Today's Usenet Usage: NNTP Traffic Characterization

Juhoon Kim | Fabian Schneider | Bernhard Ager | Anja Feldmann

{jkim|fabian|bernhard|anja}@net.t-labs.tu-berlin.de

Technische Universität Berlin/Deutsche Telekom Laboratories

Global Internet Symposium
2010-03-19

Motivation

- Maier et al. [IMC'09] report a fraction of up to 5% of the total traffic in a residential broadband access network being transferred via NNTP
- But our expectation was:
  - NNTP is almost dead
  - Only used by Internet-Oldies
- Data shows only a small number of DSL lines using NNTP (<2%)
- Why so much NNTP traffic?

[IMC'09] G. Maier, A. Feldmann, V. Paxson and M. Allman, "On dominant characteristics of residential broadband Internet traffic", in Proceedings of Internet Measurement Conference 2009

Outline

- Motivation
- Refresher on NNTP
- Methodology
- Data Sets
- Evaluation Results
- Summary

Refresher on NNTP

- Protocol of the Usenet
- Today: Usenet features provided by Forums or Mailing lists, integrated into Blogs and OSNs
- Features:
  - Post and read/download articles
  - Hierarchy of newsgroups to structure content
  - Articles similar to emails (same message format)
  - Request article by
    - Newsgroup & article number or
    - Globally unique Message-ID (newsgroup independent)
  - Limited Storage: Remove old articles as new ones arrive

Methodology

- Developed NNTP analyzer for Bro NIDS:
  - Detects NNTP traffic on any port
  - Identifies/logs NNTP commands
  - Handles single- and multi-line requests
  - Identifies/logs binary encoding method
  - Determines content type based on filename
- Run analyzer on anonymized packet traces
- Run analyzer live logging only anonymized summaries
- Focus on client⇔server communication

Data Sets

- Same vantage point as Maier et al [IMC'09]
  - Monitoring residential DSL connections at large ISP in EU
  - 3 anonymized packet level traces (24h/48h each)
  - 1 anonymized log summary covering >15 days

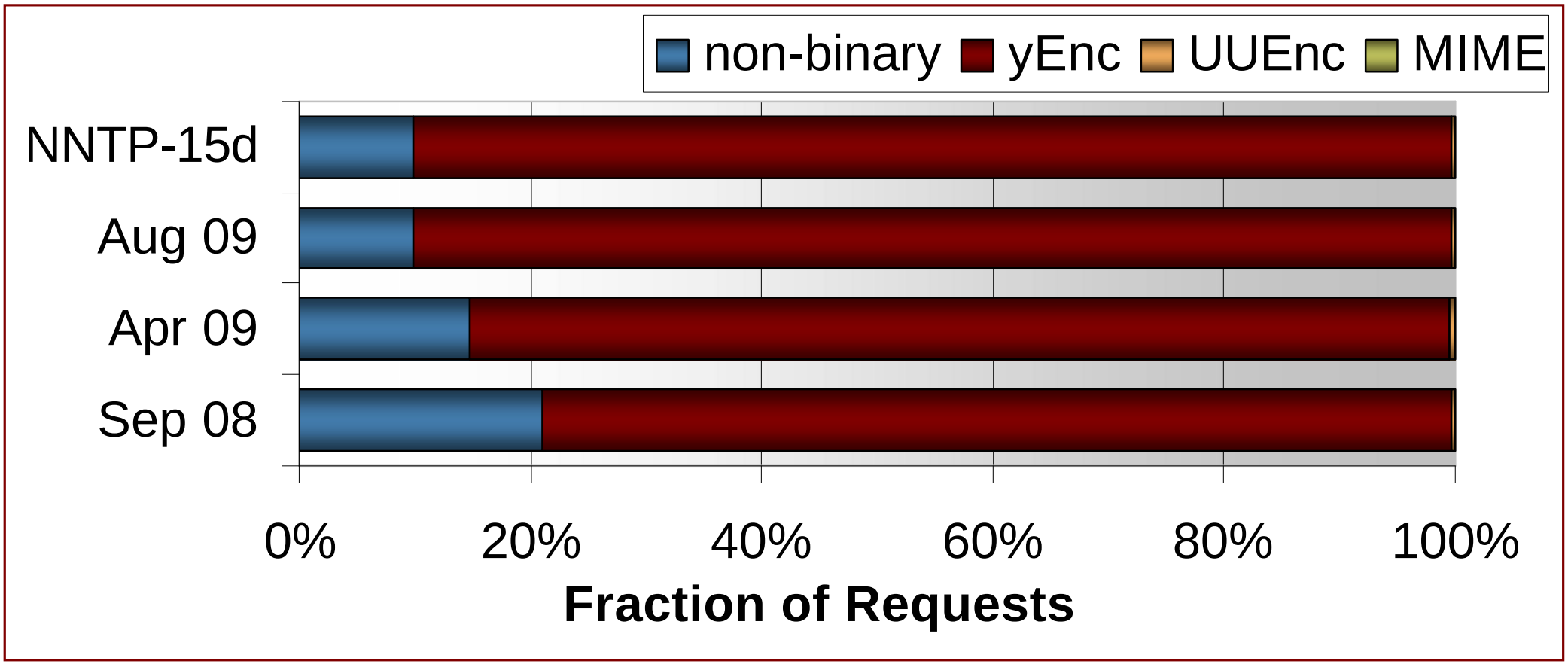
Name	Start date	Duration	Size	NNTP
NNTP-15d	Wed 5 Aug'09 3am	15¼ d	n/a	n/a
SEP08	Thu 18 Sep'08 4am	24 h	> 4 TB	5 %
APR09	Wed 01 Apr'09 2am	24 h	> 4 TB	2 %
AUG09	Fri 21 Aug'09 2am	48 h	> 11 TB	2 %

Outline

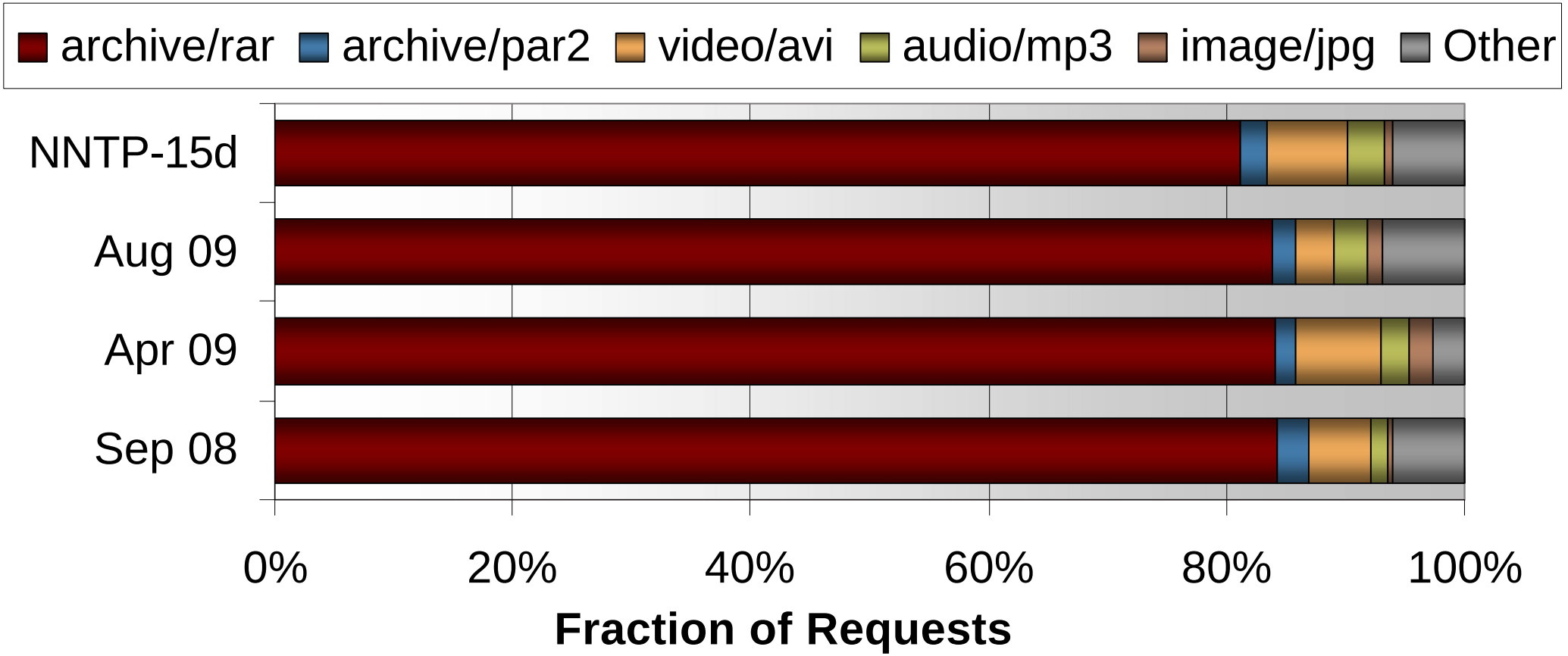
- Reminder on NNTP
- Methodology
- Data Sets
- Evaluation Results
  - Text vs. encoded binaries
  - Popular Content types
  - Identification of NNTP servers
  - Additional Observations
  - Throughput analysis
- Summary

Text vs. Encoded Binaries

99% of volume is binary encoded with yEnc



Popular Content Types



- More than 80% are RAR archives
- Multimedia types also popular (.avi, .jpg, .mp3)

Popular Content Types

Similar to content types of file-sharing services like:

- **One-Click Hosters, e.g. Rapidshare**
- **P2P networks, e.g. BitTorrent**

- More than 80% are RAR archives
- Multimedia types also popular (.avi, .jpg, .mp3)

Identification of NNTP Servers

- Findings so far:
  Mostly binary and file-sharing content
- Binary content on public NNTP servers
  ➡ Not allowed or very short retention times

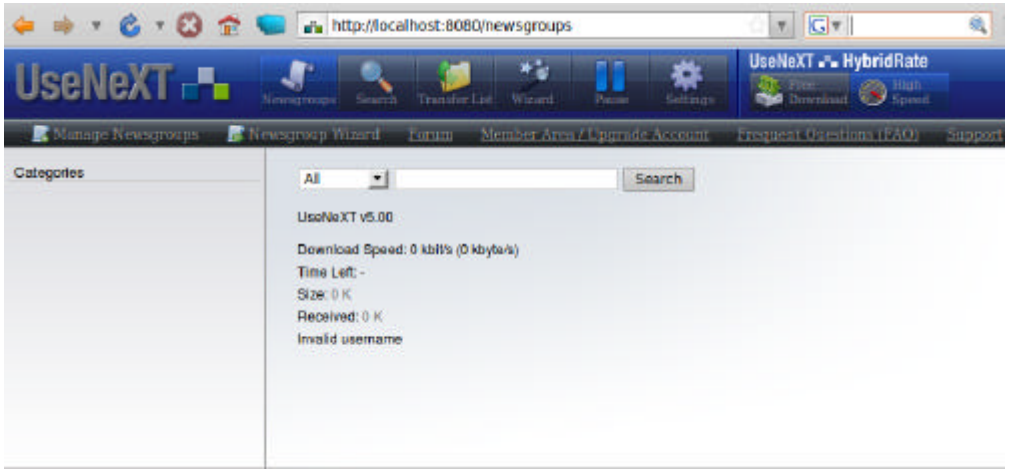
- Which servers provide this content?

➡ Binary groups are hosted on commercial servers, requiring a monthly fee to access them:
  - Serve >99% of the volume (93% of the requests)
  - Feature search function to locate content

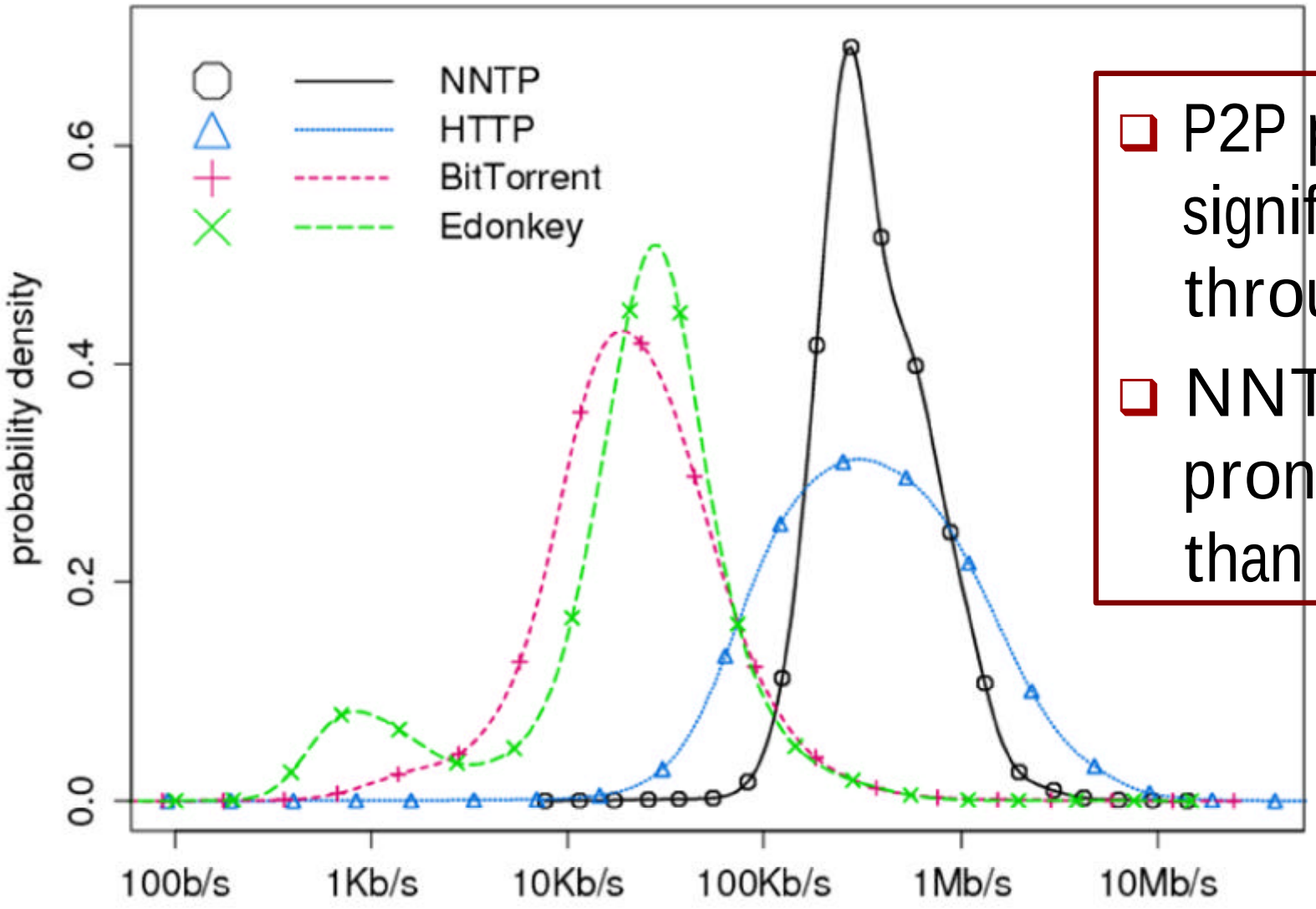
Additional Observations

- Multiple NNTP connections in parallel
- Most queries via globally unique Message-ID
- High frequency of BODY command
- Low frequency of GROUP command

➡ Likely explanation:
Specialized clients of commercial NNTP offers
  - Provide shiny GUIs (e.g. via a local Web server)
  - Specialized for file download



Throughput of File-sharing Offers



- P2P protocols significantly lower throughput
- NNTP more pronounced peak than HTTP

Throughput of flows >50kB, download direction only

Summary

- Almost all of the volume (>99%) is binary data
- NNTP carries similar content as file-sharing services
- 99% of the volume is exchanged with commercial NNTP servers that require a monthly fee to access them
- Client/server protocols achieve an order of magnitude higher throughput than P2P

➡ **NNTP is used as a commercial, high performance alternative to file-sharing services**

Questions?

Backup Slides

- NNTP command distribution
  - Frequency of commands
  - Volume of commands
- Transaction Volumes

Command Frequency

Command	SEP08	APR09	AUG09	NNTP-15d
ARTICLE	82.6 %	83.2 %	66.5 %	76.5 %
BODY	8.1 %	10.3 %	27.1 %	18.1 %
GROUP	0.9 %	1.1 %	0.4 %	0.6 %
HEAD	4.1 %	<0.1 %	<0.1 %	0.6 %
AUTHINFO	1.8 %	2.3 %	2.0 %	1.8 %
QUIT	0.2 %	0.7 %	0.2 %	0.2 %

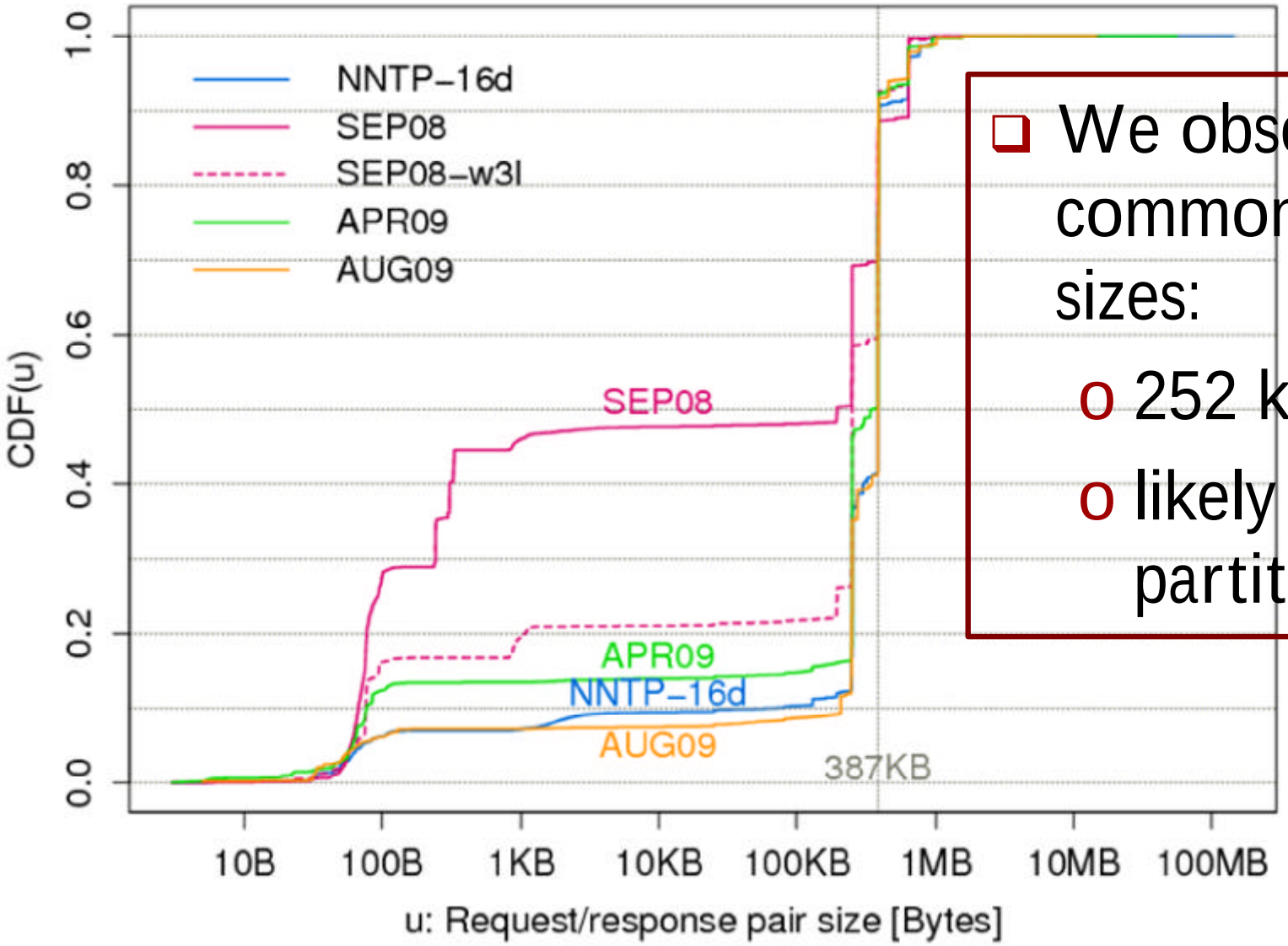
Command Frequency

In terms of volume:

`ARTICLE` and **`BODY`**

together contribute more than 99%!

Transaction Volumes



- We observe two common transaction sizes:
  - 252 kB and 387 kB,
  - likely due to file partitioning