Network Traffic Measurements

Applications to Internet Services and Security

Enrico BOCCHI
CERN, IT-Storage

GTTI – Riunione Annuale 2017
Premio per Tesi di Dottorato
Udine, 21-23 Giugno

Research Scope

Users producing and consuming content

High bandwidth access

Different types of devices

Ubiquitous connectivity

Outbound Traffic

- Photo backup on Dropbox from tablets and smartphones
- Documents editing on Google Drive
- Video publishing and sharing on YouTube

Unsolicited Traffic

- Harming users' privacy
- Threatening data confidentiality
- Exposing identities to malicious adversaries

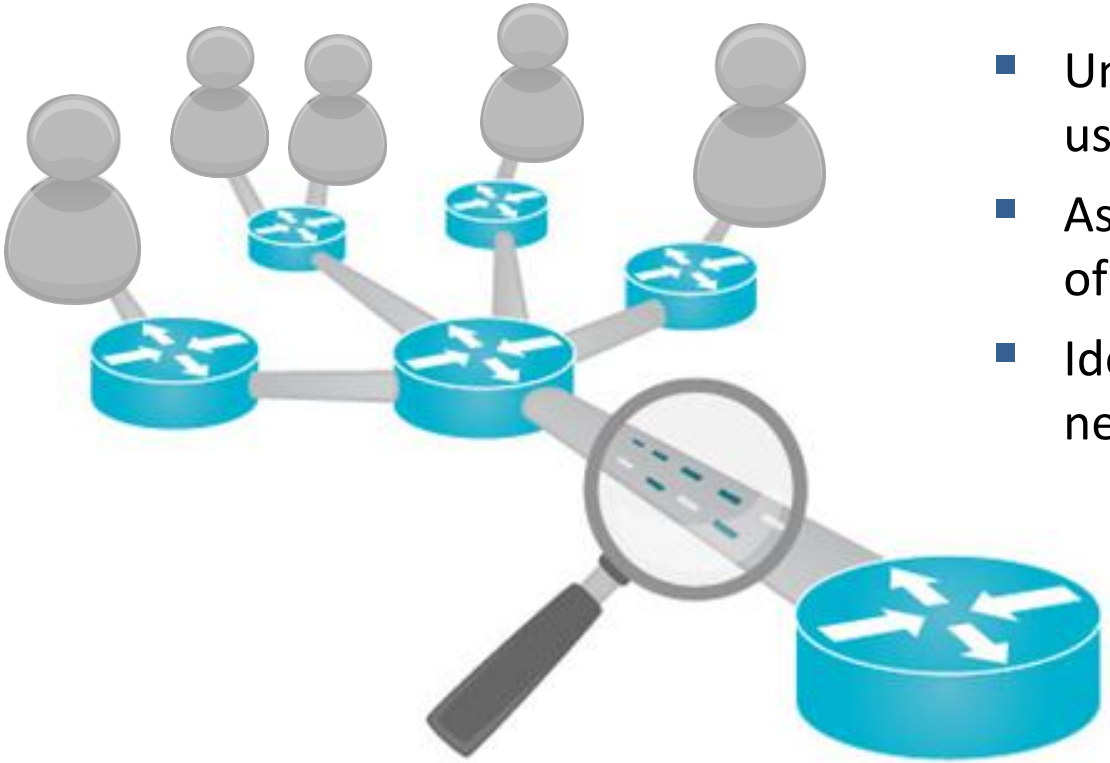
Inbound Traffic

- Web search and information retrieval
- Mobile Apps download
- Video streaming

NETFLIX hulu Rai.tv

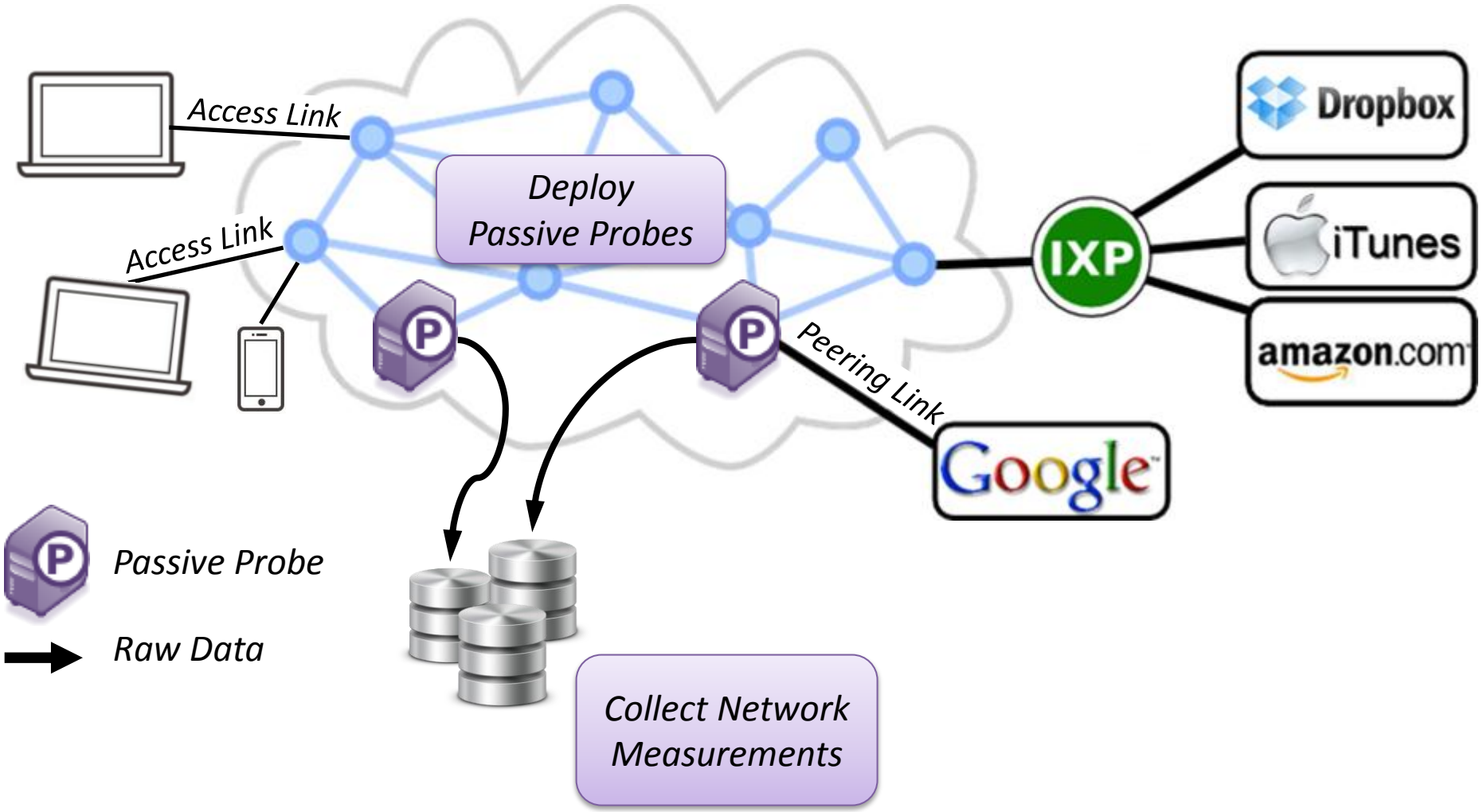
Research Scope

Communication networks collect users' traffic exchanged with websites, Internet services, cloud-based applications, etc.

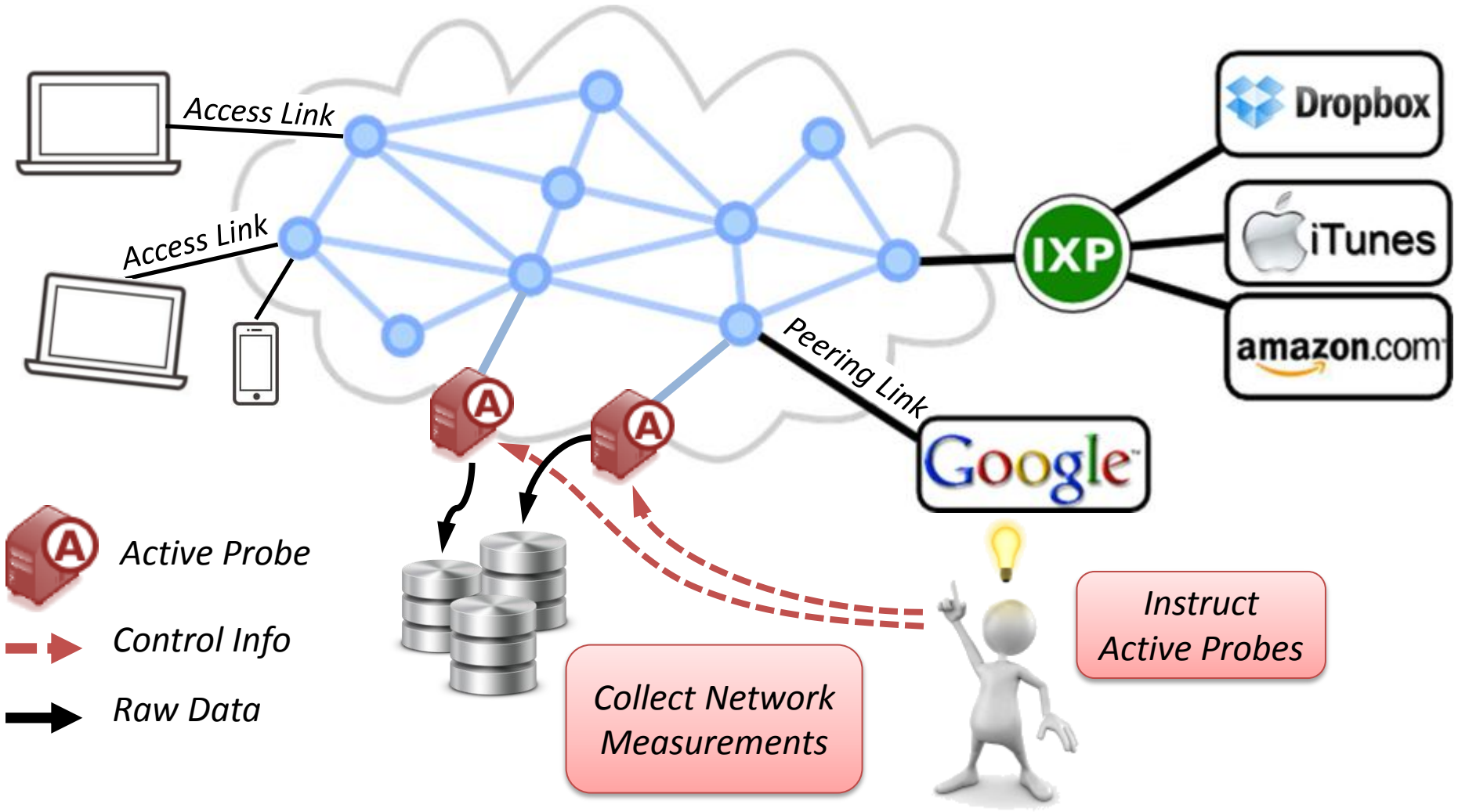


- Understand and characterize users' activities
- Assess performance of Internet services
- Identify and troubleshoot network impairments

Passive Measurements



Active Measurements



Data Processing Workflow

Data Flow

Applications

Visualization

Detection

Configuration

Analytics

Apply suitable techniques to interpret data
e.g., machine learning, statistics, data mining, etc.

Filtering

Query, filter, and aggregate data
to access the measurements of interest

Storage

Build a repository of measurements
either via passive captures or active probing

Research Topics

Research Topics

Outbound Traffic

- Photo backup on Dropbox from tablets and smartphones
- Documents editing on Google Drive
- Video publishing and sharing on YouTube

Unsolicited Traffic

- Harming users' privacy
- Threatening data confidentiality
- Exposing identities to malicious adversaries

Inbound Traffic

- Web search and information retrieval
- Mobile Apps download
- Video streaming

NETFLIX hulu Rai.tv

Research Topics

Traffic Type	Research Topic	Research Objectives	Employed Methodologies
Outbound	Personal Cloud Storage	**Monitoring and benchmarking of personal cloud storage services** 1. Understanding of typical usage 2. Benchmarking of performance 3. Detection of advanced capabilities	i. Passive Traffic Measurements ii. Active Traffic Measurements a. Large-scale Data Analysis b. Reverse Engineering c. Ad-hoc Testbed Design
Inbound	Web Quality of Experience	**Understanding differences between HTTP/1 and /2 from user standpoint** 1. Survey on QoE metrics for Web 2. Collection of 4,000+ MOS grades 3. Impact of Carrier Grade NAT	i. Active Traffic Measurements ii. Passive Traffic Measurements a. Ad-hoc Testbed Design b. Statistics
Unsolicited	Malware Detection	**Automatic detection and classification of malicious activities** 1. Characterization of malware 2. Graphical representation of security incidents	i. Passive Traffic Measurements a. Data Mining b. Machine Learning c. Data Classification

Personal Cloud Storage

Cloud Storage Services

Personal cloud storage services are customary among Internet users

- Synchronization among multiple devices
- Collaborative work and document editing
- Content sharing with friends and colleagues

Market is crowded by offers

- Providing a significant amount of free storage
- Relying on ad-hoc protocols and designs

Cloud Storage Services

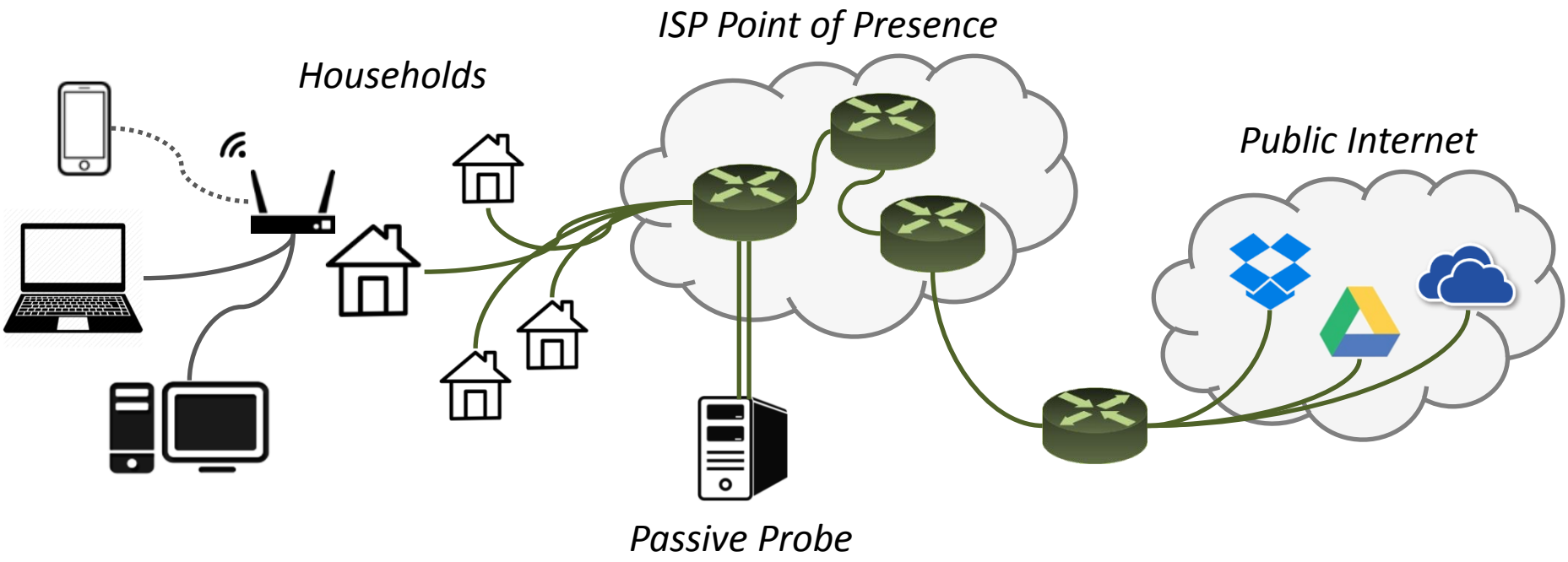
- What is the **typical usage** of these services?
  - What is the workload they have to face?
  - Are mobile devices popular?
  - Is there a dependency between workload and device type used?

Passive Measurements

- How is **synchronization** tackled?
  - Do clients implement advanced capabilities?
  - How long does it take to synchronize devices?

Active Measurements

Passive Characterization



- **1 month** of real traffic recorded
- **20,000 households**, residential customers located in Europe
- Subscribers with **ADSL** (20Mb/s) or **FTTH** (100Mb/s) lines

Usage Patterns per Device Type

Focus on Dropbox

	Households	Upload (GB)	Download (GB)	Volume (GB)
PC Client	2,196	462	980	1442
Mobile App	1,628	180	51	231
Web Interface	3,823	38	177	215
Total	**5,020**	**680**	**1,208**	**1,888**

Usage Patterns per Device Type

Focus on Dropbox

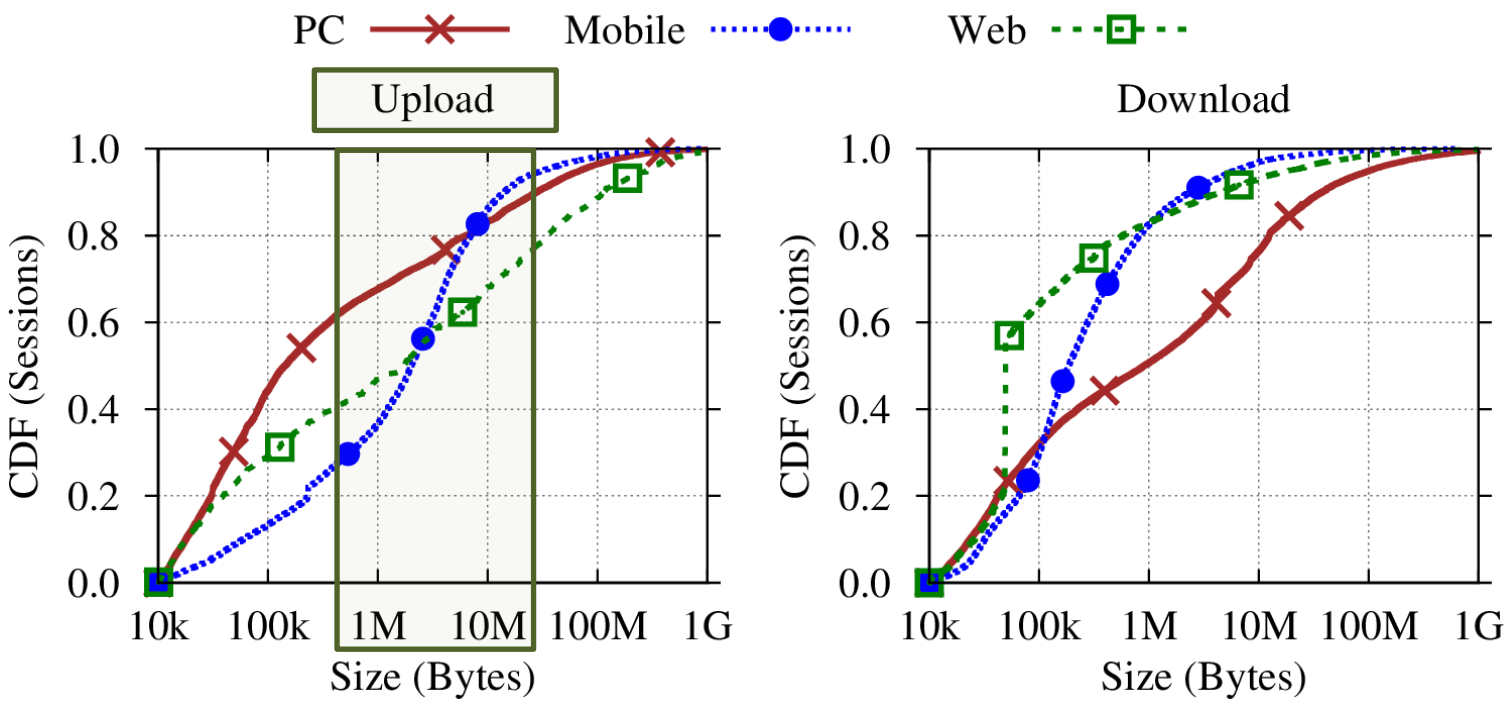
	Households	Upload (GB)	Download (GB)	Volume (GB)
PC Client	2,196	462	980	1442
Mobile App	1,628	180		
Web Interface	3,823	38		
		680		

3,581 households access the Web Interface to consume **direct links**

Mobile devices store photos and videos on the cloud

PCs have to download the content produced by other devices

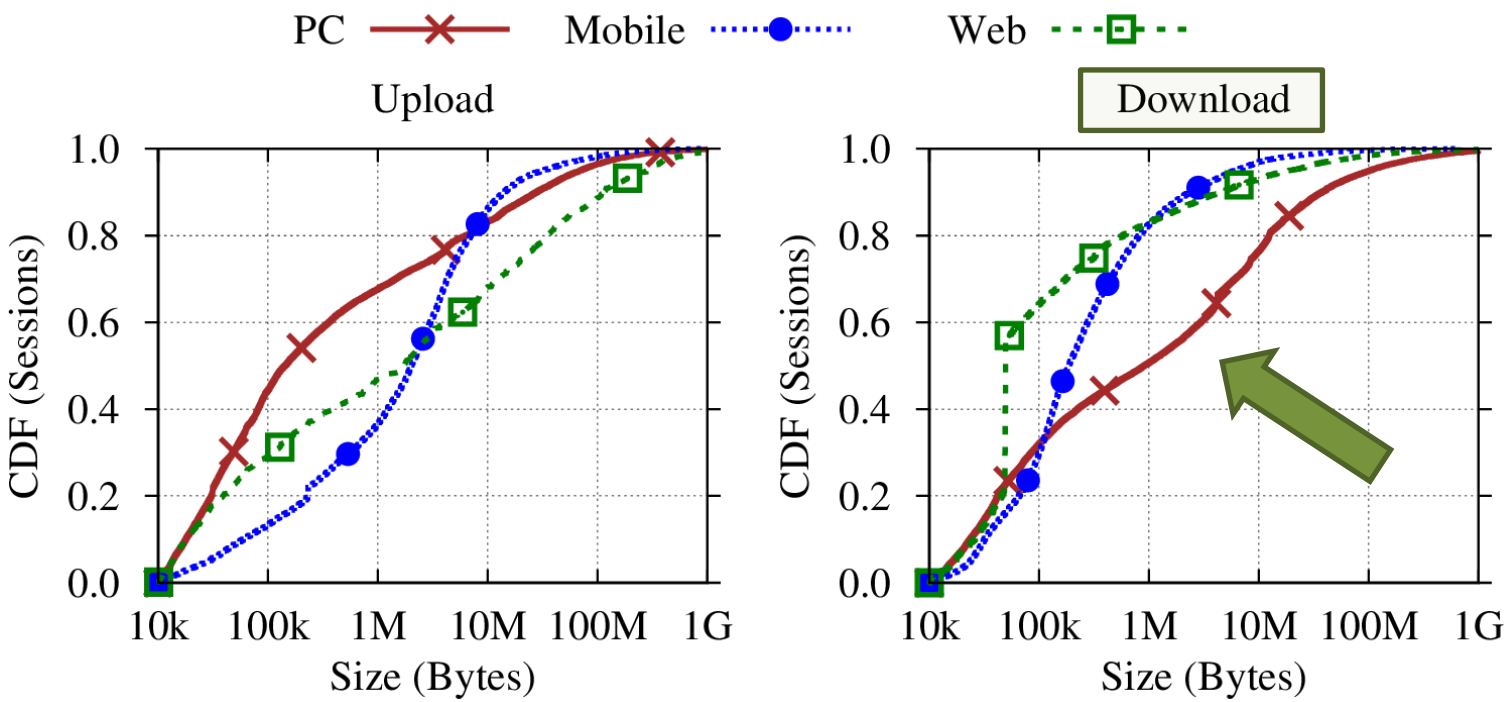
Typical Workloads



Storage sessions: Merge storage flows overlapped in time

- Uploads
  - 50% are in the 1÷10MB range for **mobile devices** – photo and video backup
  - Small for PC clients: 70% carry less than 1MB

Typical Workloads



Storage sessions: Merge storage flows overlapped in time

- Downloads
  - PC clients download more bytes – bootstrap synchronization
  - Web access CDF is biased toward 50kB – picture thumbnails

Typical Workloads

TAKEAWAY

- **Usage patterns differ** according to the device type
- PC clients dominate in download to keep the local folder synchronized
- Mobile and Web access are used for specific tasks

Cloud Storage Services

- What is the **typical usage** of these services?
  - What is the workload they have to face?
  - Are mobile devices popular?
  - Is there a dependency between workload and device type used?

Passive Measurements

- How is **synchronization** tackled?
  - Do clients implement advanced capabilities?
  - How long does it take to synchronize devices?

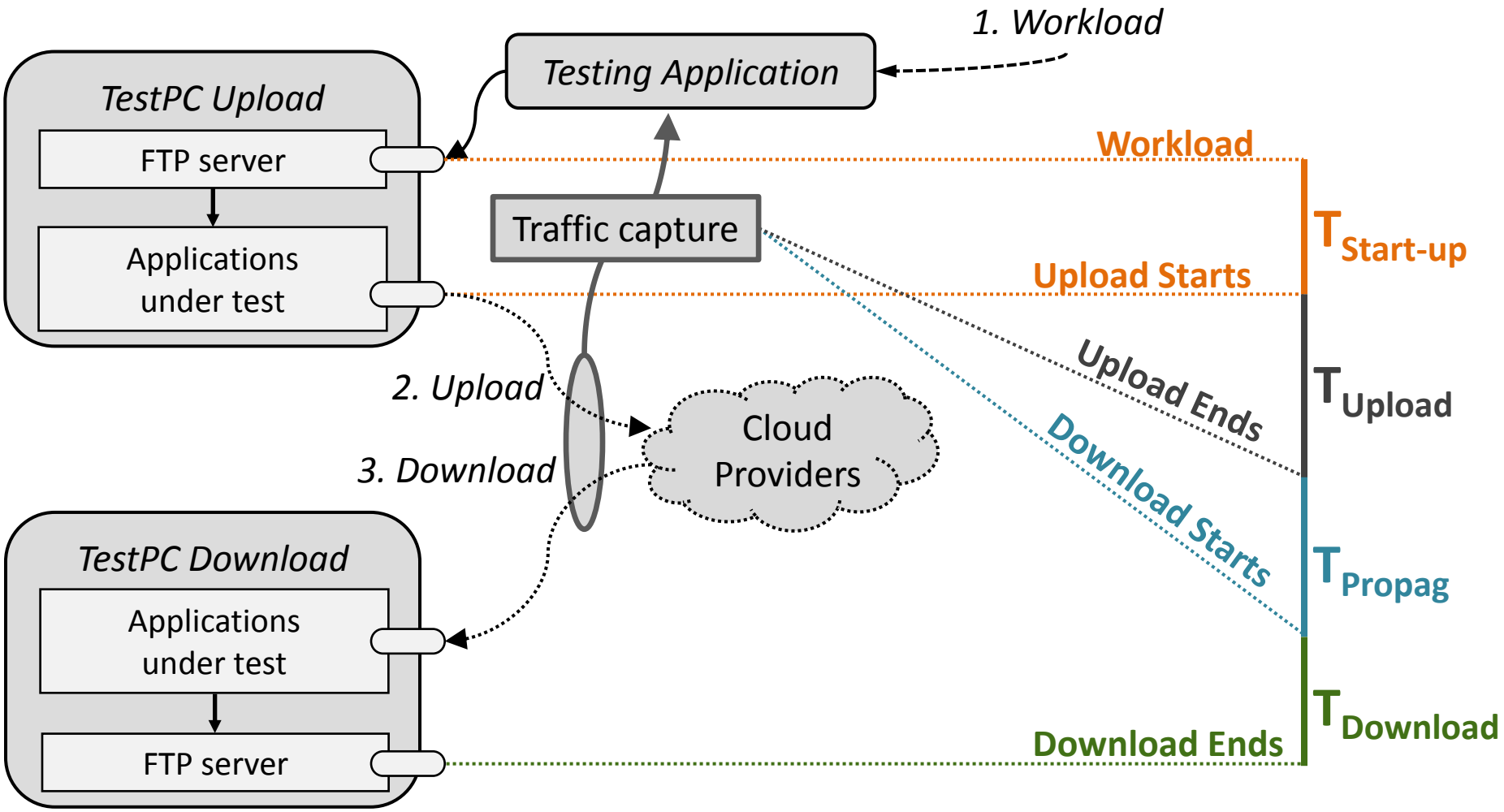
Active Measurements

Active Measurements

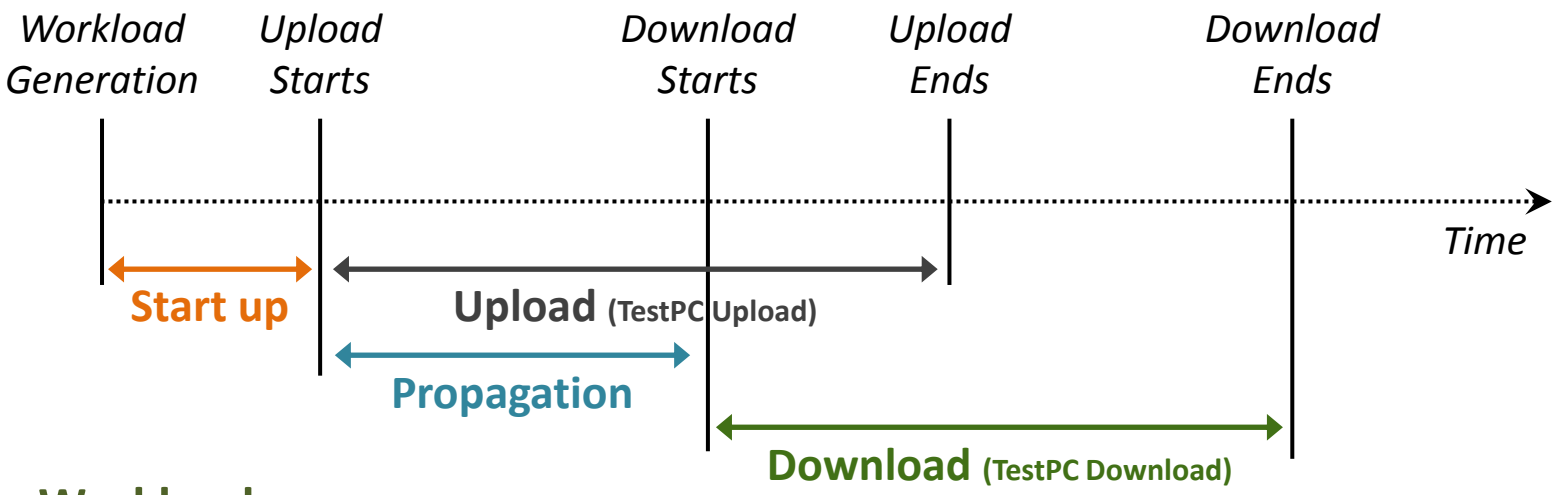
How is synchronization tackled?

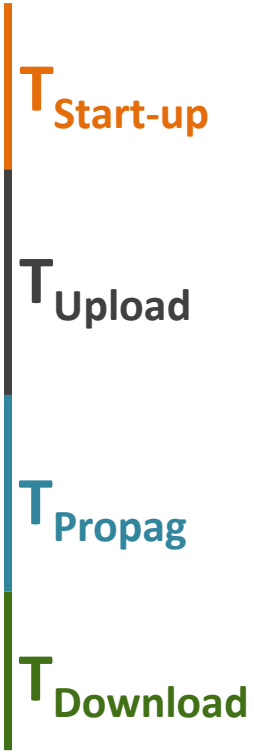
- How long does it take to **synchronize devices**?

Active Measurements



Active Measurements





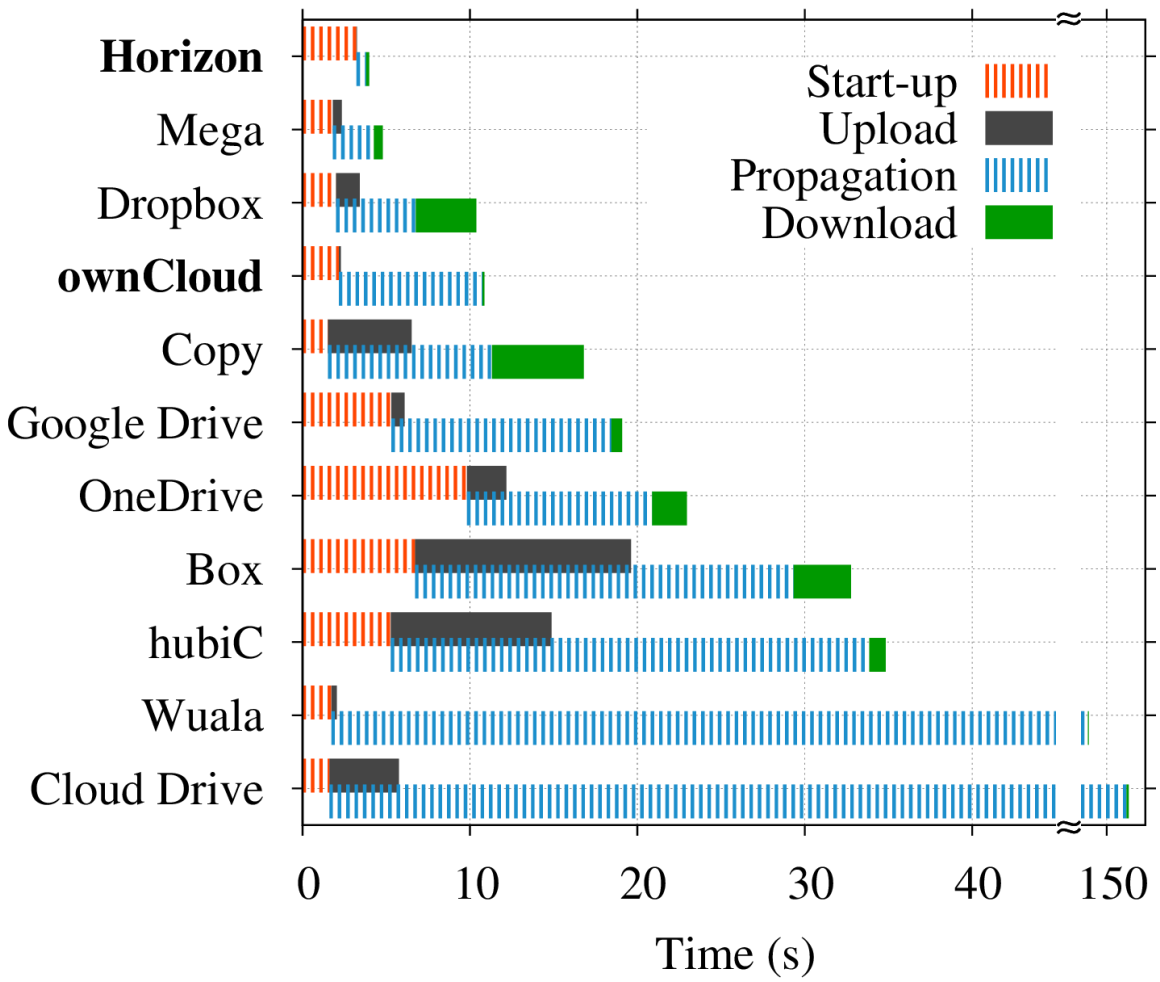
Workloads

Files	Binary	Text	Total Size	Replicas
1	1	-	100 kB	-
1	1	-	1 MB	-
1	1	-	20 MB	-
100	50	50	1 MB	-
365	194	171	87 MB	97 (5.4 MB)
312	172	140	77 MB	136 (5.5 MB)

Synchronization Delay

Several workloads

- Single file, 1MB



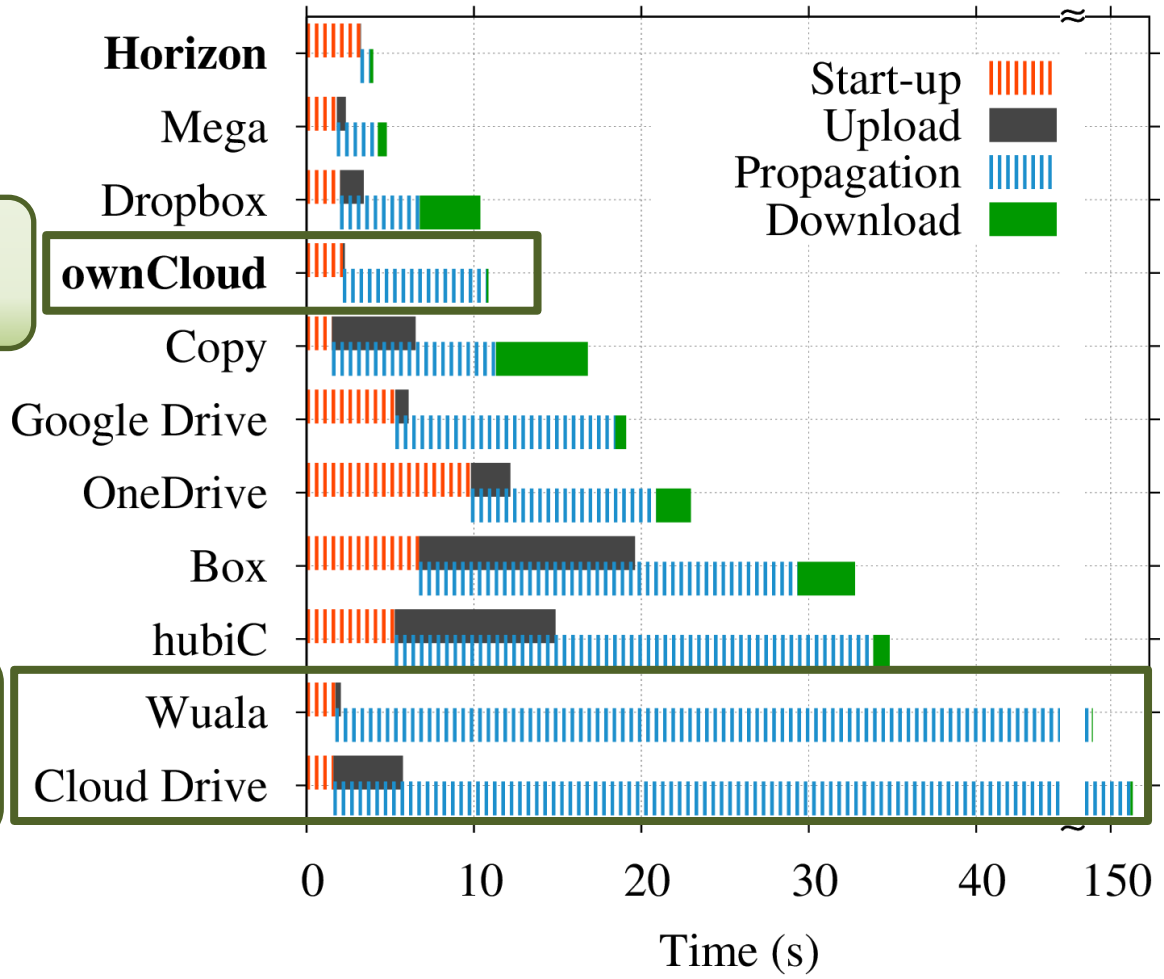
Synchronization Delay

Several workloads

- Single file, 1MB

ownCloud takes more than 10s to synchronize 1MB

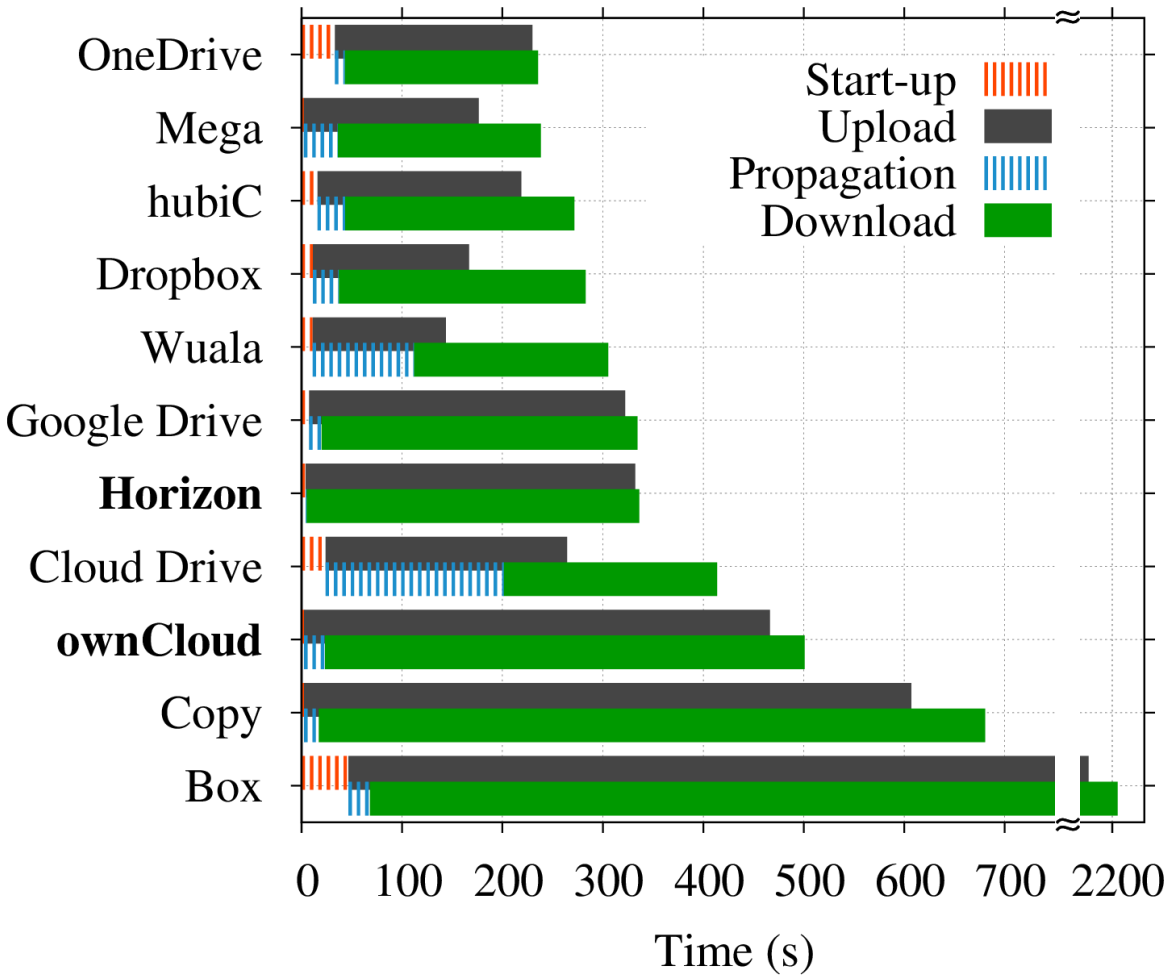
Wuala and **Cloud Drive** are severely limited by implementation choices



Synchronization Delay

Several workloads

- Single file, 1MB
- Realistic – 365 Files
  - 194 binary, 171 text
  - 87MB total size
  - 97 files replicated (5.4MB)



Synchronization Delay

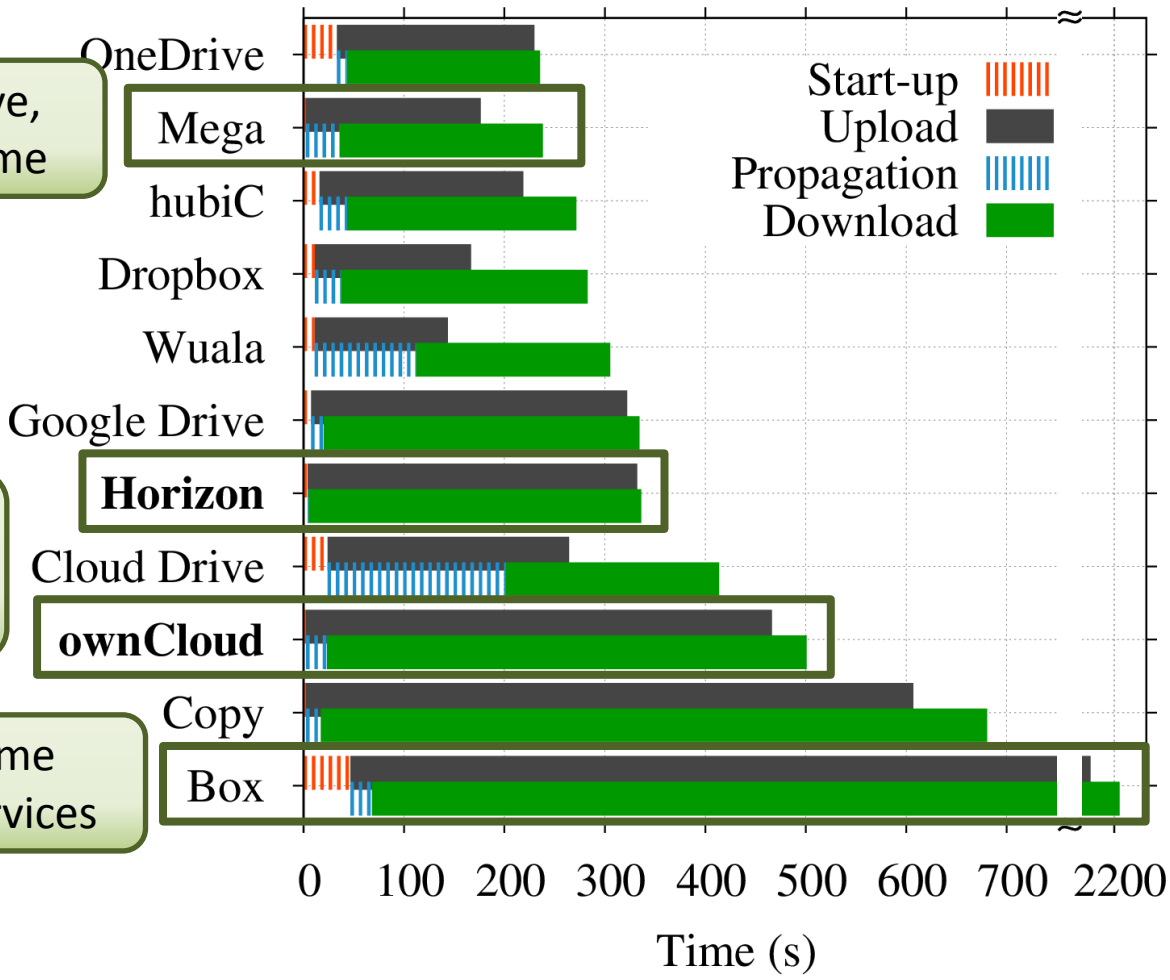
Several workloads

- Realistic – 509 Files
  - 194 binary, 171 text
  - 87MB total size
  - 97 files replicated (5.4MB)

Mega is the most reactive, with very low start-up time

Horizon and **ownCloud** perform poorly with complex workloads

Box requires 10x time compared to best services



Synchronization Delay

Several workloads

- Single file, 1MB
- Realistic – 365 Files
  - 194 binary, 171 text
  - 87MB total size
  - 97 files replicated (5.4MB)

TAKEAWAY

- Client implementation changes according to the storage provider
- Performance is impacted by (i) **client capabilities**, (ii) **protocol design**
- Strong dependency with **workload** used

Conclusions

Conclusions

- Focus on users' perceived **performance**, **experience**, **security** following a measurement-based approach
  - Identification of performance bottlenecks for service improvement
  - Quality of experience assessment for end-users
  - Automatic detection of malicious activities

- Combine diverse measurements techniques to gain **full visibility** on the network traffic
  - Active measurements through ad-hoc testbed deployment
  - Large-scale passive data collection and post-processing

- Achieve a **complete understanding** and extract knowledge from network activities leveraging various techniques

  E.g., data science, machine learning, statistics, etc.

Network Traffic Measurements

Applications to Internet Services and Security

POLITECNICO
DI TORINO

Enrico Bocchi

GTTI – Riunione Annuale 2017
Udine, 21-23 Giugno