# To The Moon and Back

## are Internet delays really that large?

**Chiara Chirichella**
**Dario Rossi**
dario.rossi@enst.fr
http://www.enst.fr/~drossi

FP7 IP project
(grant no. 318627)

# Agenda

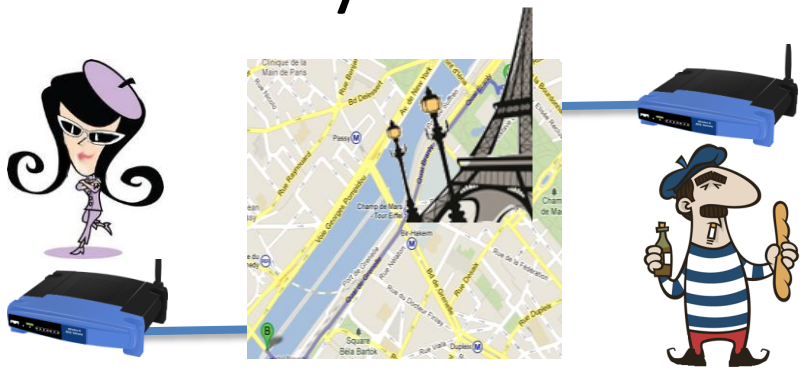
- Background
  - Bufferbloat, or the Internet is broken (again) [PAM'10,IMC'10,CACM'12,IMC'12, TMA'13,CCR'13]
  - Measuring Bufferbloat by exploiting LEDBAT [PAM'13]
- Hybrid Internet measurement campaign
  - Quantify bufferbloat (caveat of coupling bias!)
  - Root cause analysis

# Background: Bufferbloat

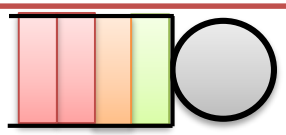
- RTT Delay between two Internet hosts ?

# Background: Bufferbloat

- RTT Delay between two closeby Internet hosts ?

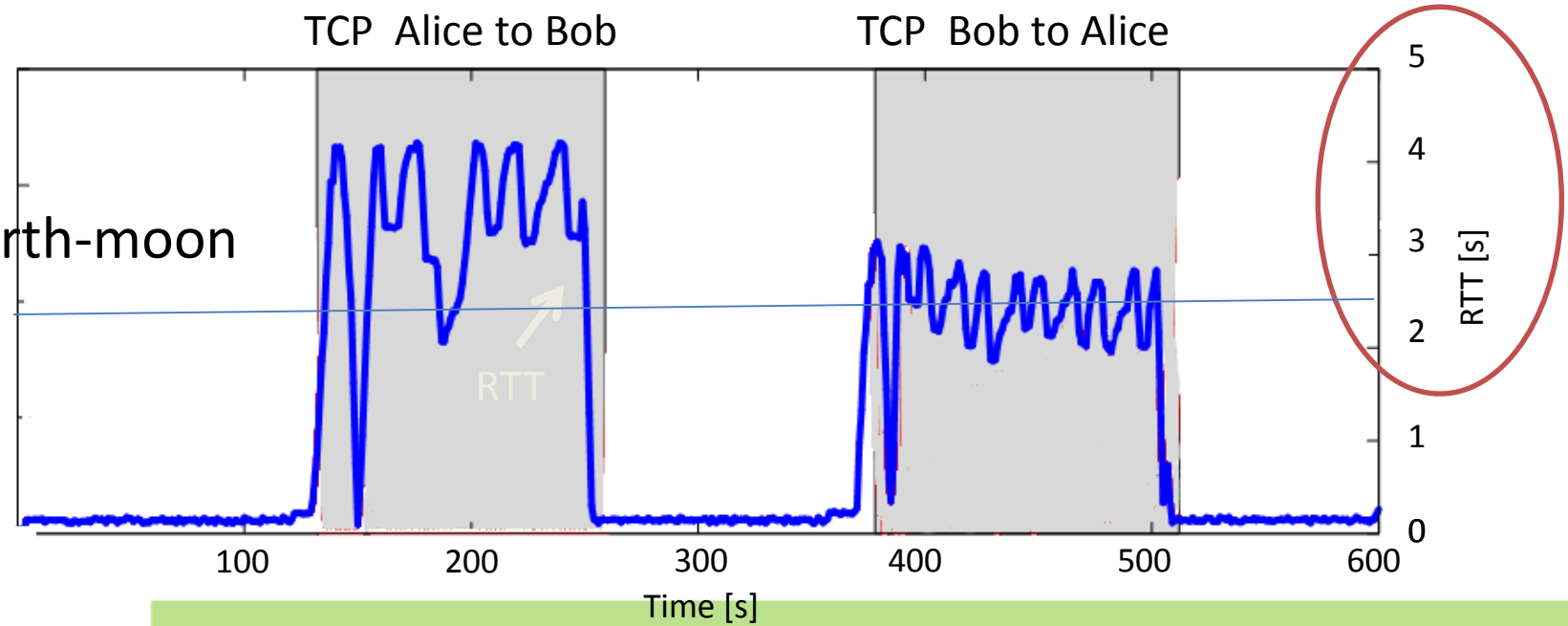


Bufferbloat!



RTT may grow to several seconds!
Nasty impact on interactive Web, VoIP, gaming traffic, etc.

RTT earth-moon




Source:
[PAM'10]

# Background: Bufferbloat

- Old issue, new interest
  - Previously work on buffers focused on the network core, where due to high speed, bufferbloat is limited --but situation is different at the network edge!
  - Small buffers not enough: cause of inefficiency, not fit for varying capacity (eg. WiFi) [TMA'13]
  - BitTorrent abandoned TCP in favor of LEDBAT due to this problem [PAM'10]
  - Moore law drove down memory costs, exhacerbate bufferbloat [CACM'12]
- Measurement work
  - Maximum bufferbloat via backlogged transfers [IMC'10],
  - Active testbed on 3/4G via ICMP RTT measurement [IMC'12]
  - Passive observation of TCP RTT variation on campus [CCR'13]

# Background: LEDBAT

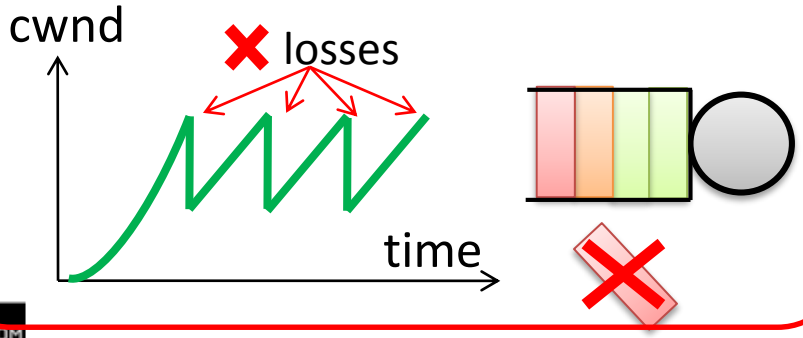
## TCP

### Detect congestion by losses

- Increment the congestion window (*cwnd*) by one packet per RTT
- Halve cwnd on loss events

### Consequences

- The buffer always fills up
- High delay for interactive apps
- Users need to prioritize traffic !




## LEDBAT

### Infer via delay measurement

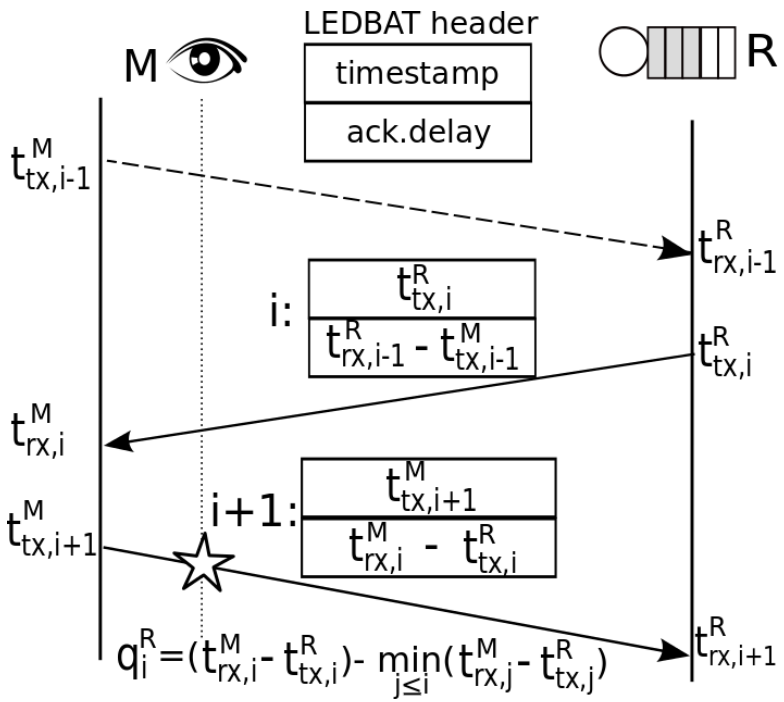
- Senders measure minimum delay
- Queuing delay = difference wr.t. min
- React with linear controller

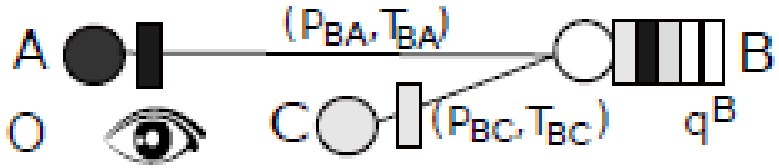
### Aim

- At most TARGET ms of delay
- Lower priority than TCP
- Do not harm interactive application
- Avoid self-induced congestion

# Background: LEDBAT

- Exploiting LEDBAT to infer Bufferbloat [PAM'13]
  - Passive traffic observation
  - Exploit LEDBAT headers (UDP header)
  - Perform same state update sender would do
  - Queuing delay = difference with respect to mimum
  - Gauge remote peer buffer even with hidden B->C traffic
  - Very accurate vs ground truth
    - Kernel-level queue
    - Application-level RTT

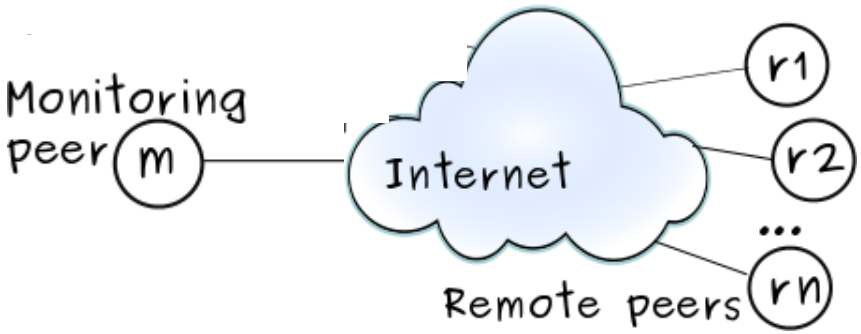


# Internet campaign

Hybrid methodology

- actively join torrents (M)
- passively analyze remote peers traffic (R1.. RN)




Dataset

- Nearly 100experiments, >10 torrents (Book, Games, Software, Video, …), 3 vantage points (FR, IT, AU)
- Available at

  http://www.enst.fr/~drossi/dataset/bufferbloat-internet

~20000 remote peers
~8000 unique LEDBAT IPs
~6000 sending >50 packets

# Internet campaign

- Quantify bufferbloat in the wild Internet
  - Caveat: coupled sampling

$$\mathrm{cwnd}_{i+1} = \max(1, \mathrm{cwnd}_i + \gamma(\tau - q_i)/\tau\mathrm{cwnd}_i)$$

- Identify root causes
  - Access type (reverse DNS)
  - Operating system (TTL fingerprint)
  - BitTorrent client (BT handshake)

| Feature | Class | Details | Cardinality |
|---|---|---|---|
| AT | Low-speed | DSL | 2660 |
| | High-speed | Cable | 245 |
| | | Fiber | 152 |
| | | *Total* | *3057* |
| | | *Unclassified* | *1144* |
| OS | Windows | WinXP/200x/Vista/7 | 5668 |
| | | Win95 | 897 |
| | Other | MacOS/Linux<2.x/BSD | 254 |
| | | Linux>2.x | 77 |
| | | *Total* | *6896* |
| BC | uTorrent | uTorrent | 3552 |
| | Other | Vuze | 345 |
| | | Transmission | 14 |
| | | Azureus | 11 |
| | | *Total* | *3937* |

# Quantify bufferbloat

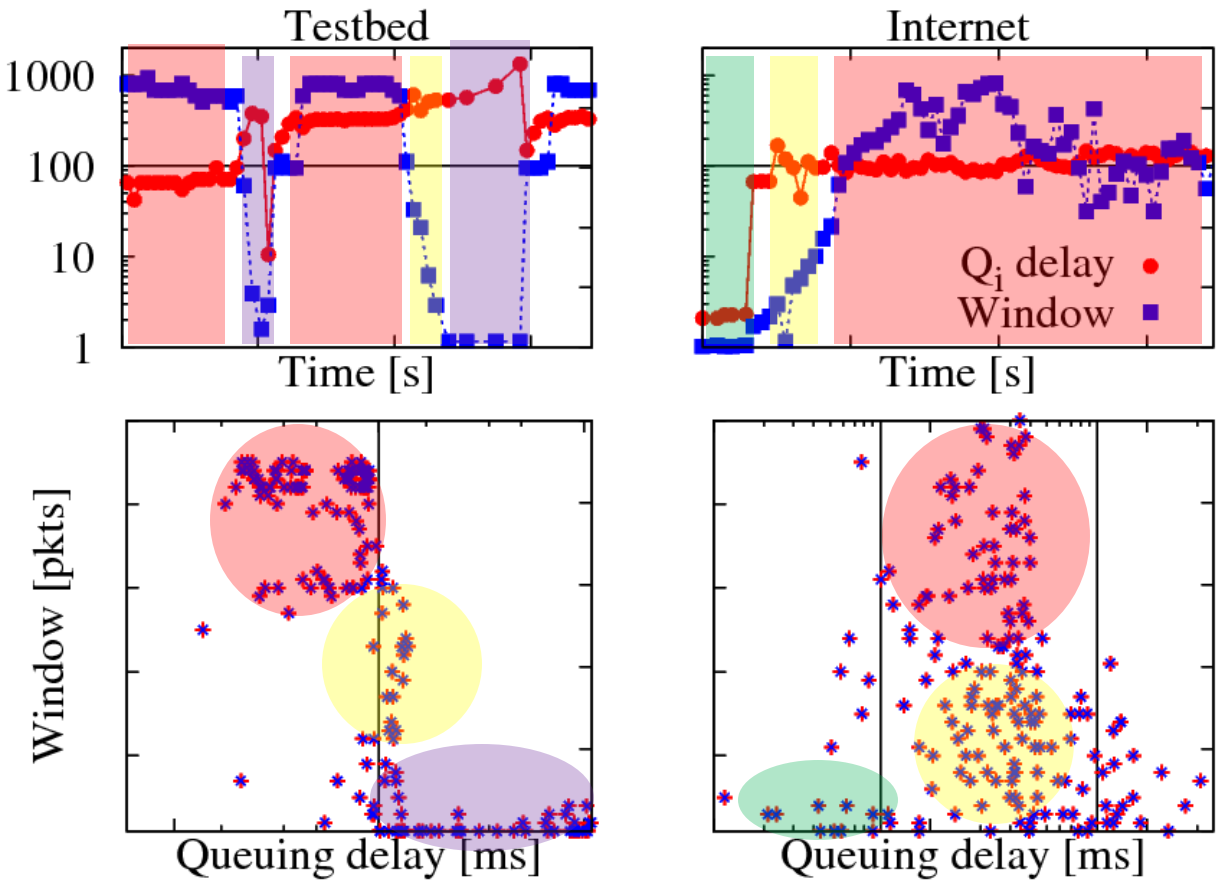
- Avoid bias due to coupled sampling
  - Bufferbloat samples = LEDBAT window,
  - LEDBAT window shrinks when queuing delay increases

$$\mathrm{cwnd}_{i+1} = \max(1, \mathrm{cwnd}_i + \gamma(\tau - q_i)/\tau \mathrm{cwnd}_i)$$

- Problem
  - Risk of *bufferbloat* understimation: few samples when the queuing delay is large
  - Risk of *target delay* overestimation: LEDBAT stabilizes around target delay if alone in the bottleneck
- Solution
  - Batch samples in short fixed-size windows $t \in [ti, t0+W]$
  - Qi = E[qi] over all packets in the window

# Quantify bufferbloat

- Coupling example




Interfering TCP traffic
Window shrinks
Underestimate queue

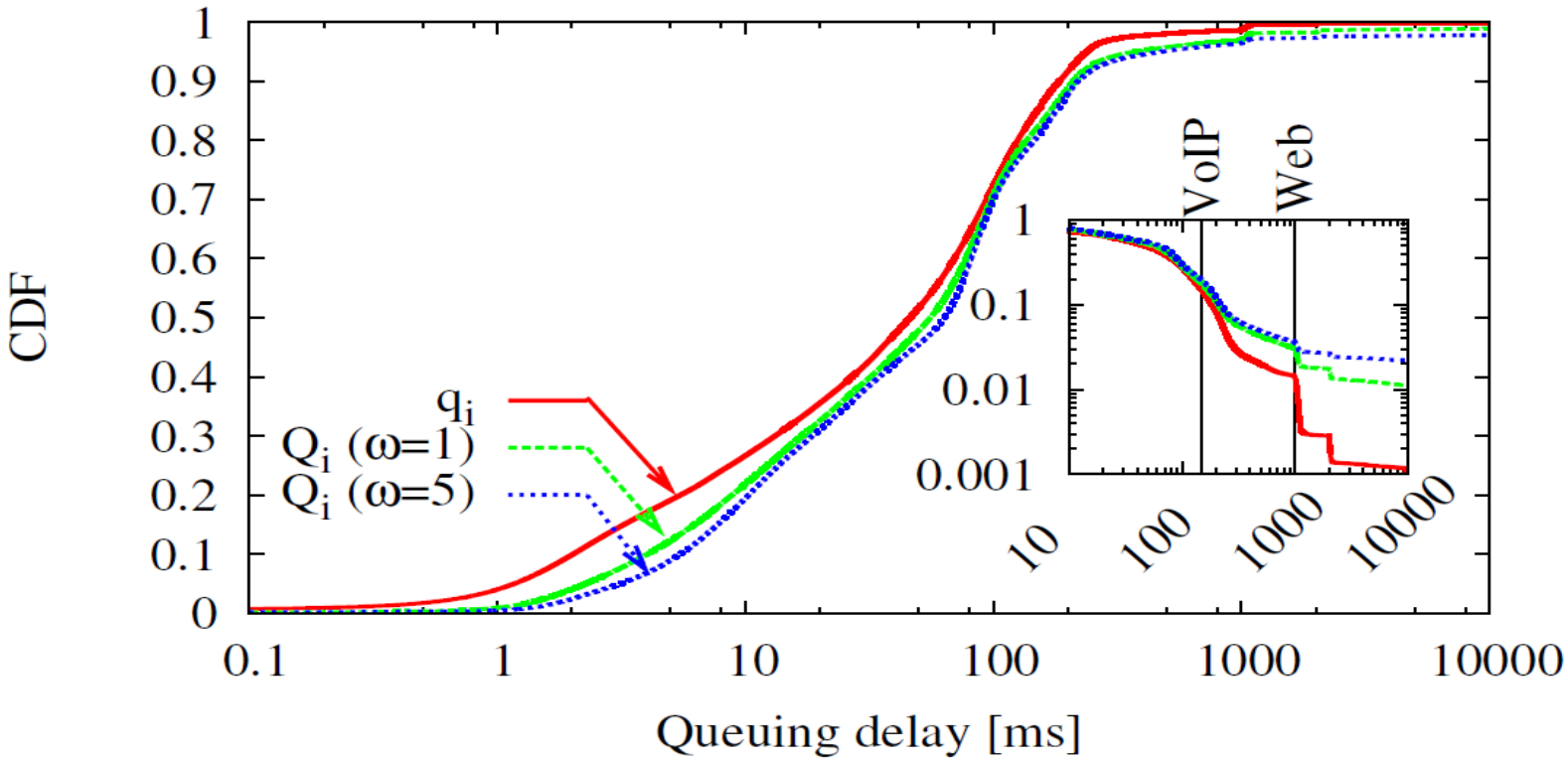
LEDBAT alone
Full window
Overestimate target

Application limited
Few samples, but low queuing

Transient periods

# Quantify bufferbloat

- Windowed estimates (Qi) correct per-packet (qi) bias
- Interactivity threshold (100ms) exceeded 10% of the times
- Bufferbloat (>1s) rare, but experienced by 1% of peers



# Root cause analysis

To protect herself from bufferbloat, user can:

- Change operator/access type (AT)
  - Better modem, smaller buffers, contracts with higher capacity, …
- Change Operating System (OS)
  - Congestion control flavor of TCP interacting traffic, timestamp precision affect LEDBAT behavior, …
- Change BitTorrent Client (BC)
  - L7 application settings: # peers, connection management policy, …

| Feature | Class | Details | Cardinality |
|---|---|---|---|
| AT | Low-speed | DSL | 2660 |
| | High-speed | Cable | 245 |
| | | Fiber | 152 |
| | | *Total* | *3057* |
| | | *Unclassified* | *1144* |
| OS | Windows | WinXP/200x/Vista/7 | 5668 |
| | | Win95 | 897 |
| | Other | MacOS/Linux<2.x/BSD | 254 |
| | | Linux>2.x | 77 |
| | | *Total* | *6896* |
| BC | uTorrent | uTorrent | 3552 |
| | Other | Vuze | 345 |
| | | Transmission | 14 |
| | | Azureus | 11 |
| | | *Total* | *3937* |

# Root cause analysis

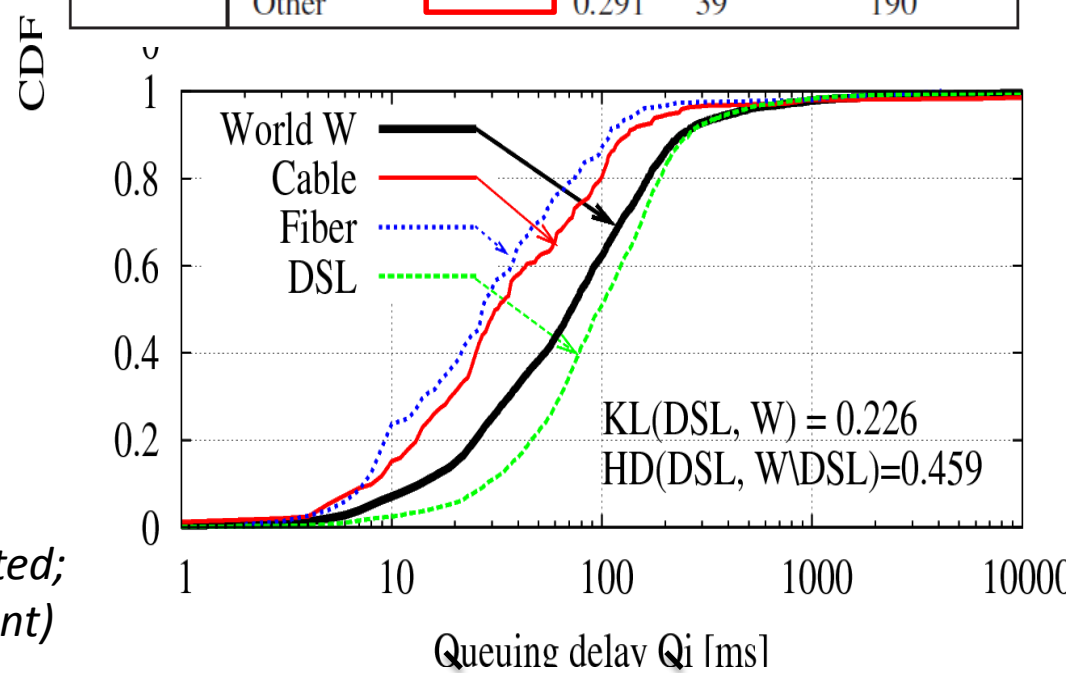
Methodology
- Breakdown peer statistics per AT, OS and BC classes
- Quantify difference per classes
- Extract samples with equal size to avoid class imbalance
- Compare to random subset for score baselines

Metrics
- Kullback-Leibler (KL) divergence
- Hellinger Distance (HD)

Results
- Access type biggest role
- OS and BC non neglibile either

*(caveat: AT, OS and BC can be correlated; e.g., all Windows peer may use uTorrent)*

| Feature | Class | HD | KL | p50 [ms] | p90 [ms] |
|---|---|---|---|---|---|
| R | Rand $A$ | 0.176 | 0.023 | 70 | 230 |
| | $W \setminus A$ | | 0.023 | 70 | 240 |
| AT | Low-speed | 0.459 | 0.227 | 94 | 240 |
| | High-speed | | 0.333 | 30 | 120 |
| OS | Windows | 0.341 | 0.141 | 55 | 220 |
| | Other | | 0.308 | 31 | 190 |
| BC | uTorrent | 0.349 | 0.135 | 77 | 220 |
| | Other | | 0.291 | 39 | 190 |



# Conclusions

- Findings
  - Bufferbloat rare but possible (1% of peers, maybe due to concurrent BitTorrent TCP transfer to legacy peers)
  - Bufferbloat primarily affected by peer access type
  - Operating system (~TCP flavors) and application (~connection management) also correlates with queuing delay magnitude
- Limits
  - Limited to hosts running BitTorrent
  - Possibly over-estimate bufferbloat (BitTorrent)
  - Possibly spatially and temporally limited view (hosts of BT swarms, when peers use BT)
  - Data intensive (RX data from remote hosts)
- Ongoing
  - Lightweight active methodology
  - Purely passive campaign

# References


**[PAM'10]** D. Rossi, C. Testa and S. Valenti, Yes, we LEDBAT: Playing with the new BitTorrent congestion control algorithm . In Passive and Active Measurement (PAM'10), 2010

**[IMC'10]** C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating the edge network. In ACM Internet Measurement Conference (IMC'10), 2010.

**[CACM'12]** J. Gettys et al. Bufferbloat: what's wrong with the internet? Communications of the ACM, 55(2):40–47, 2012.

**[IMC'12]** H. Jiang, Y. Wang, K. Lee, and I. Rhee. Tackling bufferbloat in 3G/4G networks. In ACM In ACM Internet Measurement Conference (IMC'12), 2012.

**[CCR'13]** Mark Allman. Comments on Bufferbloat, ACM SIGCOMM Computer Communication Review, 43(1), January 2013.

**[PAM'13]** C. Chirichella, D. Rossi, C. Testa, T. Friedman and A. Pescape, Remotely Gauging Upstream Bufferbloat Delays . In Passive and Active Measurement (PAM'13), Extended Abstract, 2013.

**[TMA'13]** Y. Gong, D. Rossi, C. Testa, S. Valenti and D. Taht, Fighting the bufferbloat: on the coexistence of AQM and low priority congestion control . In IEEE INFOCOM Workshop on Traffic Measurement and Analysis (TMA'13), 2013.

**[this talk]** C. Chirichella, D.Rossi, To the Moon and Back: are Internet delay really that large? In IEEE INFOCOM Workshop on Traffic Measurement and Analysis (TMA'13), 2013.


Further informations on LEDBAT and bufferbloat:

http://www.enst.fr/~drossi/ledbat