McCormick
Northwestern Engineering
Department of Electrical Engineering and Computer Science

On Dominant Characteristics of Residential Broadband Internet Traffic

G. Maier, A. Feldmann, V. Paxson and M. Allman

Presented by Zachary Bischof

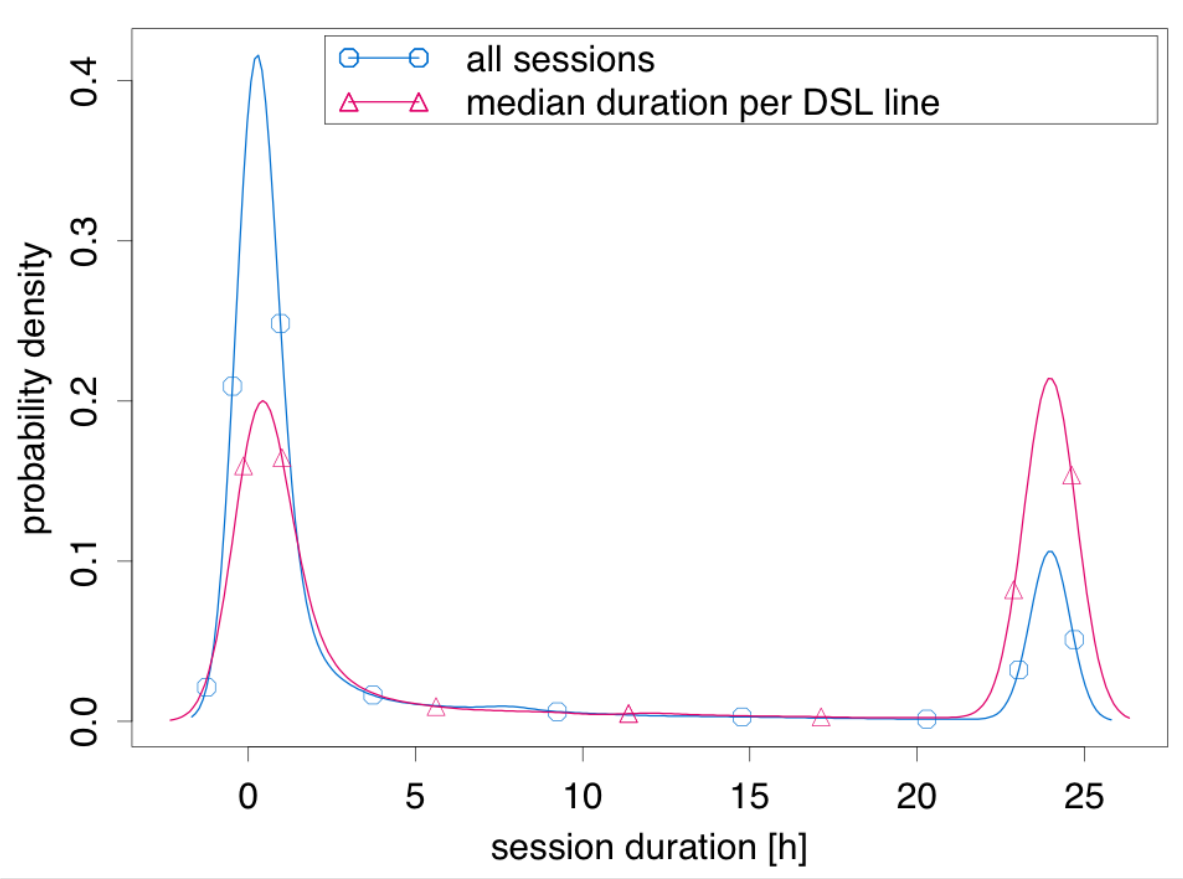
Motivation

- Residential networks are widespread
  - Differ from campus and enterprise networks
  - Not as well studied
  - No acceptable-use policies
- Focus
  - DSL customer sessions
  - Applications
  - HTTP
  - Performance characteristics

Data Sets

- Large European ISP (>10 million customers)
- Anonymized packet-level traces
  - 20,000 DSL users
  - Covering a single urban region
  - Full traces for an extended period of time
    - E.g. two months
  - Continuous monitoring

DSL Sessions

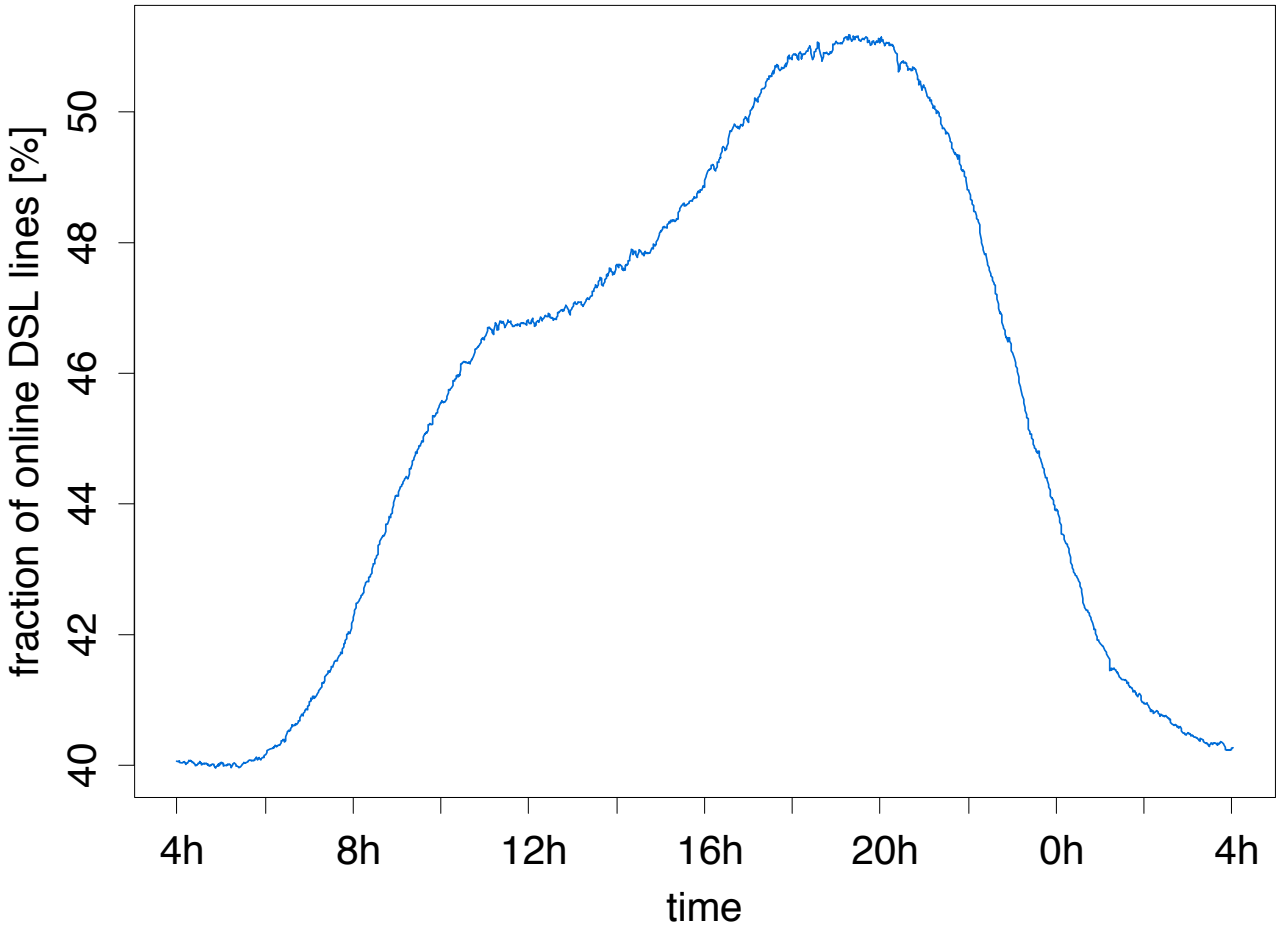


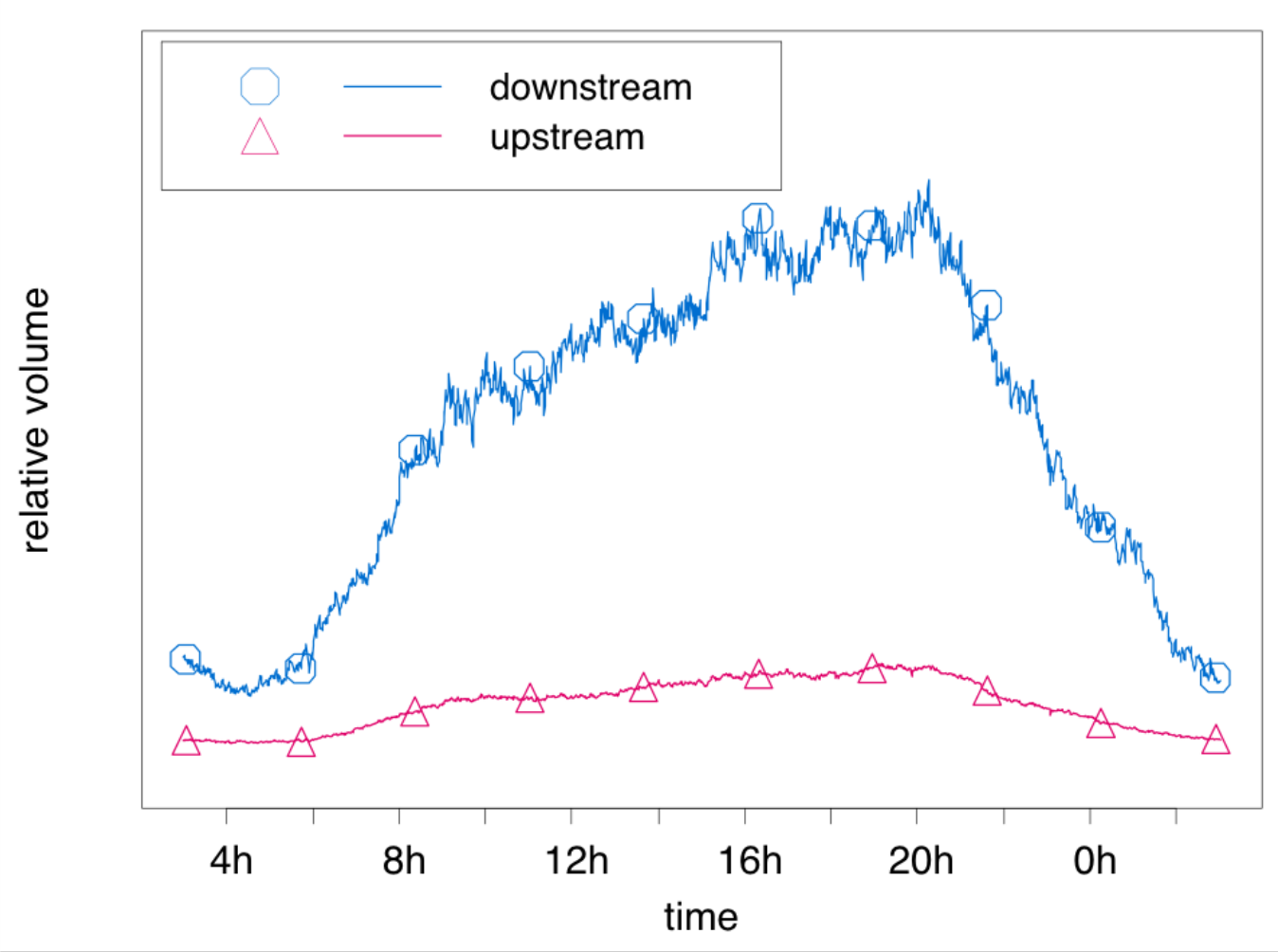
Sessions were *surprisingly* short (20-30 min)

DSL Sessions

- New IP every session
- No timeout
- No traffic shaping
- High churn
  - 50% of addresses assigned twice in a 24 hour period
  - Most sessions terminated by user
  - 5% more than 10 reassignments per day
- 40% are always online

DSL Sessions




downstream
upstream
relative volume
4h
8h
12h
16h
20h
0h
time

Application Usage

- Use Bro's Dynamics Protocol Detection
- 85% classified, 3.6% on well-known ports
- No day of week effects
- Port-based detection works well for non-P2P

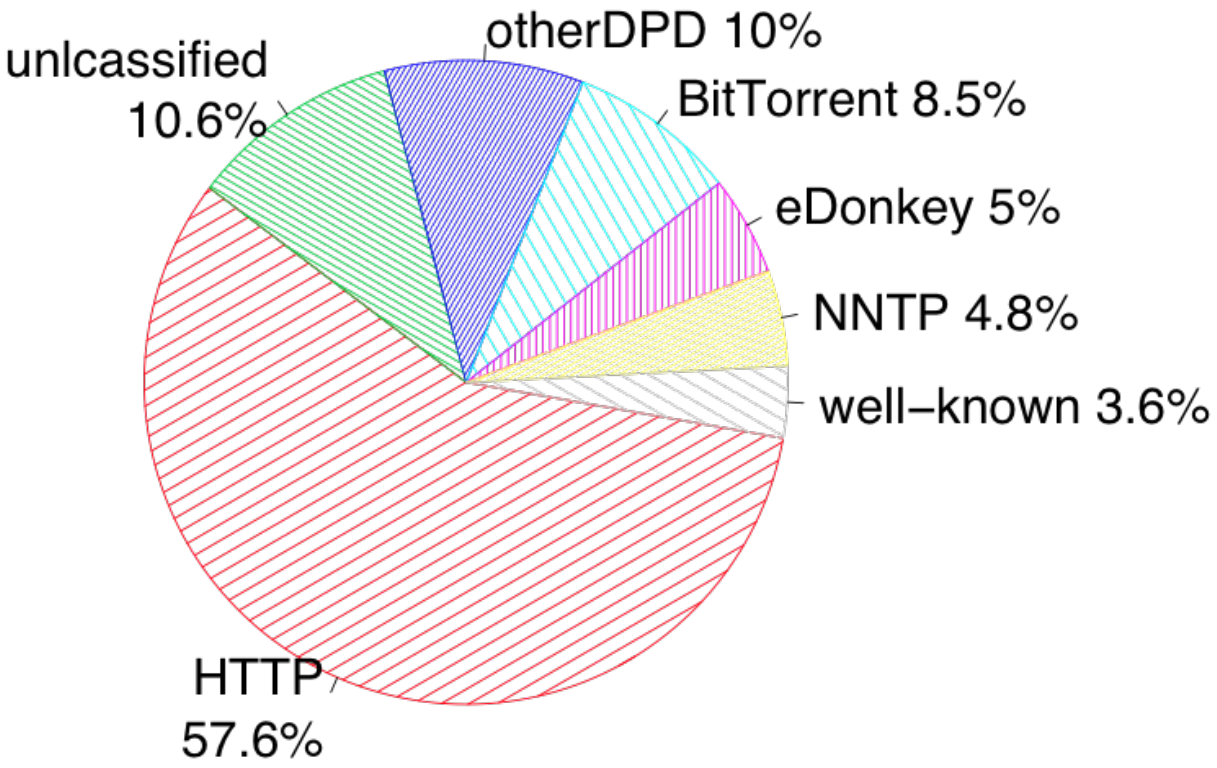
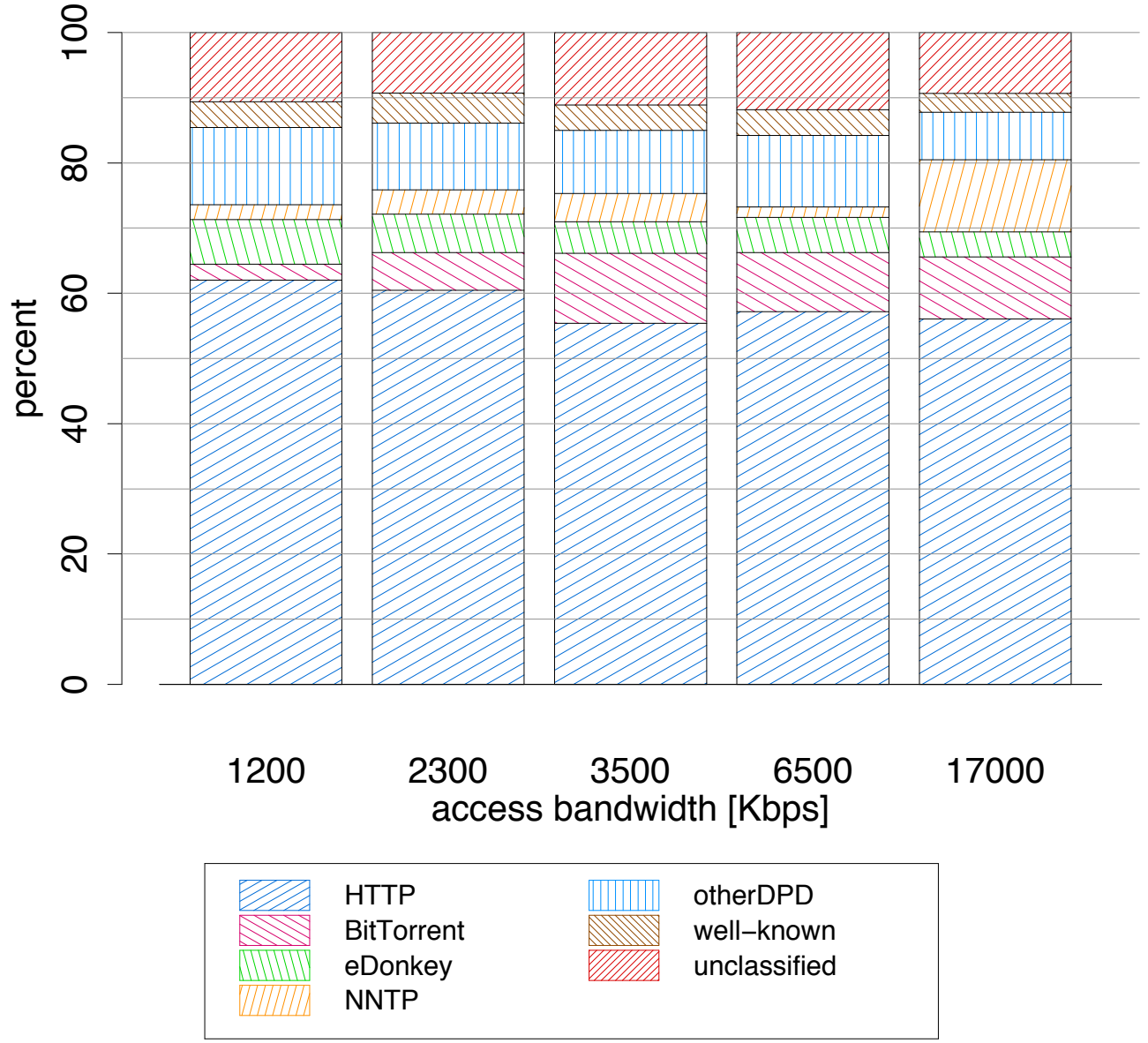


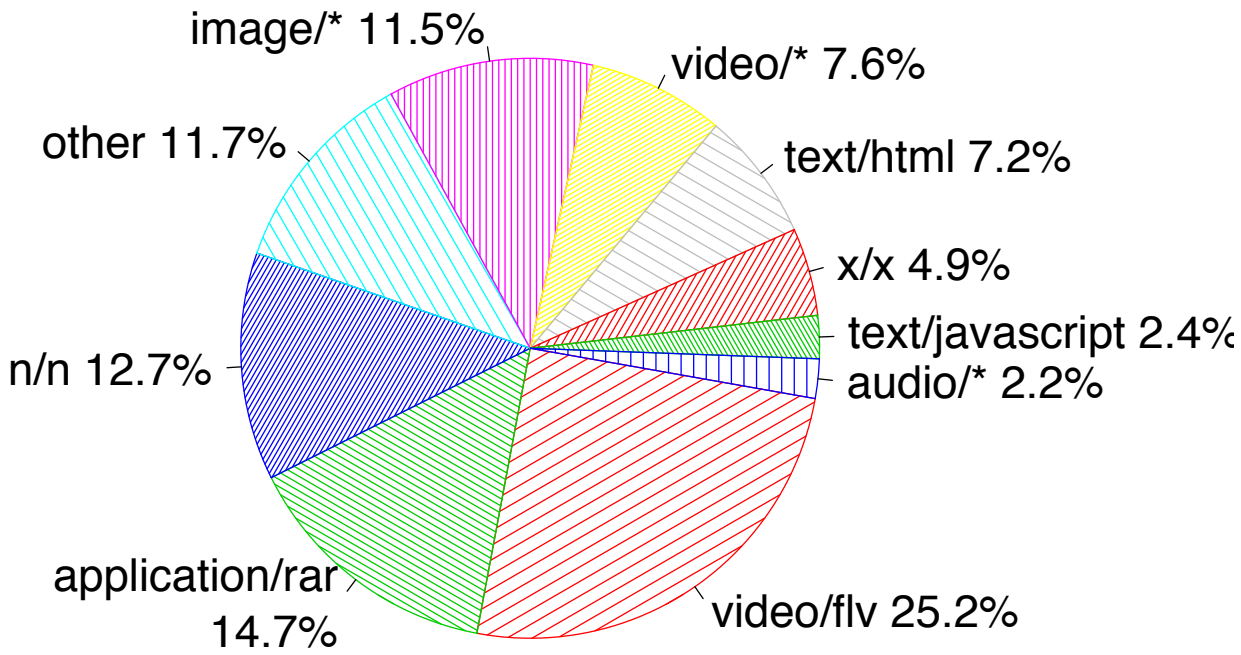
Figure 5: Application Mix for trace SEP.


percent
0
20
40
60
80
100
1200
2300
3500
6500
17000
access bandwidth [Kbps]
HTTP
BitTorrent
eDonkey
NNTP
otherDPD
well–known
unclassified

HTTP is Popular Again!

- HTTP – 25%
- File hosting – 15%
- P2P less than 14%
- Unclassified 11%

But why is HTTP popular?

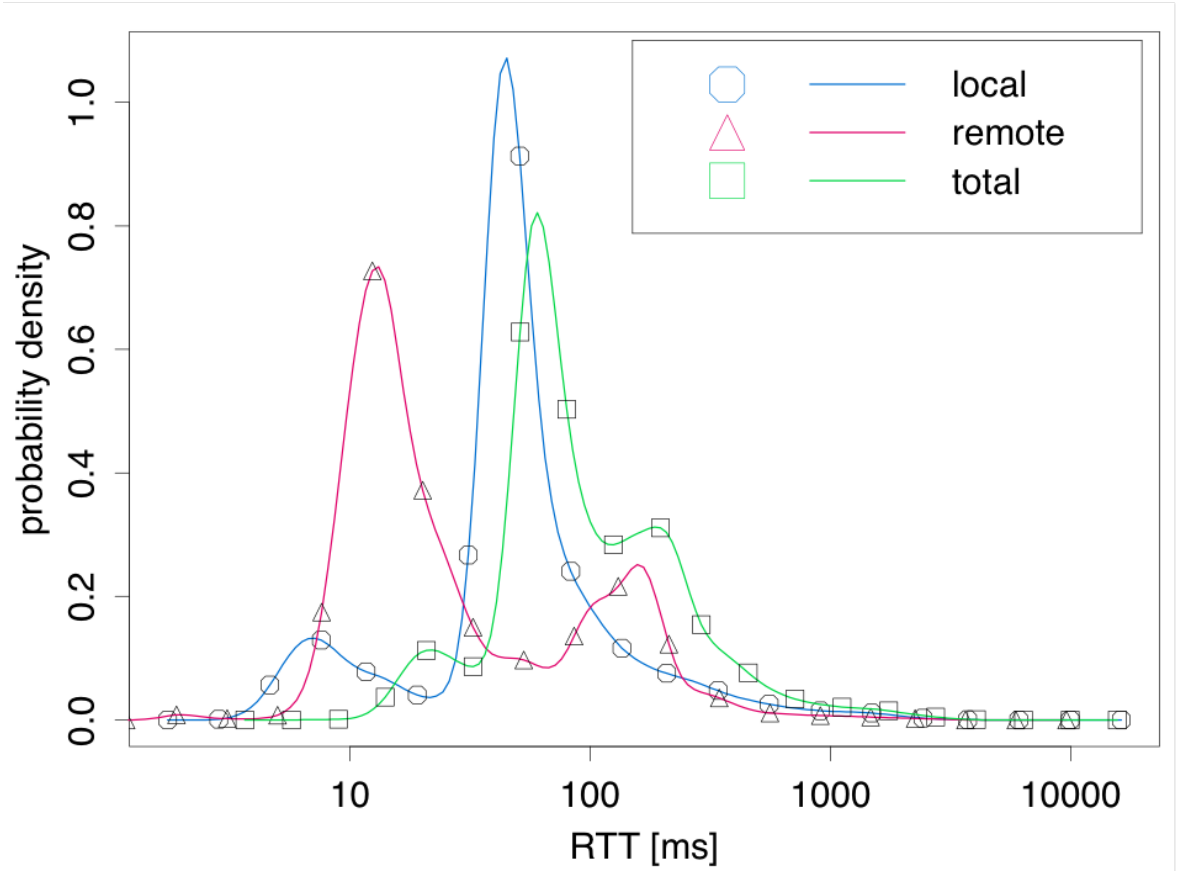


TCP Settings

- TCP misconfigurations are a common source of error
- Window scaling advertised by 50%
  - Needed when bandwidth-delay product exceeds 64K
- 21-39% advertise timestamps
- Selective Acknowledgement used by 90%
  - More efficiently ACKs received data
- No Explicit Congestion Notification

DSL Latencies

Local latency dominates (partially due to wireless)



Most people don't use all their bandwidth

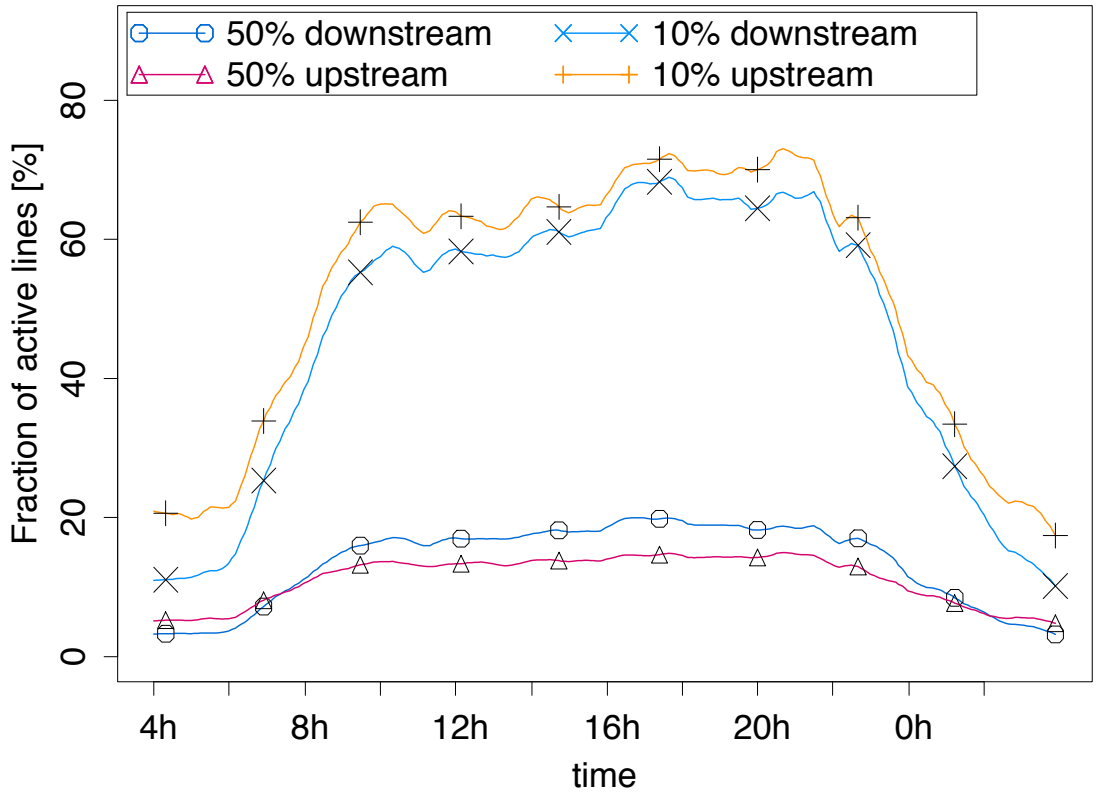
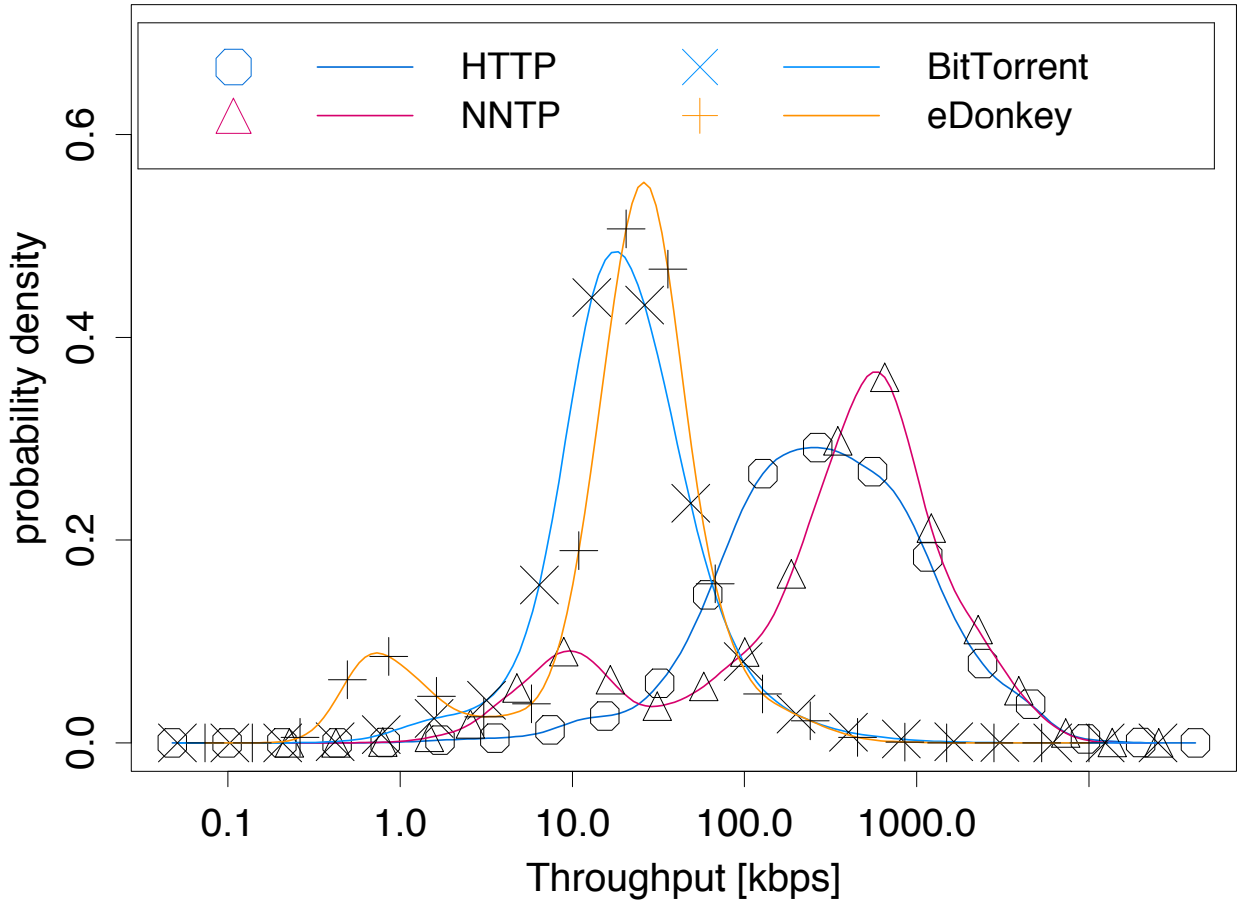


Figure 11: Fraction of active lines using 50%/10% of their available upstream/downstream bandwidth at least once per 5 minute bin (smoothed).

P2P achieved throughput is lower



Questions

- Only DSL customers (but DSL dominates in Europe)
  - How accurate is this for overall P2P usage?
- Nothing on change in application usage over time
- What is the cost of P2P traffic vs. HTTP traffic for the ISP?