



Offline Downloading in China: A Comparative Study

Zhenhua Li



Tsinghua
University

Tencent 腾讯

Christo Wilson



Northeastern
University

Tianyin Xu



Yao Liu



Zhen Lu



Tsinghua
University

Yinlong Wang



Tsinghua
University

Tencent 腾讯

lizhenhua1983@gmail.com
<http://www.greenorbs.org/people/lzh/>
Oct. 30, 2015

Outline

1 Background

2 Problem

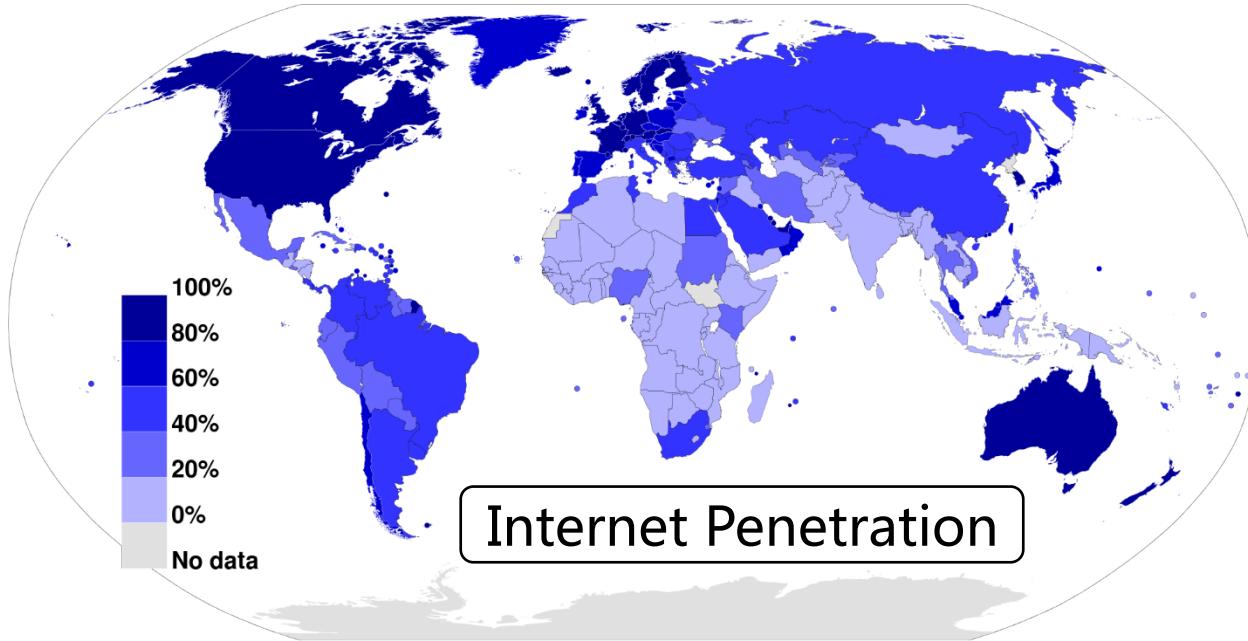
3 System

4 Workload

5 Performance

6 Optimization

Internet Access across the World



THE
**DIGITAL
DIVIDE**

Not only
penetration,
but also
quality of



Broadband: > 25 Mbps of
Download Bandwidth

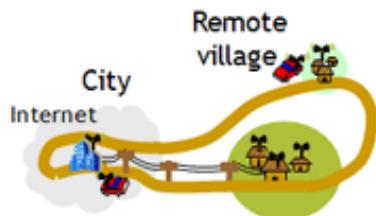


Broadband: ≈ 4–10 Mbps,
Unstable and Limited

Pains of the Developing World



Downloading large files requires high-quality network connections!



*DTN - Delay
Tolerant Networking* ?

The Case of Modern China

Promises

- ✓ 46% of China's population has come online
- ✓ World-class companies like Tencent, Baidu, Alibaba, and Sina Weibo (Microblog)



The Case of Modern China

Challenges

- Over 72% of China's Internet users have low-quality network connections

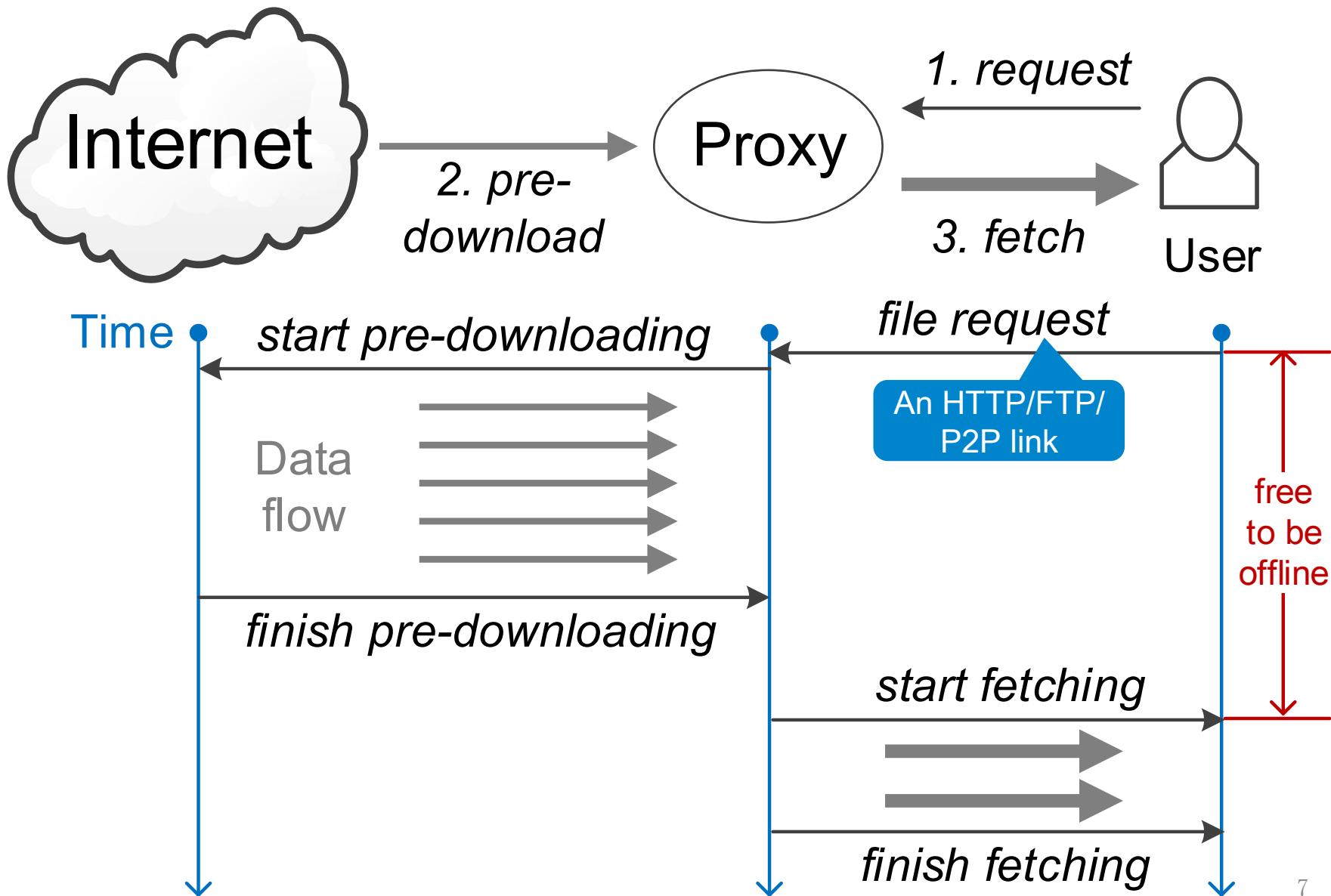
Low access bandwidth

**Unstable/
unreliable
connection**

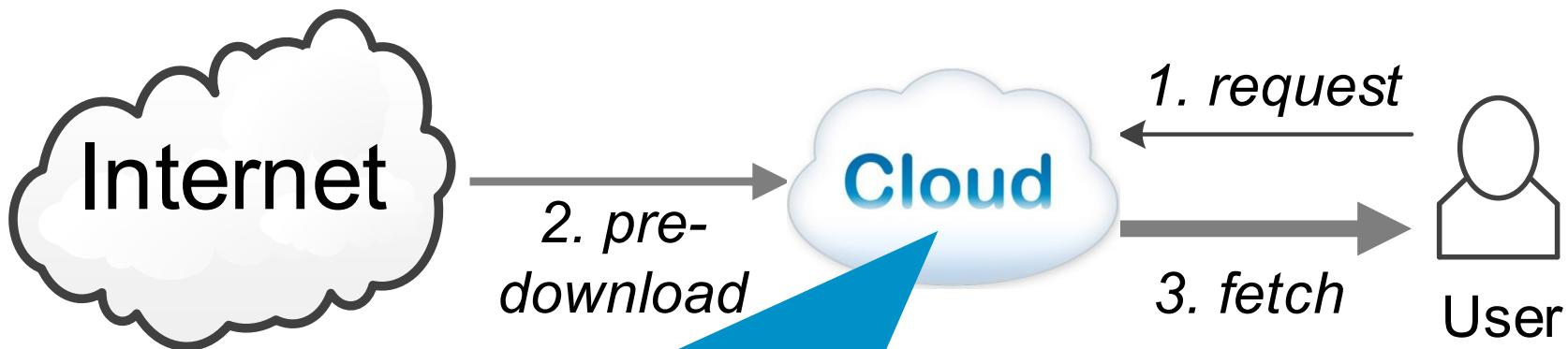
ISP barrier
(Poor inter-connectivity
between ISPs)

**Other
reasons**

“Offline Downloading” in China



Typical Implementation (1): Cloud-based



Caches PBs of files in a datacenter that is within or directly peered with the requesting user's ISP



Tencent
Xuanfeng

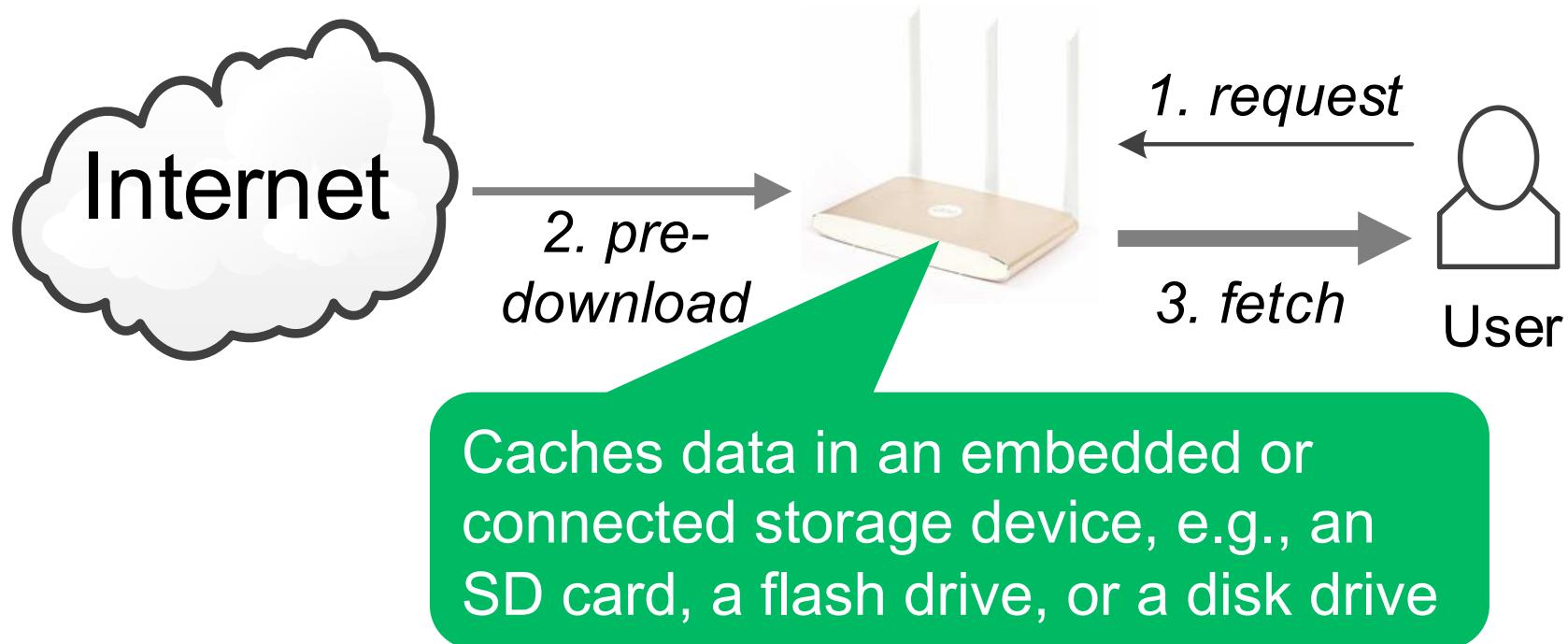


Xunlei



Baidu
CloudDisk

Typical Implementation (2): Smart AP-based



HiWiFi



MiWiFi



Newifi

Great Success in Industry



Tencent
Xuanfeng

✓ Over 30M users



Xunlei

✓ Over 80M users



Baidu
CloudDisk

✓ Over 150M users



HiWiFi

> 1.5M shipments



MiWiFi

> 2M shipments



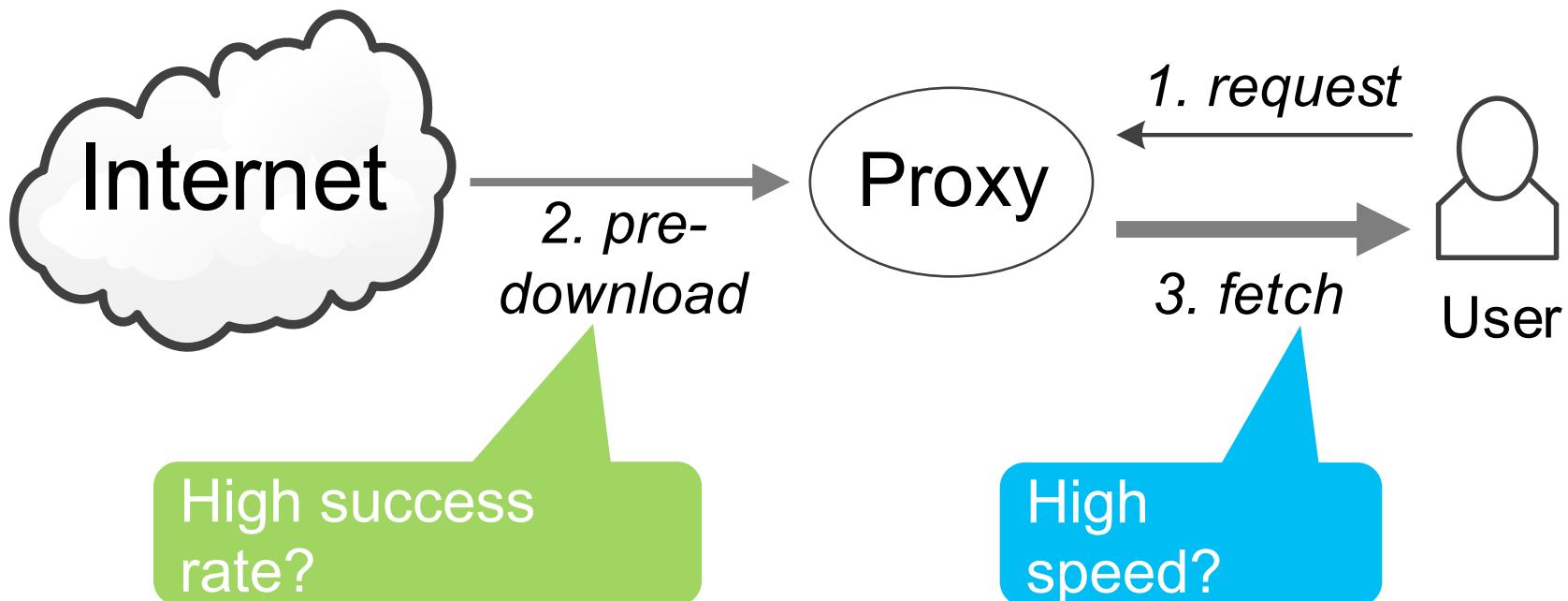
Newifi

> 0.6M shipments

2 Problem

The 1st Problem

Is offline downloading really effective in most cases?



The 2nd Problem

**Which offline downloading
approach should be selected?**



OR

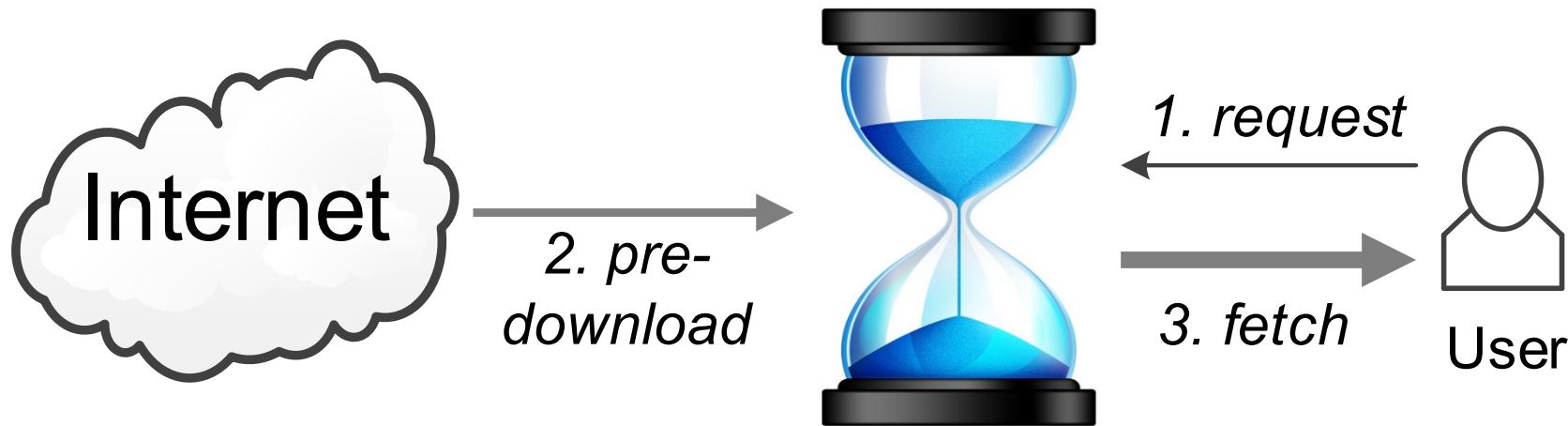


OR

Both
?

The 3rd Problem

When is offline downloading useless or even worse?



User-side
access
bandwidth

Cloud-side
service
capability

ISP barrier

Transfer
protocol

File
popularity

Hardware &
filesystem

General Problem: Selection Dilemma



Common Downloading or
Offline Downloading?

Cloud-based or Smart AP?
And which smart AP?

Our work is the first quantitative and comparative study on these problems

based on a large-scale dataset from Xuanfeng cloud and benchmark experiments of popular smart APs.

3 System



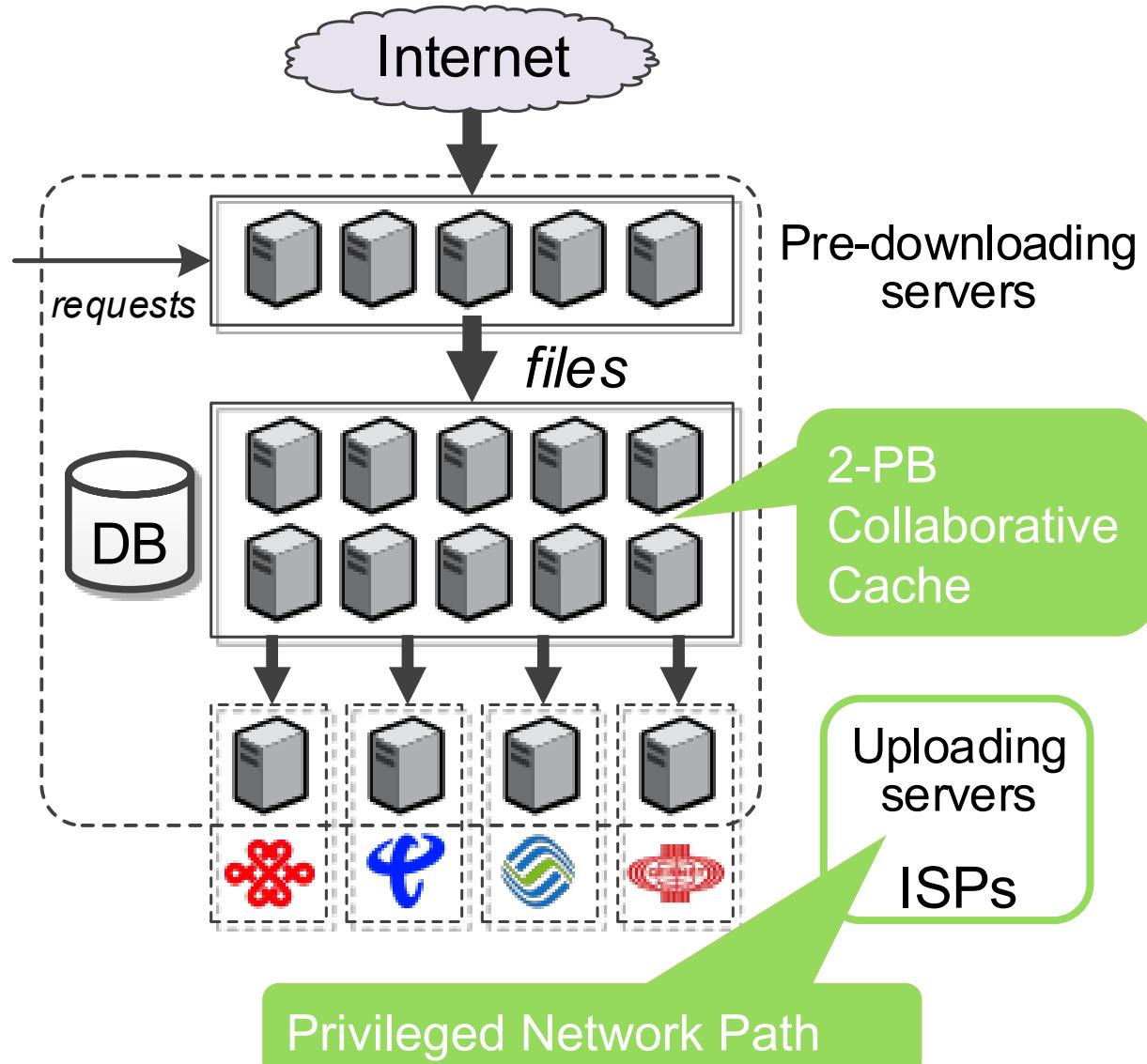
Xuanfeng Cloud



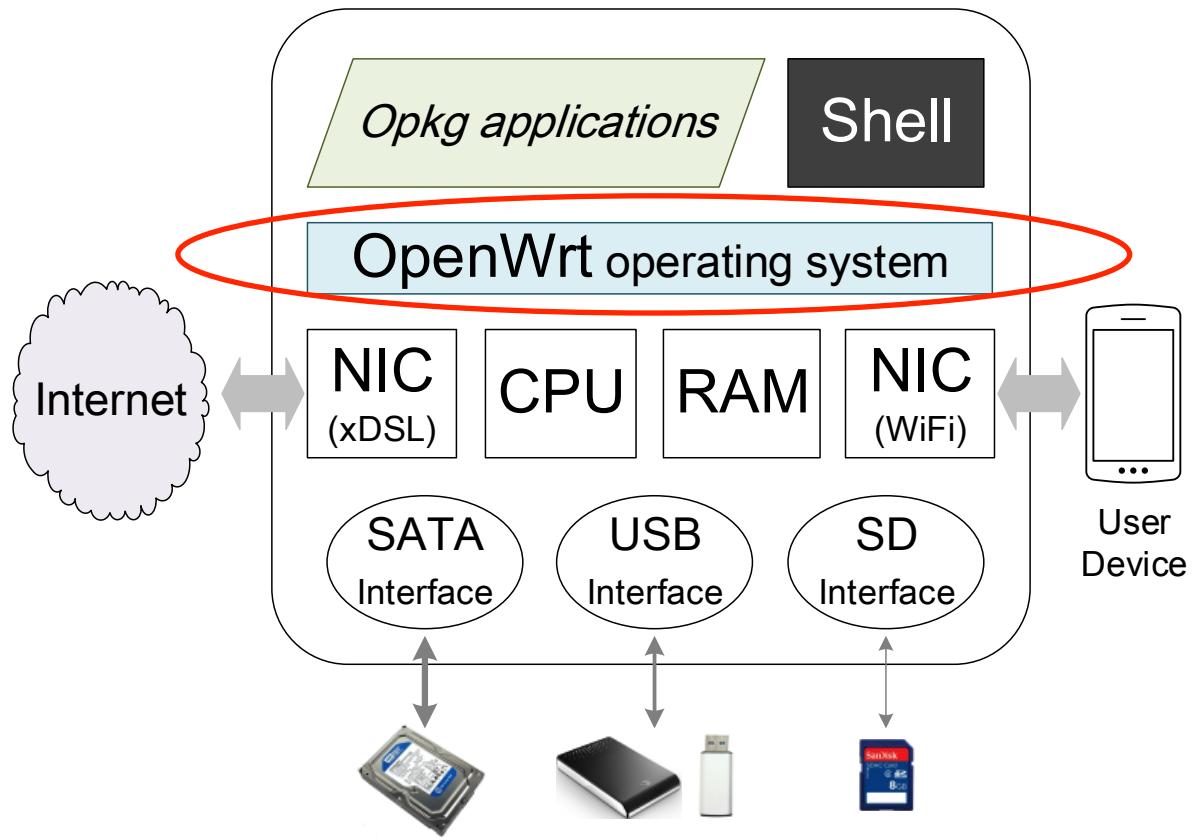
<http://xf.qq.com>



<http://lixian.qq.com/main.html>



Smart APs



Smart AP	CPU	RAM	Storage Interface (and Device)	WiFi Protocol and Channel
HiWiFi (1S)	MT7620A @580 MHz	128 MB	an SD card interface	IEEE 802.11 b/g/n @2.4 GHz
MiWiFi	Broadcom4709 @1 GHz	256 MB	a USB 2.0 interface and an internal 1-TB SATA hard disk drive	IEEE 802.11 b/g/n/ac @2.4/5.0 GHz
Newifi	MT7620A @580 MHz	128 MB	a USB 2.0 interface	IEEE 802.11 b/g/n/ac @2.4/5.0 GHz

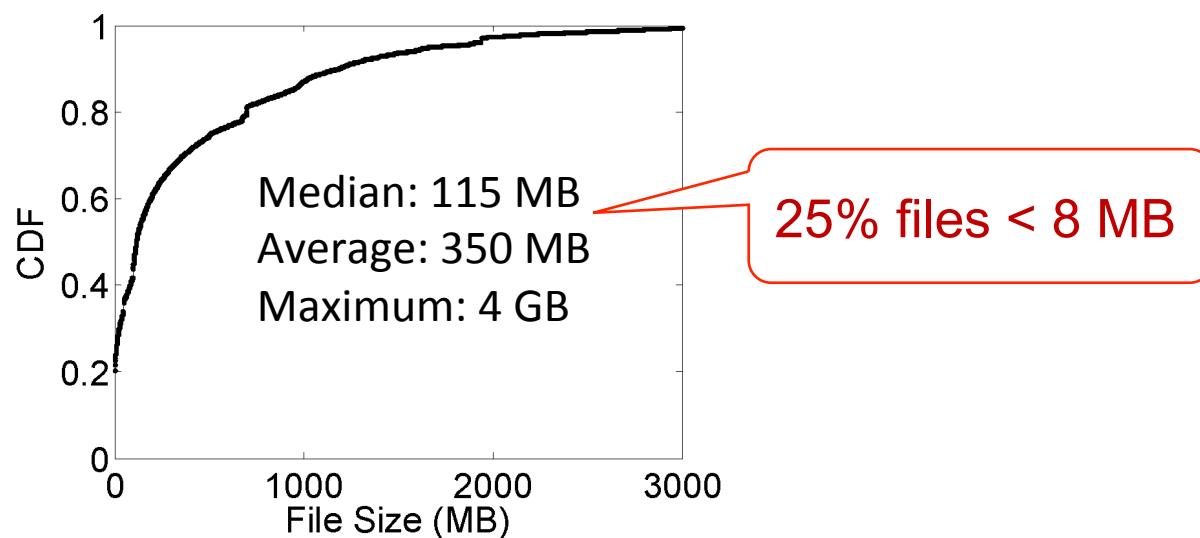
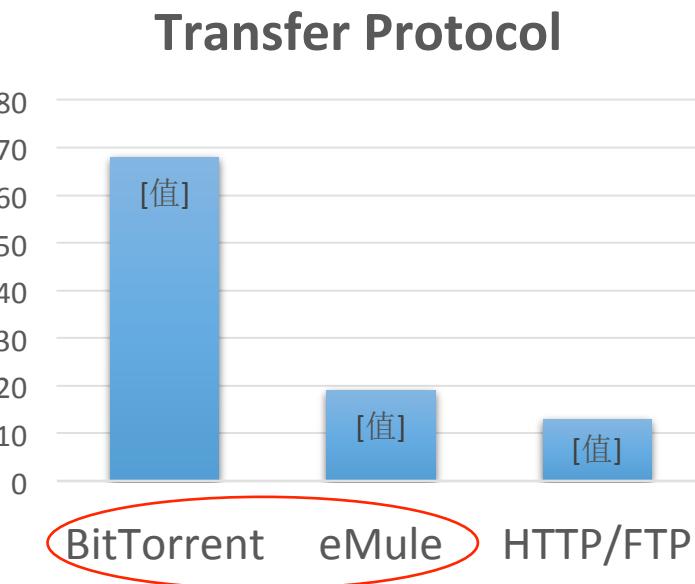
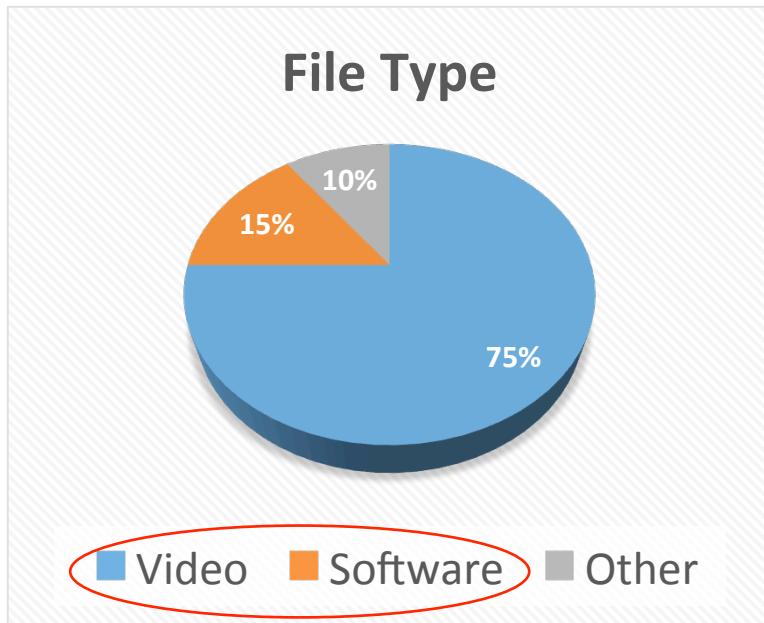
4 Workload

Xuanfeng Dataset

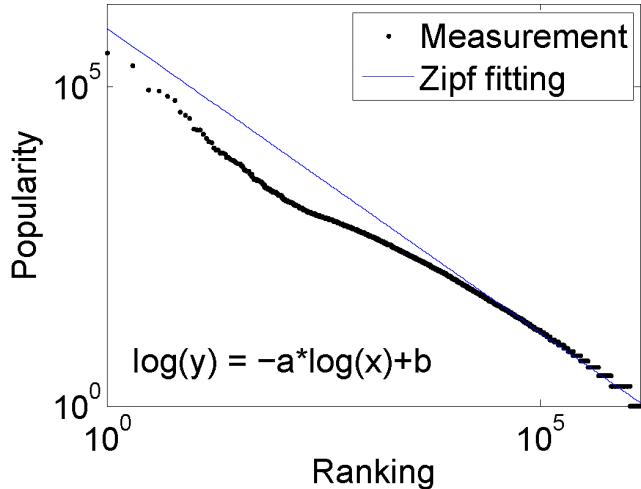
- ❑ Complete running logs during a whole week in 2015, involving 4M tasks, 0.78M users & 0.56M unique files

User Requests	Pre-downloading Trace	Fetching Trace
<ul style="list-style-type: none">• User ID• IP address• Access bandwidth• Request time• File type• File size• Original data source• Transfer protocol	<ul style="list-style-type: none">• Start time• Finish time• Acquired file size• Traffic usage• Cloud cache hit• Avg. speed• Peak speed• Success or failure	<ul style="list-style-type: none">• User ID• IP address• Access bandwidth• Start time• Finish/pause time• Acquired file size• Traffic usage• Avg. speed• Peak speed

File Type, Size & Transfer Protocol

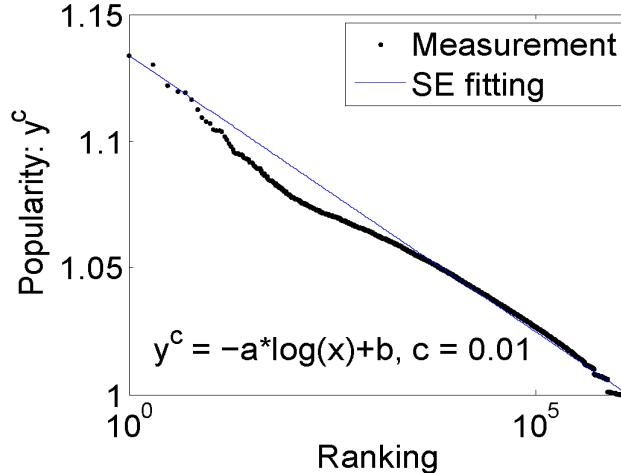


File Popularity



$$\text{Zipf: } \log(y) = -a_1 \times \log(x) + b_1$$

Zipf \approx Power law

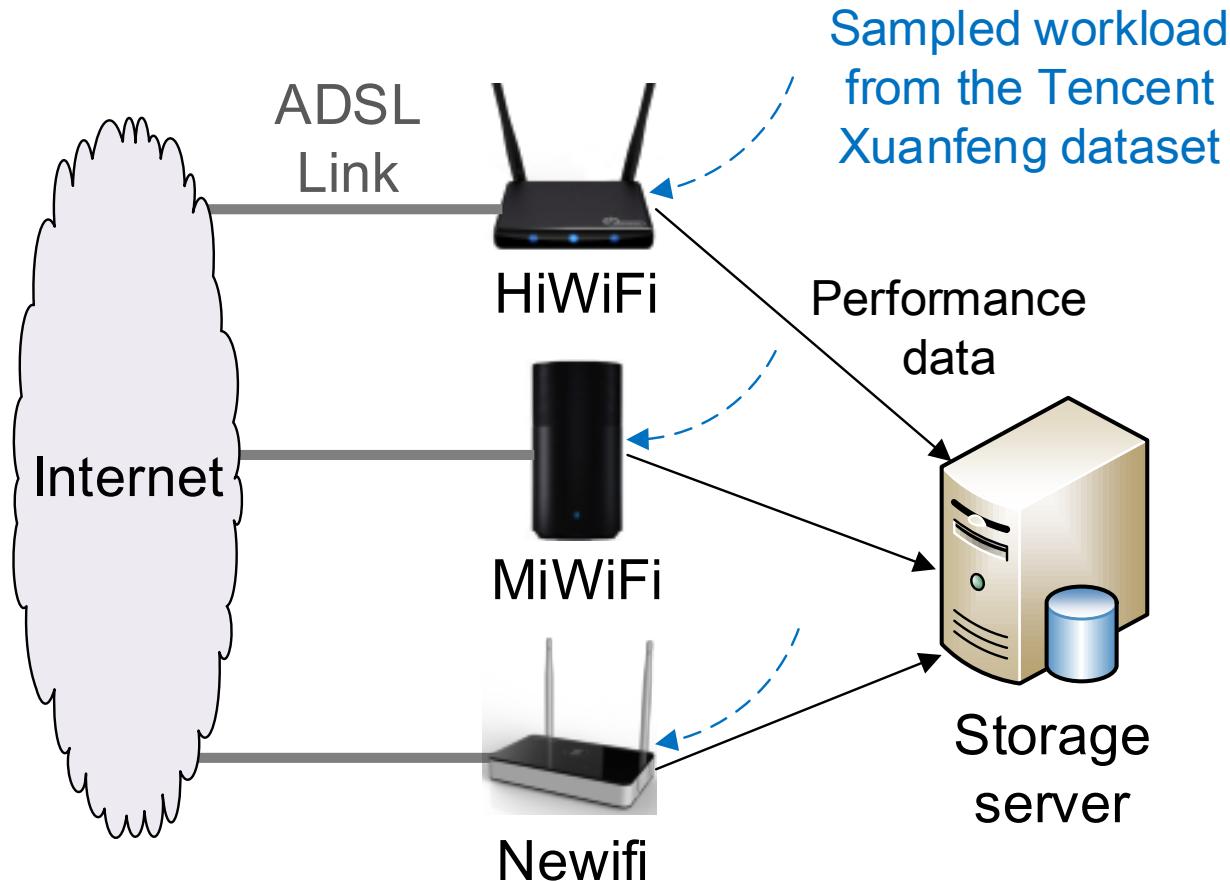


$$\text{SE: } y^c = -a_2 \times \log(x) + b_2$$

SE \approx Stretched Exponential

Matthew effect (for non-videos) +
Fetch-at-most-once effect (for videos)

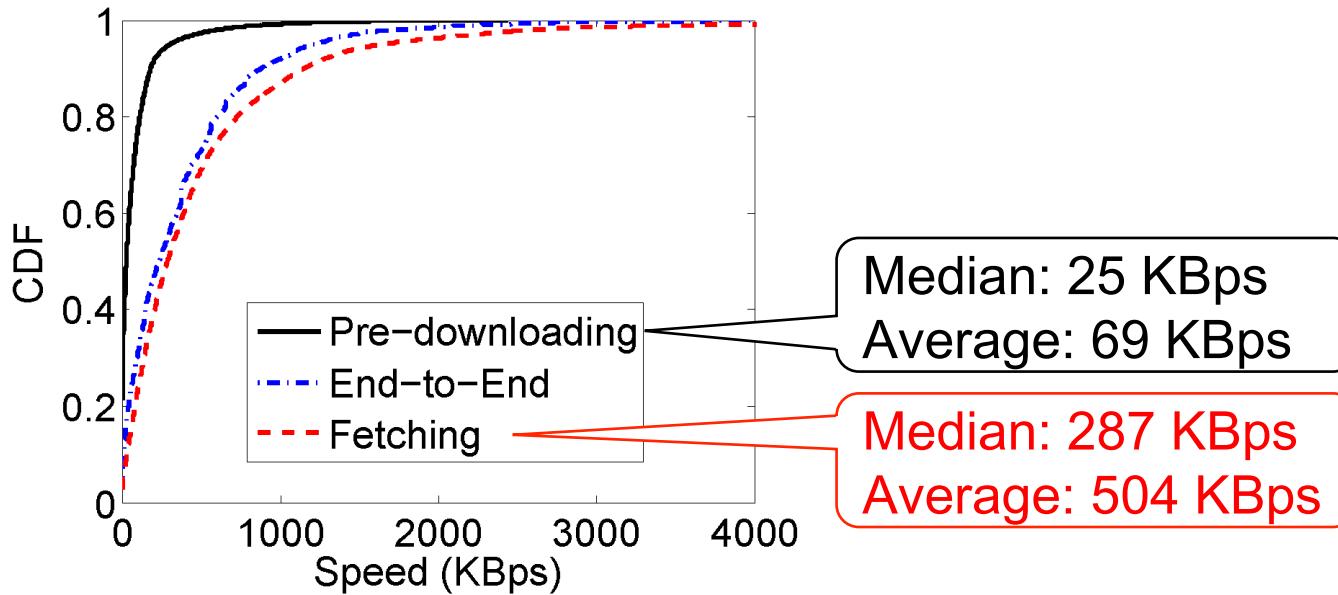
Smart APs: Benchmark



*Note: We assume that the smart AP based offline downloading systems have similar workload characteristics to Xuanfeng, since most end users are not familiar with the technical details and cannot differentiate these services. 23

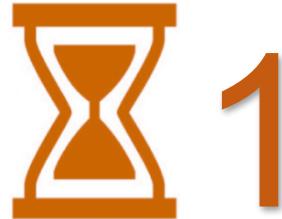
5 Performance

Xuanfeng: Pre-downloading Speed and Fetching Speed



Owing to the privileged network path, Xuanfeng significantly improves users' perceived downloading speeds by 7 – 11 times (fetching speed / pre-downloading speed)

Xuanfeng: Unsatisfactory Fetching Speed



28% of fetching speeds are **below 125 KBps**
(= **1 Mbps**, typical playback bitrate of HD videos)

9.6%
ISP barrier

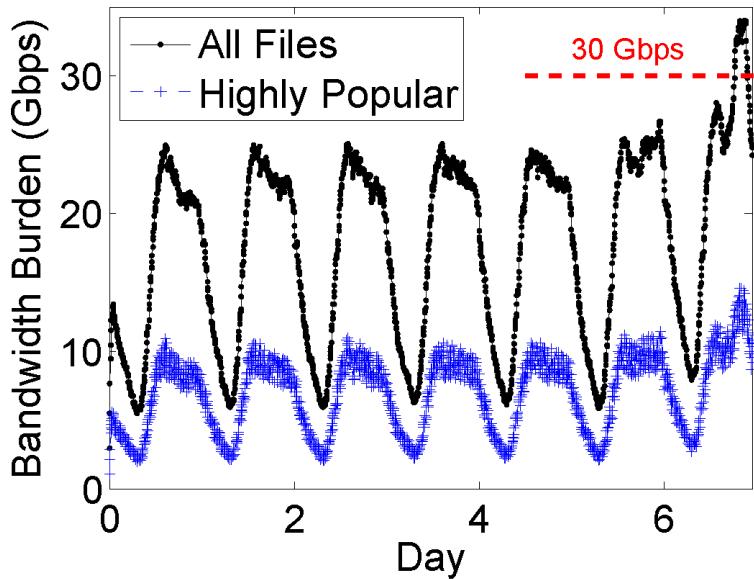
10.8%
Low user-side
access bandwidth

1.5%
Lack of cloud-side
upload bandwidth

6.1%
Unknown...

The cloud-based approach performs poorly once there is a **bandwidth bottleneck** in the privileged network path between the cloud and the user

Xuanfeng: Shortage of Cloud Bandwidth



1.5%

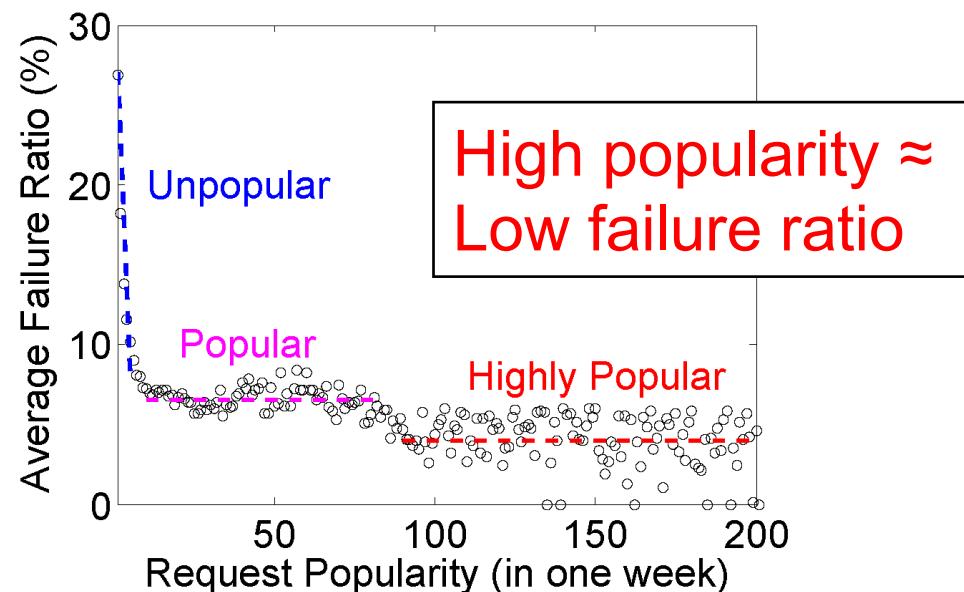
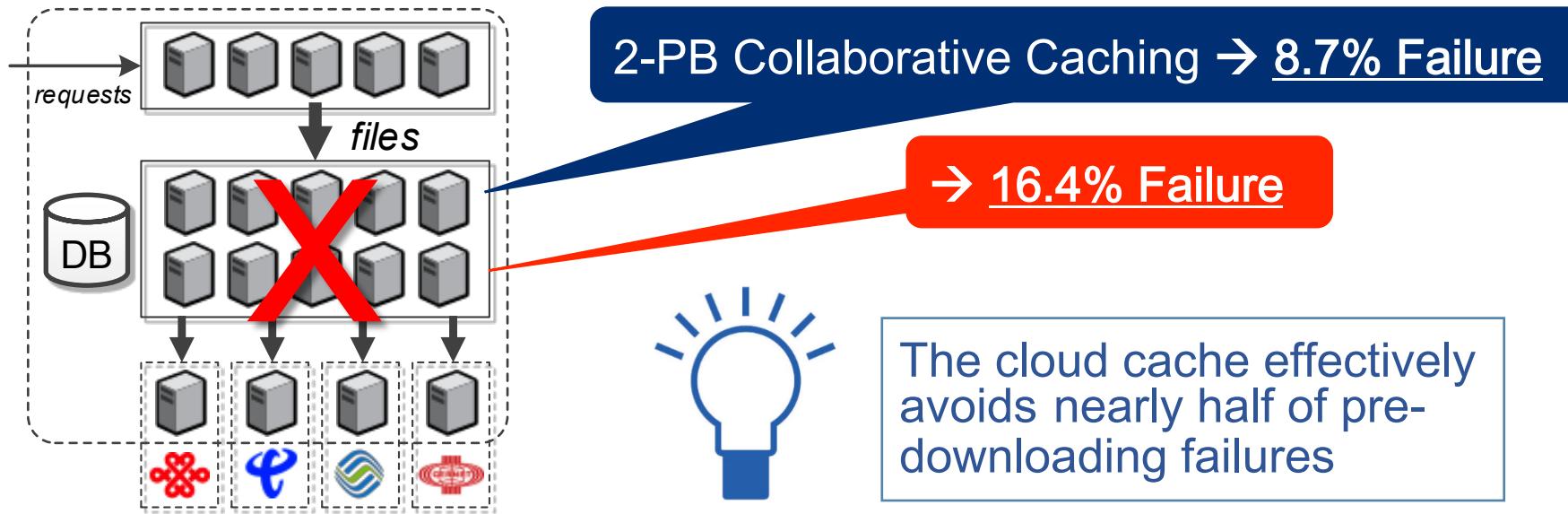
Lack of cloud-side
upload bandwidth

- ◆ 0.84% of highly popular files account for 39% of all downloads
- ◆ 87% of requested files are hosted in peer-to-peer (P2P) data swarms



The cloud is threatened by running out of upload bandwidth due to unnecessarily sending highly popular P2P files. As the user base continues to grow, the cloud will have to reject more ($>1.5\%$) fetching requests.

Xuanfeng: Pre-downloading Failure



Smart APs: Pre-downloading Failure

Failure Ratio	Xuanfeng Cloud	Smart APs
Overall	8.7%	16.8%
Unpopular files	13%	42%

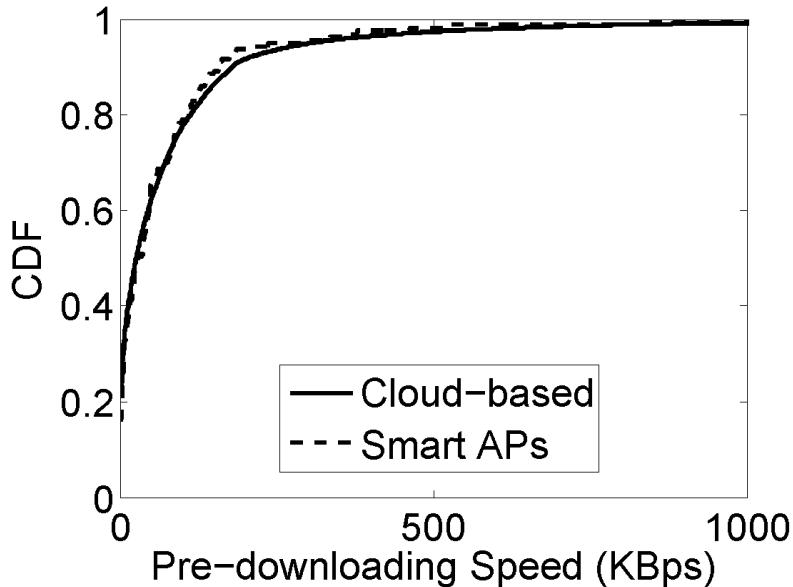
86% Insufficient seeds in a peer swarm
10% Poor HTTP/FTP connections
4% Unknown...

- ◆ 36% of offline downloading requests are issued for unpopular files



Smart APs frequently fail during pre-downloading unpopular files

Smart APs: Pre-downloading Speed



Speed	Xuanfeng Cloud	Smart APs
Median	25 KBps	< 27 KBps
Average	69 KBps	> 64 KBps



A smart AP's pre-downloading speed can be restricted by its hardware and/or filesystem, since some types of storage devices and filesystems do not fit the pattern of frequent, small data writes during pre-downloading

Smart APs: Pre-downloading Speed

Max pre-downloading speed (MBps)	FAT	NTFS	EXT4
HiWiFi + SD card	2.37	–	–
MiWiFi + SATA hard disk drive	–	–	2.37
Newifi + USB flash drive	2.12	0.93	2.13
Newifi + USB hard disk drive	2.37	1.13	2.37

NTFS is incompatible with the OpenWrt OS

USB flash drive is unsuitable
for frequent, small data writes

iowait ratio	FAT	NTFS	EXT4
HiWiFi + SD card	42.1%	–	–
MiWiFi + SATA hard disk drive	–	–	29.7%
Newifi + USB flash drive	66.3%	15.1%	55%
Newifi + USB hard disk drive	42%	9.8%	17.4%

Performance Summary

Xuanfeng Cloud	Smart APs
Bottleneck 1: Unsatisfactory fetching speed	Merit 3: Stably high fetching speed
Bottleneck 2: Shortage of cloud bandwidth	Merit 4: No cloud infrastructure
Merit 1: Effective avoidance of pre-downloading failures	Bottleneck 3: Frequent failures during pre-downloading
Merit 2: No hardware cost at the user side	Bottleneck 4: Hardware/filesystem restrictions on pre-downloading

The two approaches are subject to distinct performance bottlenecks

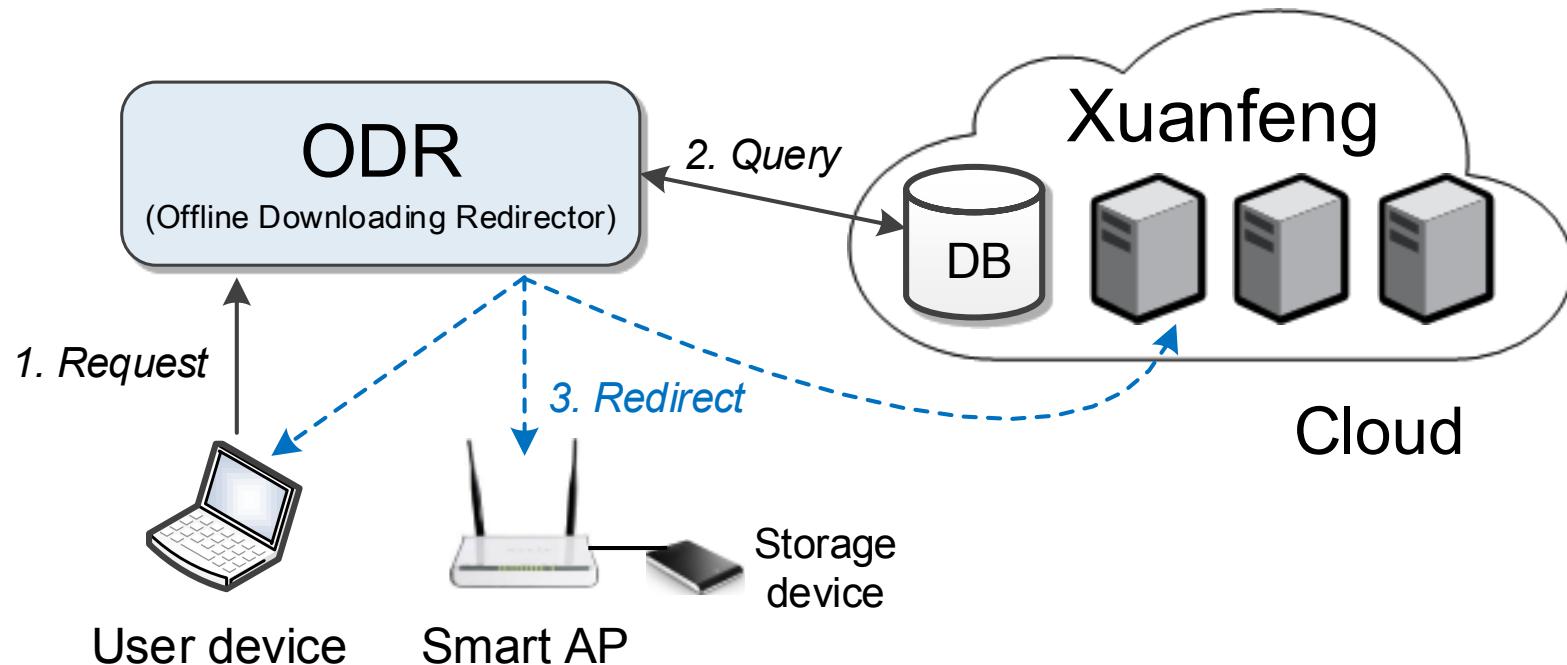


while also being complementary to each other ☺

6 Optimization

ODR Middleware

- Help users automatically select a proper (offline) downloading way



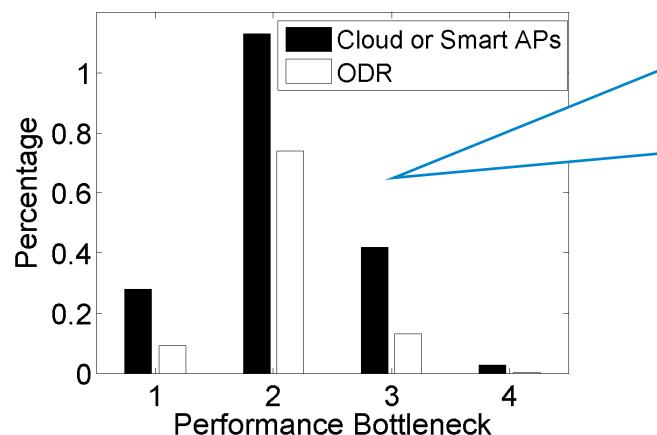
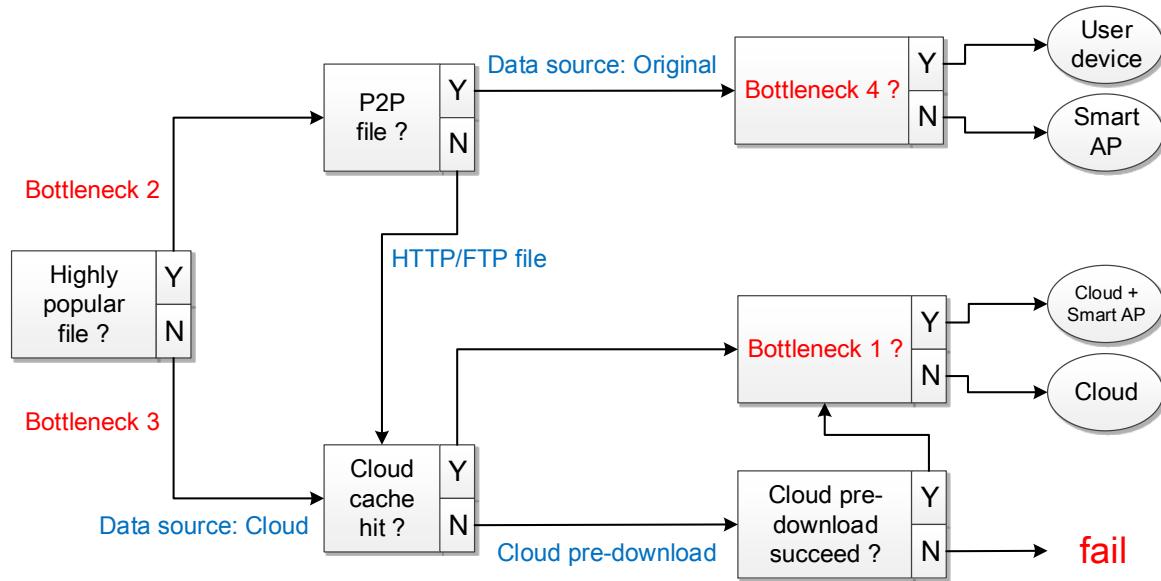
- Primary goal: minimizing the downloading time and failure ratio
- Secondary goal: minimizing the upload bandwidth burden on the cloud

ODR Web Site <http://odr.thucloud.com>



ODR Design & Performance

Heuristic rules extracted from measurement findings



Even using this coarse-grained design, we manage to remarkably reduce the current performance bottlenecks of offline downloading !

Summary

- What Offline Downloading is and Why it is popular in China
- Selection Dilemma of Offline Downloading
- Measurement findings of Xuanfeng and popular smart APs
 - ◆ The two approaches are subject to distinct performance bottlenecks, while also being complementary to each other.

We feel that offline downloading has broad applicability to other areas of the world that lack broadband penetration. By deploying offline downloading technologies, coupled with our proposed ODR middleware, the Internet experiences for users in many parts of the world can be improved.



Thank you! Any questions?



Xuanfeng is pronounced as
[ʃuæn fəŋ] or simply XF



HiWiFi



Xunlei is pronounced as
[ʃunlei] or simply XL



MiWiFi



Baidu
CloudDisk



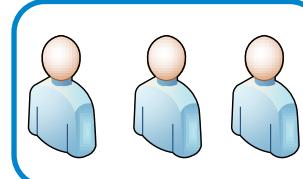
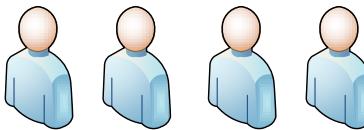
Newifi

Backup slides

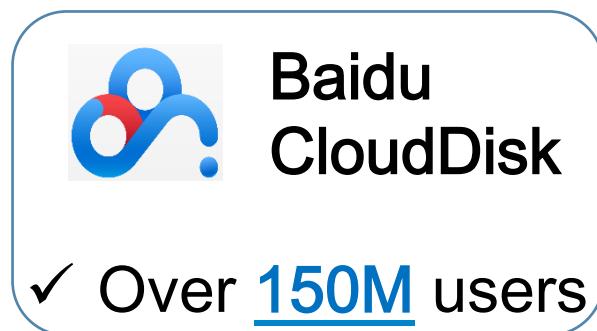
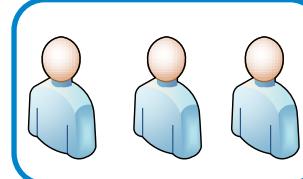
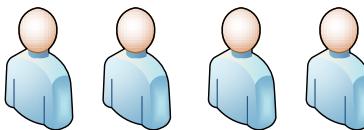
Xunlei, Xuanfeng, and Baidu CloudDisk



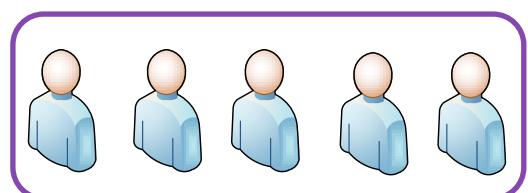
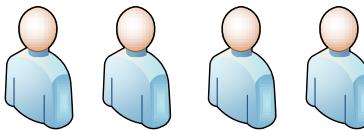
➤ Launched in 2009



➤ Launched in 2010



➤ Launched in 2012



Xunlei, Xuanfeng, and Baidu CloudDisk



Xunlei

✓ Over 80M users

Charges each user nearly \$2 per month



Tencent
Xuanfeng

✓ Over 30M users

Conditionally free --- should be the VIP user of Tencent



Baidu
CloudDisk

✓ Over 150M users

Totally free --- the user even does not need to register

Offline Downloading outside China

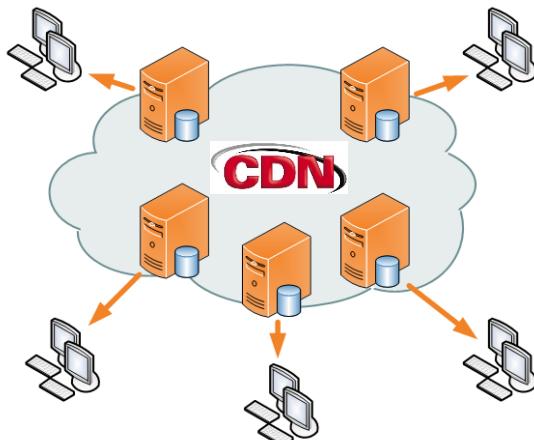
- ◆ Besides those developing countries, developed countries can also benefit from offline downloading (based) services.



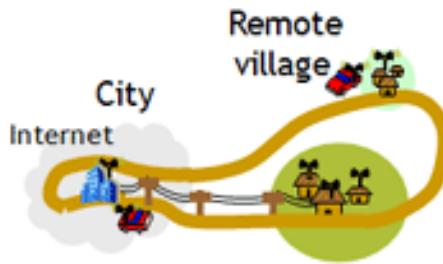
LINKSYS
SMART WI-FI ROUTERS



State-of-Art Downloading Techniques



- ❑ CDN typically only help to deliver files for content providers who pay for the service.
- ❑ The business model of offline downloading is the opposite of CDN, because it charges (or sometimes frees) its users, *i.e.*, content receivers, for better downloading experiences.

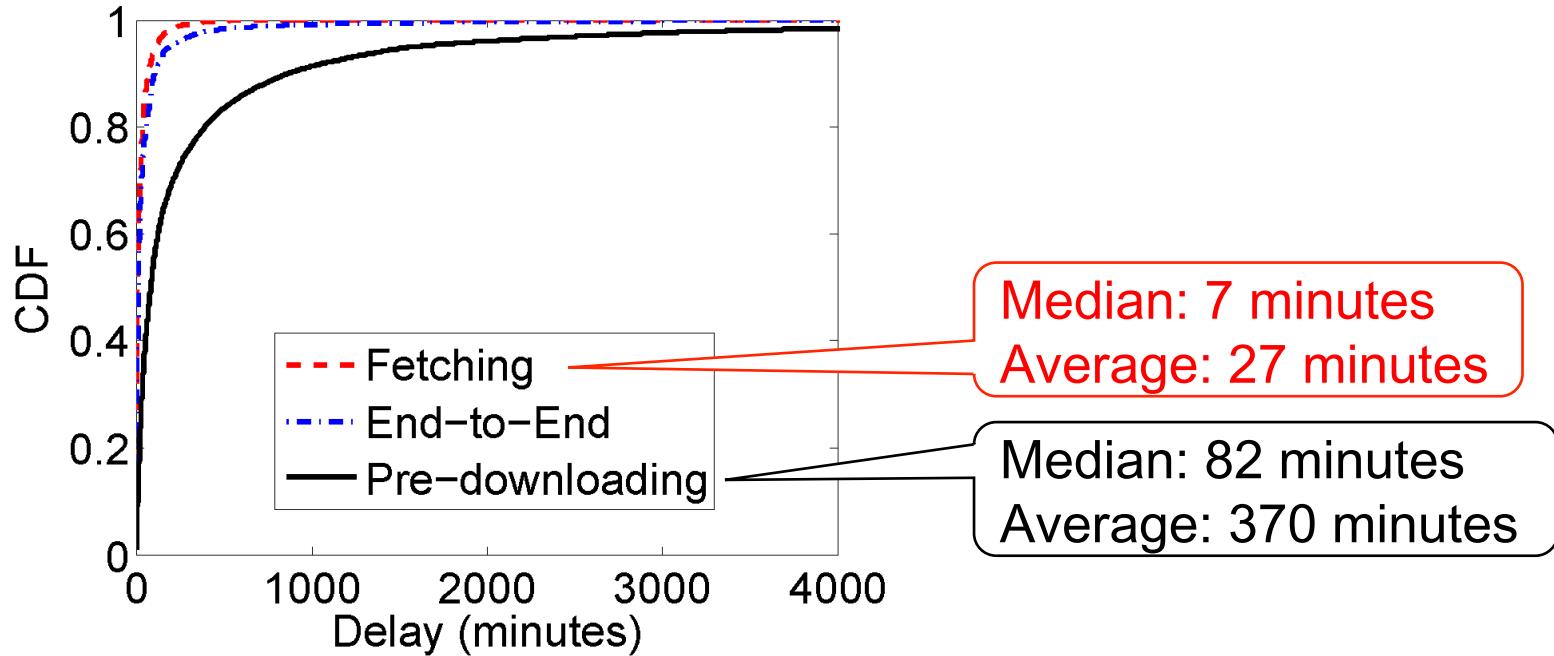


*Offline Downloading ≈
DTN + reverse-CDN*

ICN
CCN
NDN

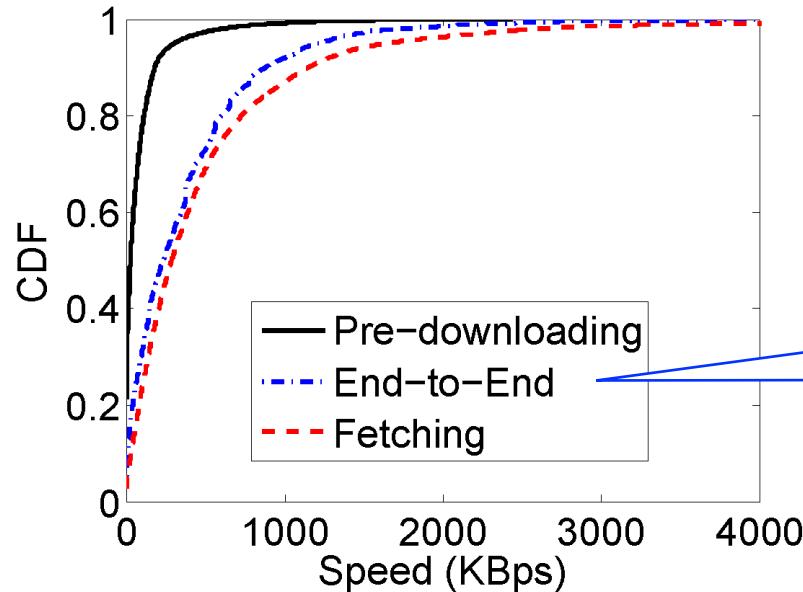
- 1) in-network storage for caching
- 2) decoupling content senders and receivers
- 3) disruption tolerance
- 4) multi-party communication through replication
- 5) mobility and multi-homing

Performance: Delay



Xuanfeng reduces users' perceived downloading delay by **12 – 14 times**

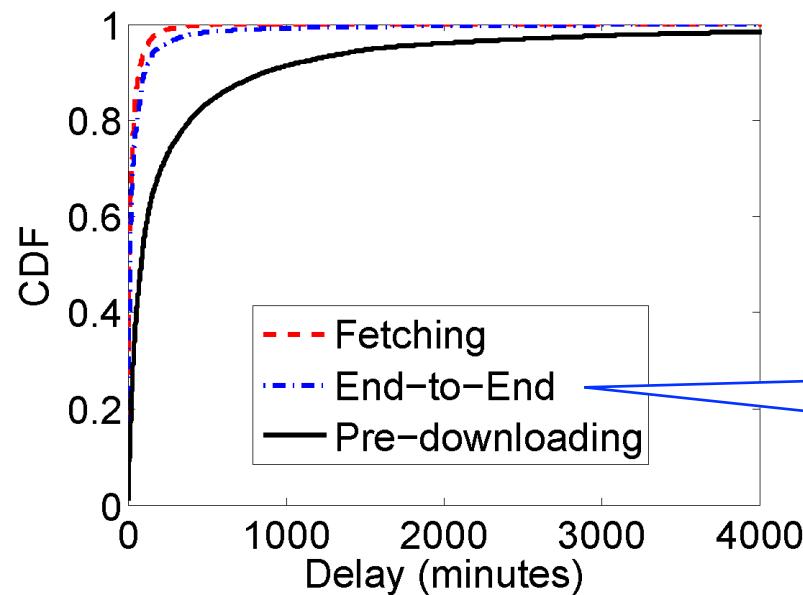
Performance: End-to-End



Median: 233 KBps
Average: 380 KBps



The collaborative cache of Xuanfeng remarkably avoids 89% pre-downloads



Median: 10 minutes
Average: 68 minutes