



On Dominant Characteristics of Residential Broadband Internet Traffic

Gregor Maier
TU-Berlin / T-Labs
`gregor.maier@tu-berlin.de`

Anja Feldmann
TU-Berlin / T-Labs

Vern Paxson
UC Berkeley / ICSI

Mark Allman
ICSI

Internet Measurement Conference
2009-11-04



Motivation

- ❑ Residential broadband access popular/widespread
- ❑ Residential traffic is "underexamined"
- ❑ Differs from well-studied campus and enterprise traffic
 - Not subject to acceptable-use policies
- ❑ Focus on broad range of **dominant** characteristics
- ❑ Questions
 - DSL session characteristics / IP-reassignment
 - Popular protocols / applications (is P2P dominating?)
 - Performance characteristics



Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



Data sets (I)

- ❑ Large European ISP (> 10M customers in total)
- ❑ Anonymized packet level traces
 - Covering >20,000 DSL customers
 - One urban area
- ❑ Overview of packet level traces
 - 14 x 90min; twice per day over 1 week in Aug 2008
 - 24hr in Sep 2008 (>4TB)
 - 24hr in Apr 2009 (>4TB)
- ❑ Bro Intrusion Detection System for analysis



Data sets (2)

- Anonymized DSL session traces
 - DSL connect / disconnect times
 - Anonymized line-card ID
 - Access bandwidth
 - Augments packet data
- Overview of session traces
 - One for each packet trace
 - 10 day in Feb 2009

Outline

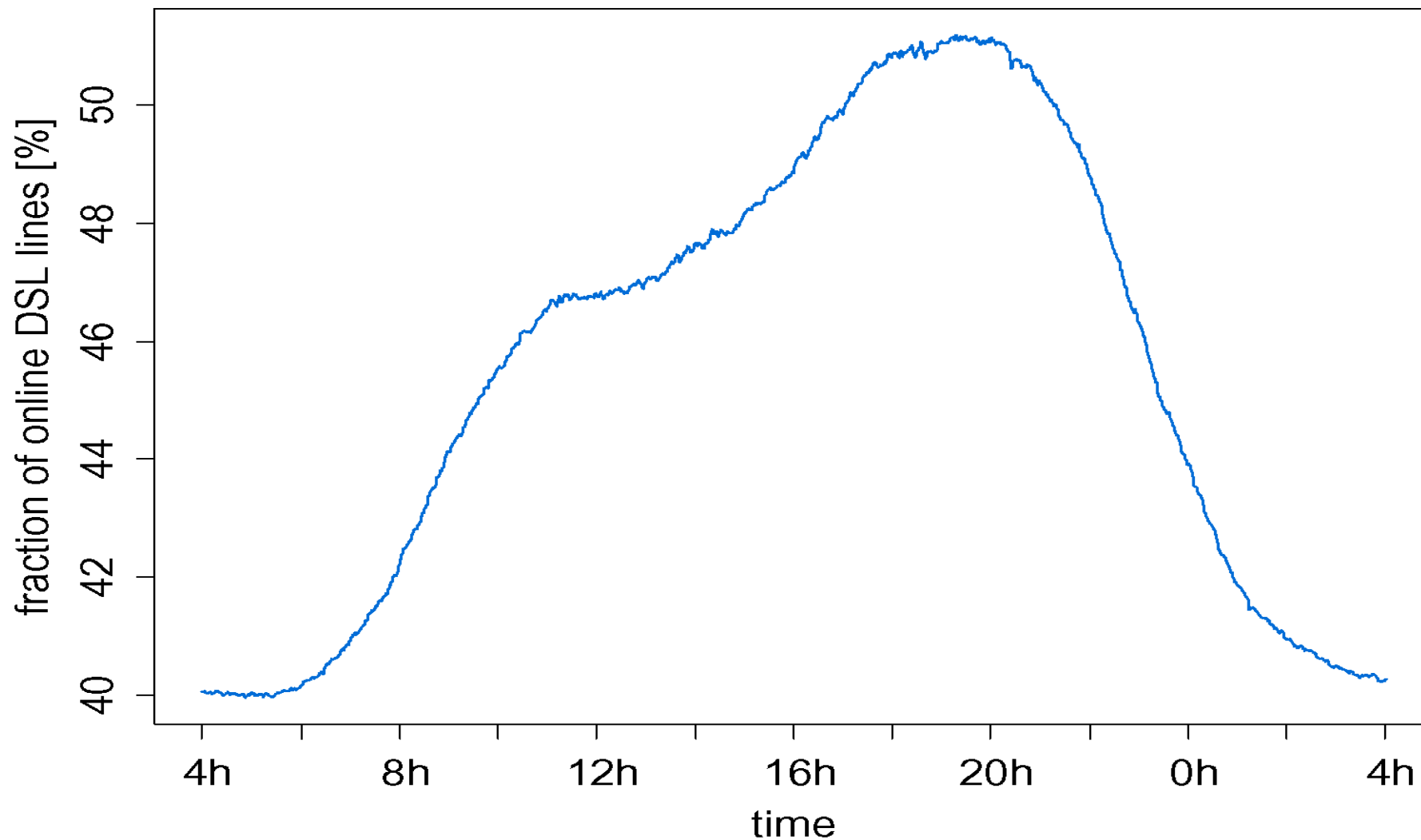
- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



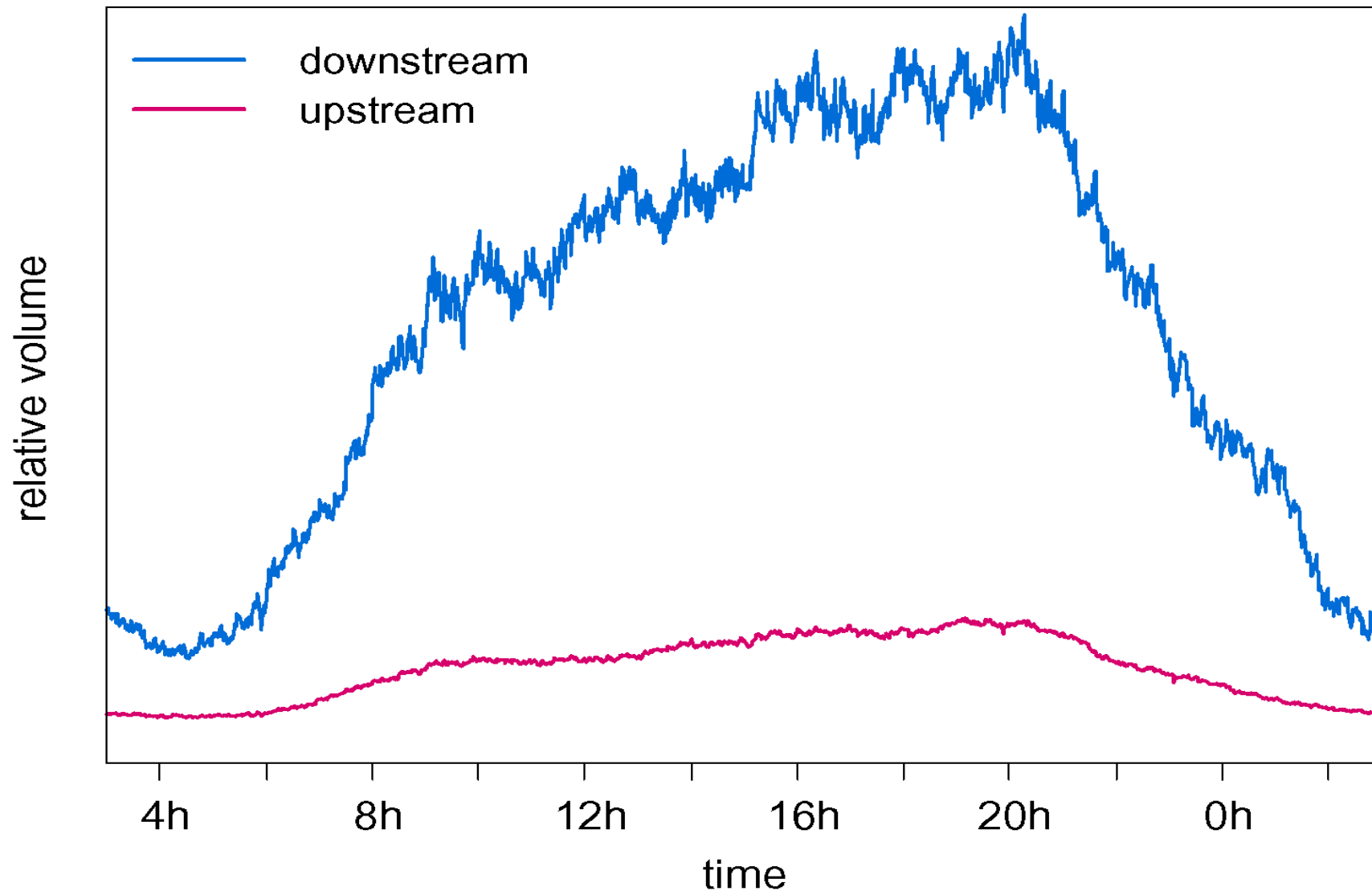
Background

- ❑ Radius is used for IP leases
 - New IP address for every session (round-robin)
- ❑ 24hr session timeout enforced by ISP
 - Home routers can immediately reconnect
- ❑ No idle timeout enforced by ISP
- ❑ No traffic shaping (e.g., P2P)
- ❑ Access bandwidths: 1–16 Mbps (downstream)

Concurrent DSL sessions

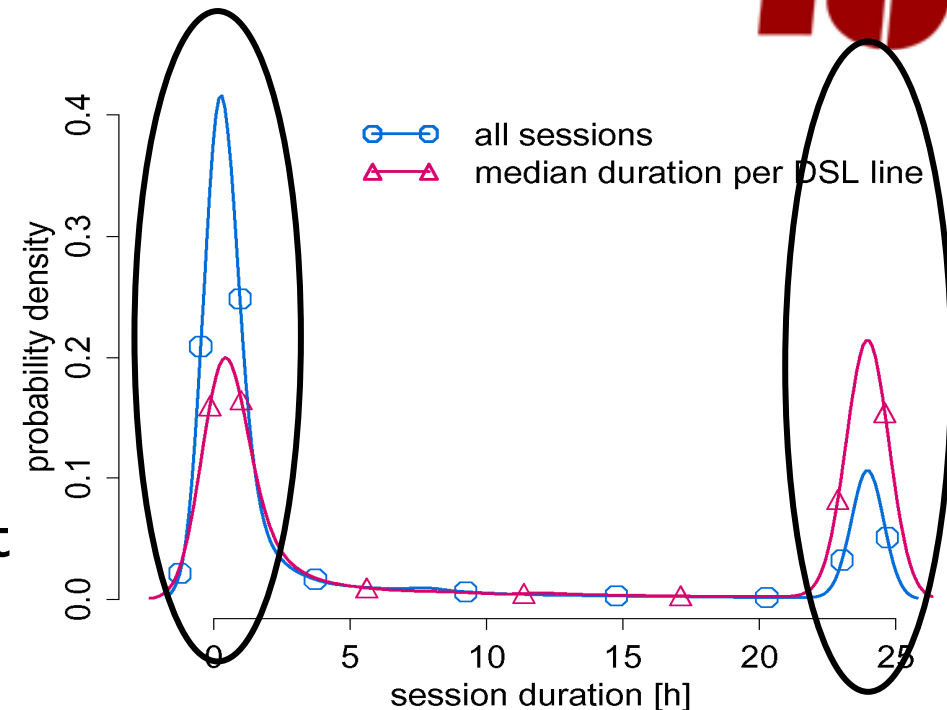


Aggregate throughput



Session results

- ❑ Only 40% of lines always-on
- ❑ Sessions are short
 - Median 20–30 minutes
 - Likely due to small default idle timeouts on home routers
 - >80% of sessions terminated by user request
- High IP address churn
 - 50% of IPs assigned twice in 24h
 - Up to 5% of IPs assigned more than 10 times





Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary

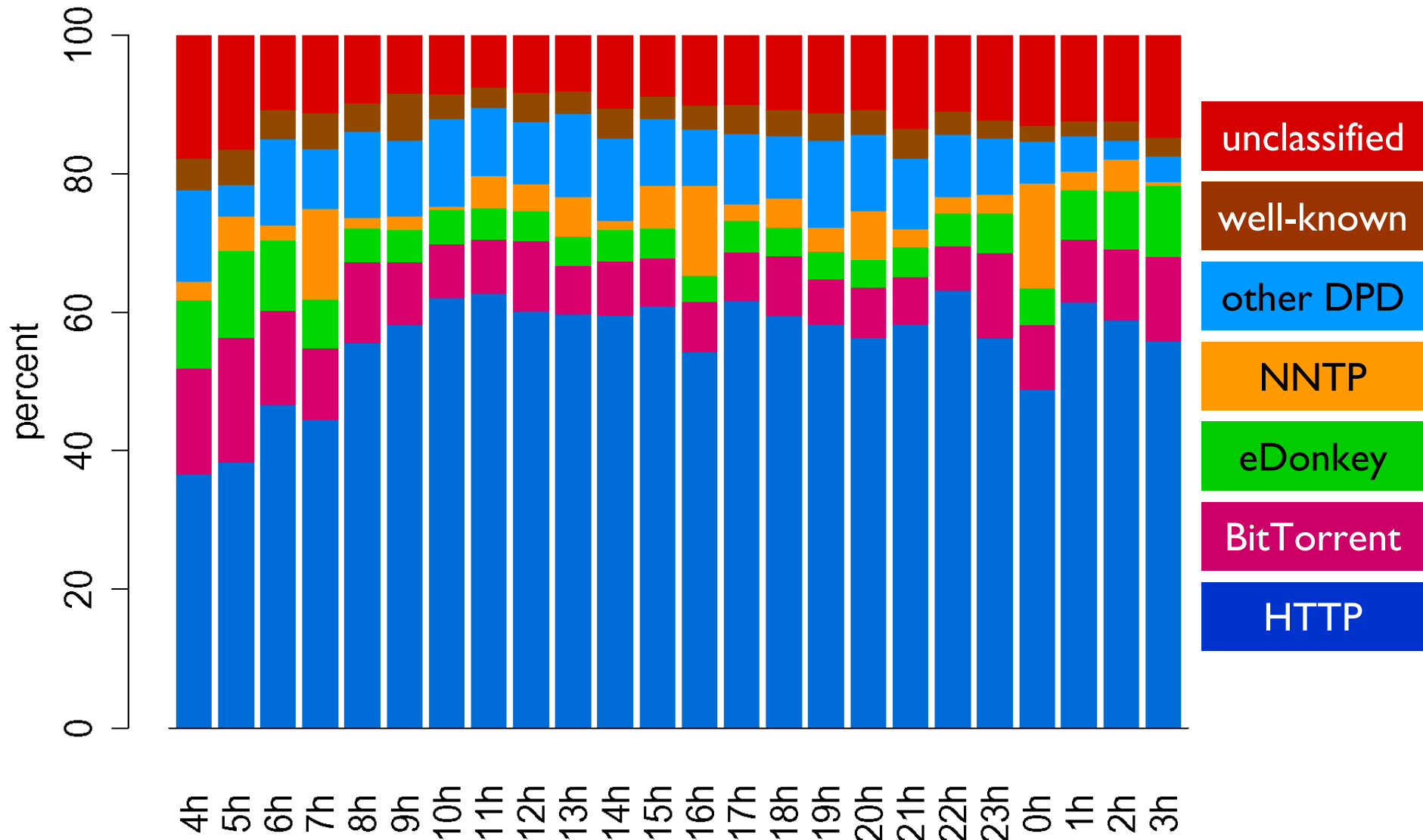


Methodology

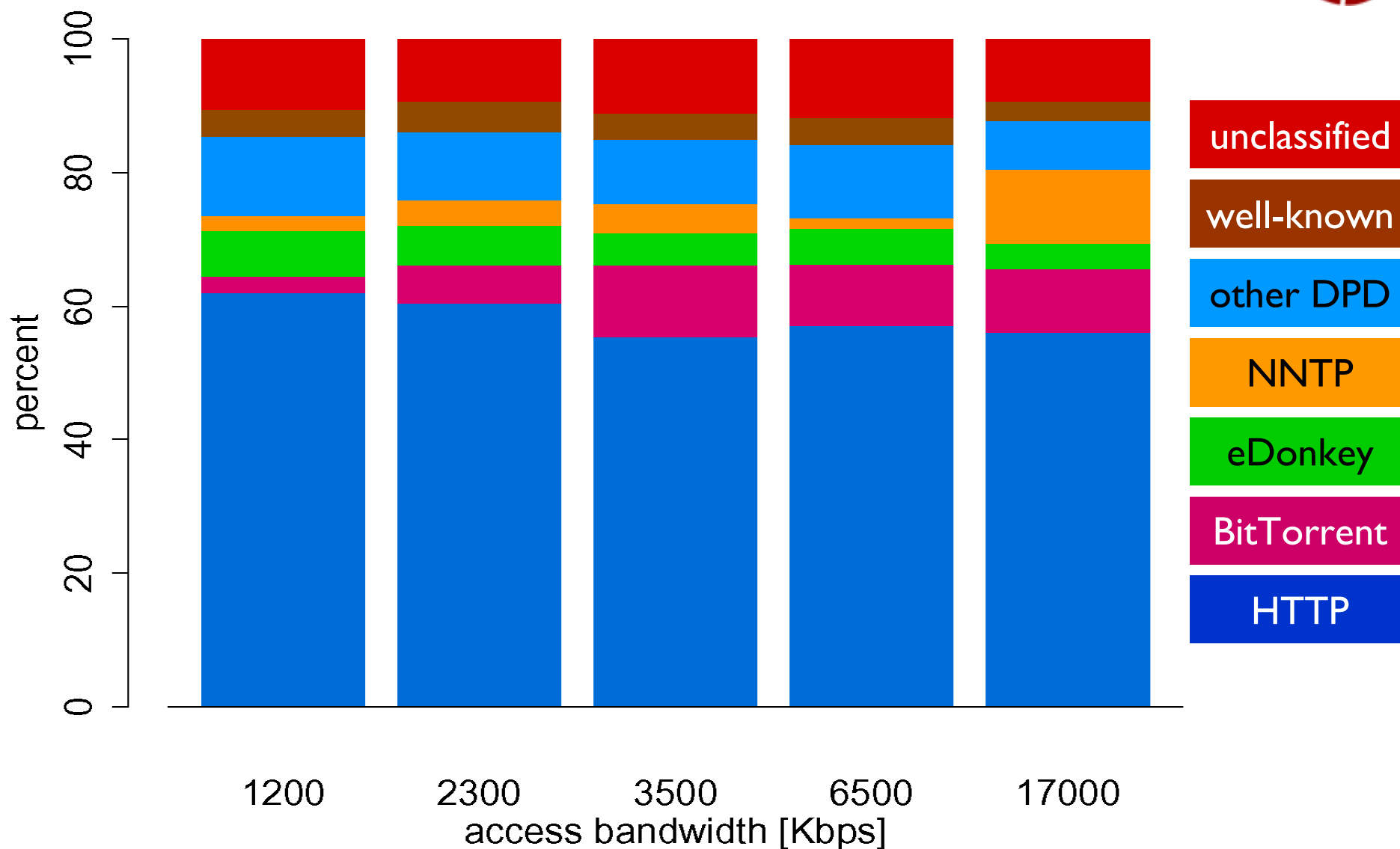
- ❑ Using Bro's Dynamic Protocol Detection (DPD)
 - Protocol semantics and/or
 - Signatures
- ❑ 85% of bytes classified, another 3.6% on well-known ports
- ❑ No dominant day-of-week effects
- ❑ Verification
 - With NetFlow data (port based)
 - With commercial DPI system at different location



Application usage per hour



Application usage per b/w





Key results

- ❑ HTTP dominates[◇]: 57% of bytes
- ❑ P2P less than 14%
- ❑ Unclassified: 11%
- ❑ Other significant protocols
 - NNTP 2–5%
 - Streaming (non-HTTP) 5%
 - Voice-over-IP 1.3%
- ❑ Port based classification works well for non-P2P protocols

[◇]Erman et al. found very similar results in contemporaneous work presented at WWW'09

Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



Motivation & methodology

- ❑ Why is HTTP so popular (again)?
 - HTTP offers popular high-volume content?
 - HTTP as transport protocol for other applications?
- ❑ Anonymized HTTP headers extracted via Bro
- ❑ Determine content-type
 - Content-type header
 - libmagic
- ❑ Second level domain (from Host header)
- ❑ User-Agent header

Key results

□ Popular content-types by volume



□ Domain popularity:

- One-click-hoster is top domain: 15% of HTTP bytes
- Video portals (using flash-video) follow

□ No significant hiding / tunneling via HTTP

➤ **HTTP dominance due to popular high-volume content**



Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



Performance/path characteristics

- ❑ Analyze TCP option usage
 - Per host
 - Per connection
- ❑ TCP loss/reordering
 - Out-of-sequence transmission
- ❑ Bandwidth utilization
- ❑ Round-trip-times (RTT)

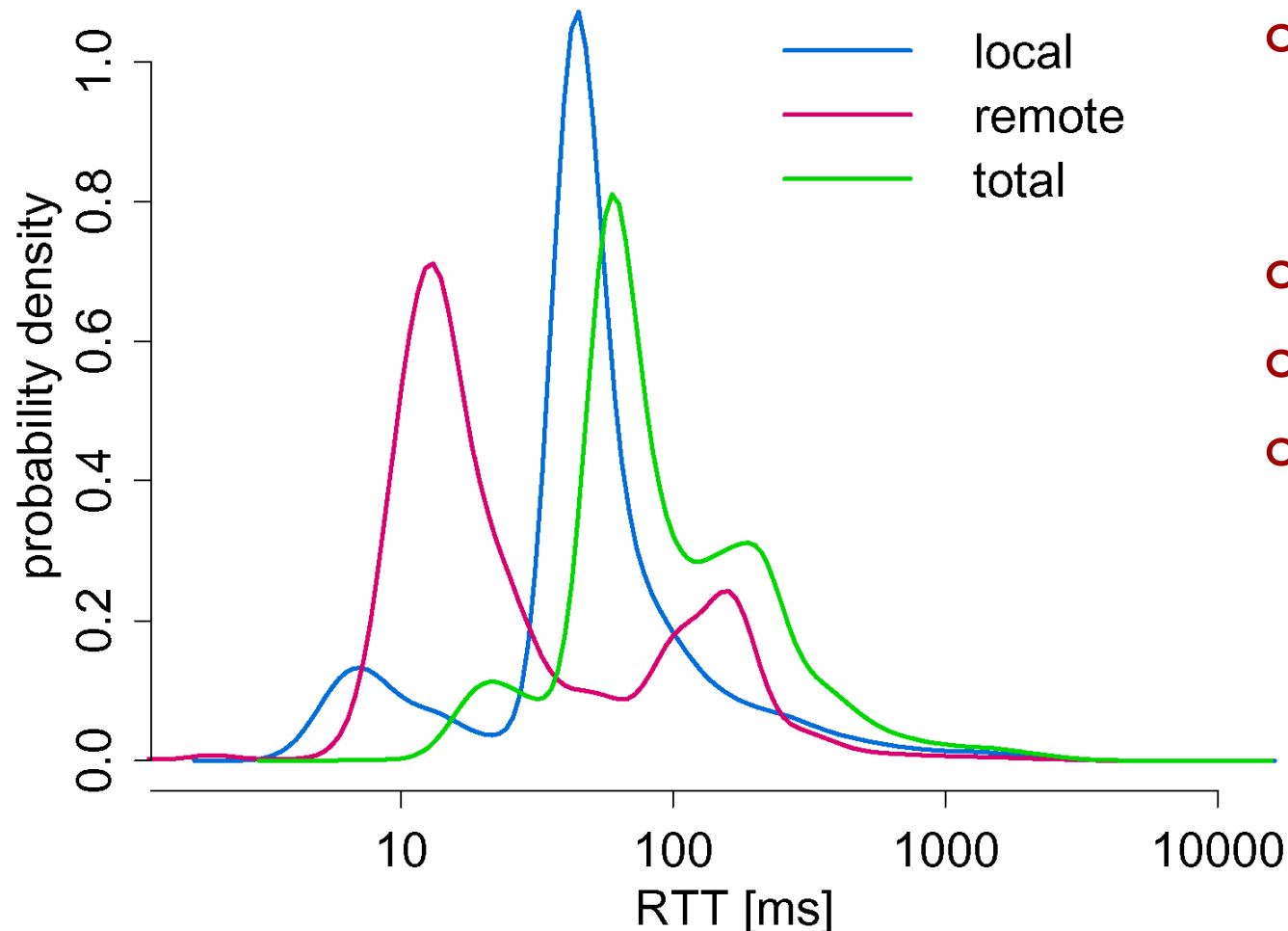


TCP options / performance

- ❑ Selective Acknowledgement (SACK) widely used (90%)
- ❑ Explicit Congestion Notification: non existent
- ❑ Window scaling
 - Approx. 50% of host advertise scaling; most non-zero
 - Maximal advertised window often 64KB
- ❑ Loss/reordering in 10% of connections
 - Corresponds with SACKs observed
- ❑ Bandwidth-delay product $>$ max. rcv. window
 - Affects 44% of connections with $>$ 50KB volume (downstream)
- ❑ Most lines only use small fraction of bandwidth

Round-trip-times (RTT)

□ Assessed during TCP handshake



- Local component dominates (DSL interleaving)
- Median: 74ms
- 99th perc: 1328ms
- Wireless equipment can cause significant delays

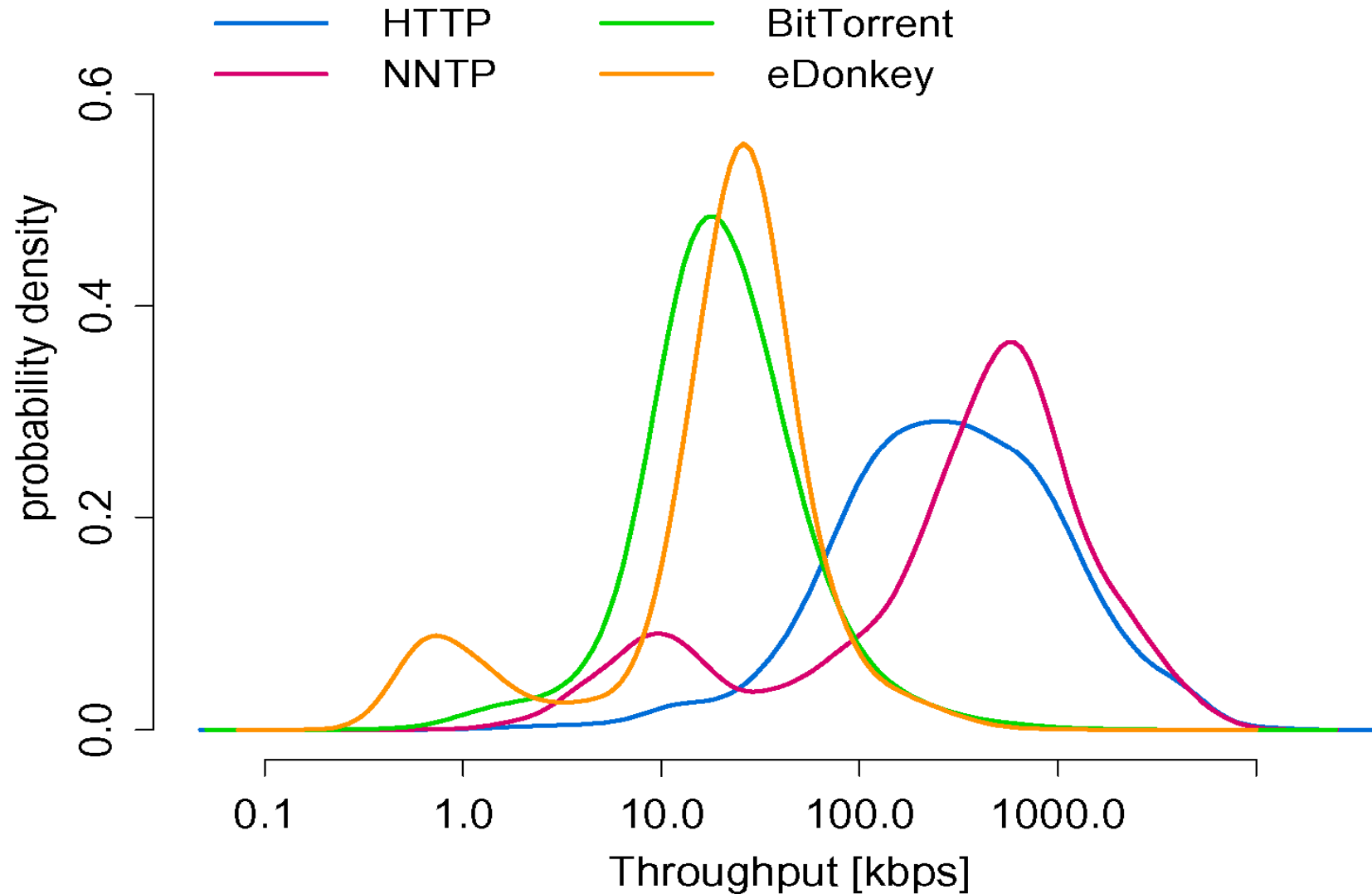


Achieved Throughput

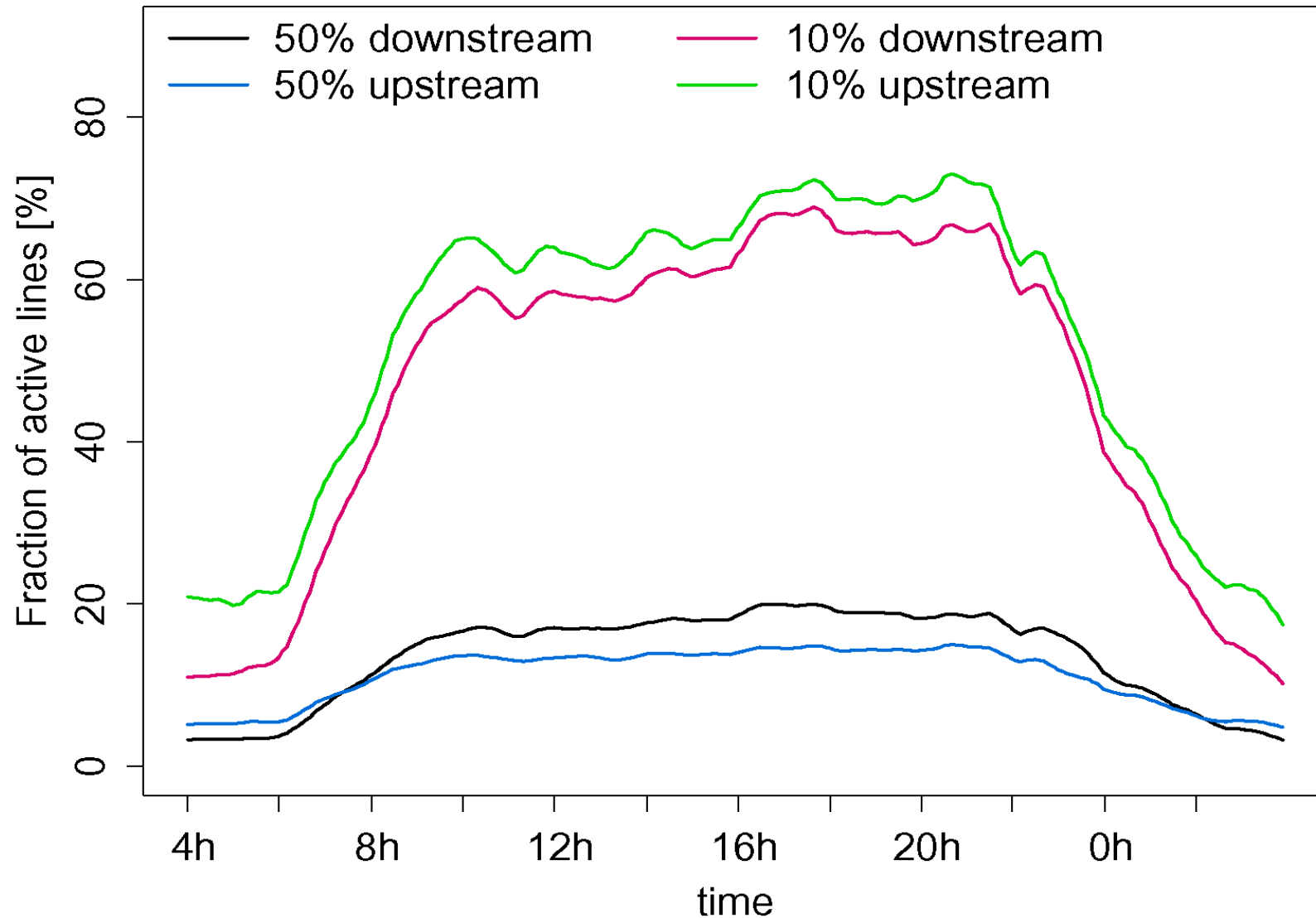
- ❑ Most lines only use small fraction of bandwidth
- ❑ Throughput by application and flow
 - HTTP, NNTP have order of magnitude higher throughput than P2P
- ❑ Mean number of parallel flows
 - P2P has 5 times as many as HTTP
 - P2P and NNTP similar



Achieved flow throughput



Link Utilization (5 min)





Outline

- Data sets
- Dominant characteristics
 - DSL session characteristics
 - Application usage
 - HTTP usage
 - Performance/path characteristics
- Summary



Summary

- ❑ High IP address churn (4% assigned > 10 times)
- ❑ HTTP dominates traffic: >57%
 - P2P only 14%
 - NNTP noticeable
- ❑ Flash-video (video portals) most popular in HTTP: >25%
 - RAR-archives (One-click-hosters): >14%
- ❑ Performance
 - DSL bandwidth in general not fully utilized
 - Selective ACK widespread
 - Window advertisements might limit performance
 - Local RTT component dominates (DSL interleaving)