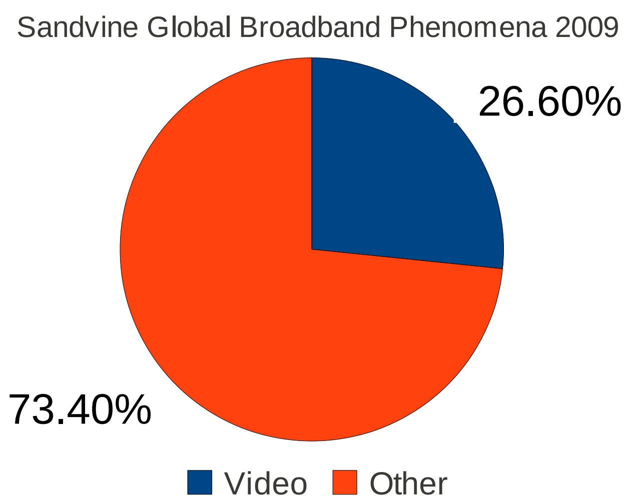
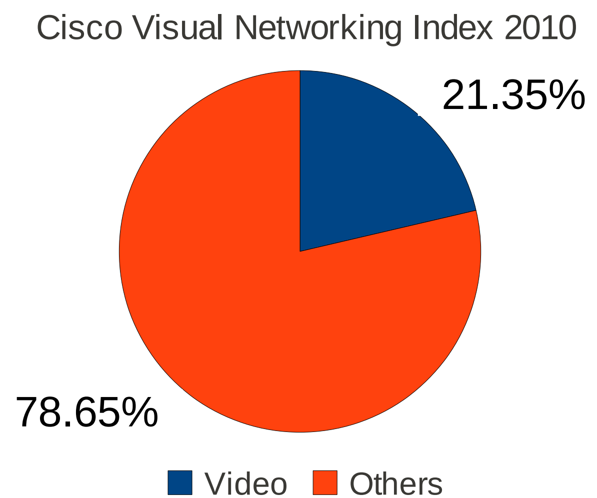
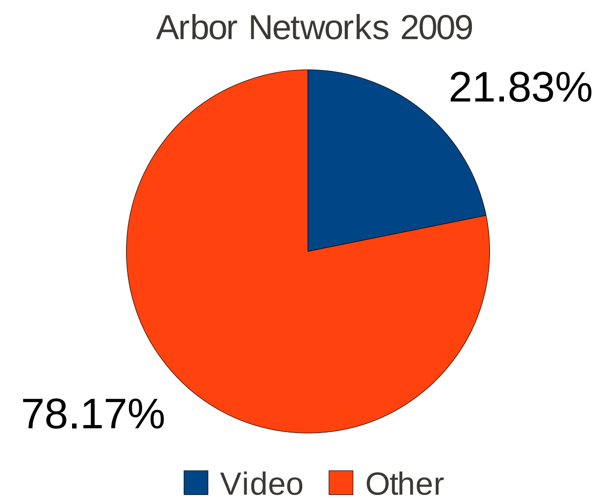
Dissecting Video Server Selection Strategies in the YouTube CDN

Ruben Torres, Alessandro Finamore, Marco Mellia, Maurizio Munafó and Sanjay Rao

Video Traffic is Dominant







- More than 20% of the Internet traffic is video
  - Popularity of video-on-demand services
  - New ways to access video (mobile devices,etc)
- Critical to characterize this new trend

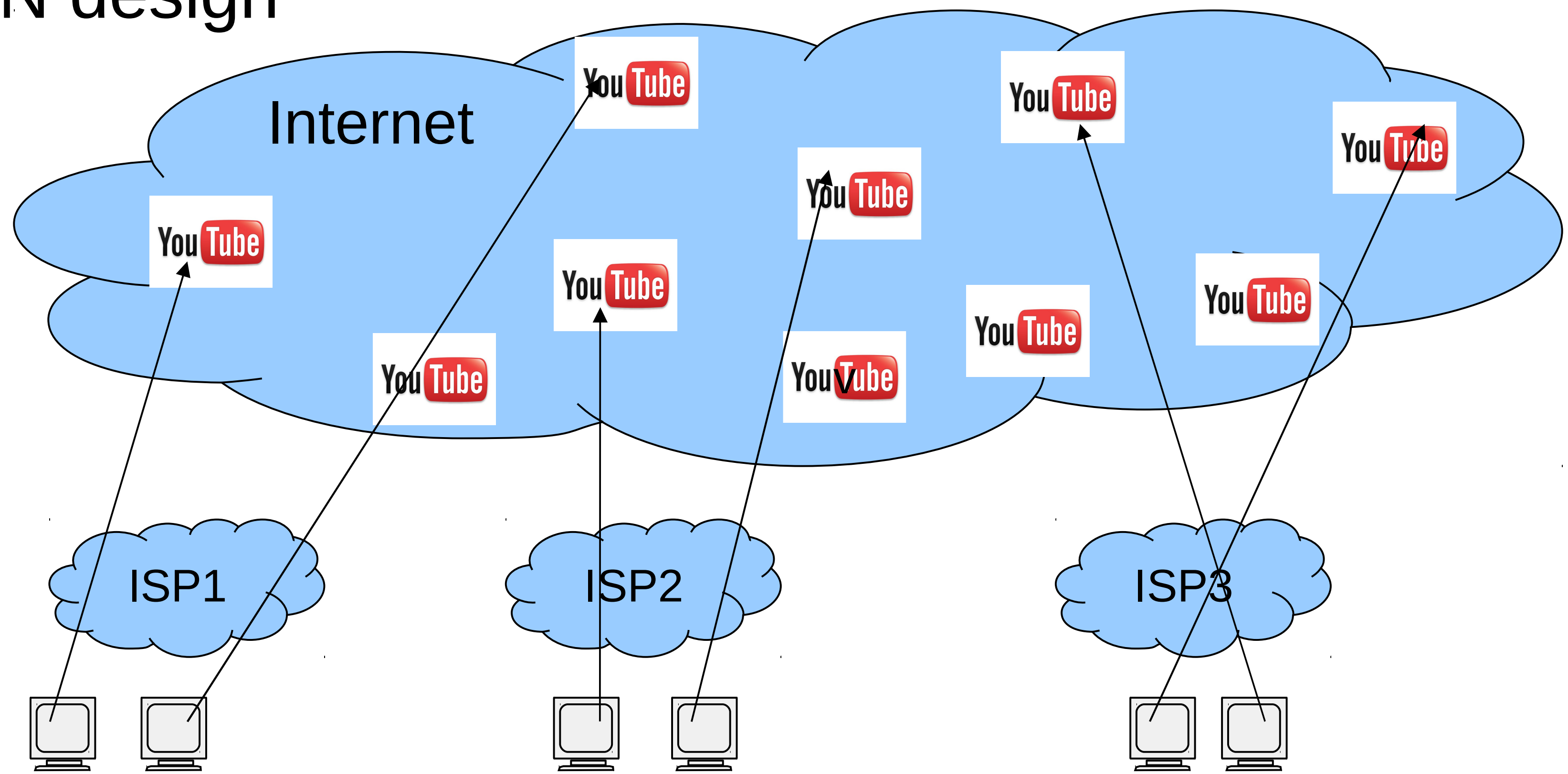
The main source of video on the Internet

- 3 billion videos viewed per day **
- 48 hours of videos uploaded every minute **
- 100's of thousand videos uploaded daily **
- 3rd most popular website after Google and Facebook [Alexa Ranking]
- 20-30% of all incoming traffic in some of our datasets

**http://youtube-global.blogspot.com/2011/05/thanks-youtube-community-for-two-big.html

Internet Video Distribution

Typical CDN design



What factors affect server selection in the YouTube CDN?

How do these factors affect user performance?

Why is this Work Important?

- **Motivation** for the design of new architectures for video delivery
- **Stepping stone** for what-if analysis:
  - Changes in DC placement, video popularity distribution or video quality:
    - How may ISP traffic patterns be impacted?
    - What is the impact on user performance?
- **Inform** ISPs of YouTube traffic patterns and user performance, which may not be obvious to operators

Contributions of this Work

- Shed some light on the mechanisms used by YouTube to direct clients to content servers:
  - DNS based load balancing
  - Application layer redirections
- Unlike prior work, we show that latency between clients and DCs plays a role in server selection
- Several other factors can play a role:
  - Load balancing
  - Popular videos causing hot spots
  - Availability of rare content
  - DNS server queried
- Show the impact server selection has on user performance

Unique Datasets from Large Networks

Dataset	YouTube Flows[x1000]	Volume[TB]	#YouTube Servers	#Clients
US-Campus	874	7.1	1985	20443
EU1-Campus	134	0.6	1102	1113
EU1-ADSL	877	3.7	1977	8348
EU1-FTTH	91	0.4	1081	997
EU2	513	2.8	1637	6552

Details

2 continents

2 University Campuses and 2 ISPs

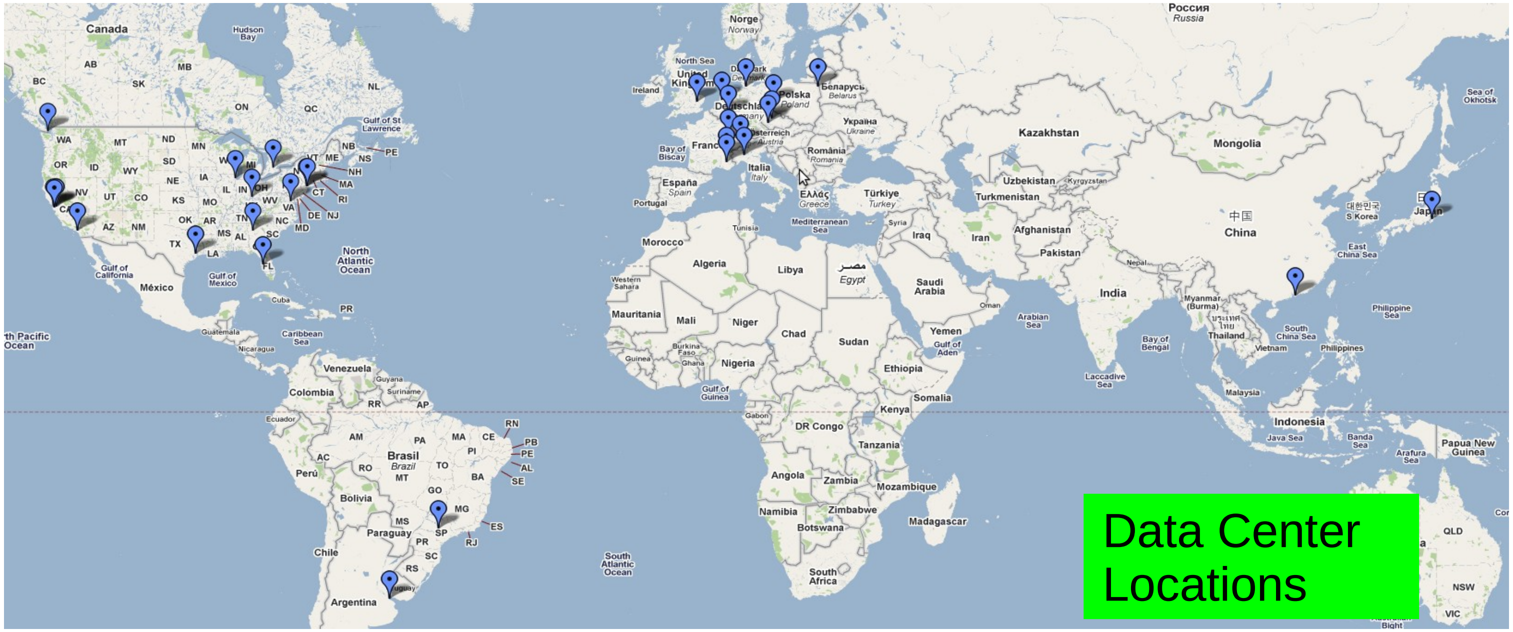
3 different access technologies

Simultaneous collection

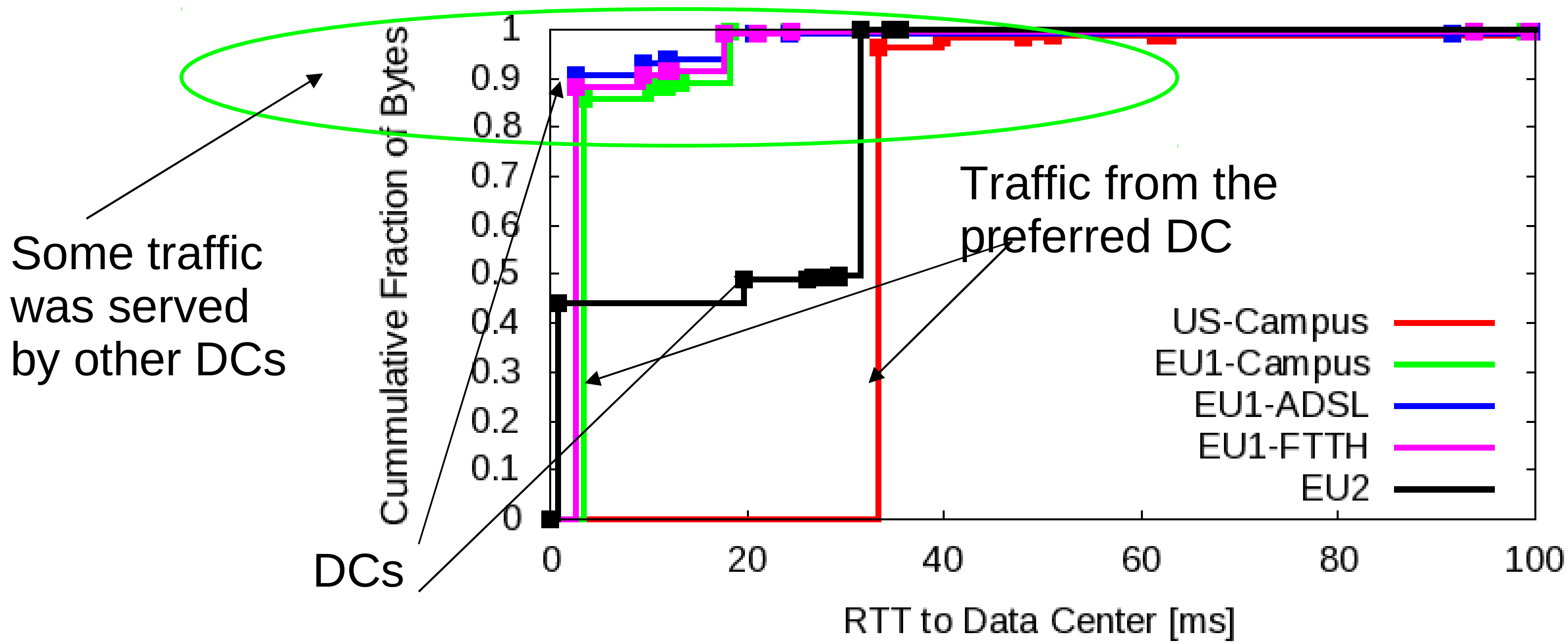
- Week-long traffic traces from 5 large networks in September 2010
- Deep Packet Inspection (DPI) with Tstat
    - Associates TCP flow with YouTube video that triggers it
    - Extracts per video information (video ID, resolution, duration, size)

Finding DC Locations

- Previously used techniques <u>do not work</u>
  - Reverse lookup on servers' IP addresses does not provide any information [IMC2010]
  - Maxmind and IP2location locate most servers in California [NOSSDAV2008]
- Active measurement-based methodology to find DC locations
  - Used CBG [ToN 2007] – Based on RTT and triangulation – Median error < 100 km



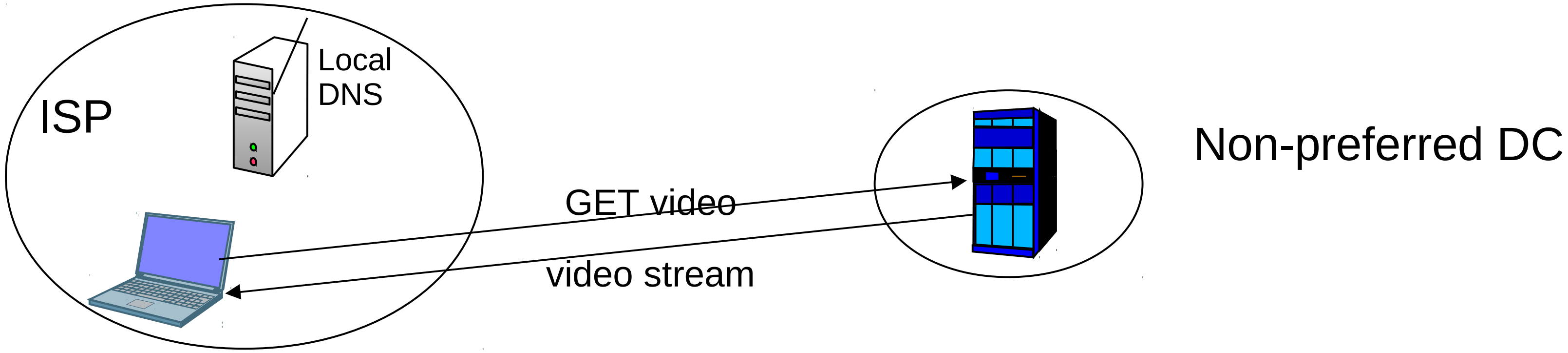
Closest DC Serves Most Videos but there are Exceptions



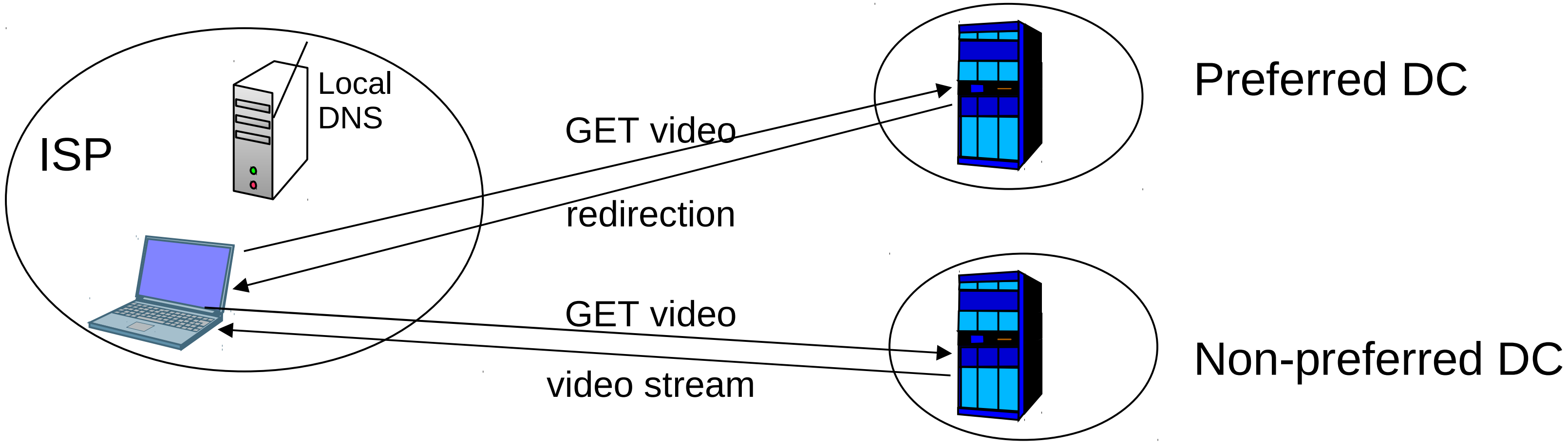
- Unlike published work [NOSSDAV'08,IMC'10], RTT does play a role on DC selection
- More interestingly, there are departures from the RTT policy:
  - For EU2, traffic splits mostly between two DCs
  - For other networks, 10%-15% of traffic comes from far DCs

Mechanisms for Accesses to Non-preferred DCs

DNS resolution sends the client to a non-preferred DC



Application-level redirection sends the client to a non-preferred DC

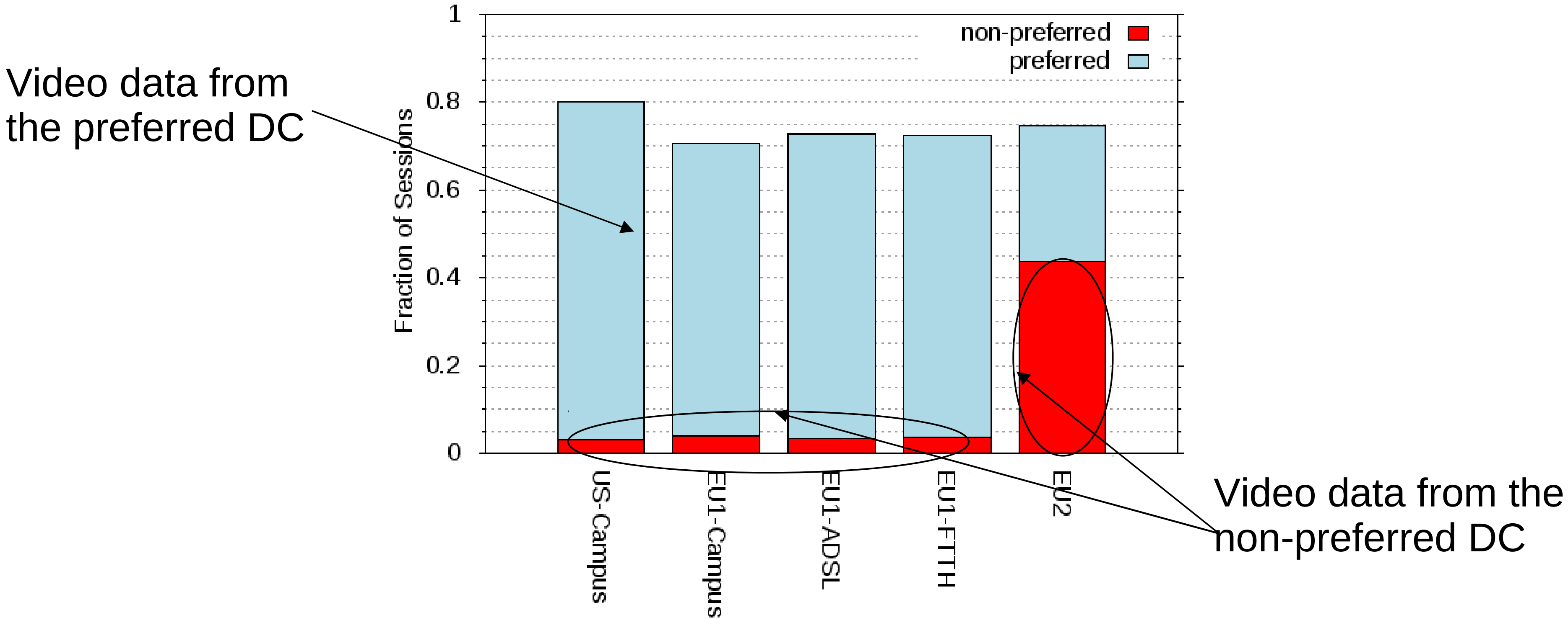


YouTube Video Sessions

- **Video Session:** group together related video flows
  - The same source IP
  - The same video ID requested
  - Interleaved by less than 1 second
    - Small interleave time to group flows triggered by the system

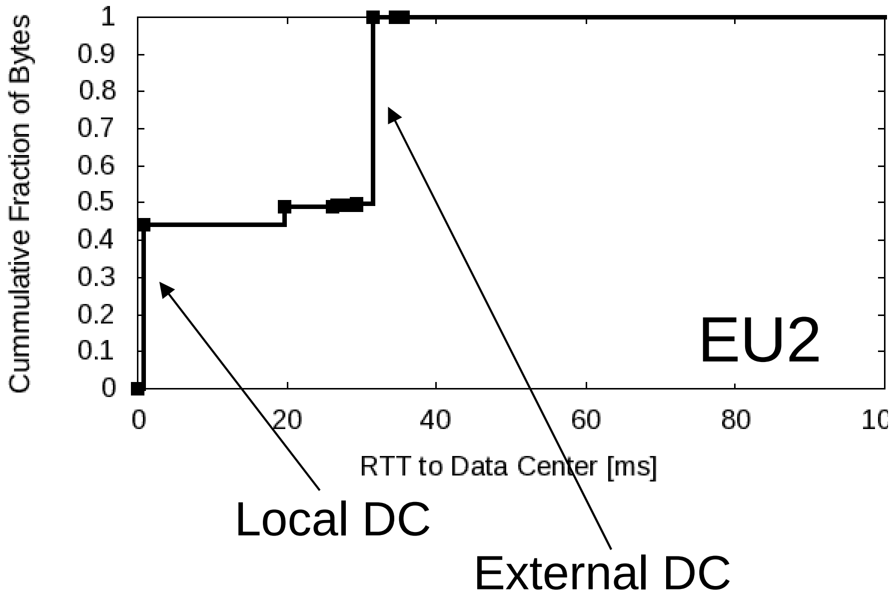
- One flow sessions **=>** DNS resolution
  - Represent 70-80% of all sessions
- Two or more flows sessions **=>** application-level redirection

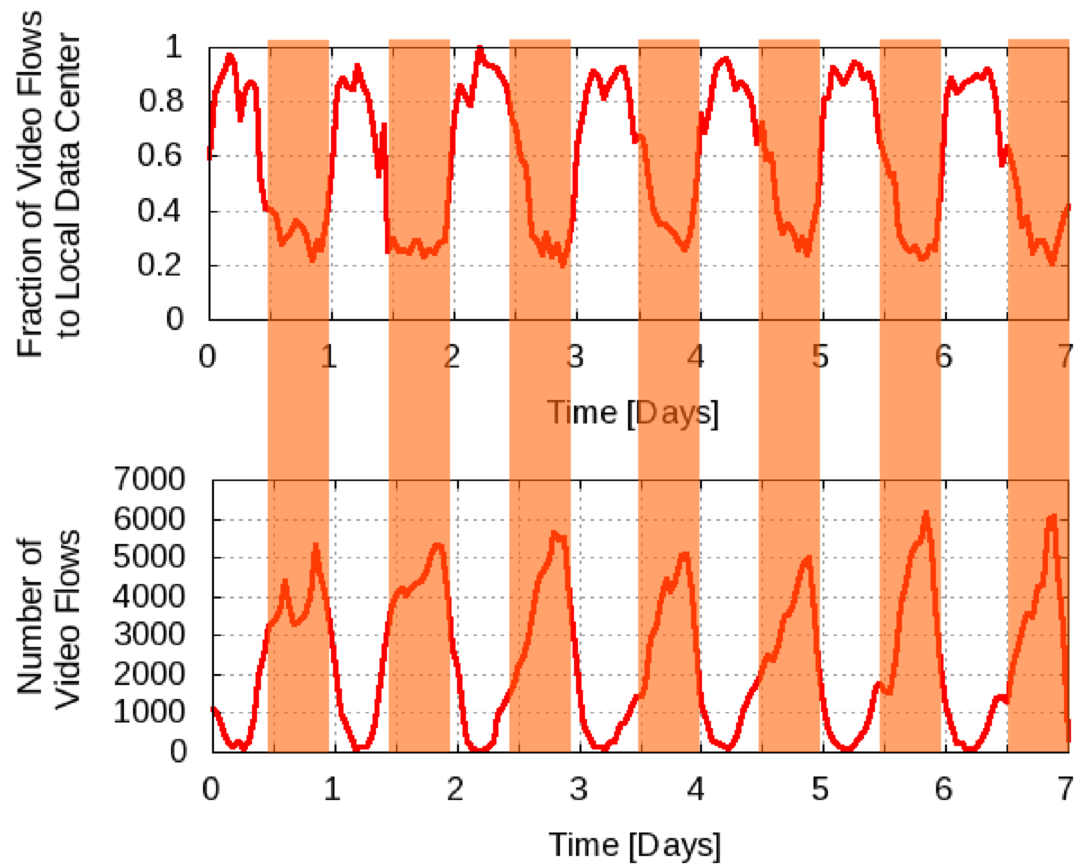
Effectiveness of DNS in Mapping Requests to the Preferred DC



- In most cases <u>the local DNS maps server names to the preferred DC</u>
- But in EU2, half of the videos downloaded from a non-preferred DC

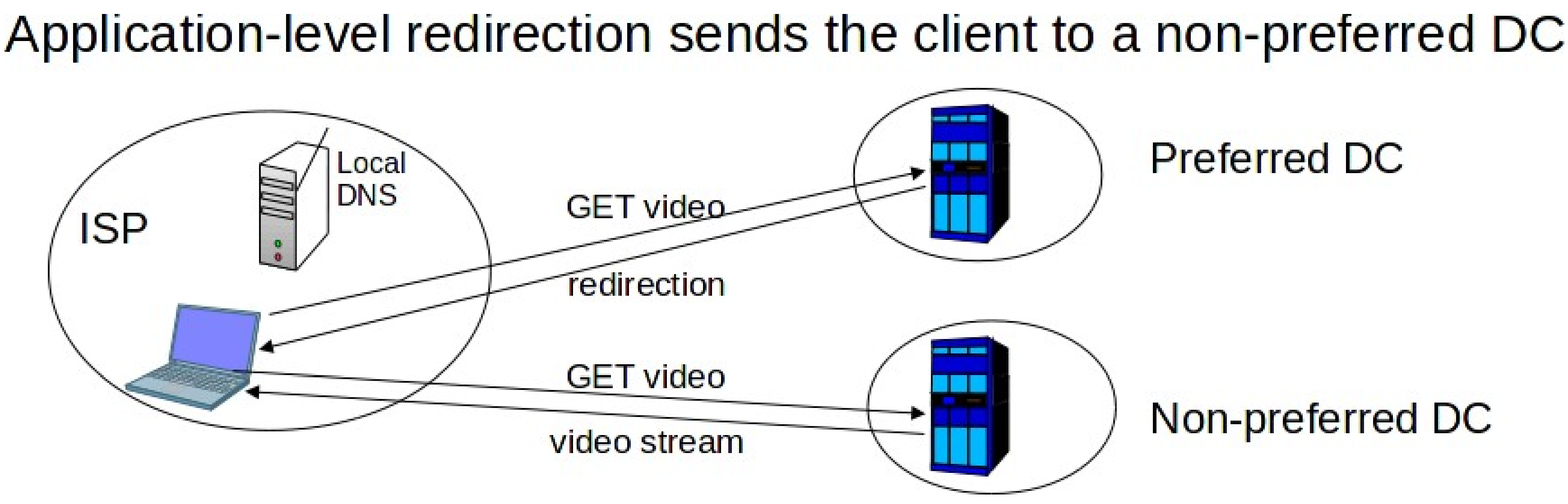
DNS-level Load Balancing





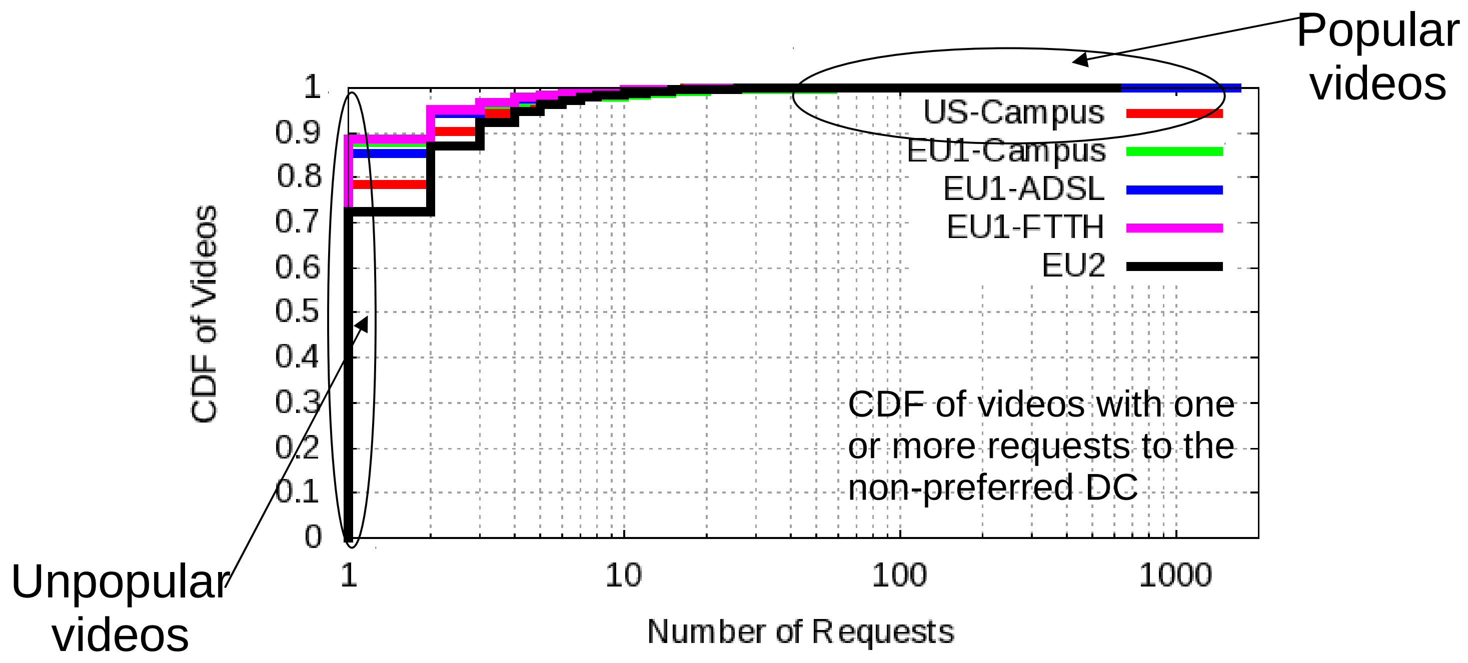
Load Balancing between the Local DC and the external DC affects server selection

Application-level Redirection



- 20-30% of sessions in this category
  - Many redirections to non-preferred DCs
- Hypothesis: Unpopular videos

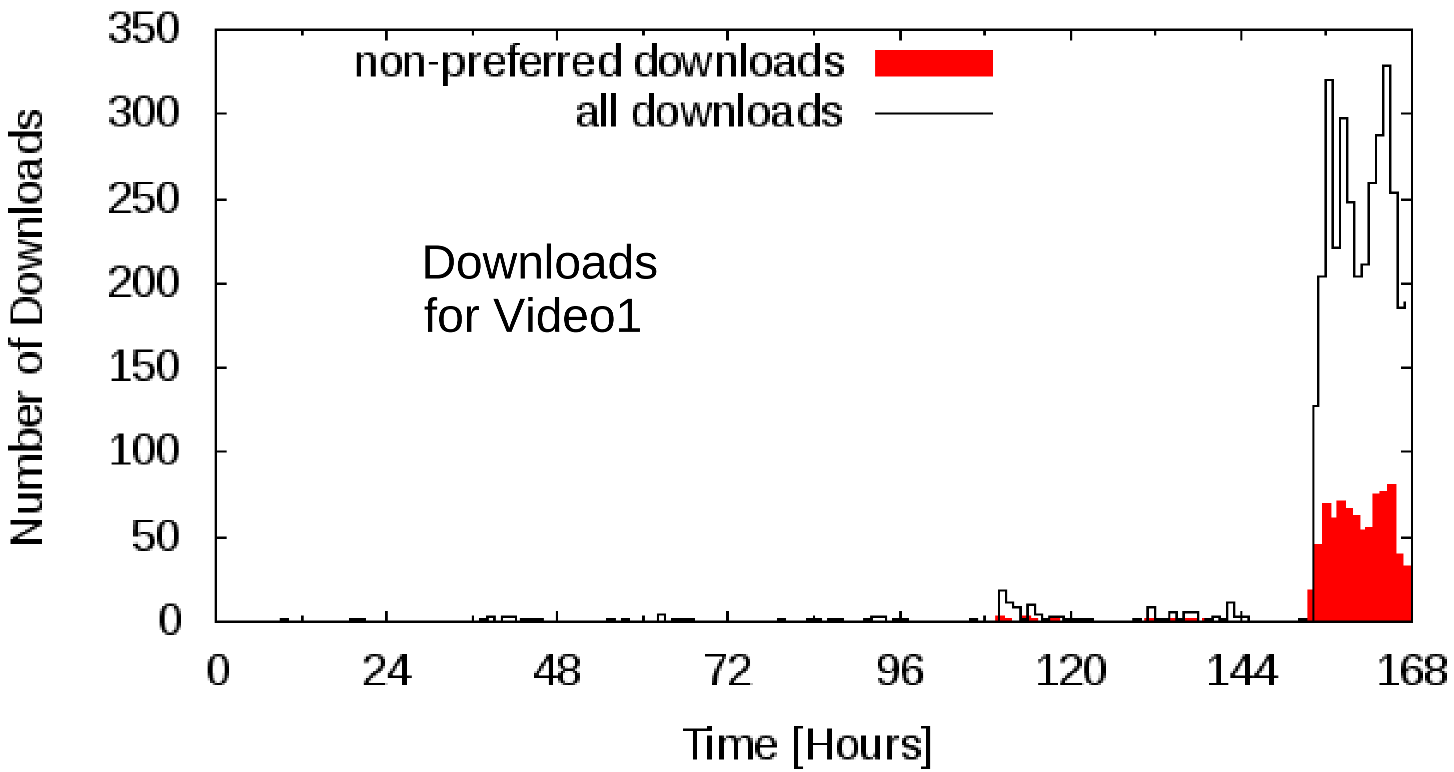
Does Video Popularity Causes Application-level Redirection?



Both popular and unpopular videos cause application-level redirections to non-preferred DCs

Popular Video Flash Crowd

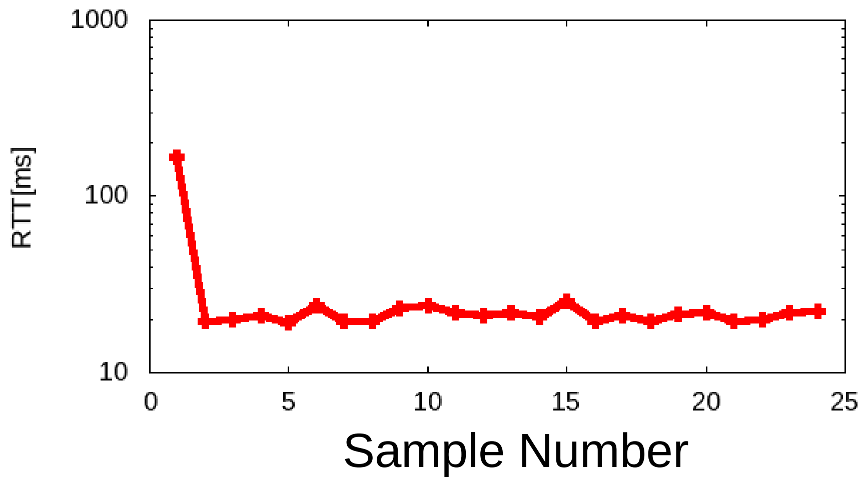
EU1-ADSL Trace



Redirection occurs due to overload of the preferred server for a popular video

Availability of Unpopular Videos

- It is not easy to identify unpopular videos in our traces
- Active experiment with Planetlab:
  - Nodes around the world download rare video (generated by us)
  - Latency measurements (RTT) from each node to the content server
  - Experiment repeated every 30 minutes for 12 hours



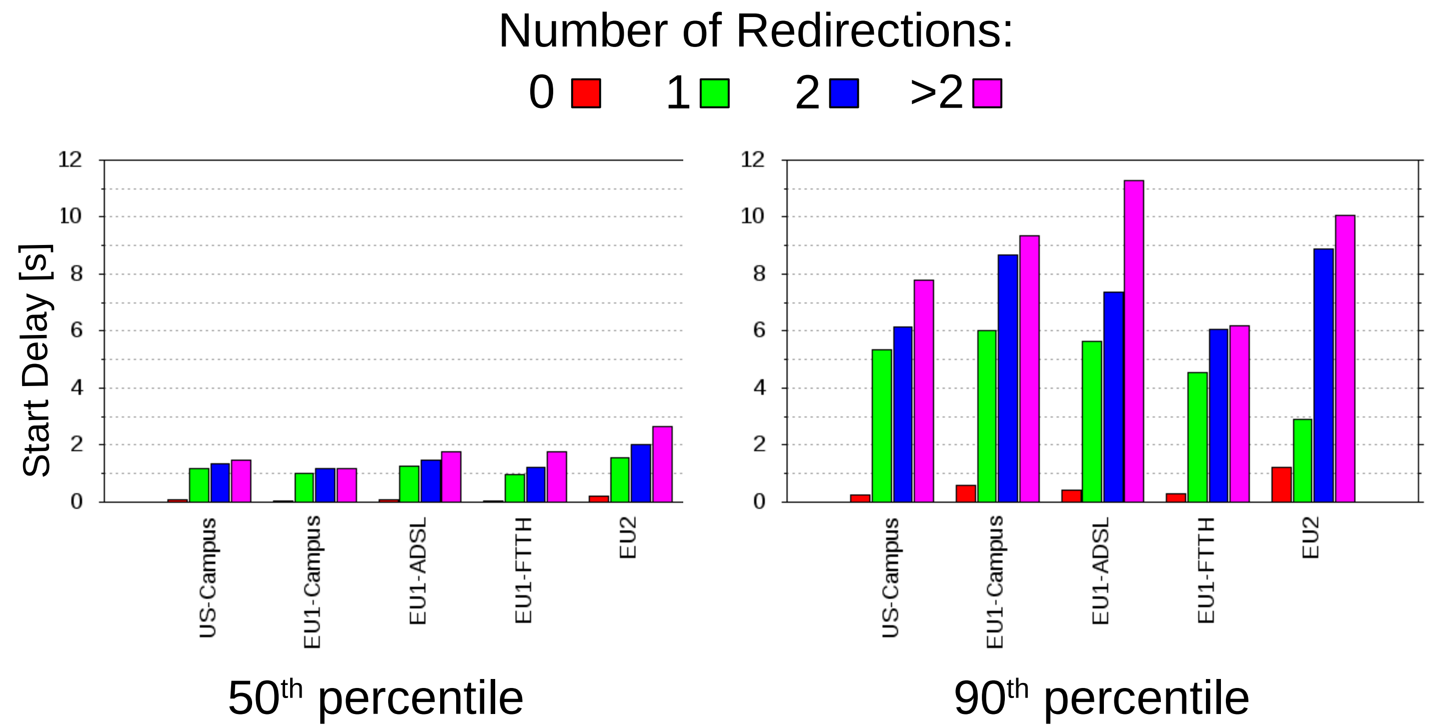
- In general, only the first access is to a non-preferred DC

Do Redirections Impact User Performance?

- Two performance metrics:

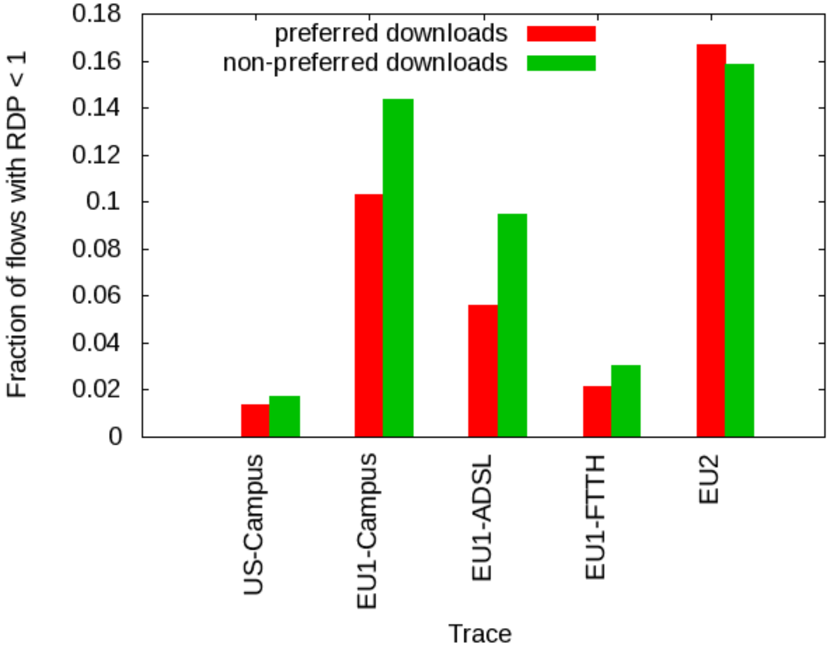
1. Startup Delay (SD) captures how long the user waits before watching the video
2. Ratio of Download rate to Playback rate (RDP)
   - RDP < 1, the video stalled

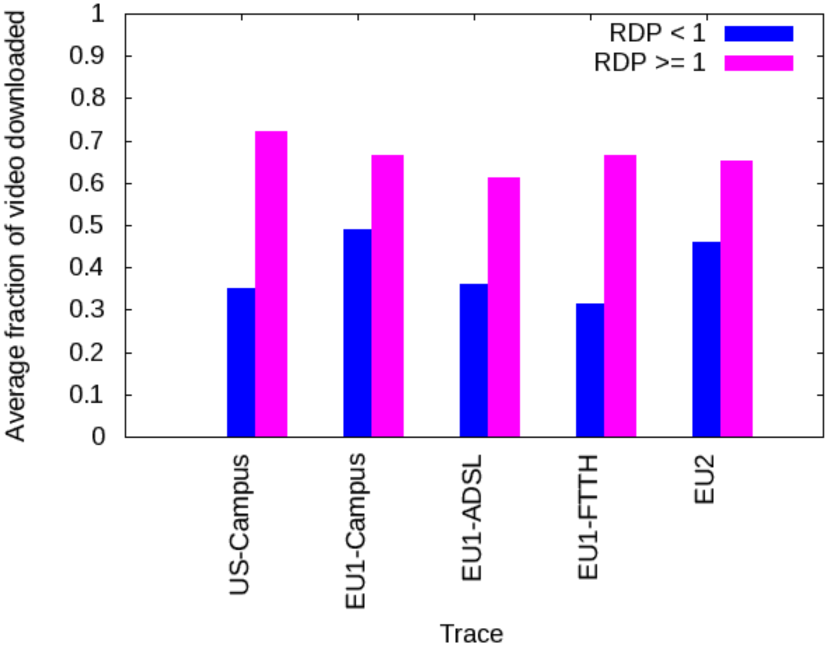
Large Start Delays on Redirections



- Without redirections, delay in the order of milliseconds
- With redirections, delay can increase by orders of magnitude, up to 10 seconds!

Users Watch Less when Videos Stall





- Downloading from non-preferred DCs may have an impact on user performance
- Users only watch 30-50% of videos with RDP<1

Summary

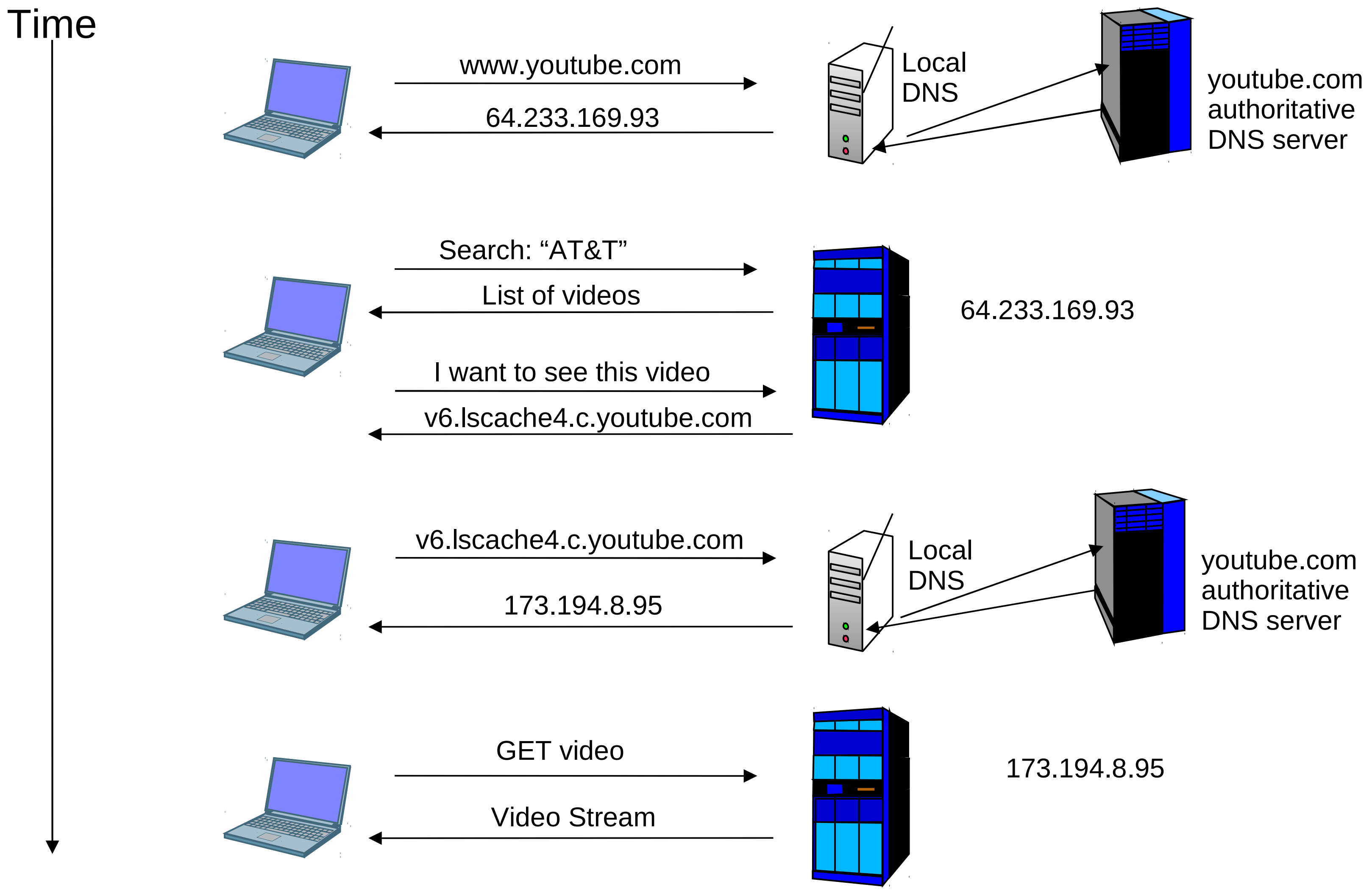
- By extensive measurement we shed light on the infrastructure deployed by YouTube
- Expose modern CDNs mechanism for redirection:
  - DNS-level redirection and application-level redirection
- Unlike prior work, we show that latency between clients and DCs plays a role in server selection
- Departing from conventional wisdom, several factors deviates from the latency policy:
  - Load balancing, DNS server queried, popular videos causing hot spots, availability of rare content
- Redirections can negatively impact user performance

Thank You!
Any Questions?

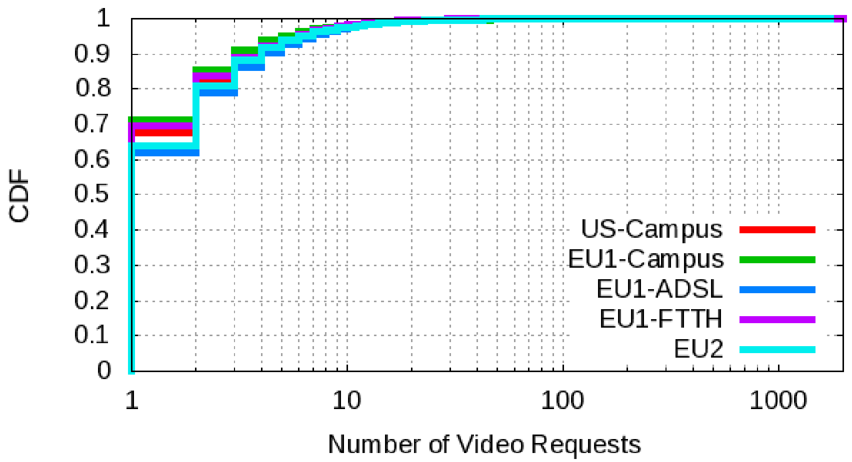
Related Work

- YouTube videos characterization studies:
  - Hill [IMC2007], Cha [IMC2007], Chen [IWQoS2008], Zink [COMNET2009]
- YouTube infrastructure characterization studies
  - Saxena [NOSSDAV2008], Adhikari [IMC2010]
- Earlier design was relatively simple
  - [NOSSDAV2008]: Most videos served from one data center (DC) located in US
  - [IMC 2010]: The DC selection is proportional to the DC size
- New design more mature and complex
  - Important to understand state-of-art systems

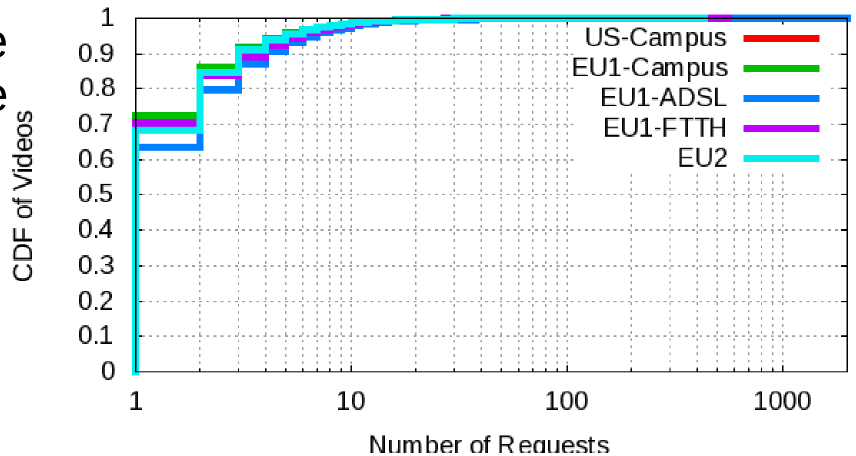
Background on YouTube



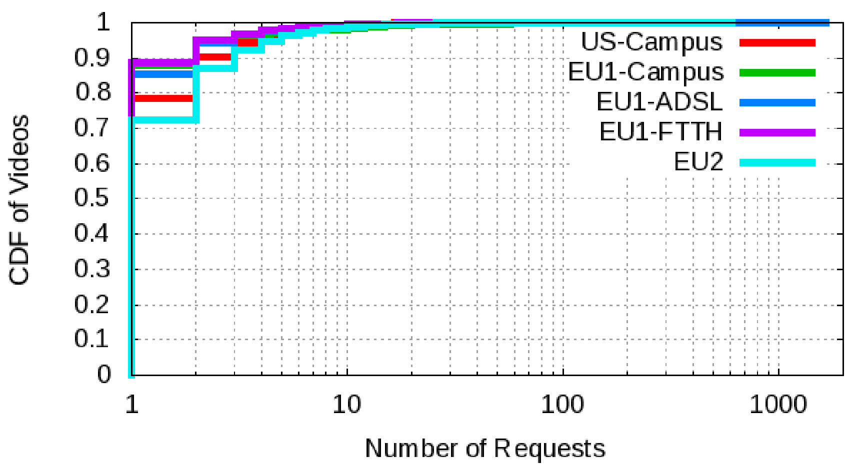
CDF of video requests



CDF of videos with one or more requests to the preferred DC

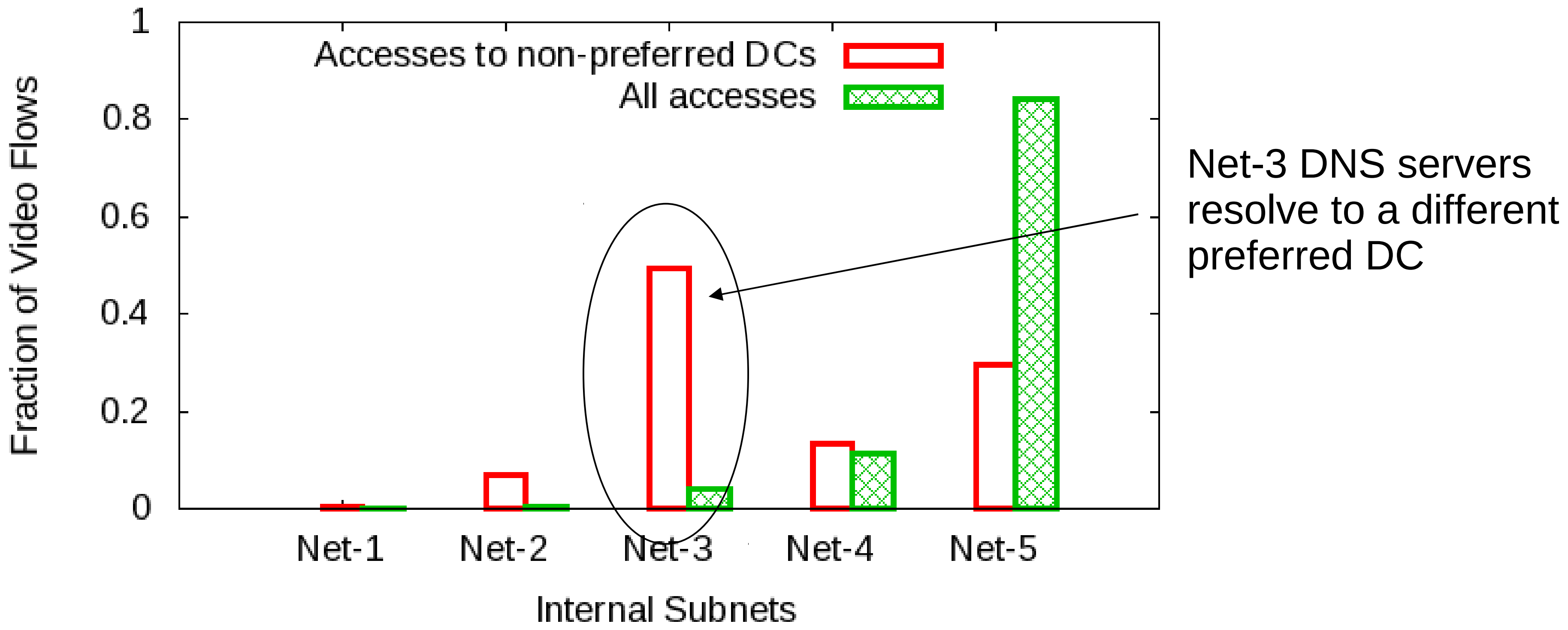


CDF of videos with one or more requests to the non-preferred DC



Variations Across DNS Servers in a Network

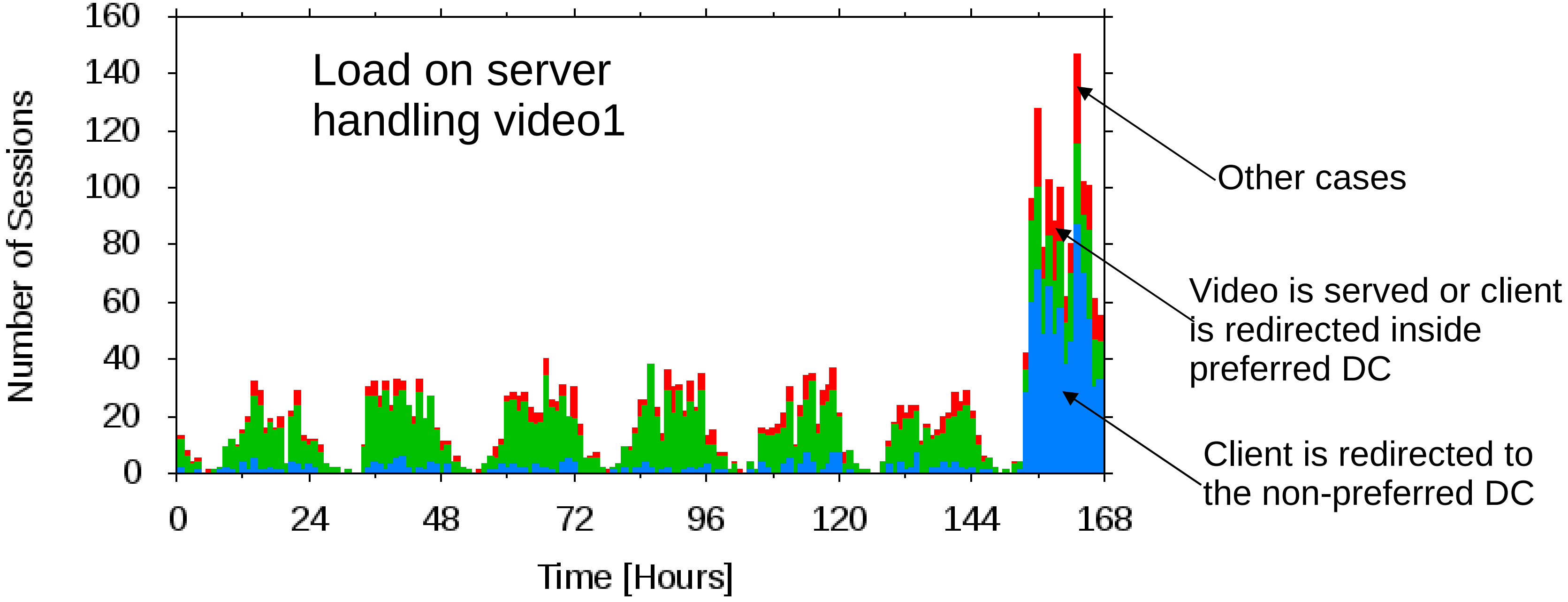
US-Campus Trace



- Depending on the internal DNS server queried, the redirection may be different

Server Flash Crowd

EU1-ADSL Trace



- Server assigned to popular video gets overloaded
- The number of redirections from this server increases because of the video flash crowd