Per Priority Data Rate Measurement in Data Plane

Habib Mostafaei **Georgios Smaragdakis**

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

TU Delft

Motivation

Measuring the data rate per priority group (PG) of flows can enhance the performance of a network.

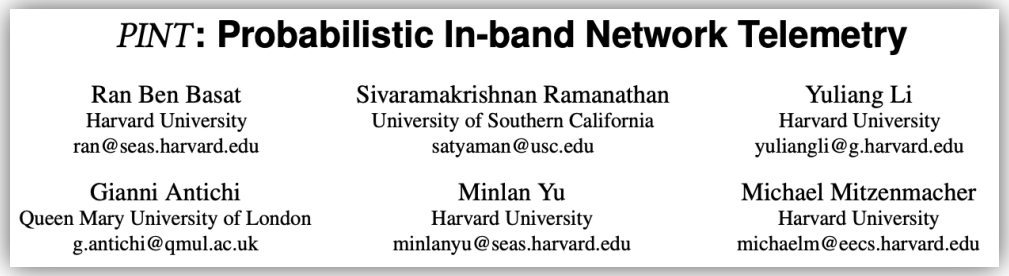
- Different applications may have different priorities
  - e.g., real time vs. backups

Previous work typically considers one class of traffic

HPCC: High Precision Congestion Control

Yuliang Li[♠♡], Rui Miao[♠], Hongqiang Harry Liu[♠], Yan Zhuang[♠], Fei Feng[♠], Lingbo Tang[♠], Zheng Cao[♠], Ming Zhang[♠], Frank Kelly[♢], Mohammad Alizadeh[♣], Minlan Yu[♡]

Alibaba Group[♠], Harvard University[♡], University of Cambridge[♢], Massachusetts Institute of Technology[♣]



***PINT*: Probabilistic In-band Network Telemetry**

Ran Ben Basat
Harvard University
ran@seas.harvard.edu

Sivaramakrishnan Ramanathan
University of Southern California
satyaman@usc.edu

Yuliang Li
Harvard University
yuliangli@g.harvard.edu

Gianni Antichi
Queen Mary University of London
g.antichi@qmul.ac.uk

Minlan Yu
Harvard University
minlanyu@seas.harvard.edu

Michael Mitzenmacher
Harvard University
michaelm@eecs.harvard.edu

Bolt: Sub-RTT Congestion Control for Ultra-Low Latency

Serhat Arslan*
Stanford University

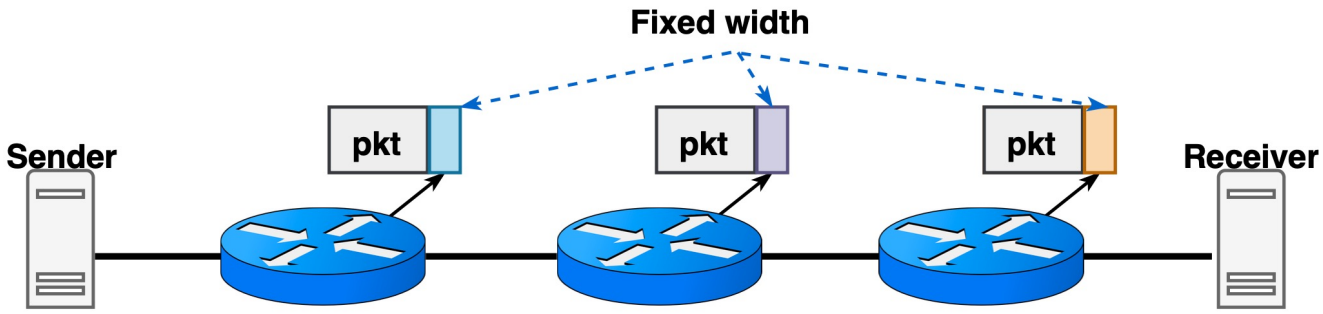
Yuliang Li
Google LLC

Gautam Kumar
Google LLC

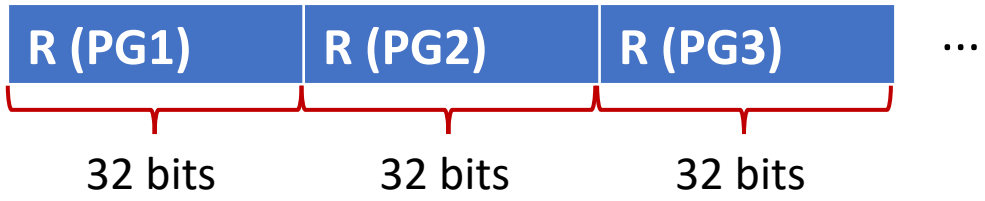
Nandita Dukkipati
Google LLC

The PrioMeter System

- Extend telemetry for end-to-end monitoring
- Meter end-to-end per PG data rate in data plane



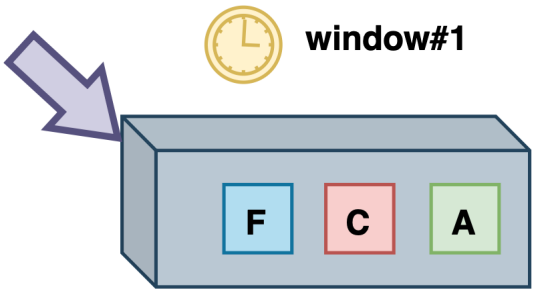
Overhead of per-PG measurement



Minimizing the overhead

- Two main components
  1. **Quantization**
  2. **Truncation**

Metering via P4 registers

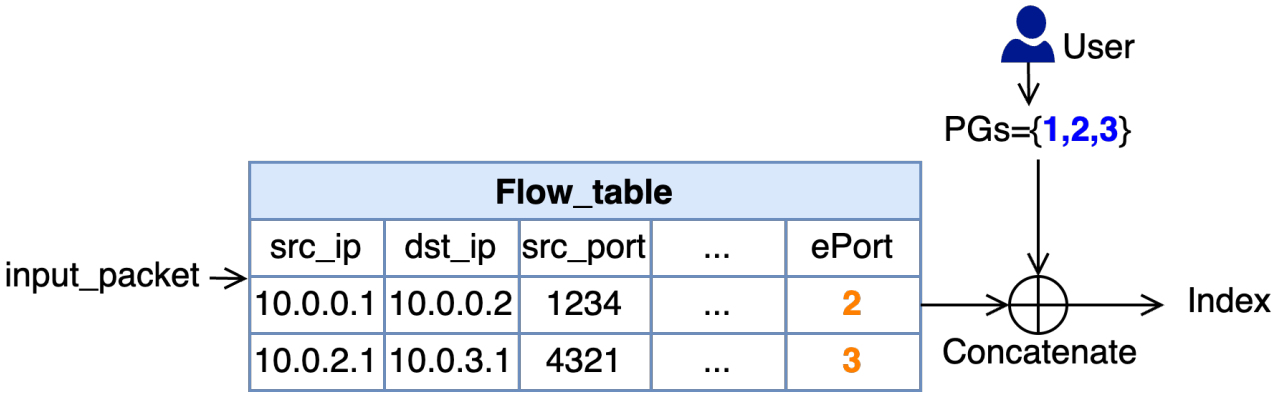


byte_reg=0, timer_reg=0, rate_reg-0

pkt	pkt_length	byte_reg	timer_reg	rate_reg
A	10	10	$T+=ts\_A$	0
C	20	30	$T+=ts\_C$	0
F	30	60	$T+=ts\_F$	0
		0	**T=ts_F**	**60**

At the end of window#1 →

Step 1: Get register index



Step 2: Quantize data rates

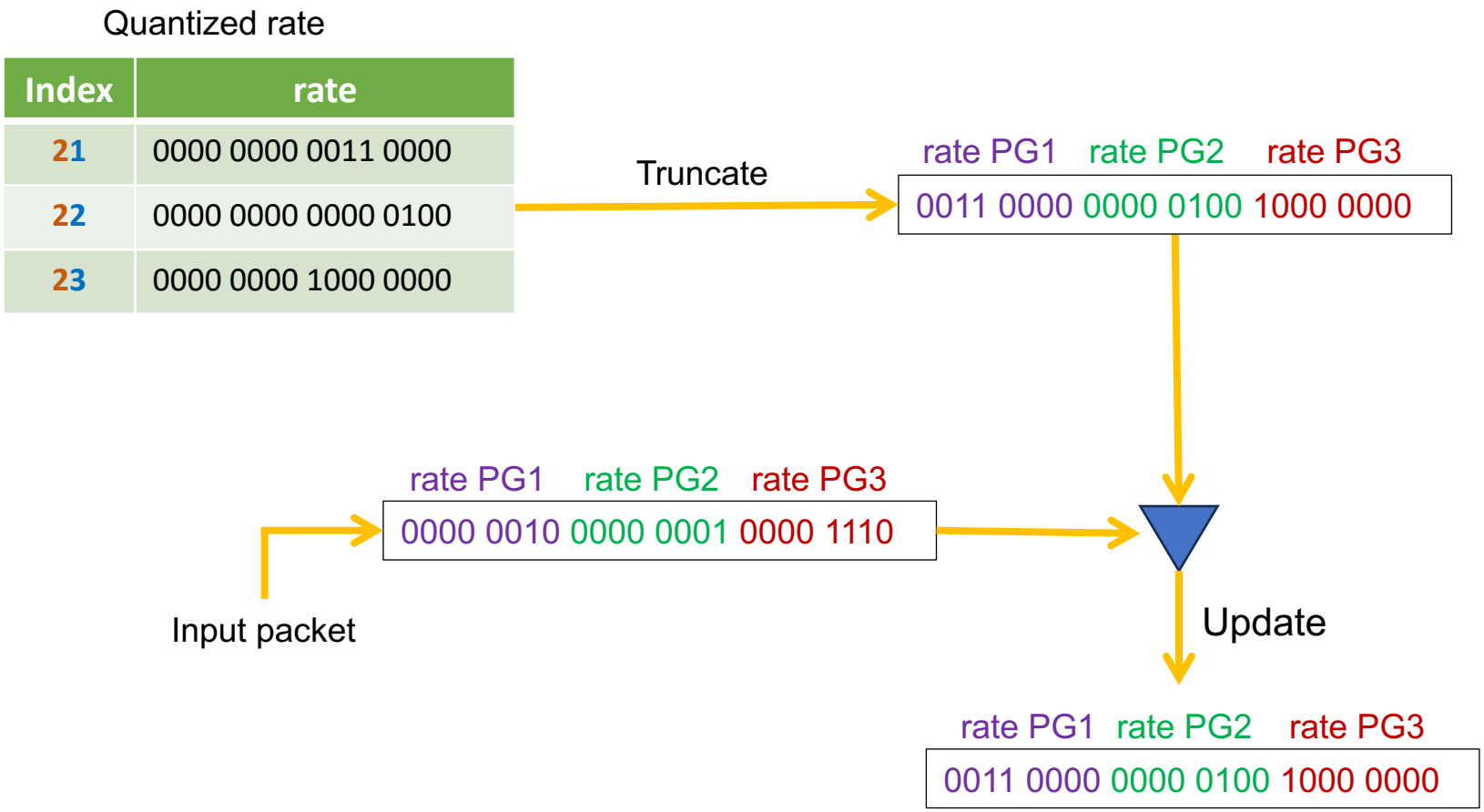
Index	rate_reg
21	0011 0000 1010 0011
22	0000 0100 1100 1100
23	1000 0000 0101 1010

QL>>8

Quantized rates

Index	rate
21	0000 0000 0011 0000
22	0000 0000 0000 0100
23	0000 0000 1000 0000

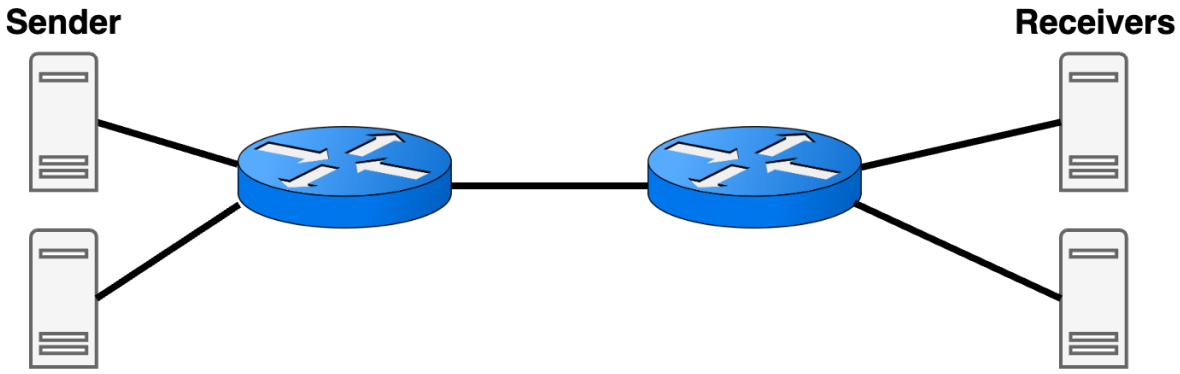
Putting all together: Truncate and append



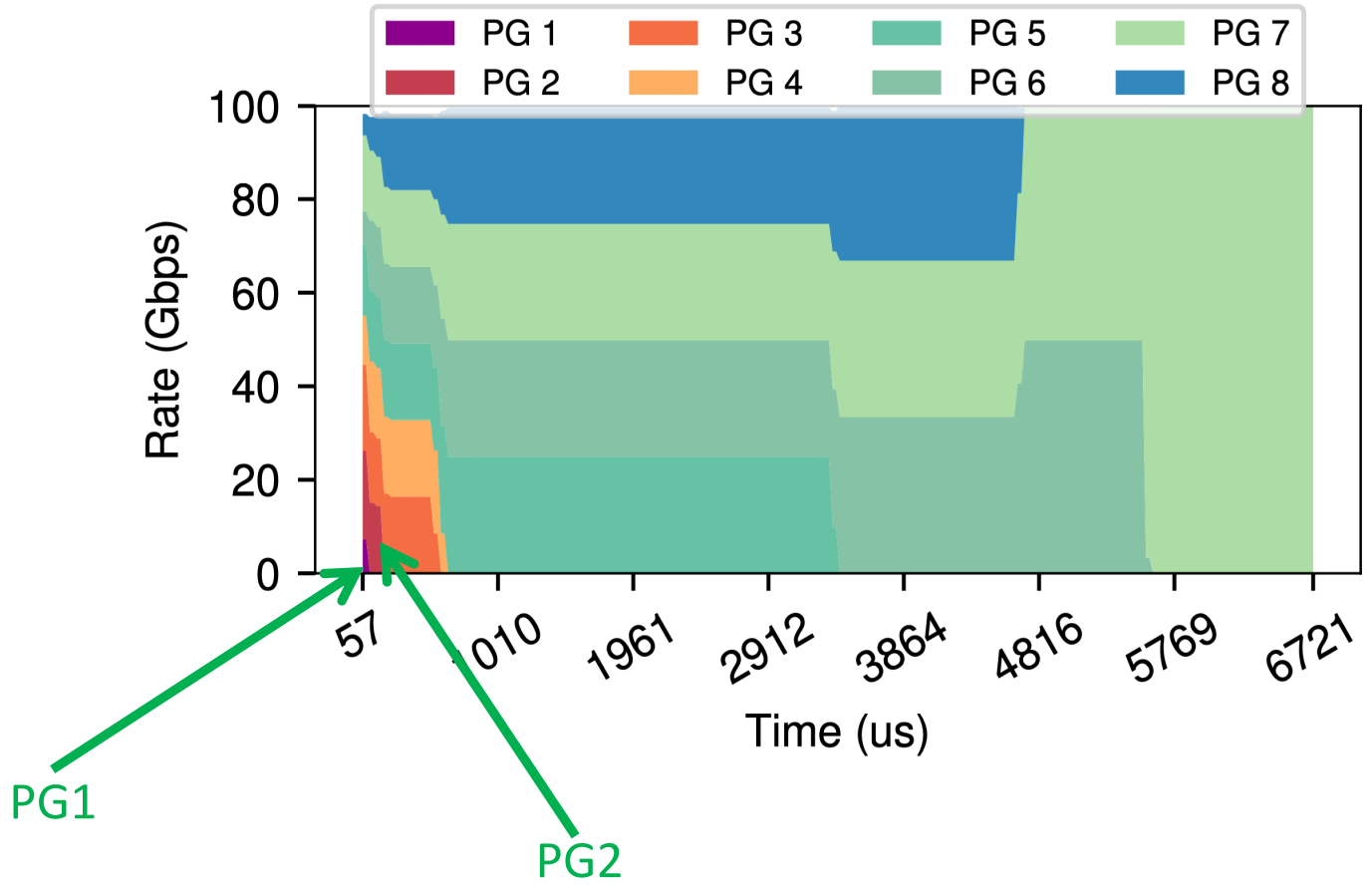
Evaluation

- Setup
  - dumbbell topology, 2 senders, 2 receivers with 100 Gbps links
  - FatTree topology, 320 servers (like HPCC)

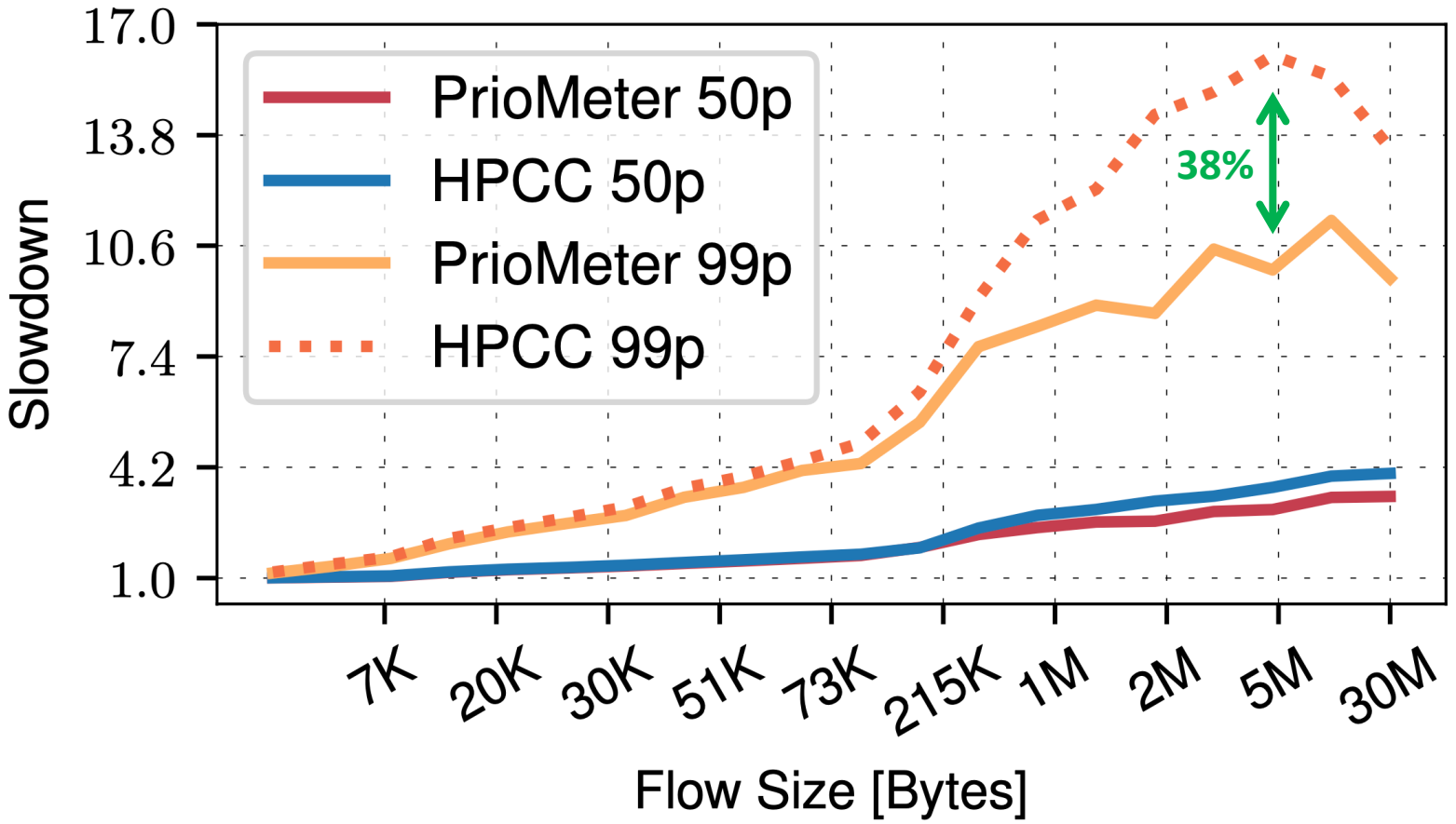
- Performance metrics
  - Per PG data rates
  - FCT slowdown



Per PG data rates



FCT slowdown (8 PGs)



Summary

- PrioMeter meters per PG data rate and minimize the overhead
  - compacting the measured data rates via **quantization**
  - **truncating** the quantized values to data rates per PG
- Results
  - up to **38% reduction** in FCT slowdown
- Tofinon implementation of PrioMeter is in progress

https://github.com/mostafaei/PrioMeter

Habib Mostafaei - h.mostafaei@tue.nl