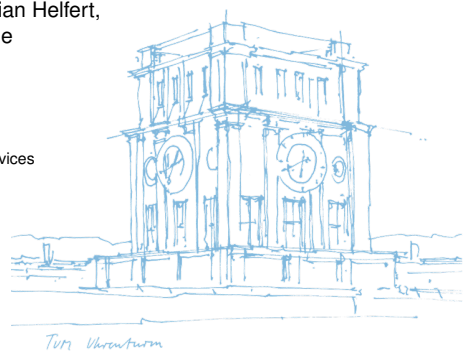
Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

On the Evolution of Internet Flow Characteristics

Simon Bauer, Benedikt Jaeger, Fabian Helfert,
Philippe Barias, Georg Carle

Tuesday 27th July, 2021

Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich



Motivation

- Ongoing evolution and emergence of technologies and services on the Internet
  - Network expansion, IoT, audio and video streaming ...
- Previous studies present methodologies to survey flow characteristics
  - E.g. Thompson et al. (1997) [1], Zhang et al. (2002) [2], Lan et al. (2006) [3]

How did characteristics of Internet flows change during the last few years?

This paper:

- Surveys the distribution and correlation of flow characteristics
- Applies different taxonomies to assess the relevance of heavy hitters

Motivation

- Ongoing evolution and emergence of technologies and services on the Internet
  - Network expansion, IoT, audio and video streaming ...
- Previous studies present methodologies to survey flow characteristics
  - E.g. Thompson et al. (1997) [1], Zhang et al. (2002) [2], Lan et al. (2006) [3]

How did characteristics of Internet flows change during the last few years?

This paper:

- Surveys the distribution and correlation of flow characteristics
- Applies different taxonomies to assess the relevance of heavy hitters

Motivation

- Ongoing evolution and emergence of technologies and services on the Internet
  - Network expansion, IoT, audio and video streaming ...
- Previous studies present methodologies to survey flow characteristics
  - E.g. Thompson et al. (1997) [1], Zhang et al. (2002) [2], Lan et al. (2006) [3]

How did characteristics of Internet flows change during the last few years?

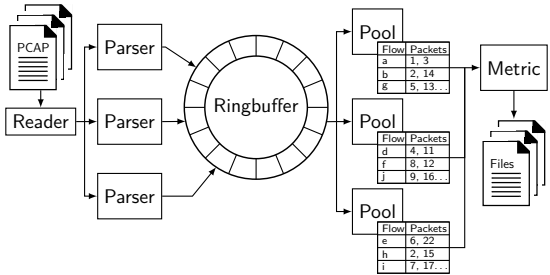
This paper:

- Surveys the distribution and correlation of flow characteristics
- Applies different taxonomies to assess the relevance of heavy hitters

Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

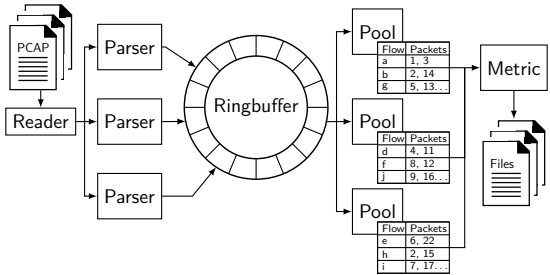
1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

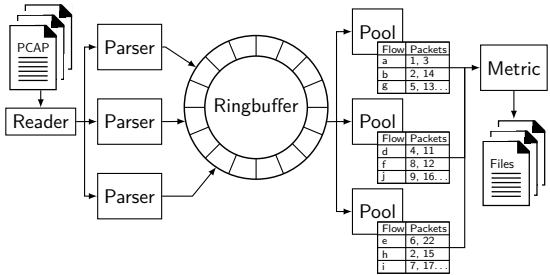
1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

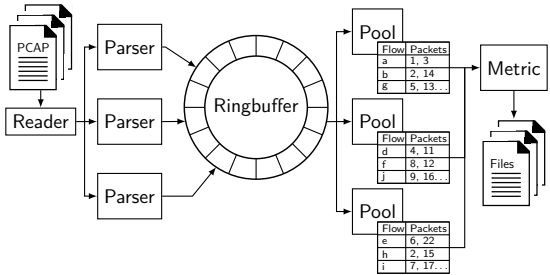
1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

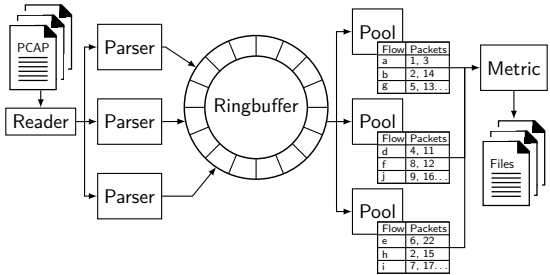
1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

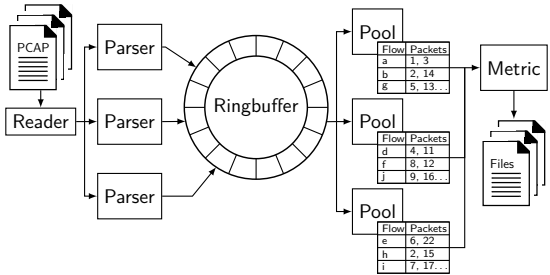
1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Analysis

- Scalable flow analysis due to parallelized packet parsing and flow aggregation implemented in Go
- Code as free and open-source [4]

1. Reader
   - Read packets from pcap
2. Parsers
   - Extract packet features
3. Ringbuffer
   - Re-order packets
4. Pools
   - Collect packet features per flow
5. Metric
   - Calculate flow characteristics



Flow Characteristics

Identifying Flows

- Based on IP 5-Tuple
- TCP:
  - Start with the TCP 3-Way-Handshake
  - Termination by observed connection tear-down or after a timeout period after last seen packet

Characteristics

- **Size:** Sum of the Layer 4 payload sizes of all packets
- **Duration:** Time interval between the first packet and the last packet
- **Rate:** Average data rate calculated by size and duration

Flow Characteristics

Identifying Flows

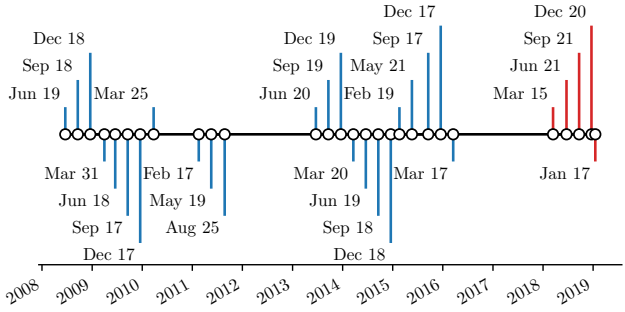
- Based on IP 5-Tuple
- TCP:
  - Start with the TCP 3-Way-Handshake
  - Termination by observed connection tear-down or after a timeout period after last seen packet

Characteristics

- **Size:** Sum of the Layer 4 payload sizes of all packets
- **Duration:** Time interval between the first packet and the last packet
- **Rate:** Average data rate calculated by size and duration

Dataset

- Traffic captures provided by CAIDA
- Anonymized IP addresses, no Layer 4 payloads
- One hour captures of 10 Gbit/s Tier-1 ISP backbone links
  - 23 traces taken in Chicago between 2008 and 2016 [5]
  - 5 traces taken in New York between 2018 and 2020 [6]
- Considered for analysis: all TCP flows longer than or equal to 100 ms

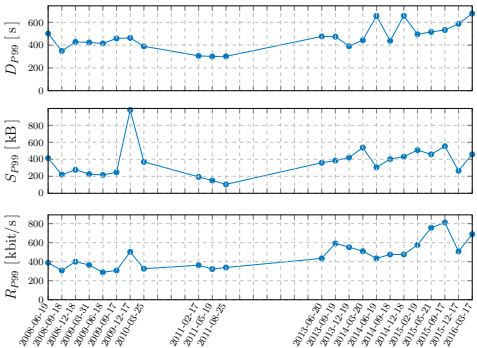


Evolution of 99th Percentiles

- Lan et al. [3] defines heavy hitters based on 99th percentiles

→ D_{P99}: increase by factor 1.5 between June 2013 and March 2016

→ R_{P99}: increase from rates around 400 kbit/s up to 800 kbit/s



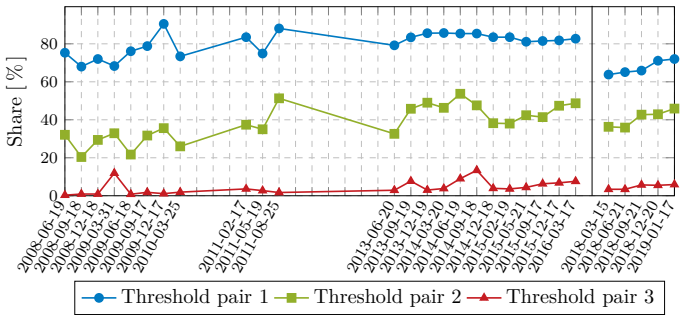
Relevance of 99th Percentiles

- How relevant are flows within the 99th percentiles of characteristics?

Flow set	Chicago Share	Chicago Bytes
D_{P99}	1,000 %	40,5 %
S_{P99}	1,000 %	89,2 %
R_{P99}	1,000 %	55,9 %
$D_{P99} \cap S_{P99}$	0,185 %	39,9 %
$D_{P99} \cap R_{P99}$	0,009 %	19,9 %
$S_{P99} \cap R_{P99}$	0,337 %	54,8 %
$D_{P99} \cap S_{P99} \cap R_{99}$	0,009 %	19,9 %

Evolution of Big-Fast Flows

- Two-two taxonomy based on threshold values for flow size and flow rate according to Zhang et al. [2]
- Thresholds:
  - Pair 1: 100 kB and 10 kB/s
  - Pair 2: 1 MB and 100 kB/s
  - Pair 3: 10 MB and 1 MB/s



Conclusion and Future Work

Summary:

- Significant increase of the 99th percentiles of flow duration and rate
- Large significance of heavy hitters regarding the share of transmitted bytes
- Increasing relevance of (very) big - (very) fast flows during the past years

Not included in this talk:

- Analysis of distribution of flow characteristics over time
- Correlation analysis

Bibliography

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

[5] CAIDA, “Passive monitor: equinix-chicago.” http://www.caida.org/data/monitors/passive-equinix-chicago.xml.

[6] CAIDA, “Passive monitor: equinix-nyc.” http://www.caida.org/data/monitors/passive-equinix-nyc.xml.