Traffic Profiles generated by Peer-to-Peer (P2P) Networking

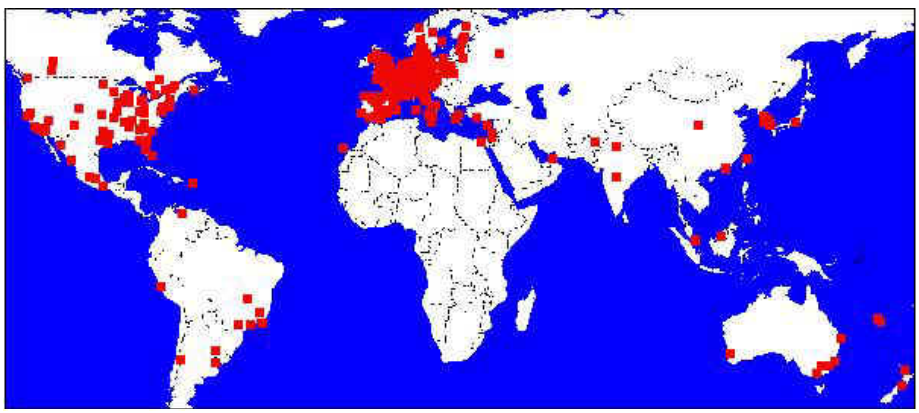
Gerhard Haßlinger, T-Systems TZ ENPS, Deutsche-Telekom-Allee 7,
D-64295 Darmstadt; E-Mail: gerhard.hasslinger@t-systems.com

- P2P Applications and Effects on Broadband IP Access
- Smoothing of Traffic Variability due to P2P Protocols
- Traffic Paths in P2P Overlays
- Is Caching an Option for P2P ?
- Conclusions on P2P from the ISP's View

Efficiency of Peer-to-Peer Solutions

- P2P is an efficient communication scheme to distribute large data volumes avoiding bottlenecks of servers
- P2P networking makes vacant capacities of each user available: storage, computation & transmission capacity
- Connecting many small systems via P2P often yields higher performance than powerful single server systems (seti@home)
- Networks of global size are established for file sharing (music, video, software) and VoIP; even the most popular protocols seem to be initially developed by small teams or one person

P2P File-Sharing: Fast distribution of large files



Example: Harry Potter III early propagation after 2 hours on May 28th 2004 (Source: *www.itic.ca/DIC/News/archive.html*)
Availability & replication of data as needed, driven by demands

P2P Protocol Functions for File Sharing

- **Search- & Download Phase:** Data search via distributed schemes or servers, super nodes; downloads by pure P2P data exchange
- P2P establishes **Application Layer Routing Functions**
  Routing message overhead and possible cross layer inefficiency
  Choice of sources for download independent of IP network structure
- **Segmentation:** Small data chunks improve performance & reliability
  - Multi source transfers in parallel (BitTorrent: ~ 5; eDonkey: many) with a TCP connection per data chunk; flexible up-/download rates
  - Interrupted downloads can recover without repeating correctly received chunks; identification based on unique hash values
- **Tit-for-tat:** Parallel up-/download enforced, as soon as a node has data chunks available; upstream bottleneck for symmetric P2P traffic

Emerging P2P services and prerequisites

- Online Gaming and e-Learning Applications
  - Software downloads via P2P e.g. with BitTorrent build in
  - P2P support of online activity (VoIP, video channel, real time data)
- Content distribution: More frequent and larger software updates
- P2P solutions for small communities; cooperative work tools
- Extension to wireless and mobile environment
- Prerequisites:
  - Many security aspects are still unsolved in popular P2P networks
  - Appropriate payment schemes for converging telecommunication & broadcast TV market to be introduced especially for P2P

Peer-to-Peer traffic in ISP networks

- Since the millennium Internet traffic growth is driven by
  - deployment of residential broadband access (DSL) and
  - peer-to-peer applications (mainly file-sharing)
- P2P traffic generates a smooth background load:
  Transfers of large GByte files often over hours or days
  ⇒ Smoothing effect on short term & daily traffic profiles
  and on traffic matrices due to uniform distribution of data
- P2P traffic characteristics is favourable
  but makes increase in traffic volume difficult to predict
  ⇒ High risk for planning & upgrade processes in IP networks

Measurement of P2P Traffic and Activity

- A decreasing part of P2P traffic can be identified on network and transport layer (P2P standard ports → HTTP, dynamic)
- Application layer & flow based analysis is more precise; approaches for popular protocols are available e.g. via Linux firewall and by manufacturers of measurement equipment
- To be updated for each new or modified P2P protocol; Unique & complete online classification is expensive and depends on protocol mechanisms to disguise themselves
- Analysis of P2P network structures is possible by participating and monitoring/crawling of content, nodes and transfer activity

P2P share of Internet traffic measured in Europe

Application Mix due to TCP Port Measurement	Deutsche Telekom 2004[1]	France Telecom 2003[2]	France Telecom 2004[3]	CacheLogic at a European Tier-1 Provider 2004[4]
eDonkey	60 %	38 %	~ 54.5%	~ 20 %
FastTrack	6 %	8 %	~ 1%	~ 10 %
BitTorrent	?	?	~ 3.5%	~ 16 %
Other Peer-to-Peer	4 %	4 %	~ 1%	~ 10 %
All Peer-to-Peer	70 %	50 %	~ 60%	~ 56 %
HTTP	10 %	15 %	?	~ 12 %
Other (non-P2P or unknown)	20 %	35 %	?	~ 32 %

The majority of traffic is often generated by P2P

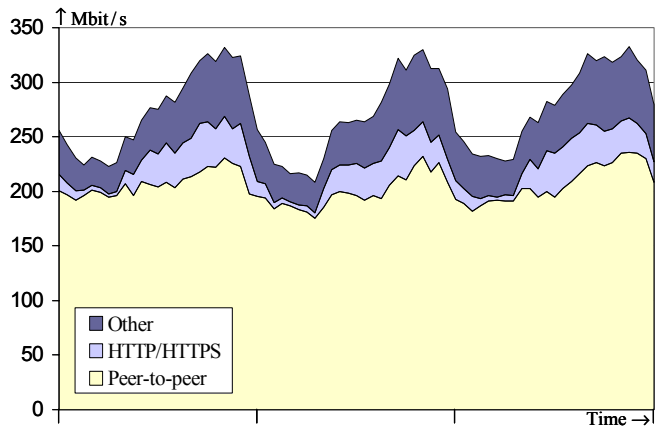
P2P share seems to decrease a bit in last years

Preference for protocols varies (BitTorrent and Gnutella are still increasing)

Sources [1]: **www.eurescom.de/public/projects/P1500-series/p1553 (2005)**
[2]: **N. Azzouna, F. Guillemin: Experimental analysis of P2P applications on traffic in commercial IP networks. Europ. Trans. on Telecom. 15 (2004) 511-522**
[3]: **L. Plissonneau, J. Costeux and P. Brown, Analysis of Peer-to-Peer traffic on ADSL, Passive & active measurement workshop, Boston (2005)**
[4]: **www.cachelogic.com (2005)**

Typical traffic profile in the IP backbone over 3 days

Smoothing effect of P2P on daily traffic profiles due to measurement in the backbone in 2003



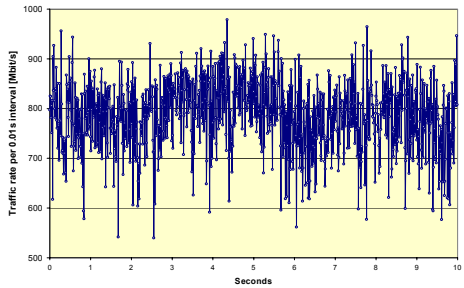
$$\frac{\text{Peak Rate}}{\text{Mean Rate}}:$$

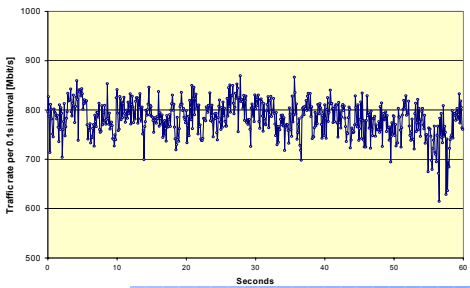
- > 1.5 before 2001
- ~ 1.25 in 2003; is increasing since 2003

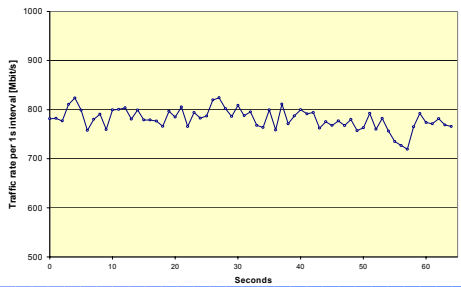
Traffic Measurement in multiple time scales

Evaluation of a traffic traces in 0.01s (→), 0.1s (↓) & 1s (↘) intervals on broadband access networks (2.5 Gbit/s link)

Variability is decreasing on larger time scales





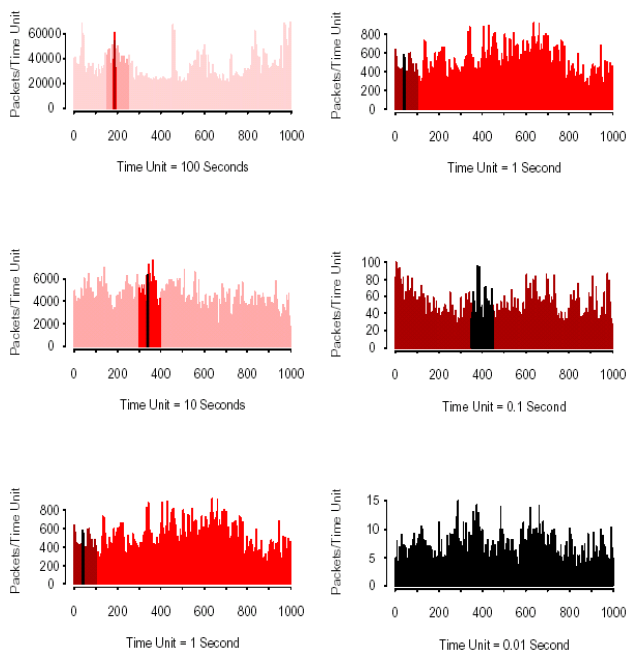


Comparison to Long-range Dependent, Self-similar Traffic

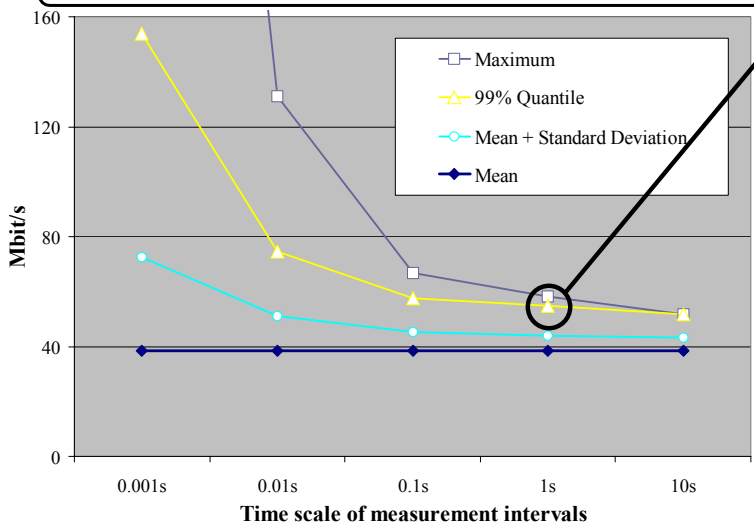
Source:

W. Leland et al.

IEEE/ACM Trans. on Networking (1994)



Variability in different time scales measured for a traffic trace of a 39Mb/s MPLS flow



99% quantile of traffic samples in 1s time scale

proposed by Hans v.d. Berg et al.* for dimensioning;

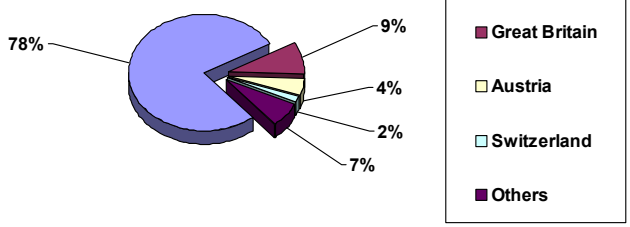
observed to depend on the mean μ by $\mu + k\sqrt{\mu}$ with $k \approx 1$; μ in Mb/s

*** Hans v.d. Berg et al:**
QoS-aware bandwidth provisioning for IP network links, Computer Networks 50/5 (2006) 631-647

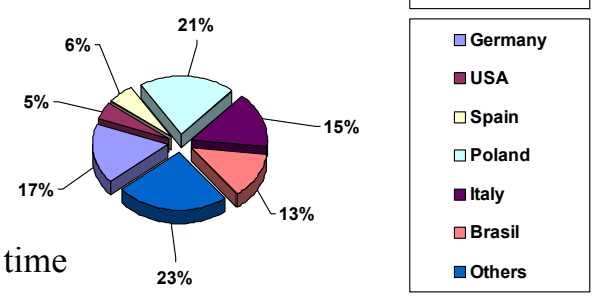
P2P Source Locations: Communities due to language

Distribution of eDonkey sources for download to German destination:

Downloads of German content



Downloads of English content: Seems to depend on server locations which may change over time



Caches for P2P Traffic

- P2P data is suited for caching, but often bypasses usual web caches
  High cache hit rate expected for file sharing; no out of date problems
- Caches can replace the upstream path of P2P transfers
  most relevant for limited bandwidth (mobile, asymmetric access)
- Transparent P2P caching offered by network analysis systems:
  Online detection of P2P downloads for most popular protocols ...
- eDonkey option for transparent caching since October 2004:
  20-25% of the peers make use of the eDonkey caching option
  ~ 8% of data can be downloaded from caches in DT's IP platform
- Does eDonkey option decrease or increase the traffic load ?
- Problems with illegal data in caches persist !

P2P Traffic Profiles
EUROVIEW Würzburg, 1 .Aug. 2006

Conclusions

- P2P generates a relevant share of the Internet traffic since 2000
  P2P is a factor in the penetration of residential broadband access
  P2P makes prediction of traffic growth more difficult
- Smoothing effects of P2P are relevant for the variability of short term and daily profiles and in traffic matrices; Traffic profiles for residential broadband access differ from Ethernet traffic
- More free as well as business services to use P2P networking
- Main advantages: Simple overlay architecture to offer global services at low cost even without own network infrastructure; Better performance than client-server for several purposes

G. Haßlinger
T-Systems