Traffic Patterns in Peer-to-Peer-Networking

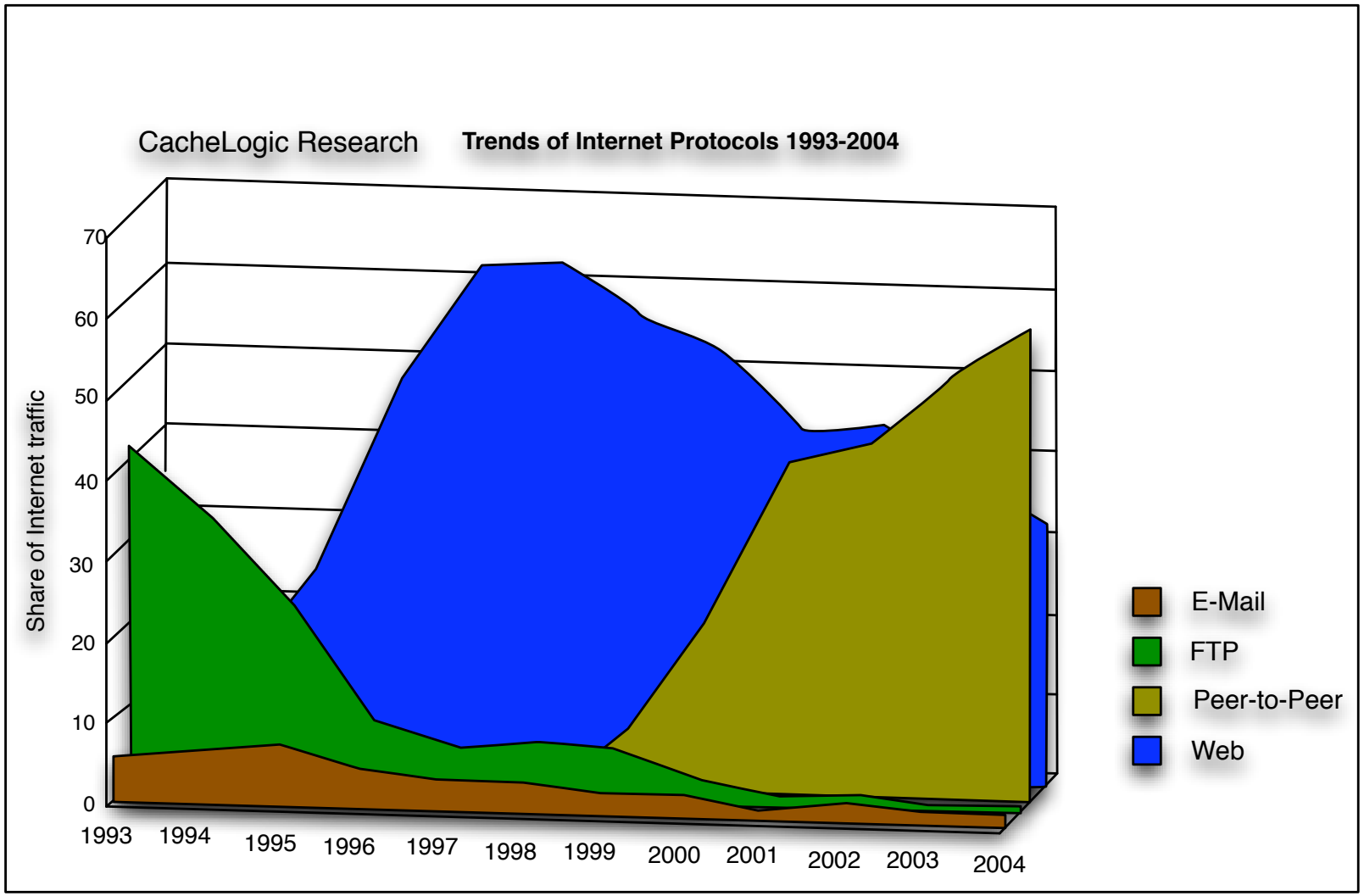
Christian Schindelhauer

joint work with
Amir Alsbih
Thomas Janson

to be presented at ITA

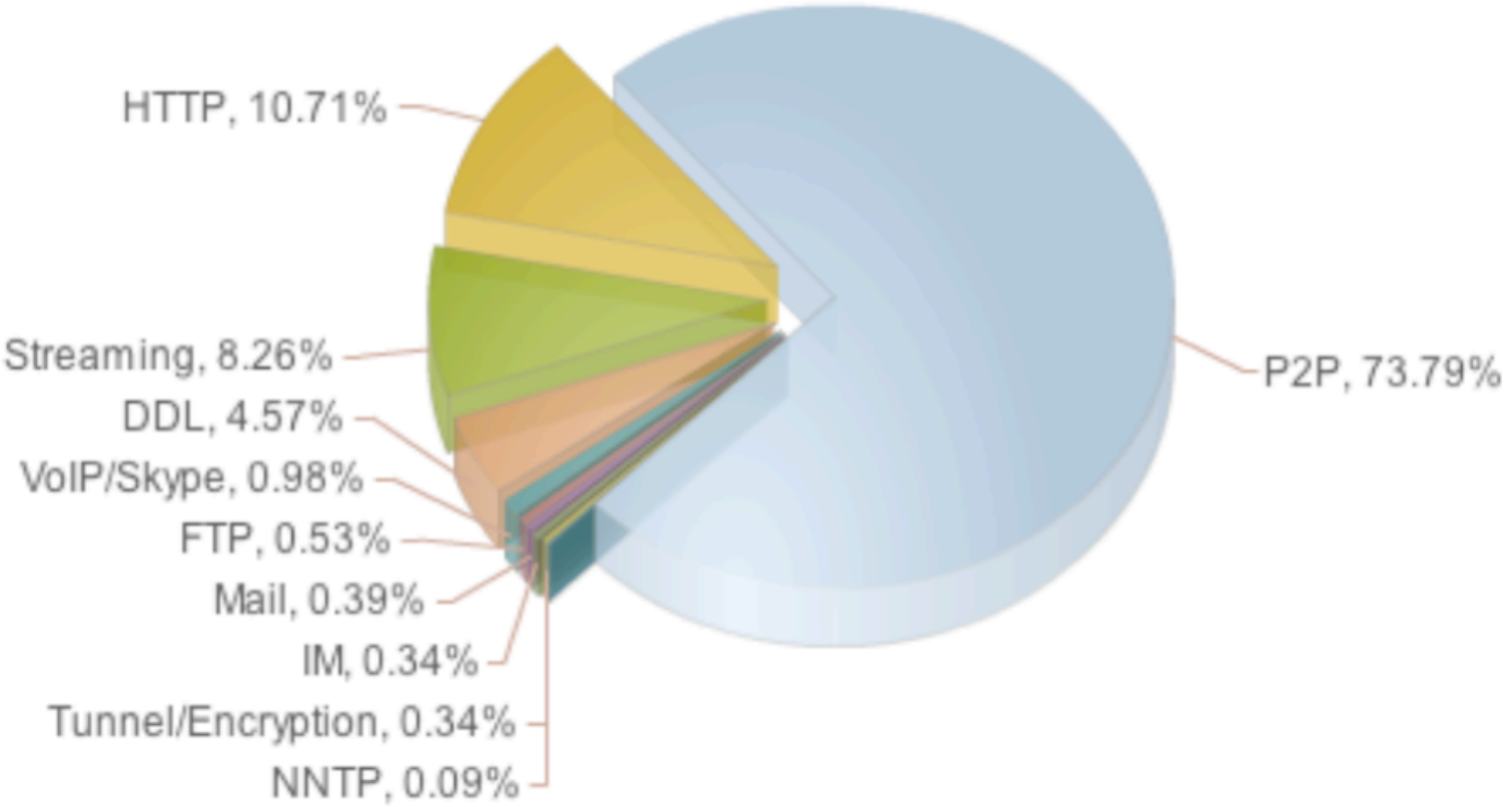
Albert-Ludwig University Freiburg
Department of Computer Science
Computer Networks and Telematics

Global Internet Traffic Shares 1993-2004



CoNe Freiburg

P2P Share Germany 2007

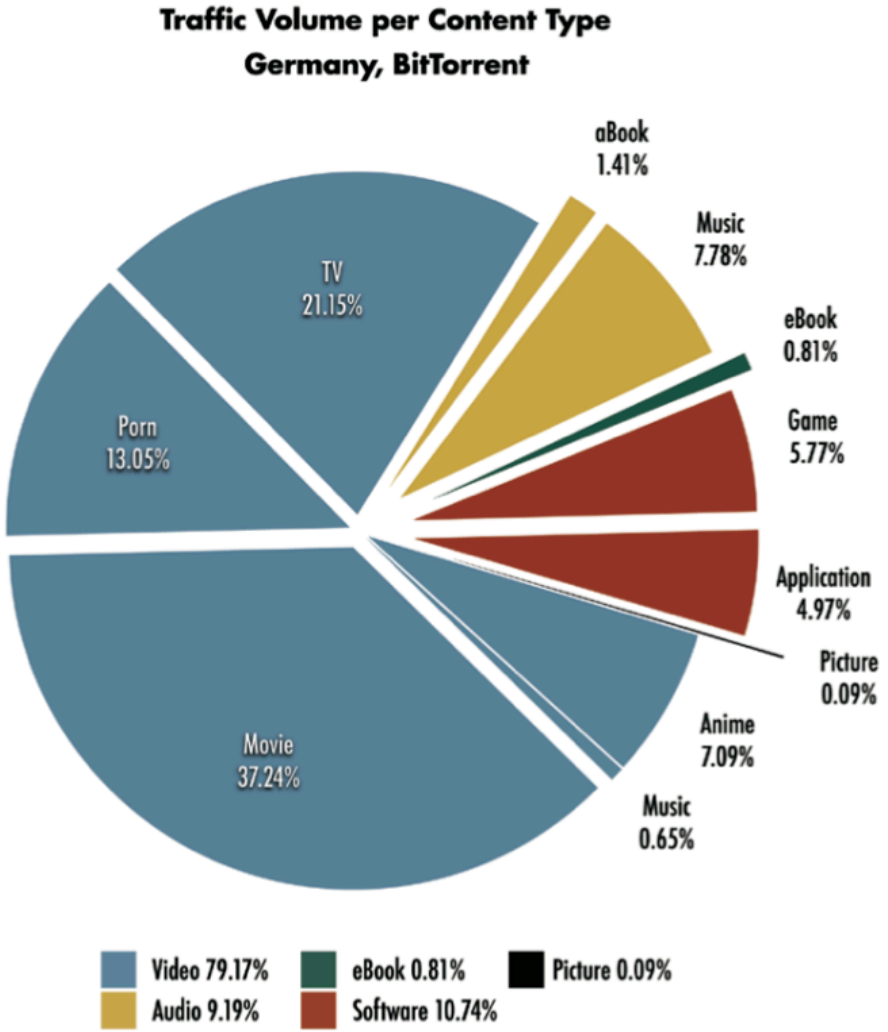


Quelle: Ipoque 2007

UNI FREIBURG

CoNe Freiburg

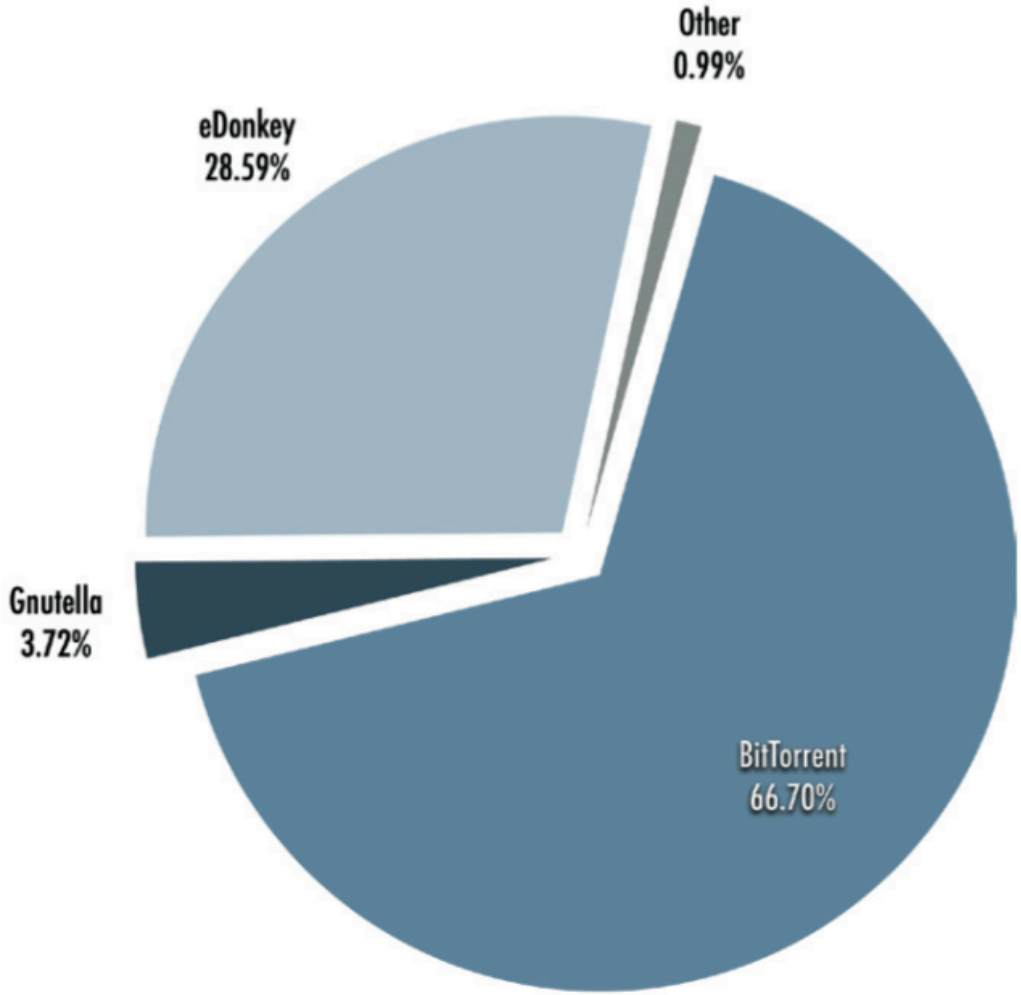
What Germans Download 2007 by Volume



Quelle: Ipoque 2007

CoNe Freiburg

P2P Systems Germany 2007 by Volume



Quelle: Ipoque 2007

CoNe Freiburg

Motivation

- Peer-to-Peer-Networks
  - Quality depends on user behavior
  - High churn rates
  - Egoistic users
- Only a small number of independent studies of Internet traffic
- We analyze the complete traffic of 20,000 users in August 2009 of a German digital cable TV based Internet provider.
  - Traffic was centrally monitored
  - Type classification by deep packet inspection
  - Looked at BitTorrent traffic

UNI FREIBURG

CoNe Freiburg

Network Monitoring

- Network monitoring systems
  - installed in recent years
  - allow to monitor the behavior of each user.
- Motivation
  - new governmental regulations
  - detection and prevention of
    - Internet fraud
    - denial-of-service attacks
    - spam mailers
    - phishing attacks,
    - criminal conspiracies,
    - forbidden contents
    - copyright violations.
- ISPs are (usually) not the juridical target
  - are required to uphold an infrastructure, which allows law enforcement to take action in such cases.

UNI FREIBURG

Background of the Study

- Our study
  - limited access to anonymized user data
  - gathered by a network monitoring systems using deep packet inspection (DPI)
- Main product of „our“ ISP: digital cable TV
  - thousands of German households
  - byproduct they also offer telephone and Internet service
- German households are connected via DSL
  - rural area the bandwidth is rather low
  - urban areas high data rates
- Mobile phone carriers providing GPRS, EDGE and HSDPA gain in traffic.
- Digital TV cable is a stable market
  - Installation of the necessary infrastructure is expensive
  - Television is still important media of Germans
  - No open market for digital cable TV.
  - Cable TV users extend their contracts to include Internet service because of low prices and high bandwidth rates.

CoNe Freiburg

Internet over Digital TV Cable

- Each user needs a digital cable modem
  - encodes and decodes the data traffic
- Throughput rates range from 32-100 MBit/s download
  - DSL traffic: 2 to 16 kBit/s
  - HSDPA ending at 7.2 MBit/s
- No network bottleneck
  - measured traffic behavior directly reflects the users wishes.
- Ideal opportunity
  - What do Internet users want?
  - How long are users online?
  - How much data do users download or upload?
  - What are the network services they use?

UNI FREIBURG

CoNe Freiburg

BitTorrent and „Friends“

- BitTorrent
  - most successful peer-to-peer network protocol
  - BitTorrent encourages to upload data using incentives
- Several BitTorrent clients deviate from the original protocol
  - BitTyrant
    - achieves a download gain up to 70 percent
    - strategic selection of peers in the swarm
  - BitThief
    - free riding client
    - allows downloading without any upload
    - achieve higher download rates than the official client.

UNI FREIBURG

CoNe Freiburg

Bittorrent

- Bram Cohen
- Bittorrent is a real (very successful) peer-to-peer network
  - concentrates on download
  - uses (implicitly) multicast trees for the distribution of the parts of a file
- Protocol is peer oriented and not data oriented
- Goals
  - efficient download of a file using the uploads of all participating peers
  - efficient usage of upload
    - usually upload is the bottleneck
    - e.g. asymmetric protocols like ISDN or DSL
  - fairness among peers
    - seeders against leeches
  - usage of several sources

UNI FREIBURG

CoNe Freiburg

Bittorrent: Coordination

- Central coordination
  - by tracker host
  - for each file the tracker outputs a set of random peers from the set of participating peers
    - in addition hash-code of the file contents and other control information
  - tracker hosts to not store files
    - yet, providing a tracker file on a tracker host can have legal consequences
  - Is often replaced with a decentralized peer-to-peer network
- File
  - is partitions in smaller pieces
    - as describec in tracker file
  - every participating peer can redistribute downloaded parts as soon as he received it
  - Bittorrent aims at the Split-Stream idea
- Interaction between the peers
  - two peers exchange their information about existing parts
  - according to the policy of Bittorrent outstanding parts are transmitted to the other peer

UNI FREIBURG

CoNe Freiburg

Bittorrent
Part Selection

- Problem
  - The Coupon-Collector-Problem is the reason for a uneven distribution of parts
    - if a completely random choice is used
- Measures
  - Rarest First
    - Every peer tries to download the parts which are rarest
      - density is deduced from the comunication with other peers (or tracker host)
    - in case the source is not available this increases the chances the peers can complete the download
  - Random First (exception for new peers)
    - When peer starts it asks for a random part
    - Then the demand for seldom peers is reduced
      - especially when peers only shortly join
  - Endgame Mode
    - if nearly all parts have been loaded the downloading peers asks more connected peers for the missing parts
    - then a slow peer can not stall the last download

UNI FREIBURG

CoNe Freiburg

Bittorrent Policy

- Goal
  - self organizing system
  - good (uploading, seeding) peers are rewarded
  - bad (downloading, leeching) peers are penalized
- Reward
  - good download speed
  - un-choking
- Penalty
  - Choking of the bandwidth
- Evaluation
  - Every peers Peers evaluates his environment from his past experiences

UNI FREIBURG

CoNe Freiburg

Bittorrent Choking

- Every peer has a choke list
  - requests of choked peers are not served for some time
  - peers can be unchoked after some time
- Adding to the choke list
  - Each peer has a fixed minimum amount of choked peers (e.g. 4)
  - Peers with the worst upload are added to the choke list
    - and replace better peers
- Optimistic Unchoking
  - Arbitrarily a candidate is removed from the list of choking candidates
    - the prevents maltreating a peer with a bad bandwidth

UNI FREIBURG

CoNe Freiburg

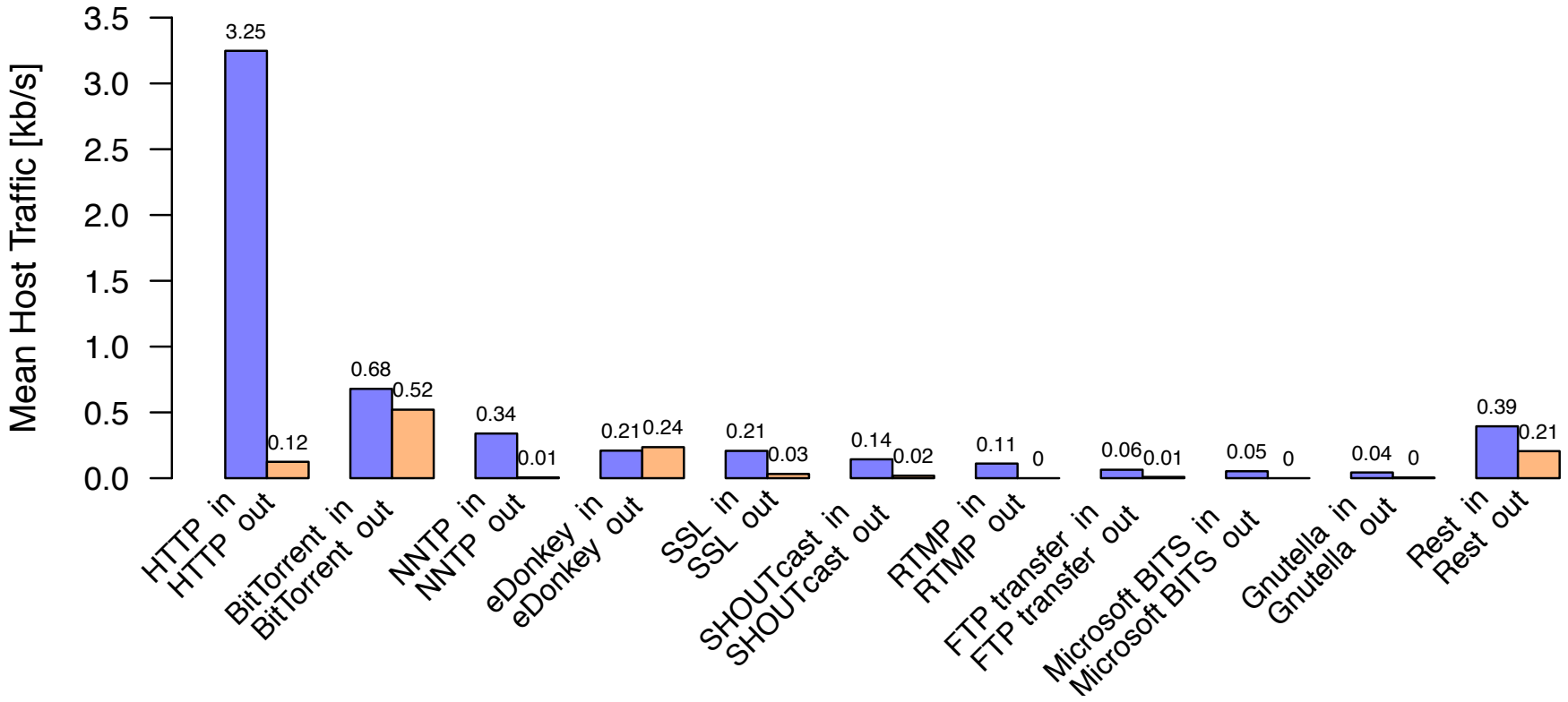
Deep Packet Inspection

- Internet Service Provider
  - deep packet inspection system for analyzing the type of traffic
- Using heuristics
  - Analyze the first few packets to identify a protocol
    - Assumption further data exchange over the connection (IP socket) belongs to the same protocol.
  - Only protocol headers of the first few packet are inspected
  - Applications without encryption can be identified this way
- Encrypted protocols
  - can only identified by version numbers and other unencrypted information
  - Up to 20 packets have to be inspected
  - User data cannot be processed

UNI FREIBURG

CoNe Freiburg

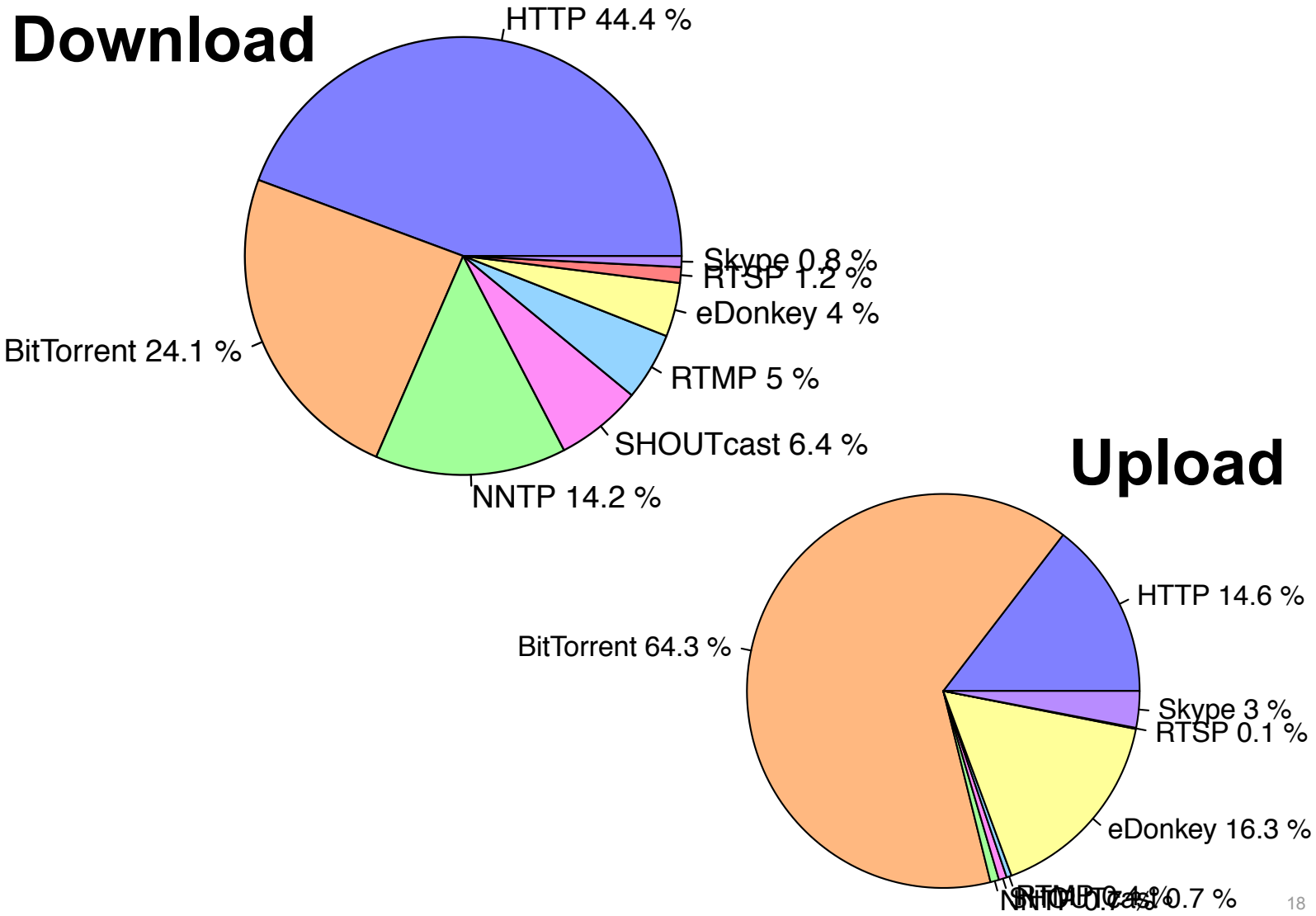
Traffic Types



UNI FREIBURG

CoNe Freiburg

Internet Traffic of a Geman ISP August 2009



CoNe Freiburg

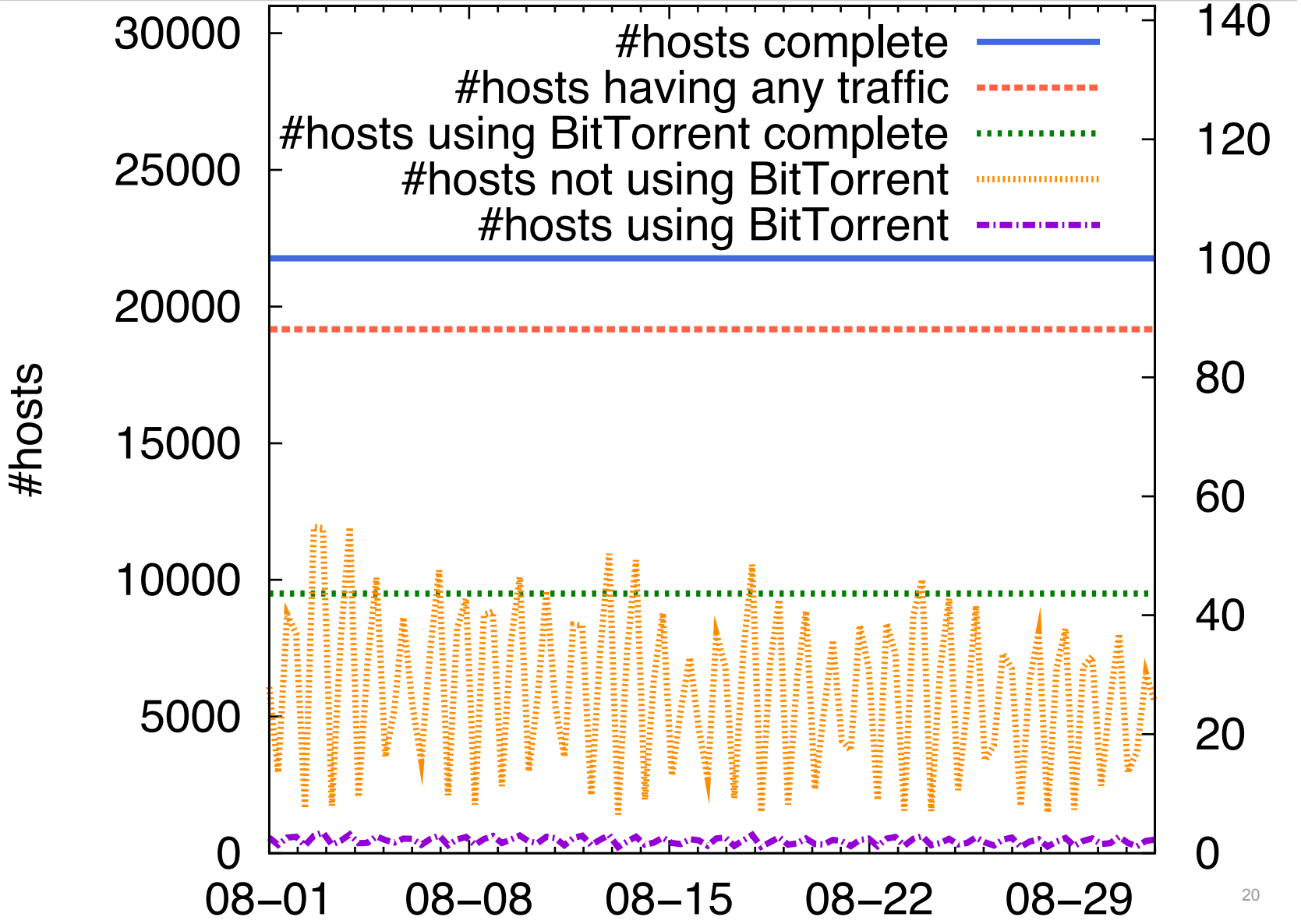
Our Data Source

- After 15 minutes the DPI systems
  - reports the number of incoming and outgoing bytes for each protocol for each user.
  - rollected in log files.
  - We have received the data without IP addresses
    - replaced by anonymized IDs integer
- For each interval of 15 minutes over a month
  - we know for each anonymized user the number of open connections
  - the incoming and outgoing overall traffic
  - the incoming and outgoing unencrypted BitTorrent traffic
  - the sum of HTTP traffic of all users.
- We have received the sum of overall traffic in this month for each host for each service type.

UNI FREIBURG

Overvie of all Traffic in August 2009

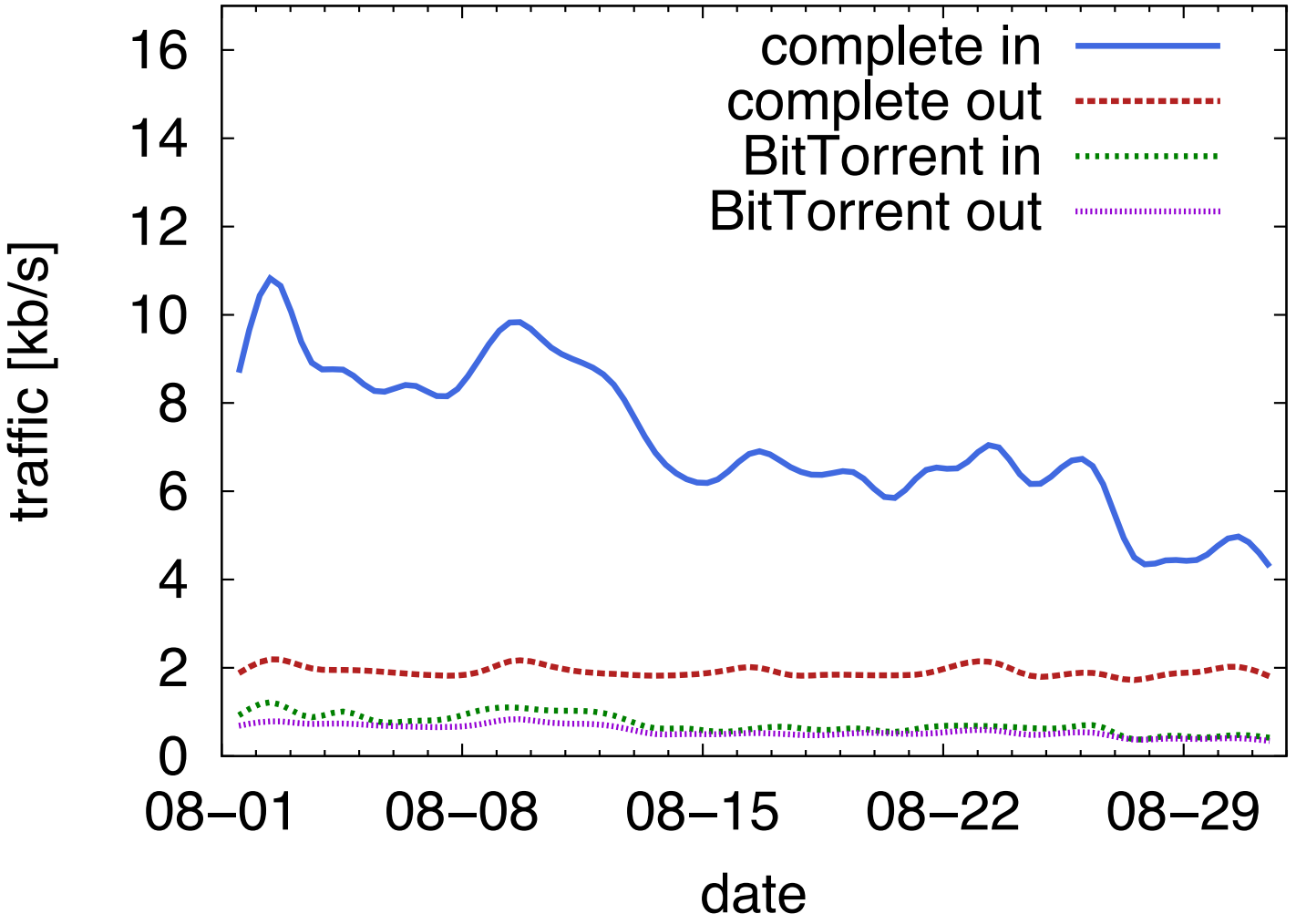
CoNe Freiburg



UNI FREIBURG

CoNe Freiburg

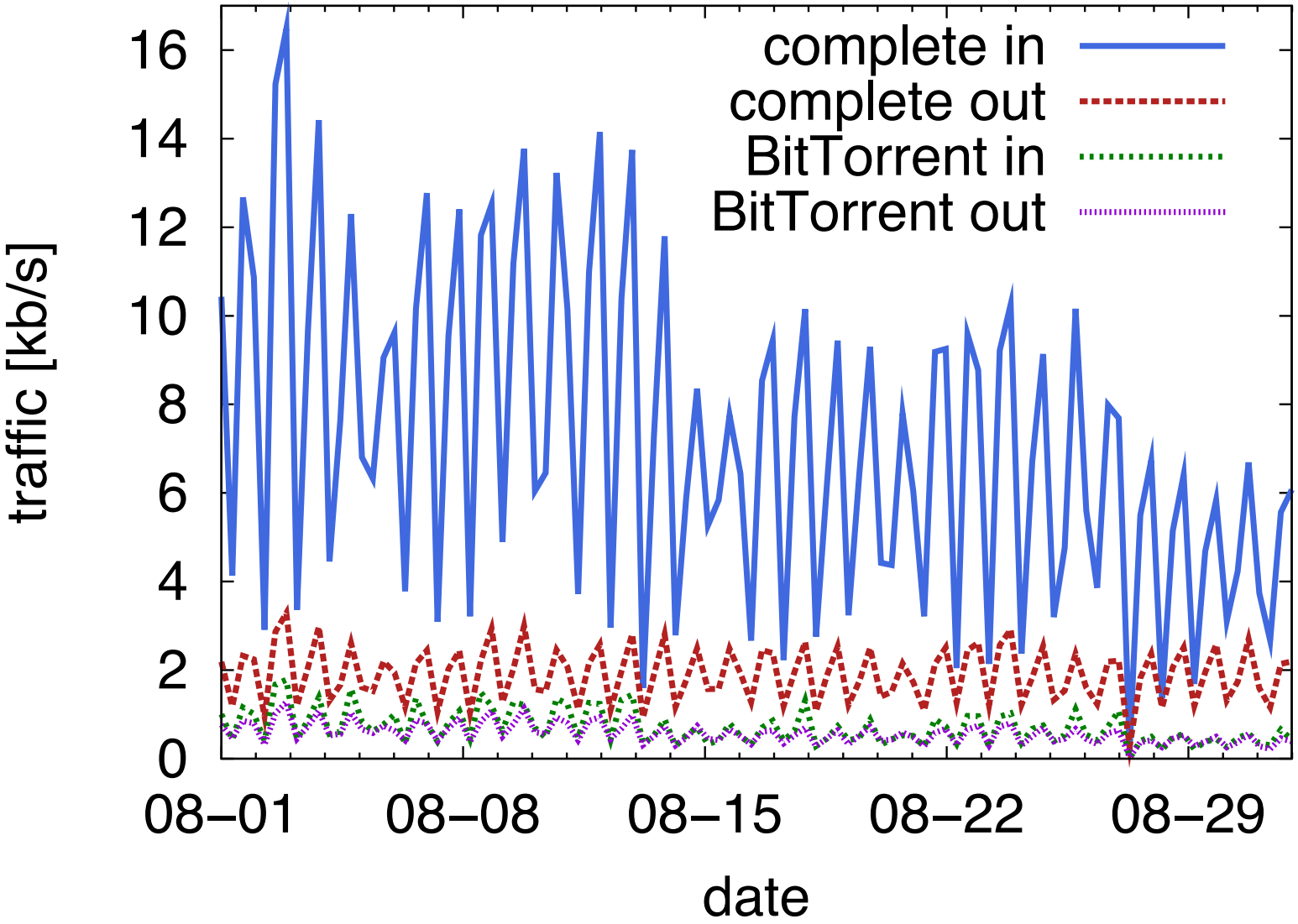
Overall Traffic Devolopent in August averaged over days 2009



UNI FREIBURG

CoNe Freiburg

Measured Traffic



UNI FREIBURG

CoNe Freiburg

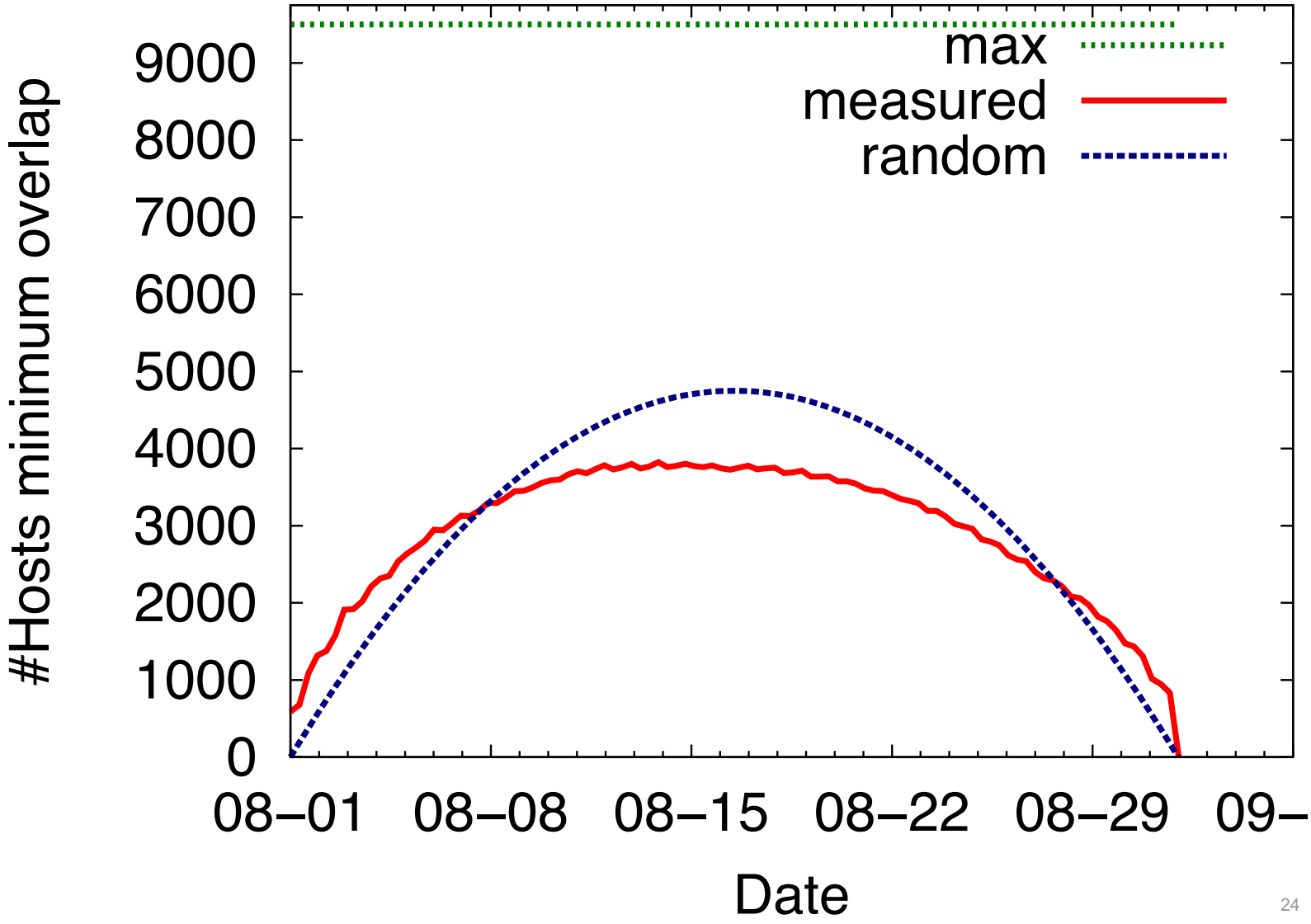
Shortcomings of Data Set

- Identification of each user by the IPv4 address is not completely reliable
- No reconnection every 24 hours
  - unlike other ISPs
  - IPv4 address of a network user remains the same until the modem is rebooted
- Possible reasons for a modem reboot are
  - hardware reset
  - disconnecting of the modem
  - power outage.
- Error types
  - user occurs under several IP addresses
    - leads to an overestimation of users.
  - different user might reuse a free IP address
    - ISP assured us that IPv4 addresses are rarely reused
- Look at the intervals when an IP address is used and count the number of such simultaneous time intervals.
  - This number gives us a lower bound of the number of distinct users

UNI FREIBURG

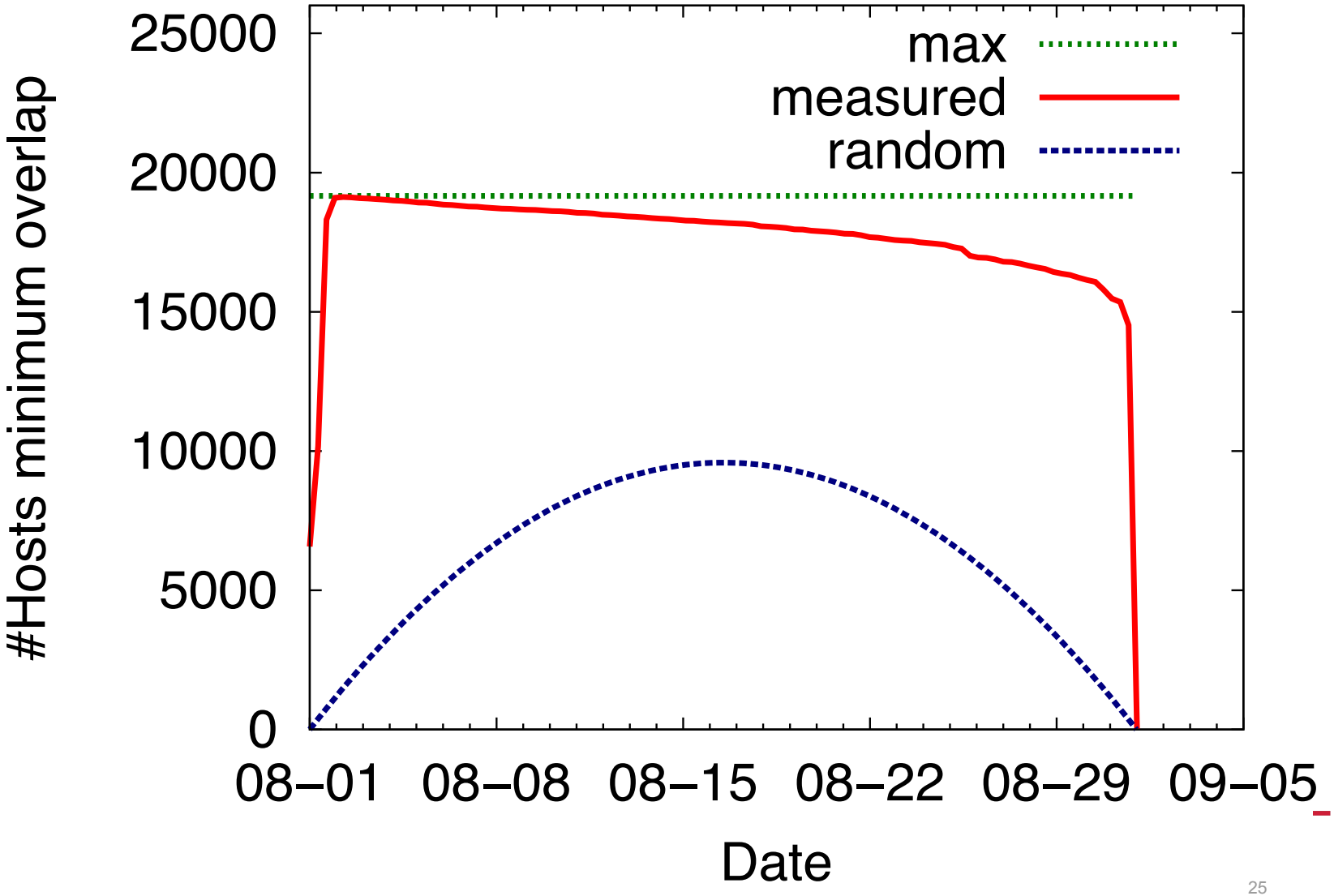
CoNe Freiburg

Overlapping Technique
BitTorrent Traffic only



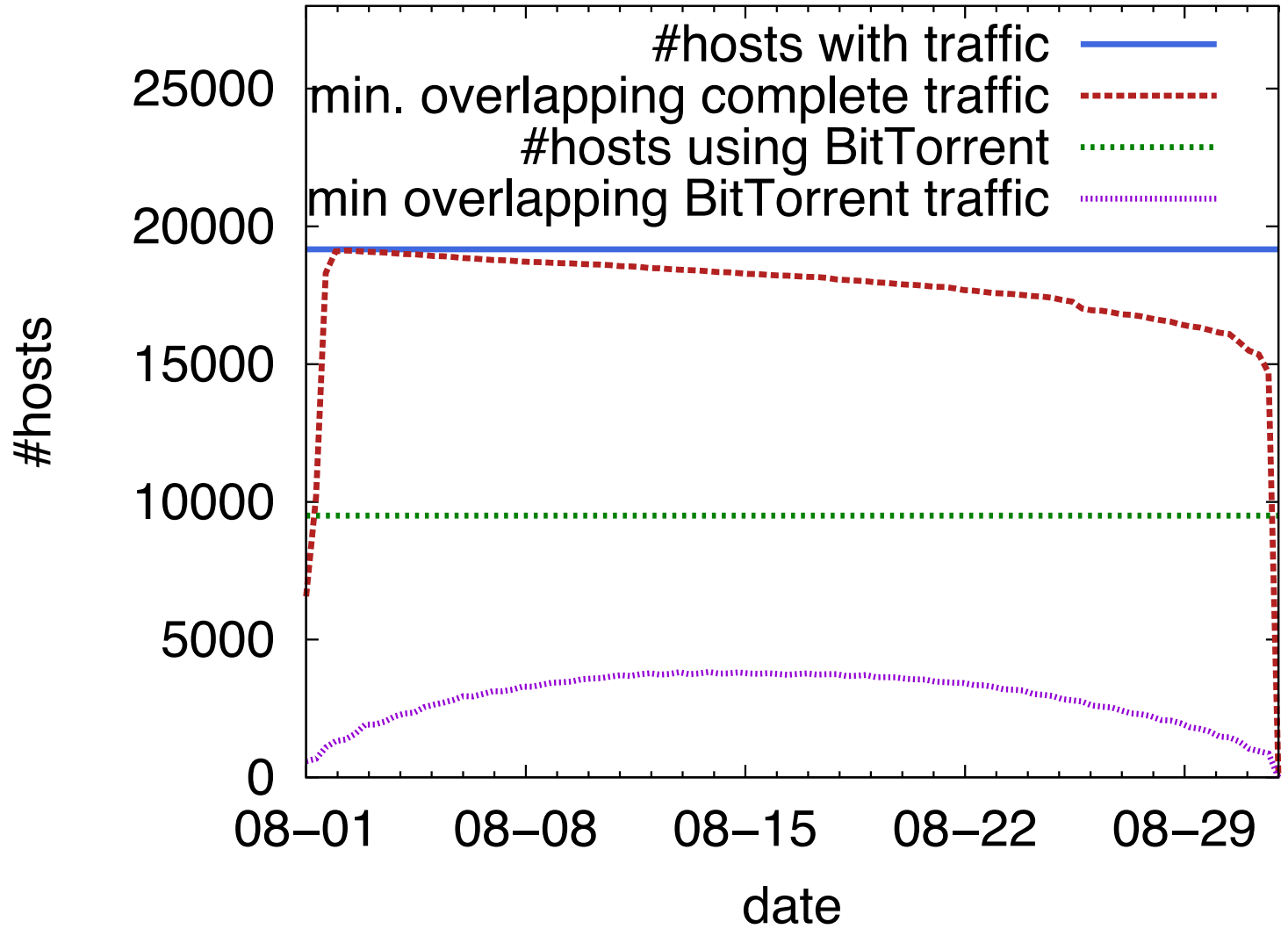
CoNe Freiburg

Overlap with Internet Traffic



CoNe Freiburg

Comparison



UNI FREIBURG

CoNe Freiburg

BitTorrent Traffic

- Scatterplots for Up/Download Traffic
- BitTorrent and other traffic not related
- Remember: correlation coefficient

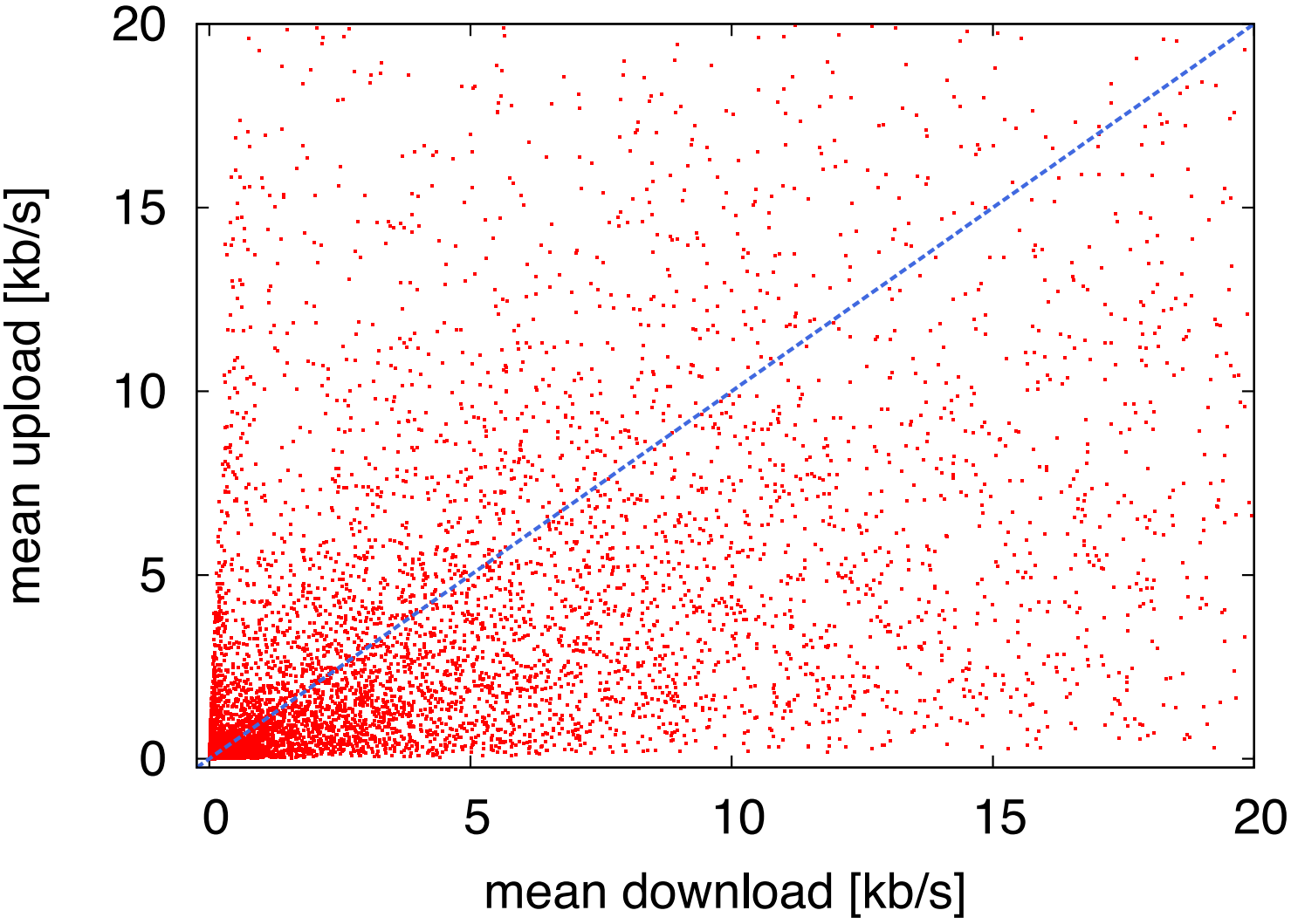
$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

UNI FREIBURG

CoNe Freiburg

Scatterplott - BitTorrent Traffic

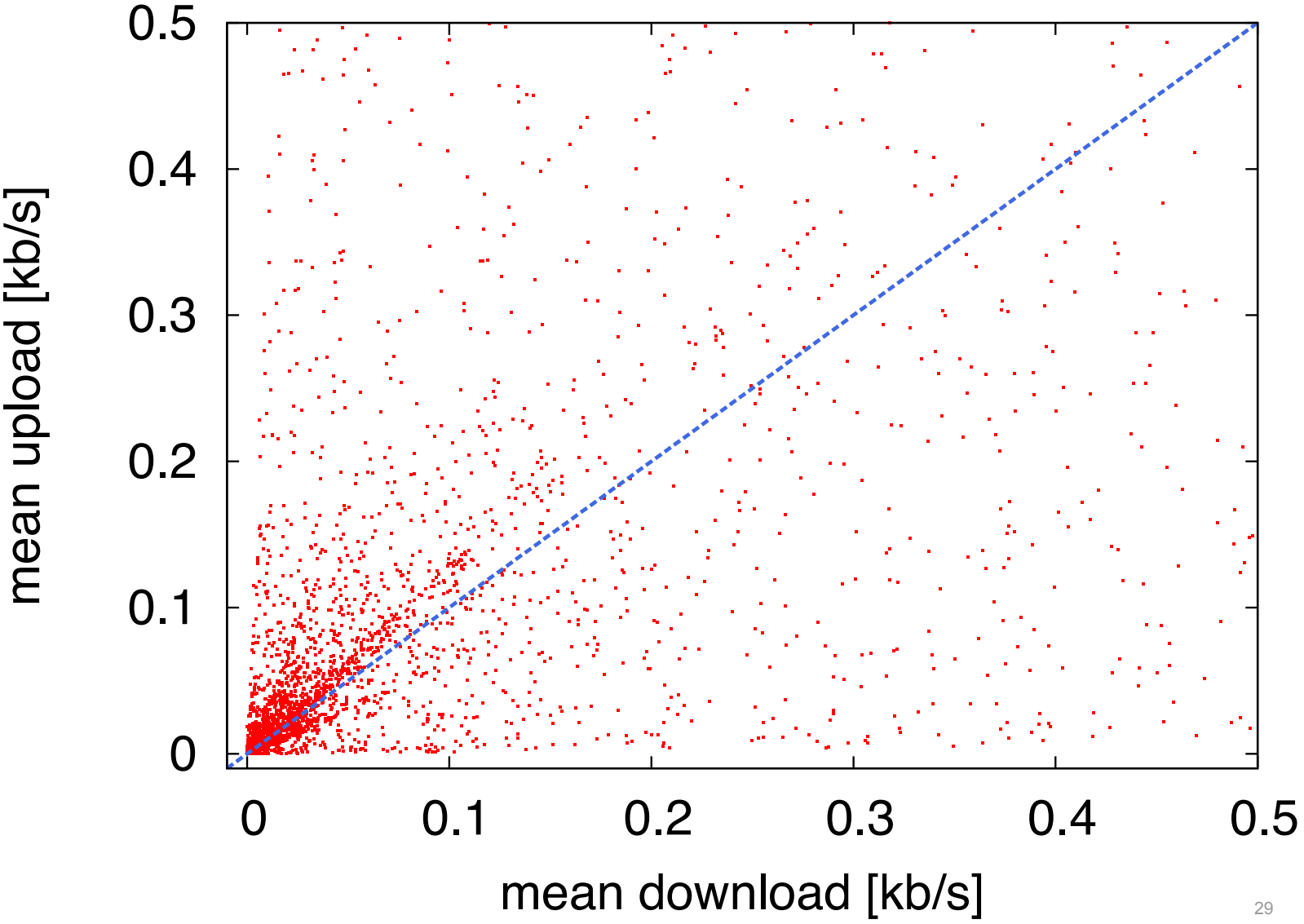
- correlation coefficient: 0.58



UNI FREIBURG

CoNe Freiburg

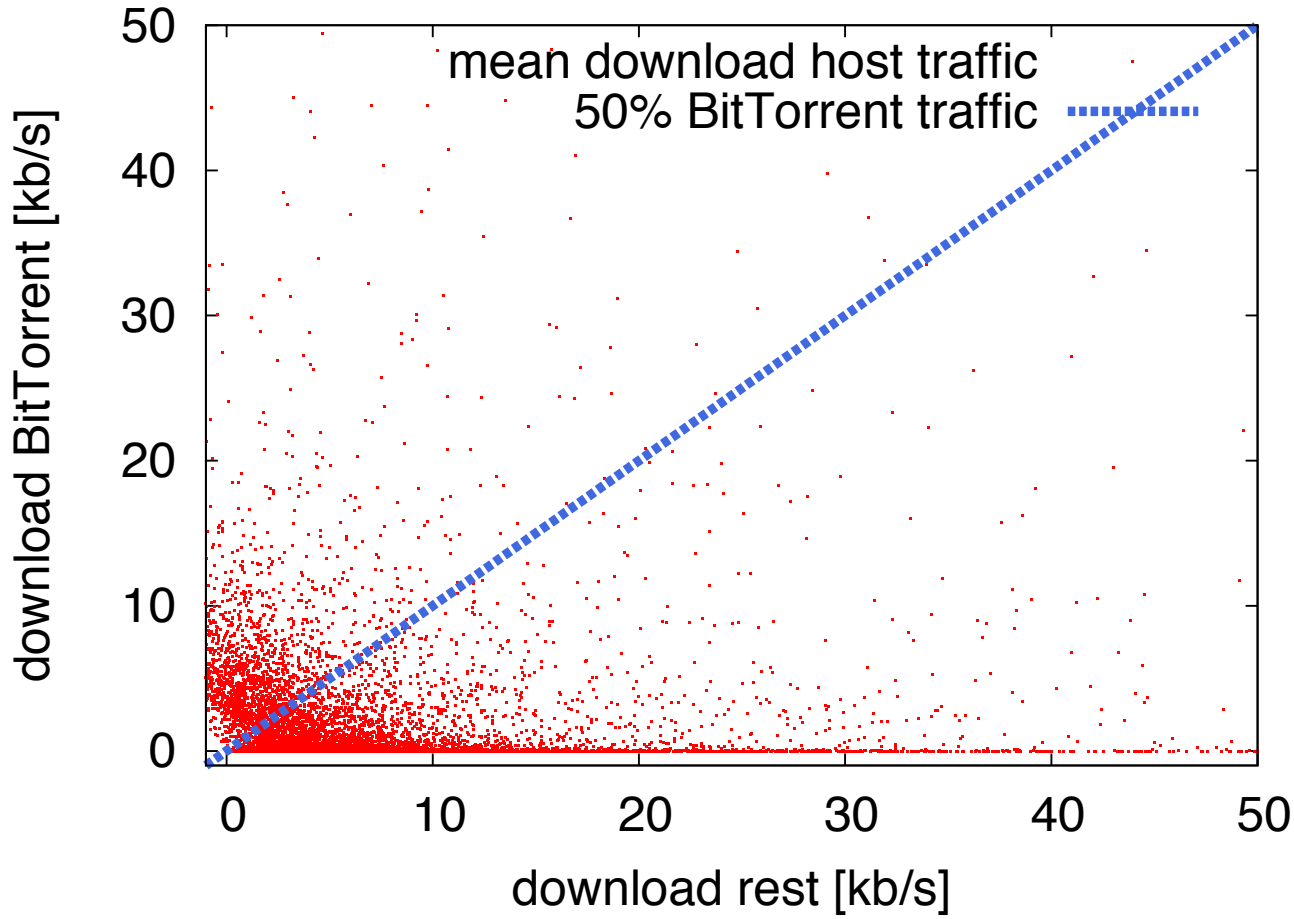
Scatterplott - BitTorrent Traffic (Zoom in)



UNI FREIBURG

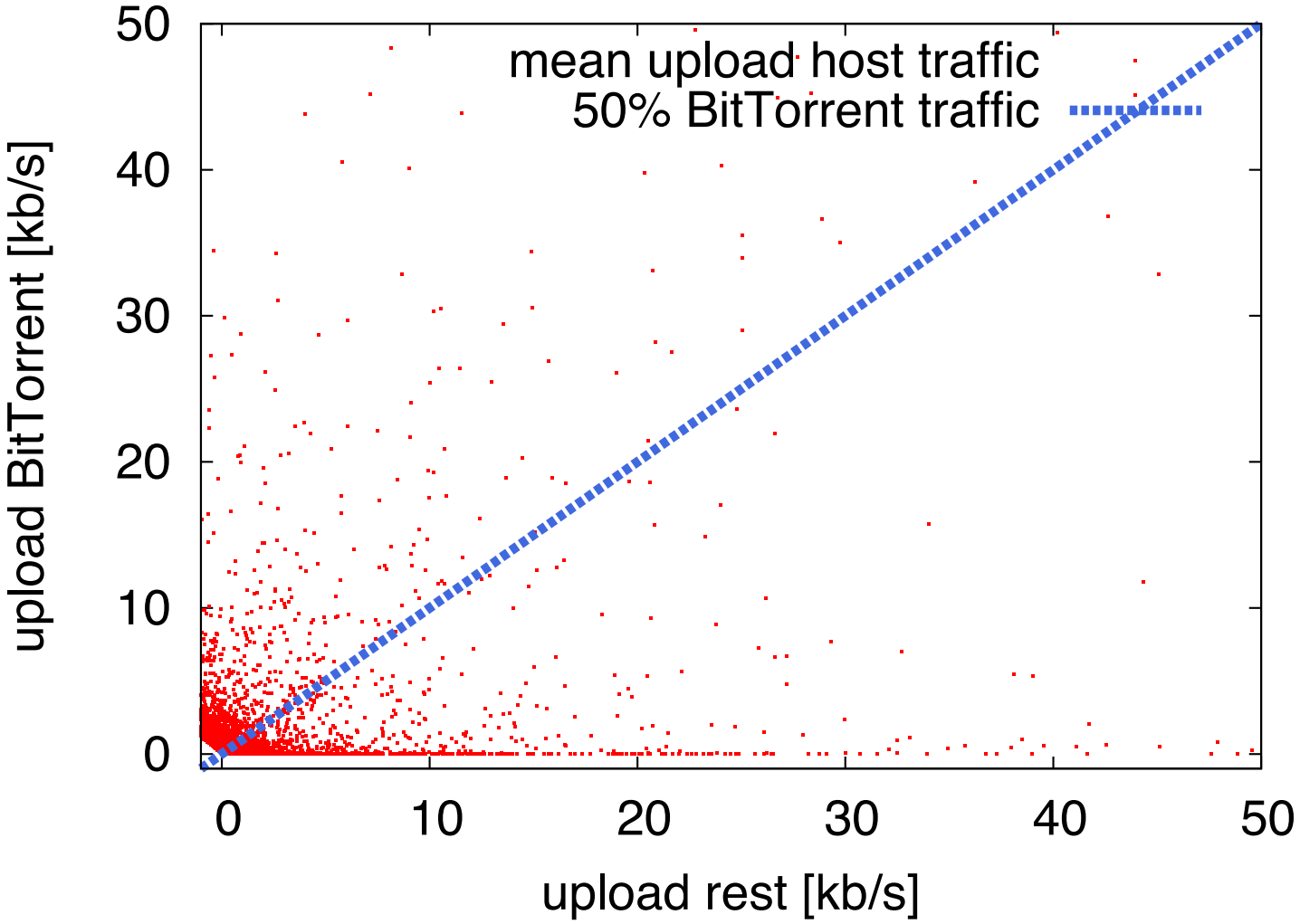
BitTorrent Traffic versus Other Traffic Downstream

- correlation coefficient: −0.38.



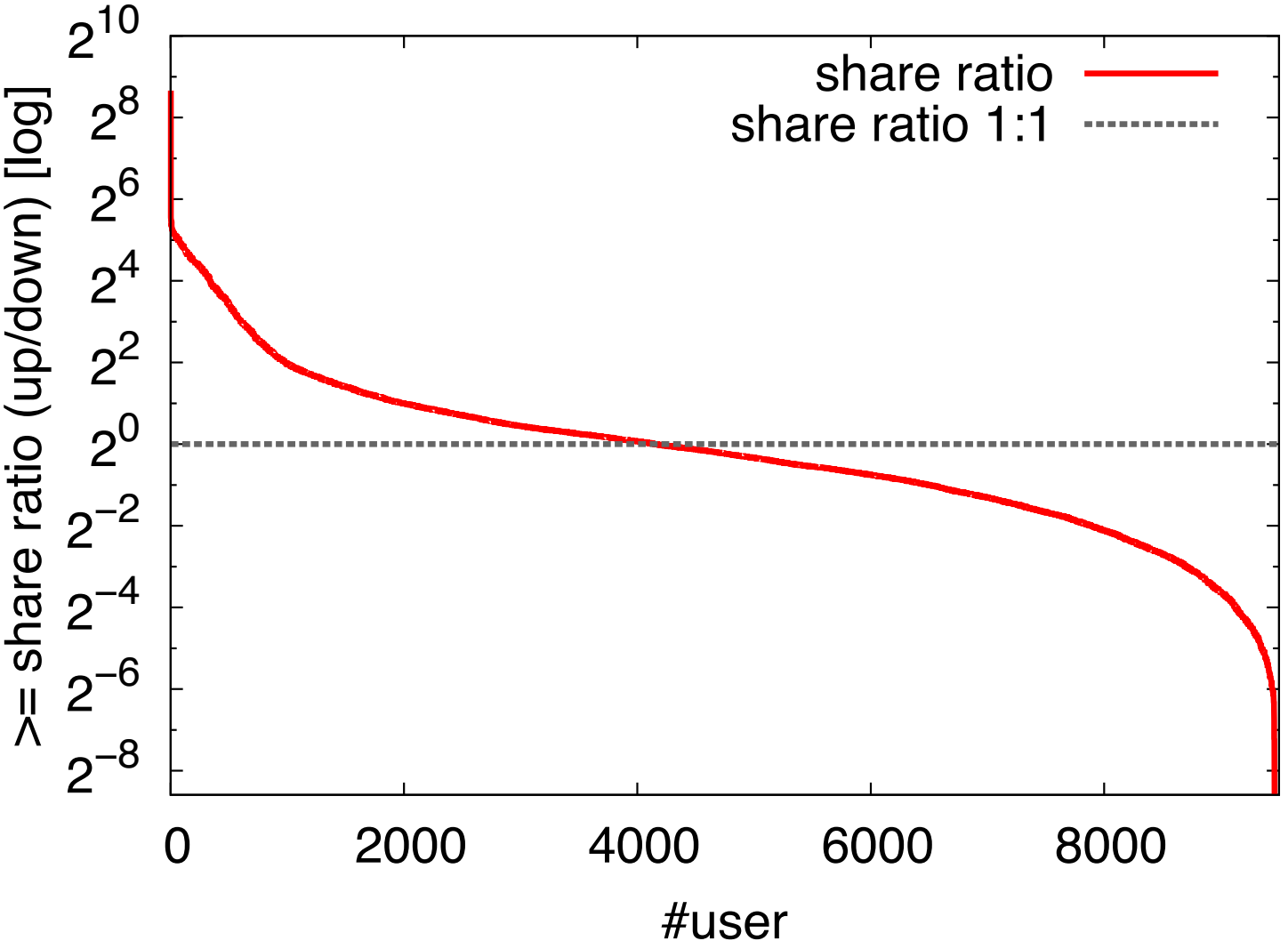
Upload BitTorrent versus Other Traffic Upload

- correlation coefficient −0.53



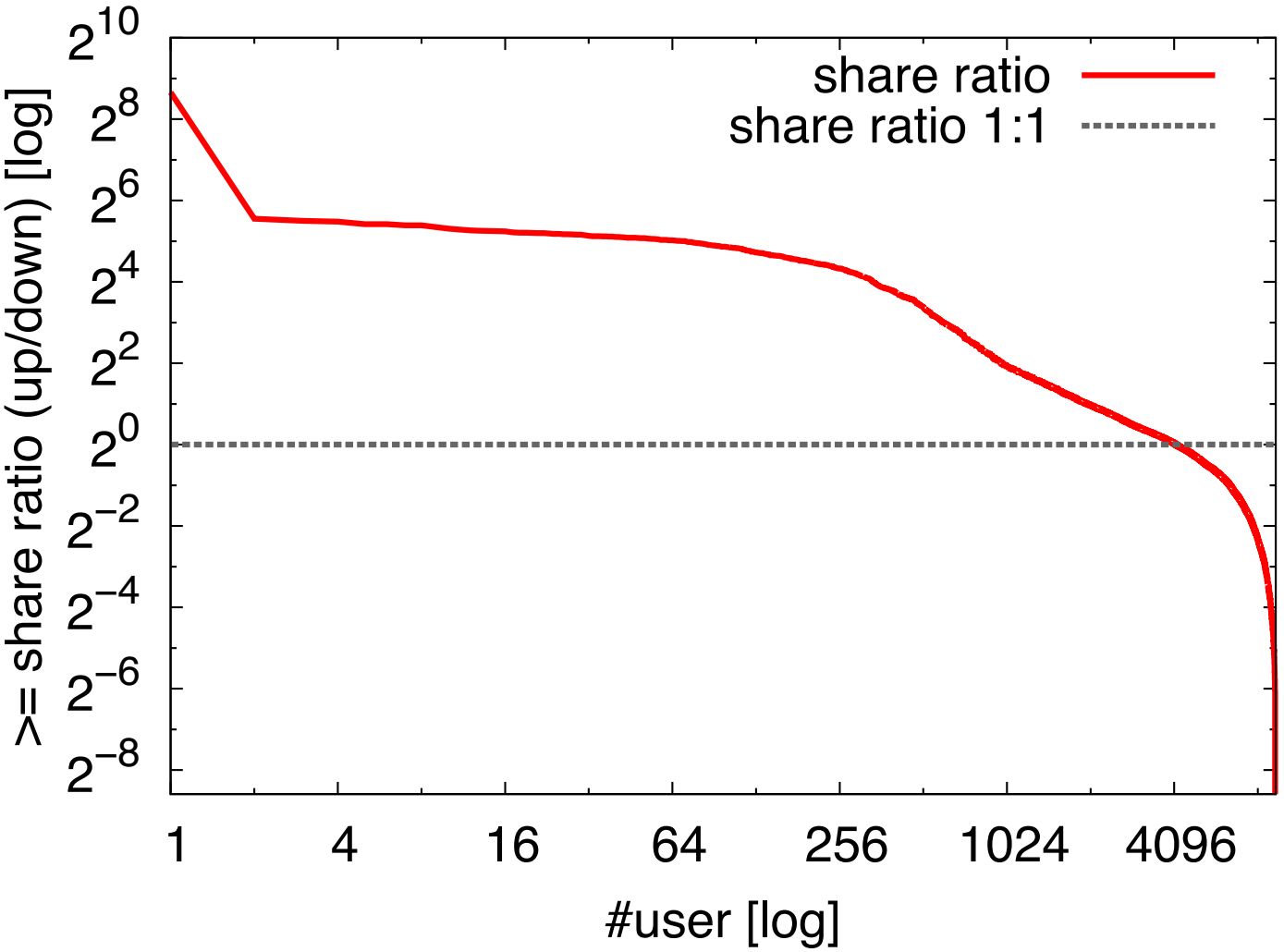
CoNe Freiburg

Upload / Download Distribution



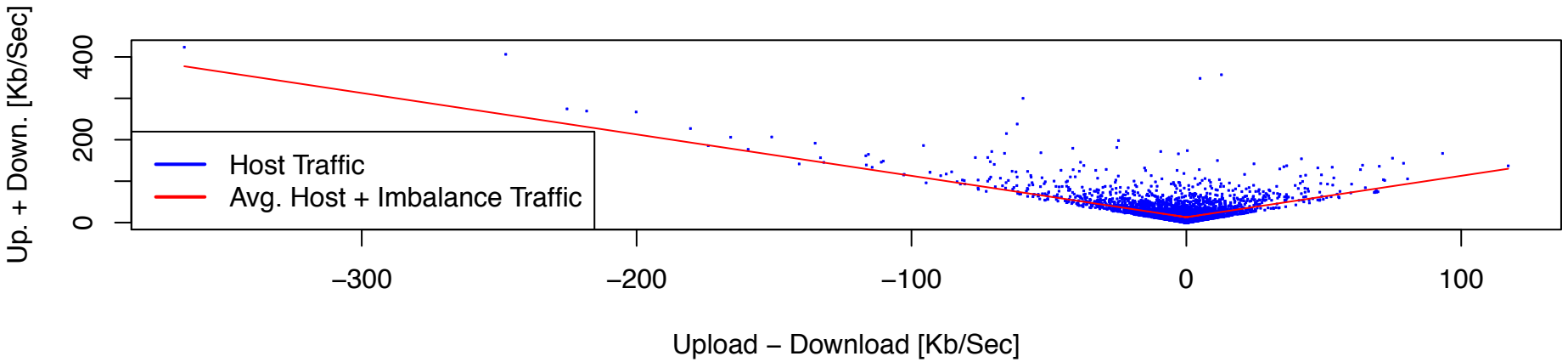
CoNe Freiburg

Upload / Download Distribution (Log-Log-Plot)



UNI FREIBURG

Traffic Difference versus Traffic Sum



CoNe Freiburg

Share Difference of Torrent Traffic

- From scatterplot: no sharp distribution for BitTorrent traffic.
- Difference of download and upload traffic is a piecewise power law (Pareto) distribution

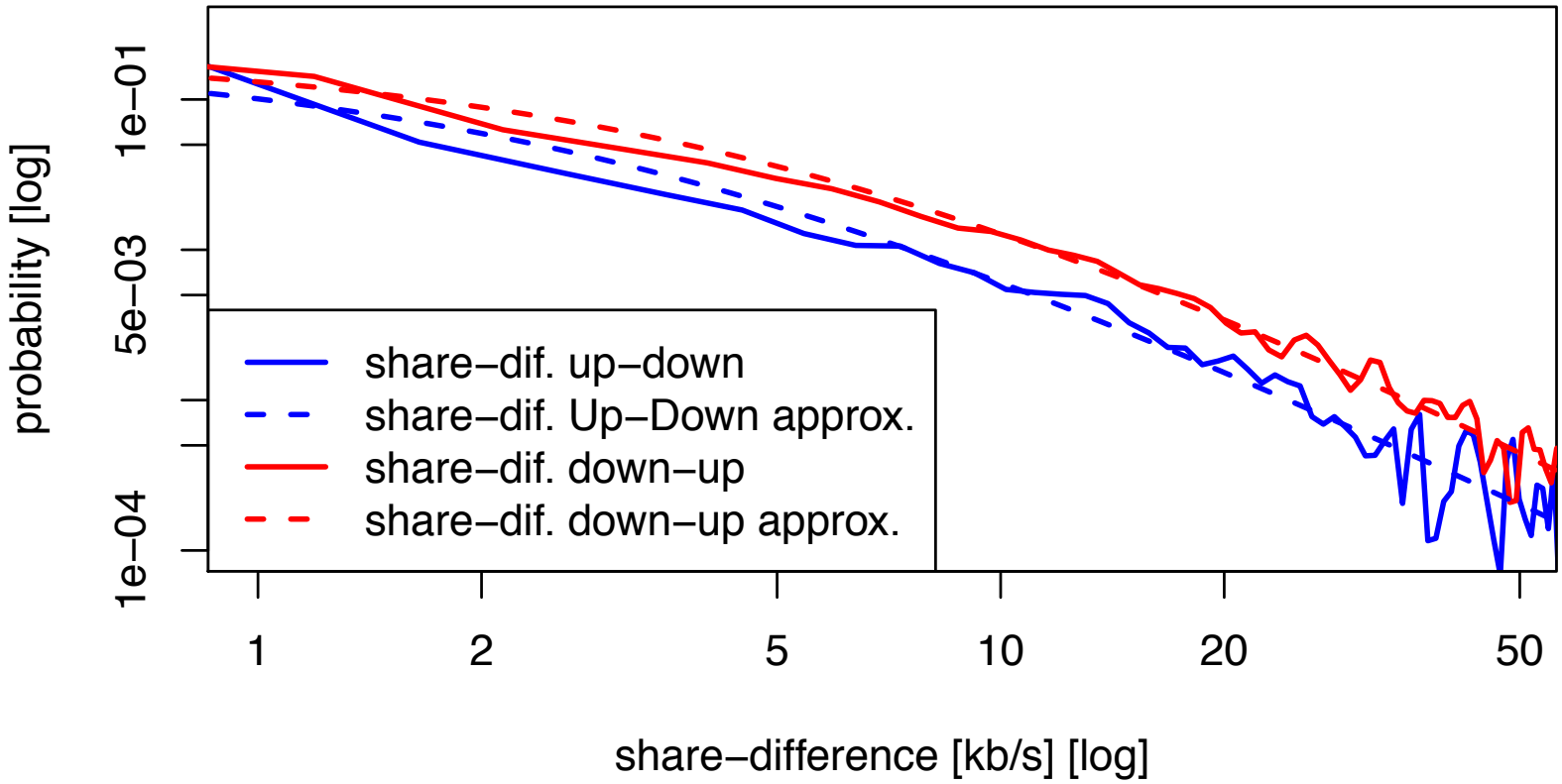
$$P_{d-u}\left[\text{share-difference } x\right] \approx \begin{cases} 0.68 \cdot (x + 1.1)^{-2.04} & \text{for } x \geq 0, \ (\sigma = 0.0008) \\ 4.33 \cdot (3.07 - x)^{-2.33} & \text{for } x < 0 \ (\sigma = 0.0006) \end{cases}$$

- Explanation: maybe the power law distribution of the overall BitTorrent upload and download?

UNI FREIBURG

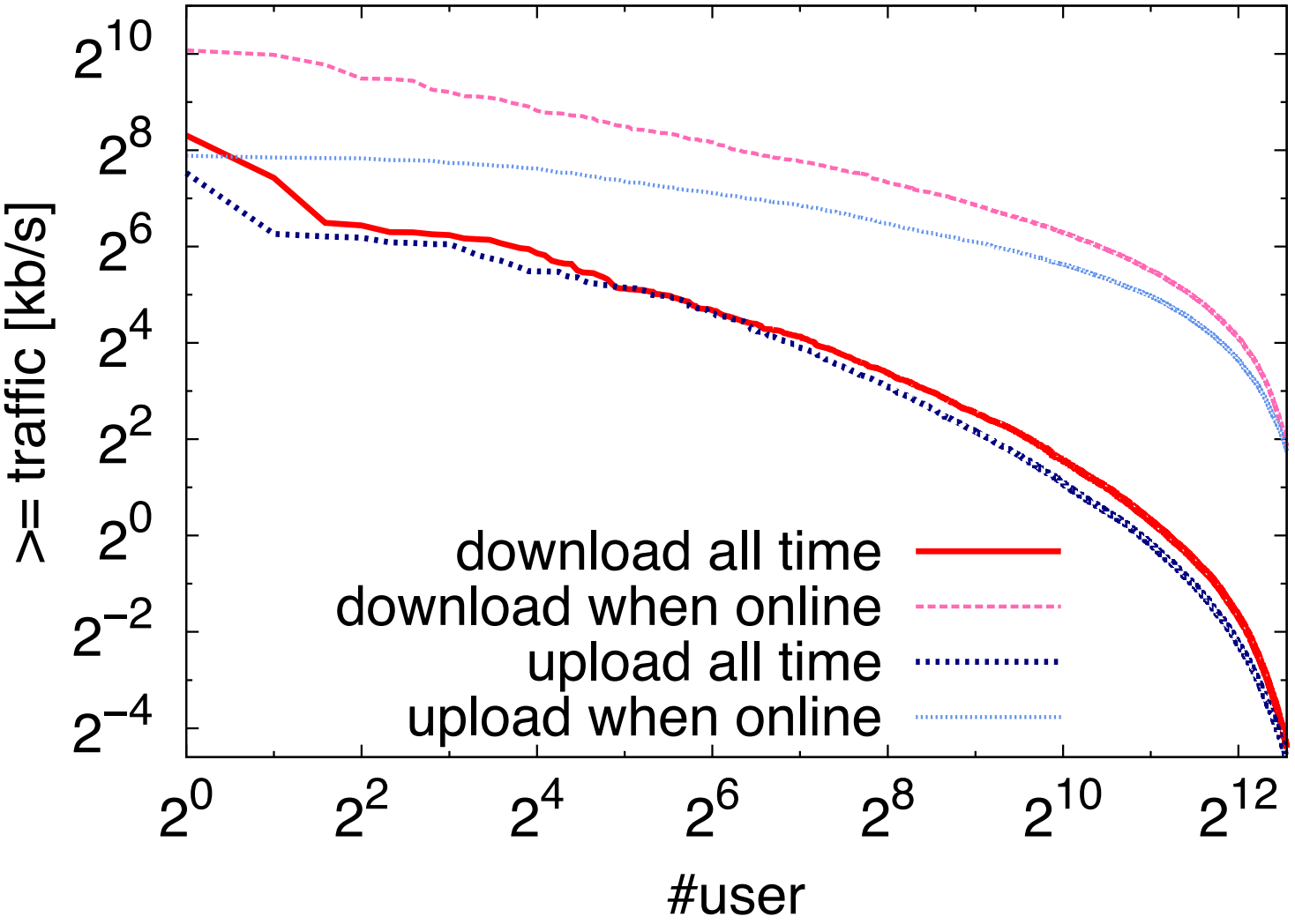
CoNe Freiburg

Share Difference of Torrent Traffic



CoNe Freiburg

Overall Traffic Distribution



CoNe Freiburg

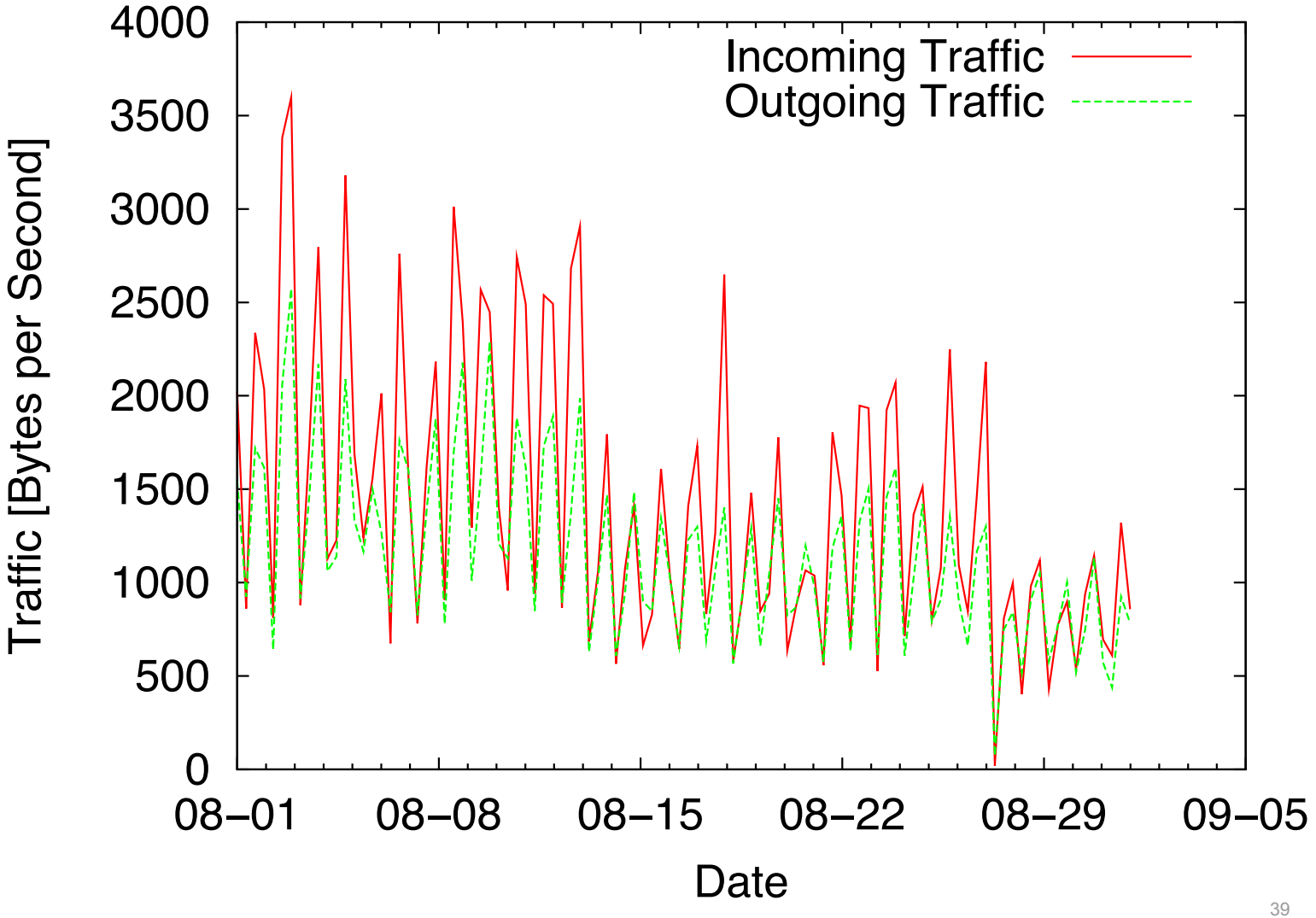
Periodicity

- Obviously there is a perodicity in the data
- New idea:
  - Look at Fourier Transformation
  - And normalized by frequency to receive the „energy“ level
  - and verify with averaged plots

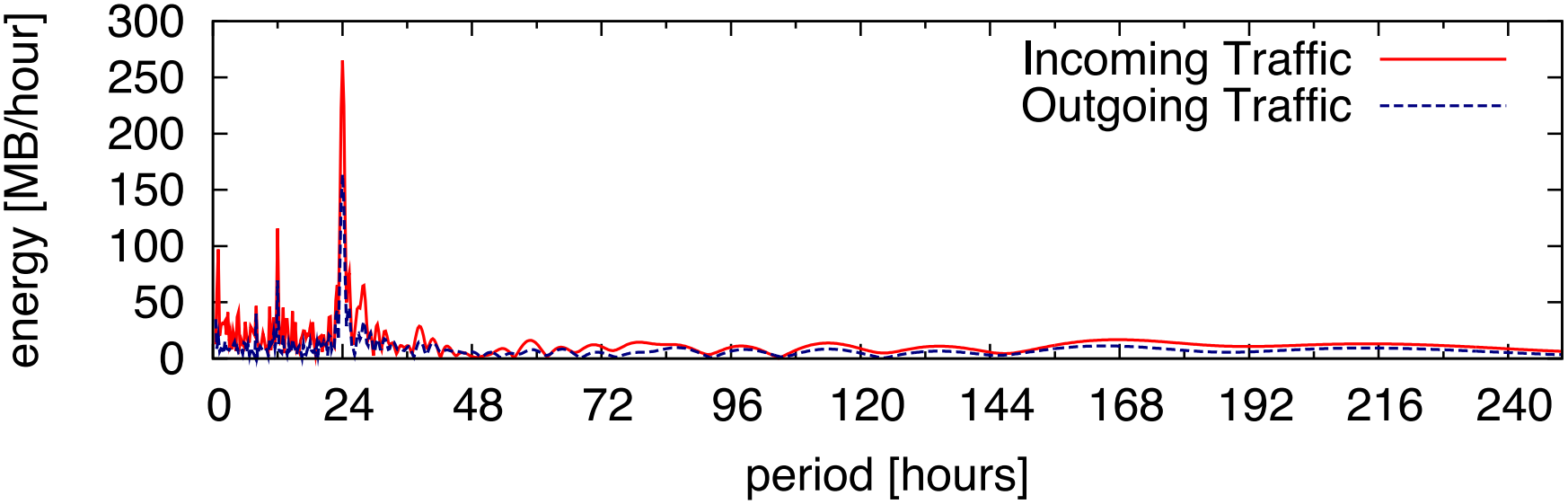
UNI FREIBURG

CoNe Freiburg

Average Traffic per Host

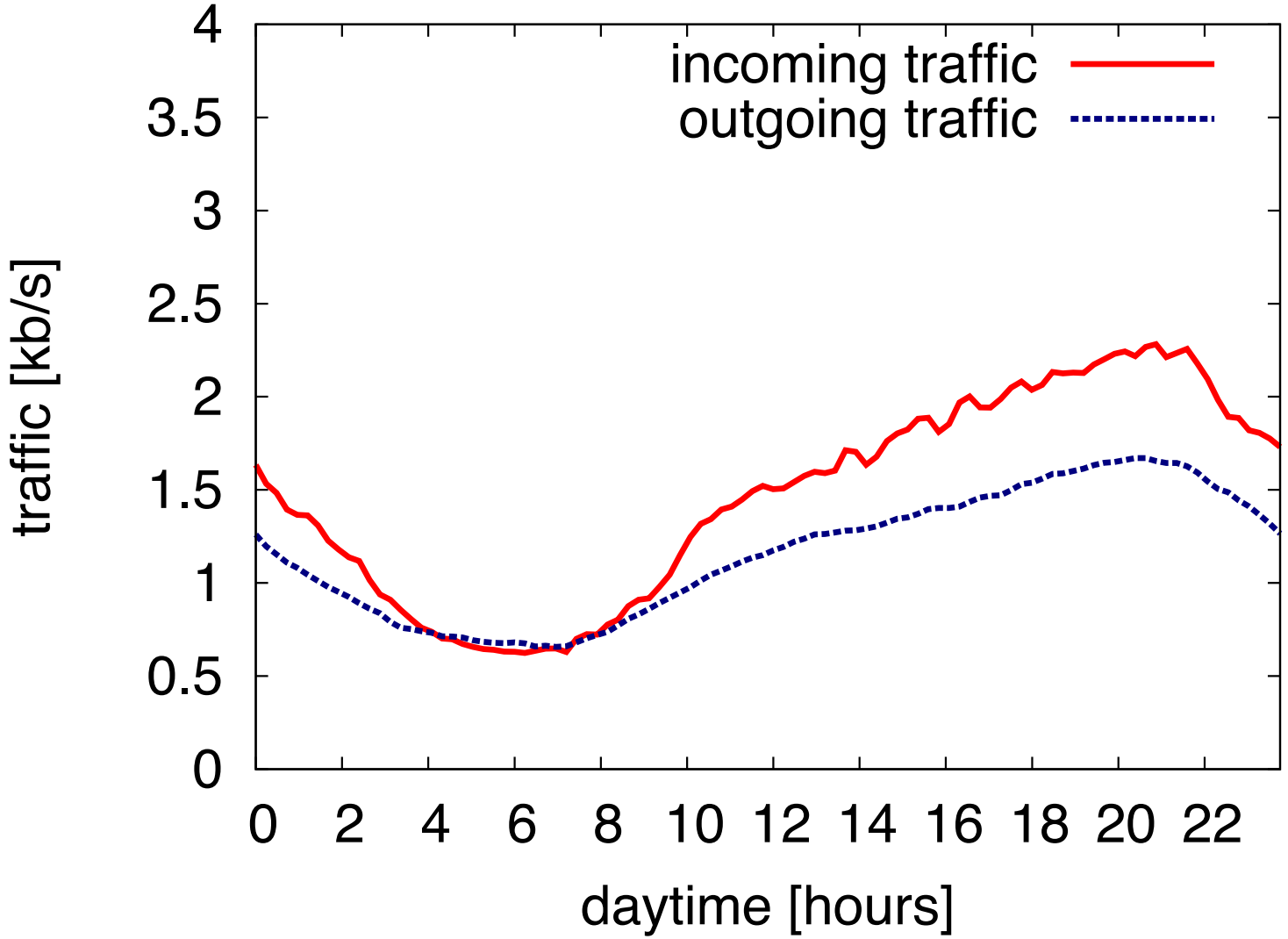


Fourier Transformation of Traffic Pattern



CoNe Freiburg

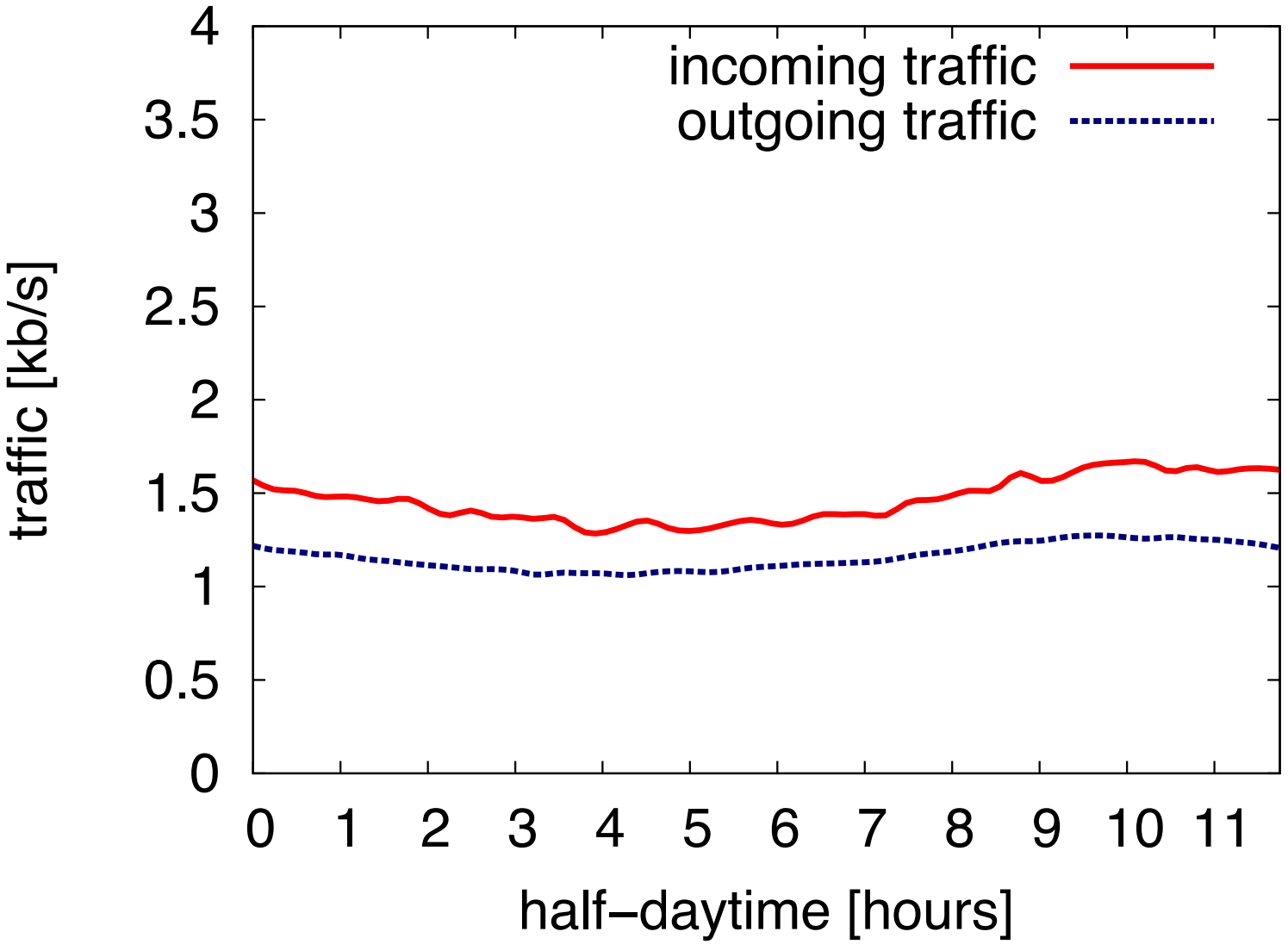
Traffic Average per Hour of the Day



UNI FREIBURG

CoNe Freiburg

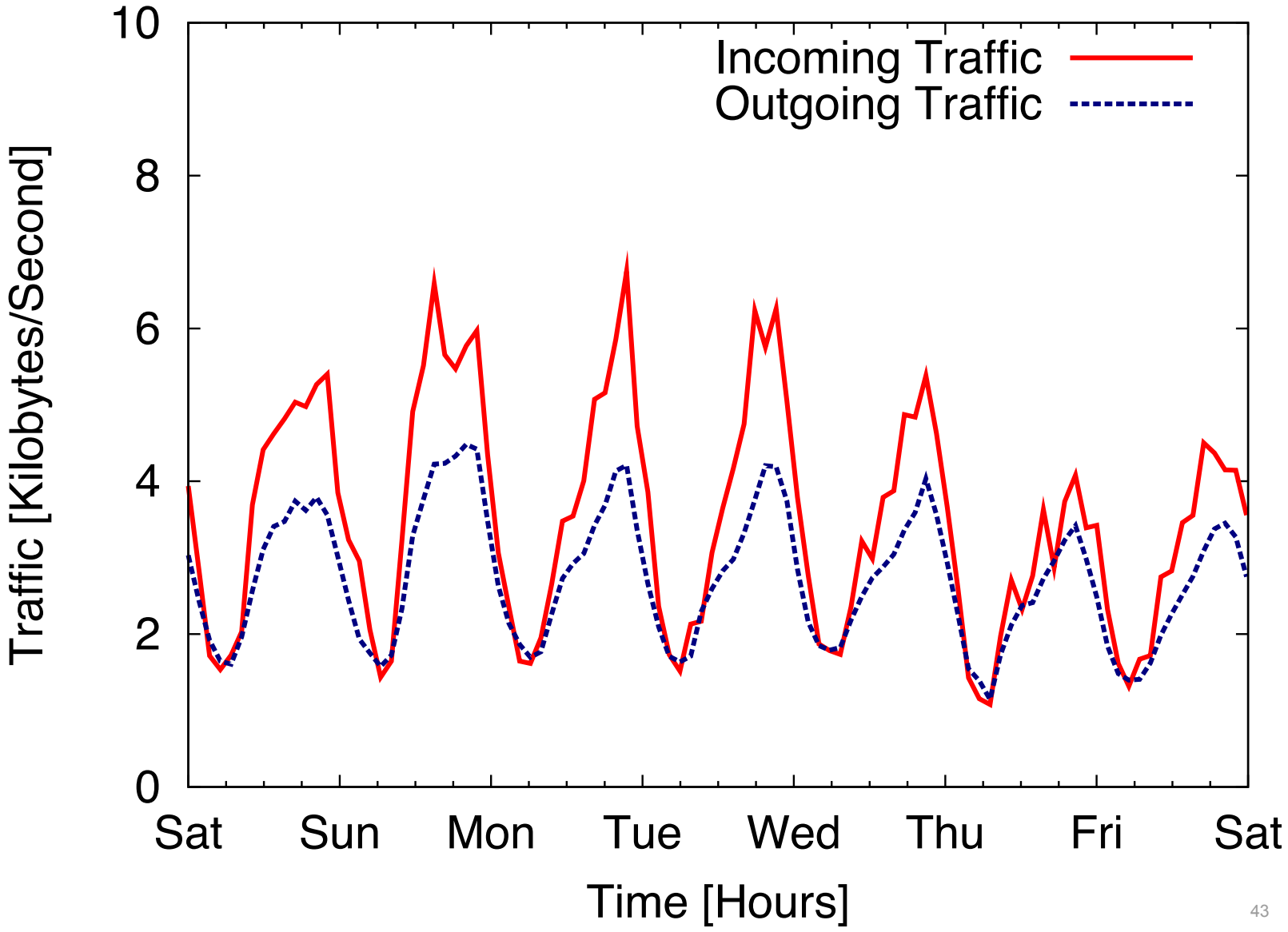
Is there a 12h Periodicity?



UNI FREIBURG

CoNe Freiburg

Traffic Averaged per Weekday



CoNe Freiburg

How Long are Users Online

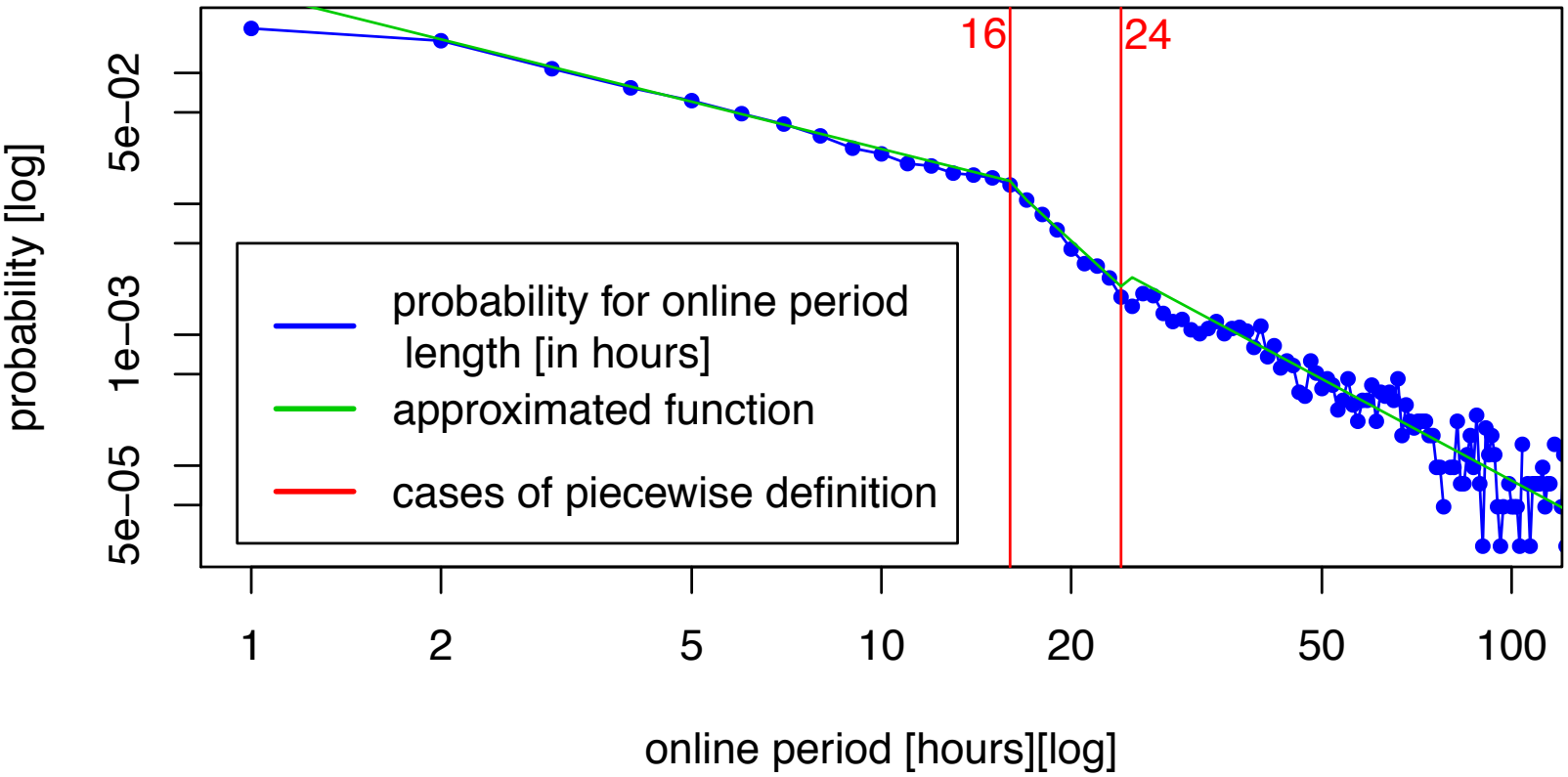
- Online period
  - continuous amount of time when users are using HTTP
- Online times
  - sum of periods over a day/week/month

$$P\left[\text{online period } t\right] \approx \begin{cases} 0.18 \cdot t^{-0.82} & \text{for } t \geq 16, \ (\sigma = 0.013) \\ 2782 \cdot t^{-4.40} & \text{for } 16 < t \leq 24, \ (\sigma = 0.00006) \\ 11 \cdot t^{-2.58} & \text{for } t > 24 \ (\sigma = 0.000015) \end{cases}$$

UNI FREIBURG

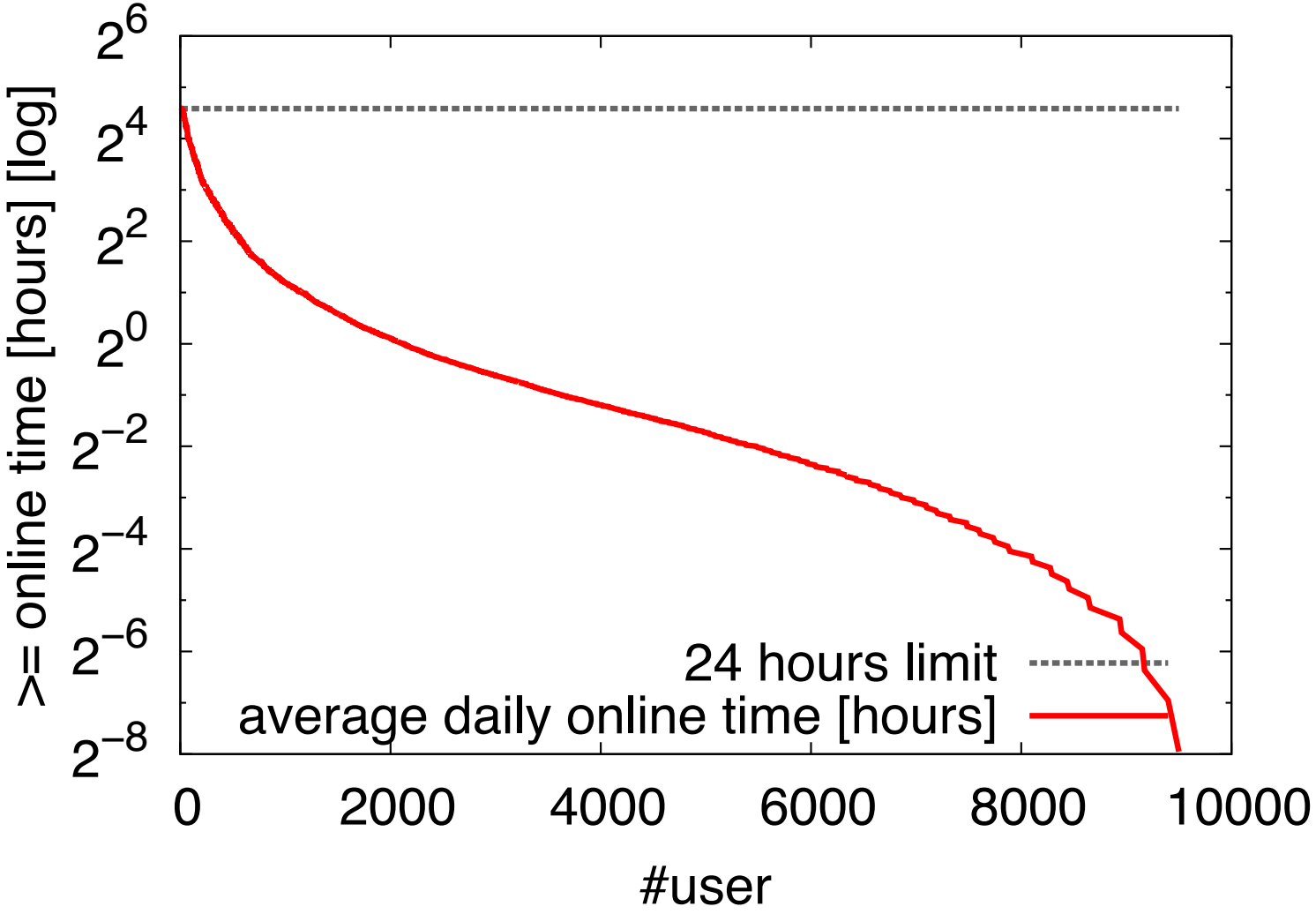
CoNe Freiburg

Distribution of Online Periods



CoNe Freiburg

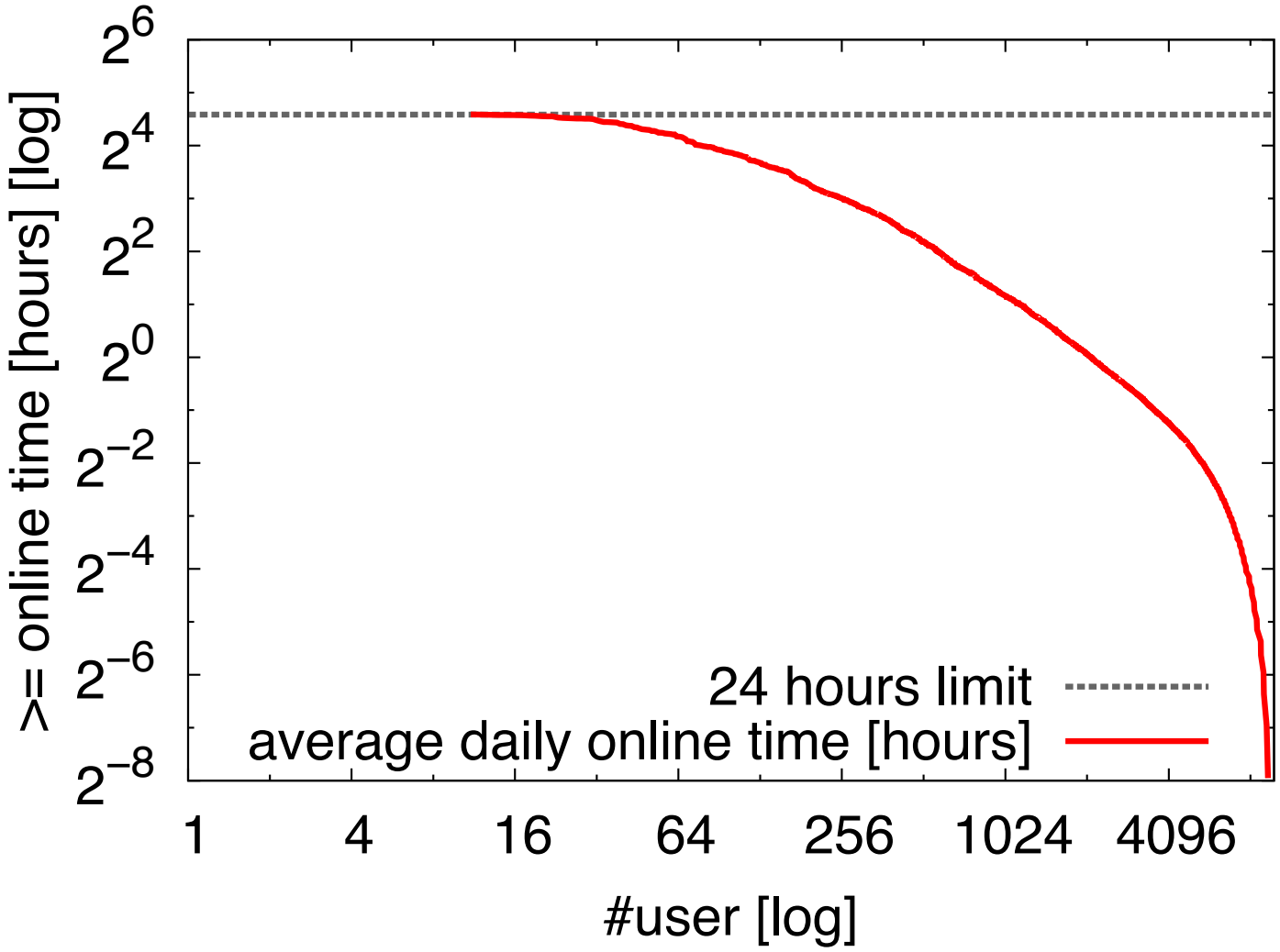
Online Times per Day (Lin-Log-Plot)



UNI FREIBURG

CoNe Freiburg

Online Times (log log Plot)



CoNe Freiburg

Conclusions

- Analysis of web traffic of 21,766 hosts of an Internet service provider (ISP) in Germany
- Emphasis BitTorrent traffic August in 2009
- 50% used BitTorrent
- At most 40% of BitTorrent users online at the same time
- Many users participate in this peer-to-peer network only for some short time periods
- Most Internet traffic is HTTP

UNI FREIBURG