrict correlation analysis to longer flows, i.e., flows longer than 1 second, 5 seconds, and 30 seconds, and observe slight differences between different flow lengths. We apply the same filters to our dataset to compare results, while we describe results for flows longer than 5 seconds in the following. We observe that correlation coefficients do not indicate specific trends over time. Therefore, we calculate the average of correlation coefficients and calculate standard deviation as a measure for dispersion. On

Table 2: Correlations of flow characteristics.

Traces	Corr.	5 seconds Avg.	5 seconds Std.	30 seconds Avg.	30 seconds Std.
Chicago	D & R	-0.0943	0.116	-0.1058	0.075
	D & S	0.2947	0.117	0.1798	0.092
	S & R	0.8847	0.0169	0.8783	0.014

average, we find a weak negative correlation between duration and rate (avg.: -0.09) and a weak positive correlation between duration and size (avg.: 0.29). We measure a relatively large standard deviation for both combinations. We find a very strong correlation between size and rate, with an average coefficient of 0.88 and small standard deviation. Such a strong correlation of size and rate was observed by earlier studies of correlations between flow characteristics [2, 3]. We find slightly less significant correlation between duration and size, respectively, size and rate, for flows longer 30 seconds, as shown in Table 2.

4.2 Comparison to Recent Traces

As our dataset for long-term analysis of flow characteristics ends in 2016, we compare our findings to a more recent dataset taken in 2018 and 2019 as described in Section 3. In the following, we focus on the differences and similarities between the findings for both datasets. CDFs of measured flow characteristics show similar distribution patterns between both datasets. CDFs for New York traces indicate larger shares of slower TCP flows and larger shares of small TCP flows. This observation is confirmed by the 99th percentiles of size and rate, which are nearly a magnitude smaller than 99th percentiles measured for the Chicago dataset. We observe more bytes transmitted by R_{P99}. For example, flows in R_{P99} for the more recent dataset on average transmit 68.0% of bytes (55.9% for Chicago traces). Flows in the intersection of the 99th percentiles of all characteristics averagely transmit 31.4% of all TCP bytes, while the intersection only includes 0.005% of all TCP flows. Further relations between intersections observed for the Chicago dataset, as described in Section 4.1, also apply to results for the more recent New York dataset, as shown in Table 1. Regarding the share of data transmitted by big-fast flows, we find slightly smaller shares for traces taken in New York than for the more recent traces within the Chicago dataset, as shown in Figure 6. This can be traced back to the observed larger shares of small flows and slow flows in the CDFs of New York traces. Average correlation coefficients only show minor differences between both datasets. The correlation between size and rate of TCP flows decreases from 0.88 to 0.83 for flows longer than 5 seconds, and from 0.87 to 0.80 for flows longer than 30 seconds. Further correlation coefficients of the New York dataset are almost within the range of one standard deviation measured accordingly for Chicago traces.

5 RELATED WORK

The field of traffic characterization and classification is a frequently addressed topic. However, detailed analysis of the evolution of flow characteristics is lacking. Our work is closely related to Zhang et al. [2], who analyzed Internet flow rates and their root causes in 2002. We confirm significantly dominating transmission of bytes by big-fast flows and approve findings regarding the correlation of TCP characteristics. Lan et al. [3] use the flow characteristics considered in this paper and consider burstiness of flows as a further characteristic. Authors classify flows as heavy hitters for duration, size, and rate based on threshold values, respectively based on the 99th percentiles of flow characteristics, which we also survey in this paper. As Zhang et al. [2] and our analysis of correlations, Lan et al. find strong correlation between TCP size and rate. Beside their characterization, Heavy Hitters are object of study regarding their identification and detection [12–14]. Burstiness of CAIDA traces is studied by Lazarou et al. [15]. A long-term study of Internet flow rate limitations was conducted by Araujo et al. in 2014 [16]. Flow rates are studied in details for cellular networks by Zhang et al. [11], considering duration, size, and rate of flows and the limiting factor behind such rates. Further studies focus on the characterization of networks [17], traffic patterns regarding daytimes [1, 18], and the deployment of network protocols and their characteristics [19]. An extensive study of residential broadband Internet traffic characteristics is presented by Maier et al. [20].

6 CONCLUSION

This paper presented a long-term study of Internet flow characteristics and introduced a tool for highly scalable flow analysis, which is published as free and open source [5].

We find that the 99th percentiles of TCP flow duration and TCP flow rate show significant change between 2008 and 2016, while different intersections between 99th percentiles show significant relevance of such heavy hitters. The share of bytes transmitted by big-fast flows and very big-very fast flows increases over time, while the share of such flows remains very small. Measured correlations stay constant over time and confirm findings from former studies.

ACKNOWLEDGEMENTS

This work was supported in part by the German Research Foundation (project ModANet under grant no. CA595/11-1) and by the German-French Academy for the Industry of the Future.

REFERENCES

[1] K. Thompson, G. J. Miller, and R. Wilder, "Wide-area internet traffic patterns and characteristics," *IEEE network*, vol. 11, no. 6, pp. 10–23, 1997.

[2] Y. Zhang, L. Breslau, V. Paxson, and S. Shenker, "On the characteristics and origins of internet flow rates," in *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '02. New York, NY, USA: ACM, 2002.

[3] K.-c. Lan and J. Heidemann, "A measurement study of correlations of internet flow characteristics," *Computer Networks*, vol. 50, no. 1, pp. 46–62, 2006.

[4] CAIDA, "The caida ucsd anonymized internet traces - 2008-2016," 2019. [Online]. Available: http://www.caida.org/data/passive/passive_dataset.xml

[5] F. Helfert, P. Barias, S. Bauer, and B. Jaeger, "Scalable flow analysis in go," https://github.com/uncatchable-de/scalable-flow-analyzer, 2021.

[6] google/gopacket: Provides packet processing capabilities for go. [Online]. Available: https://github.com/google/gopacket

[7] gopacket - godoc. [Online]. Available: https://godoc.org/github.com/google/gopacket#hdr-Fast_Decoding_With_DecodingLayerParser

[8] xxhash - extremely fast non-cryptographic hash algorithm. [Online]. Available: http://cyan4973.github.io/xxHash/

[9] CAIDA. Passive monitor: equinix-chicago. [Online]. Available: http://www.caida.org/data/monitors/passive-equinix-chicago.xml

[10] ——. Passive monitor: equinix-nyc. [Online]. Available: http://www.caida.org/data/monitors/passive-equinix-nyc.xml

[11] Y. Zhang, Å. Arvidsson, M. Siekkinen, and G. Urvoy-Keller, "Understanding http flow rates in cellular networks," in *2014 IFIP Networking Conference*. IEEE, 2014, pp. 1–8.

[12] Z. Zhang, B. Wang, and J. Lan, "Identifying elephant flows in internet backbone traffic with bloom filters and lru," *Computer Communications*, vol. 61, pp. 70–78, 2015.

[13] R. Harrison, S. L. Feibish, A. Gupta, R. Teixeira, S. Muthukrishnan, and J. Rexford, "Carpe elephants: Seize the global heavy hitters," in *Proceedings of the Workshop on Secure Programmable Network Infrastructure*, 2020, pp. 15–21.

[14] A. Pekar, A. Duque-Torres, W. K. Seah, and O. Caicedo, "Knowledge discovery: Can it shed new light on threshold definition for heavy-hitter detection?" *Journal of Network and Systems Management*, vol. 29, no. 3, pp. 1–30, 2021.

[15] G. Y. Lazarou, M. S. Alam, and J. Picone, "Measuring the variability of caida internet traffic traces," in *2016 19th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2016, pp. 1–6.

[16] J. T. Araújo, R. Landa, R. G. Clegg, G. Pavlou, and K. Fukuda, "A longitudinal analysis of internet rate limitations," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 1438–1446.

[17] P. Velan, J. Medková, T. Jirsík, and P. Čeleda, "Network traffic characterisation using flow-based statistics," in *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2016, pp. 907–912.

[18] A. Salihu, M. Shefkiu, and A. Maraj, "Characteristics and temporal behavior of internet backbone traffic," *International Journal of Business and Technology*, vol. 6, no. 3, pp. 1–8, 2018.

[19] D. Murray, T. Koziniec, S. Zander, M. Dixon, and P. Koutsakis, "An analysis of changing enterprise network traffic characteristics," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*. IEEE, 2017, pp. 1–6.

[20] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, 2009, pp. 90–102.