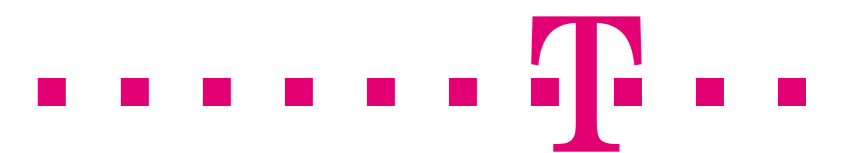


Cacheability of Bulk Content for ISPs

Bernhard Ager · Fabian Schneider · Anja Feldmann



MOTIVATION

- File sharing is responsible for a large fraction of Internet traffic
- File sharing applications often use P2P protocols, HTTP, or NNTP [1]
- HTTP caching used to be popular; however, efficiency decreased due to cookies, dynamic content, and AJAX
- Recent studies of Gnutella, Fasttrack, and YouTube suggest potential for caching

⇒ It is time to reexamine!

SUMMARY

Investigating potential of caching for three file sharing applications: BitTorrent, NNTP, and a Direct Download Service (DDS, based on HTTP)

- Caching does not appear to be beneficial for NNTP and DDS
- Promising results for BitTorrent: Cache efficiencies from 35 % to 77 % are possible depending on the location of the cache
- Potential usable via caching or peer selection (IETF ALTO Working Group)
- Increased caching potential for P2P due to:
 - Ability to peek into the past
 - Larger local user base compared to NNTP and larger observable user base by spotting non-local users
 - (Maybe) higher temporal density of downloads compared to HTTP

So far: Caching only worthwhile for BitTorrent!

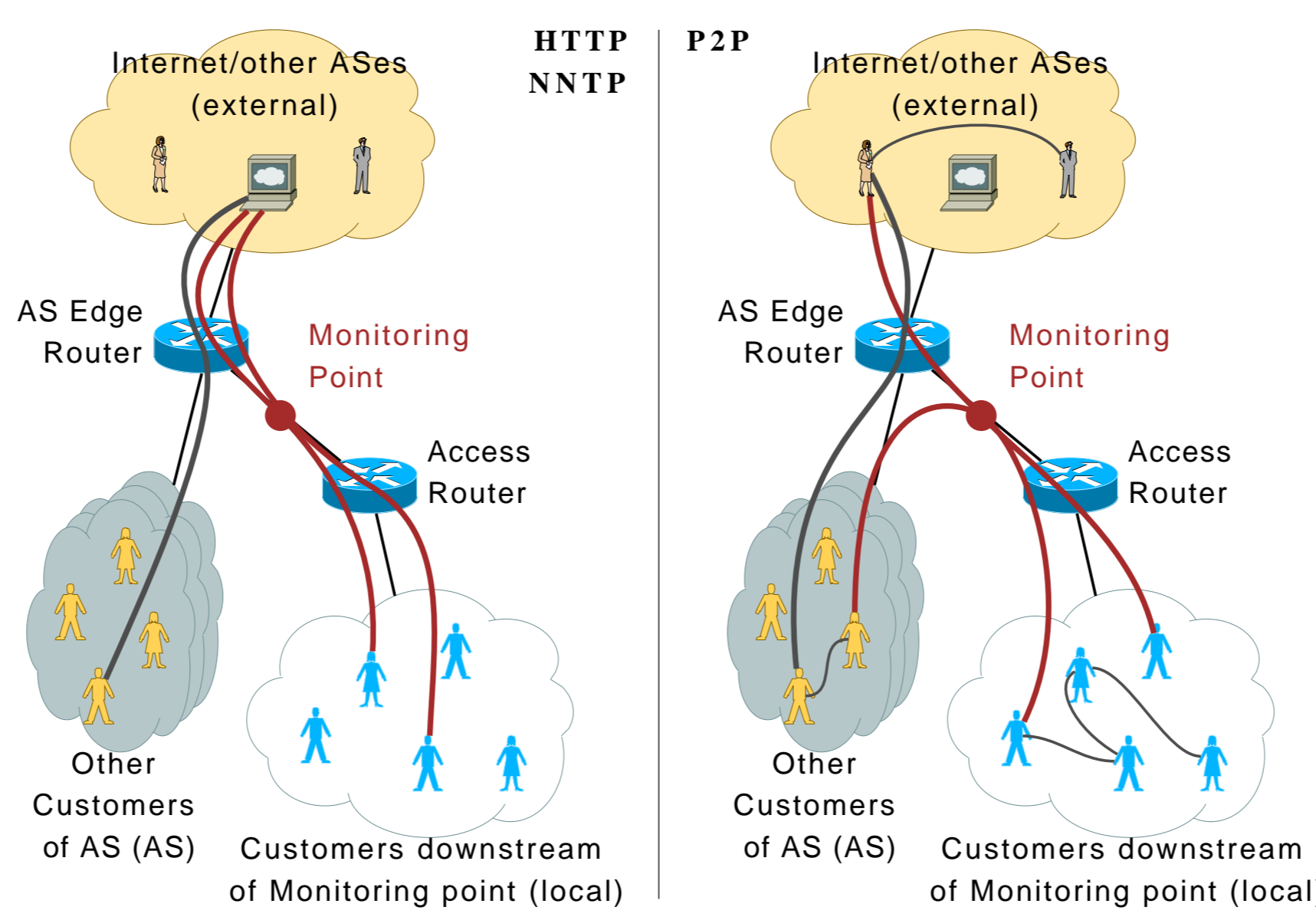
Data

- Anonymized 24h packet trace of several 10,000 residential ISP customers
- Bro IDS [2] for trace processing, using new analyzers for NNTP and BitTorrent
- Cache efficiencies calculated based on content size (bytes) and number of duplicates

Coverage

- **HTTP and NNTP:** only downloads from the local network population observable
- **BitTorrent:** interested peers outside our local network are also covered

APPROACH



Application Usage

Application	Traffic	Users
BitTorrent	9 %	2 %
DDS	6 %	6 %
NNTP	5 %	1 %

BitTorrent Traffic Patterns

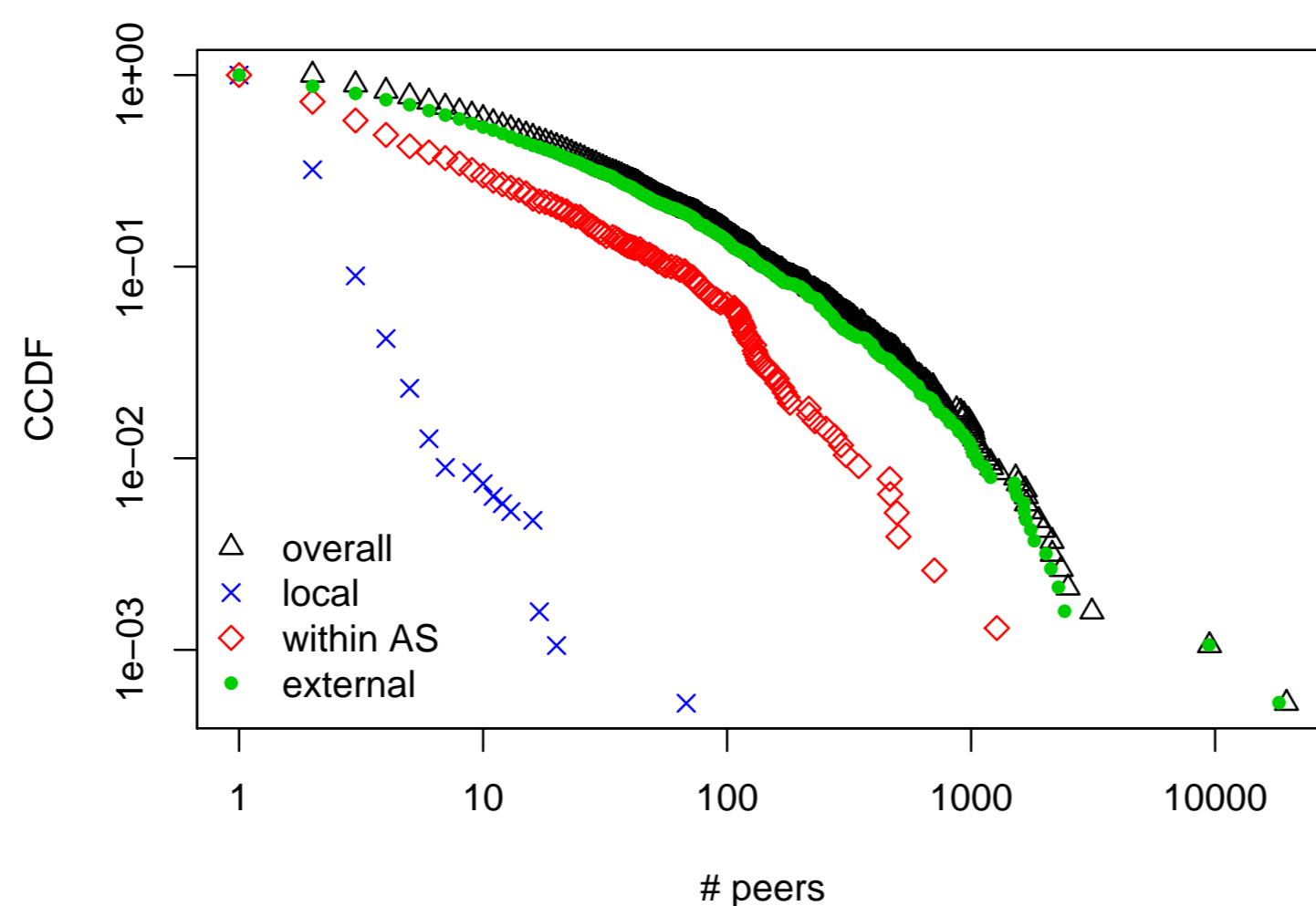
POP to/from	AS	extern
Upload	13.1 %	86.9 %
Download	9.1 %	90.9 %
Up/down ratio	0.771	0.508

- Most traffic to AS-external peers
- External peers: Lower up/down ratio

BITTORRENT EVALUATION

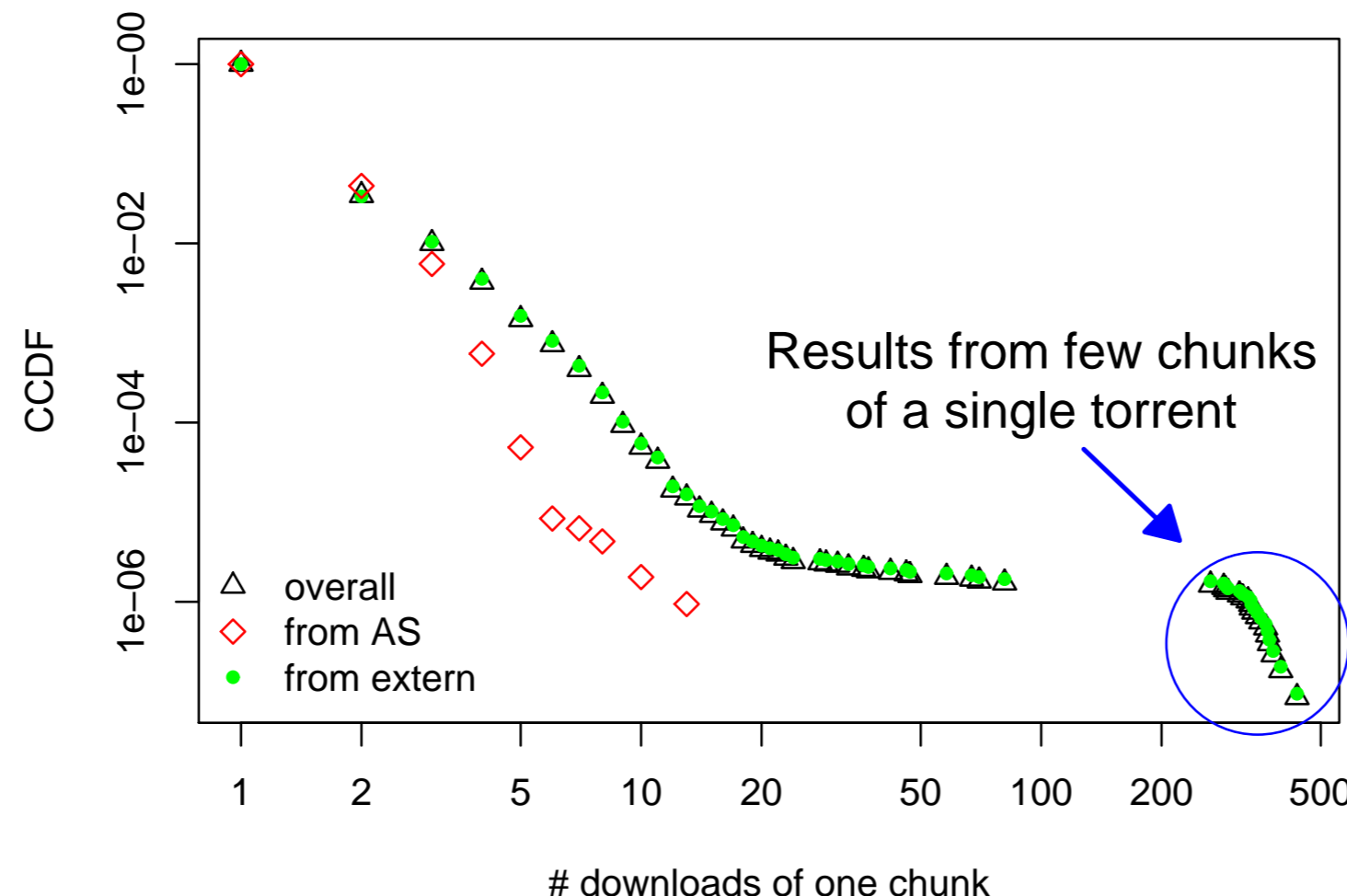
Cacheability Potential

- Metric: Peers interested in a torrent
- Small number of torrents with only one peer
- Torrents with around a thousand peers within the AS (a hundred locally)
- Caching potential of 96 % within AS and 41 % locally



Actually observed duplicated downloads

- Number of downloads per chunk
- 96 % of chunks downloaded exactly once
- Cache efficiencies of 10 % within AS and 5 % locally
- Explanation: Many peers already completed their downloads



Predicting Cache Efficiency Including Previous Downloads via Bitfields

- Bitfield messages yield information on completed downloads
- **Caching efficiency now at 77 % within AS and 35 % locally**

HTTP/NNTP EVALUATION

- Duplicates identified via article ID (NNTP) or URL (DDS)
- No cache-control headers, cookies, or GET parameters considered ⇒ upper bound on cacheability

RESULTS

Application	Cacheable Objects	Cacheable Volume
NNTP	2 %	0.03 %
DDS	9 %	0.7 %
BitTorrent (Chunks, local)	5 %	5 %
BitTorrent (incl. bitfields, local)	35 %	35 %
BitTorrent (incl. bitfields, AS)	77 %	77 %

FUTURE WORK

- Extend measurement periods to mitigate short-term effects
- Extend analysis to all HTTP traffic, not only DDS
- Other P2P applications, e. g., eDonkey

REFERENCES

- [1] MAIER, G., FELDMANN, A., PAXSON, V., AND ALLMAN, M. On dominant characteristics of residential broadband internet traffic. (to appear) In *Proc. ACM IMC* (2009).
- [2] PAXSON, V. Bro: A system for detecting network intruders in real-time. *Computer Networks* 31, 23–24 (1999), <http://www.bro-ids.org>.