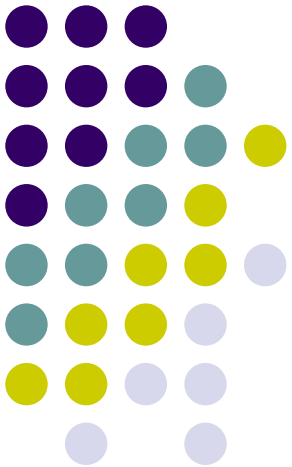
Analyzing Video Services in Web 2.0
A Global Perspective

Mohit Saxena

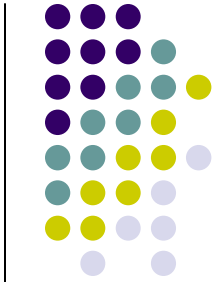
Umang Sharan

Sonia Fahmy

PURDUE
UNIVERSITY



Video Traffic Today

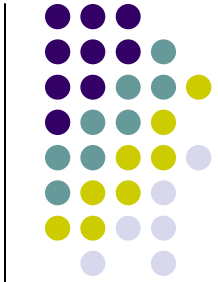


- Nearly 8 hours of video uploaded on YouTube every minute.
- All video will be high definition soon, which is 7-10 times more bandwidth-hungry.
- Video expected to account for 80% of all Internet traffic by 2010!

-- Source: AT&T's VP address on Web 2.0, April 2008.

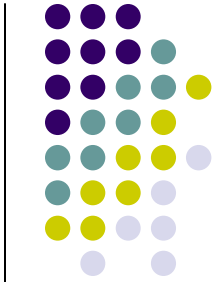
-- NANOG's discussion thread, April 2008.

Now and Then



- Traditional video streaming
  - Real-time streaming servers.
  - IP multicast and push protocols.
- Web 2.0 pseudo-streaming
  - Generic Web servers - YouTube, Dailymotion, etc.
  - HTTP/TCP media streams - progressive download and playback.
- HDTV, hybrid CDN-P2P networks
  - BitTorrent DNA.

Video Services Today



3 popular video services: YouTube, Dailymotion, Metacafe.

YouTube
- 70 million unique visitors per month.
- 100 million videos watched per day.
- 45 Tera Bytes of video data (2006).

Similar trends for Dailymotion and Metacafe.

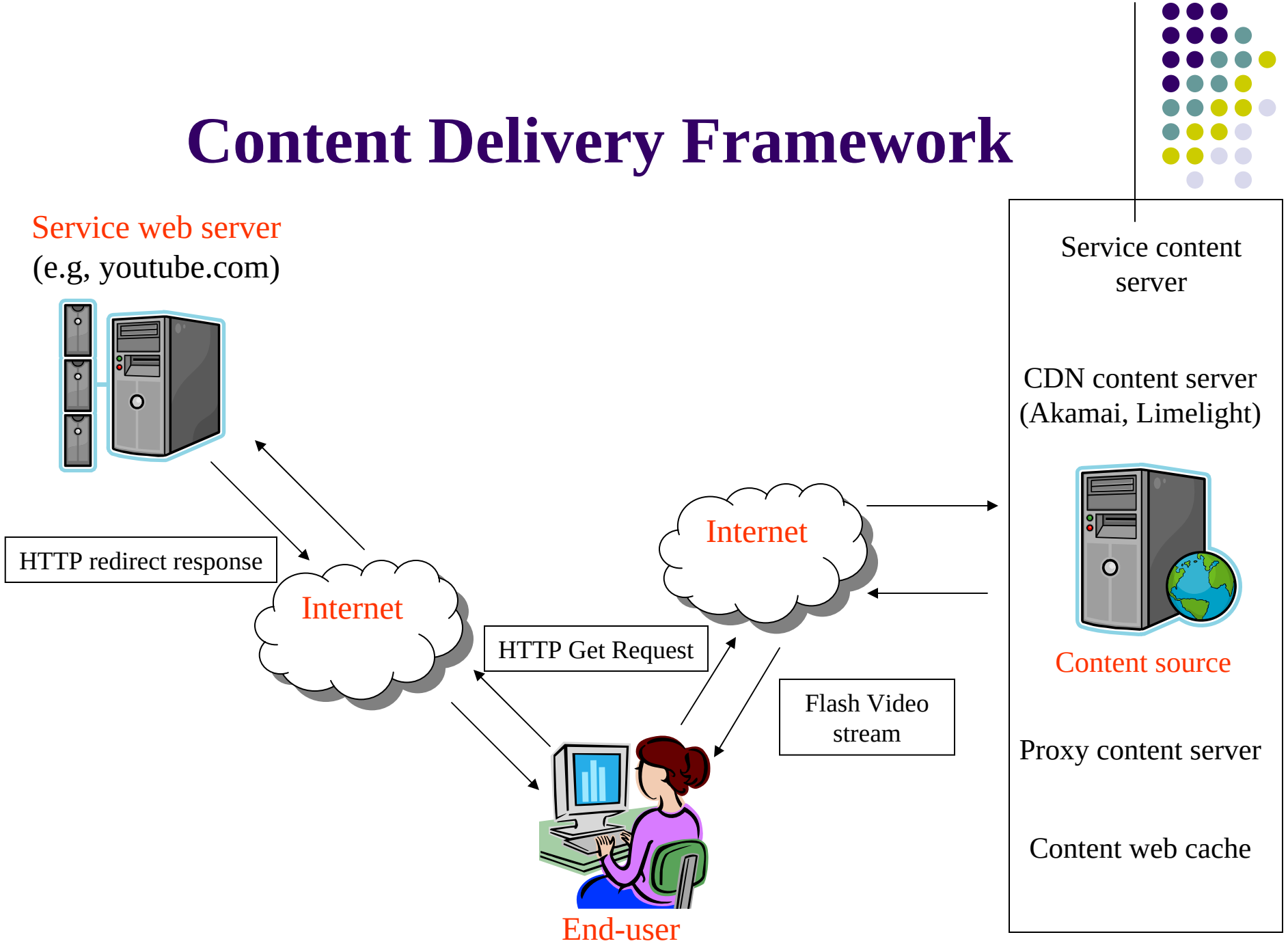
Two highly ranked websites w.r.t. traffic volume.

How do these services manage so much data?

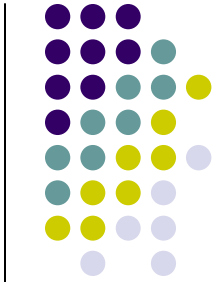
What storage and delivery models do they use?

How do they provide good end-user performance globally?

Content Delivery Framework

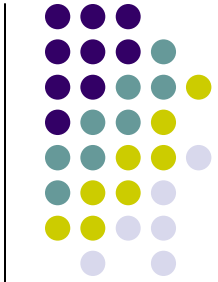


Goals and Methodology



- PlanetLab active measurements and experiments for 23 days.
- Objective:
  - Infer and compare their storage and content delivery models.
  - Investigate the variation in delay to serve video content to the end-user based on:
    1. Client's geographical location.
    2. Video characteristics - Age and Popularity.

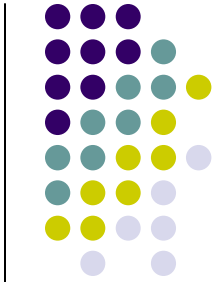
Experiments – Phase 1



- Infer and compare storage and delivery models
  - Crawled these services from differnet PlanetLab nodes, to gather:
    - Meta-information – Age and popularity of the content.
    - Geographic locations of host content servers.
- Dataset (Phase 1)

	YouTube	Dailymotion	Metacafe
# videos crawled	245,247	137,936	177,156
# distinct content servers recorded	2,405	1,252	94

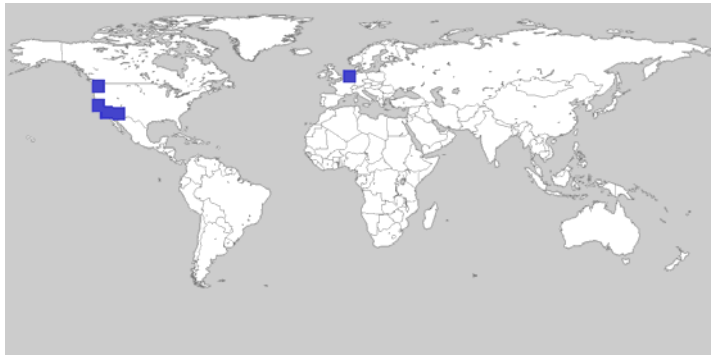
Storage and Content Delivery Frameworks



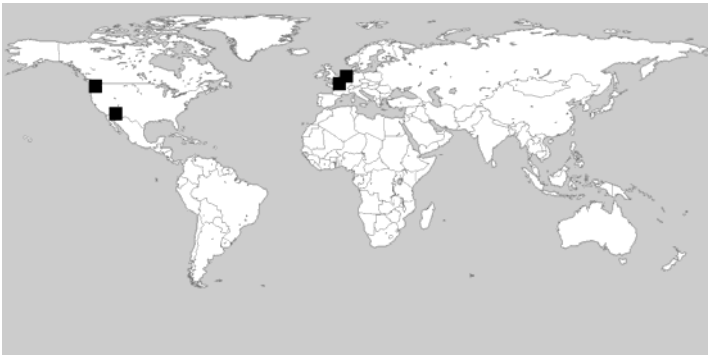
- YouTube
  - 77% YouTube content servers (San Mateo, California).
  - 22% Google Web Caches (Mt. View, California).
  - 1% CDN servers (Limelight).
- Dailymotion
  - 85% Dailymotion Proxies and content servers (France).
  - Rest – CDN servers (Limelight).
- Metacafe
  - All content – CDN servers (Akamai, Level3).

Geographic Distribution of Content Sources

YouTube



Dailymotion



Metacafe

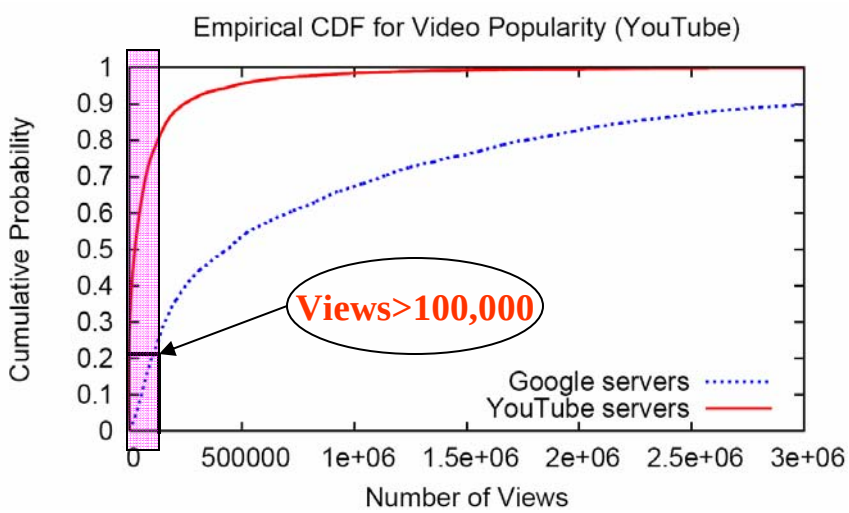
Metacafe *pushes all* content to CDN servers.

Criteria of selective pushing ?

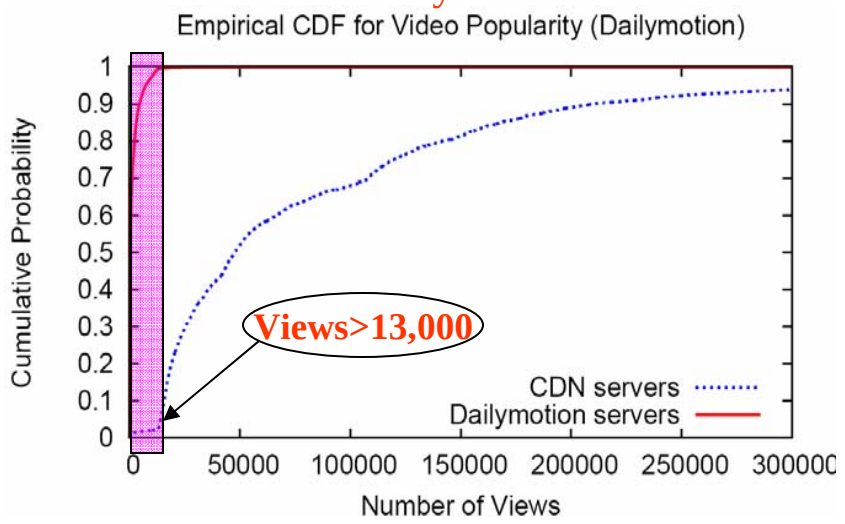
YouTube and Dailymotion *selectively push* content to different locations.

Selective Pushing: Popularity Model

YouTube



Dailymotion



Video distribution with popularity (# views)

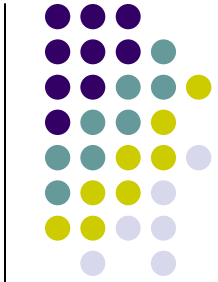
80% of YouTube videos on Google servers have views > 100,000 (high viewership).

Median # views of videos on YouTube servers is less than 23,000.

97% of Dailymotion videos on CDN servers have views > 13,000 (high viewership).

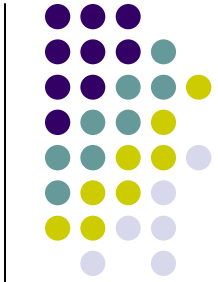
Most of the videos on Dailymotion proxies have views < 4,500.

Content Pushing: Age Model



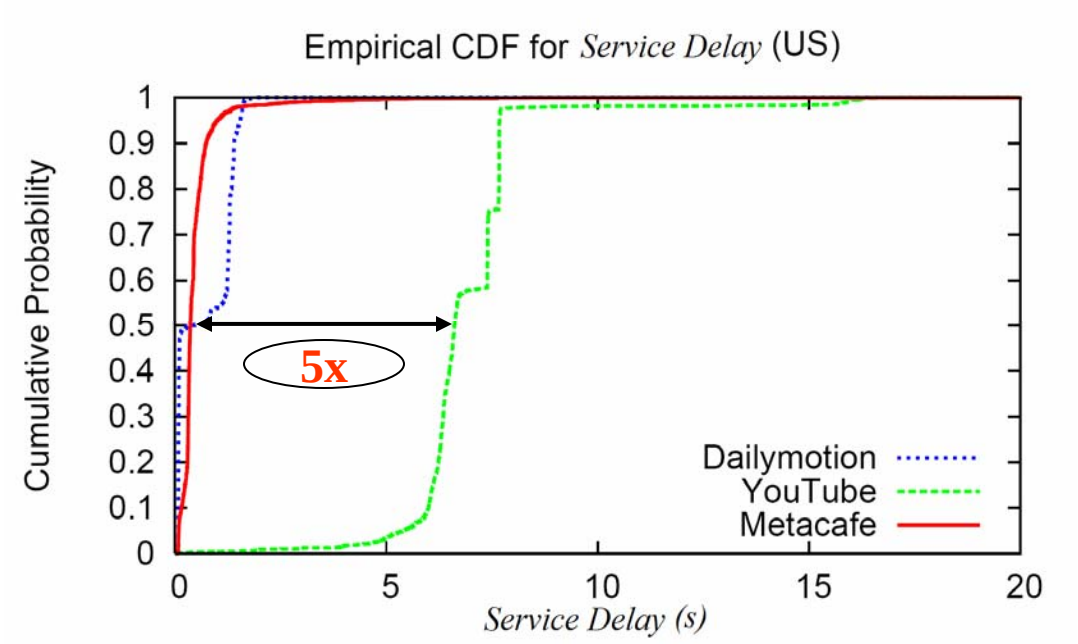
- Analyzed “most-recently” uploaded video category.
- Trends for YouTube
  - Total number of videos: 7,775
  - 7,681 on YouTube servers and 94 on Google’s domain.
  - 73% of these found on Google servers have a high viewership.
  - Popularity dominates age for selective pushing of content.
- Similar trends for Dailymotion.

Experiments – Phase 2



- Service Delay Analysis
  - Average time taken to fetch 1 Mbytes of Flash Video (FLV) stream.
  - Objectively indicates the performance of FLV streaming, without including the overheads for decoding and rendering.
- Dataset
  - Number of videos for analysis:
    - 16,118 (YouTube); 12,474 (Dailymotion); 15,919 (Metacafe).
  - Rich content diversity:
    - Age - 4 minutes to 3 years.
    - Views - ~0 to 60 million.
    - Run time duration - Few seconds to 10 hours.
- Clients: PlanetLab nodes
  - Distributed across USA, Brazil, Europe, India, China, Japan.
  - Sites across educational and commercial networks.
  - Different bandwidth limits.

Service Delay Analysis



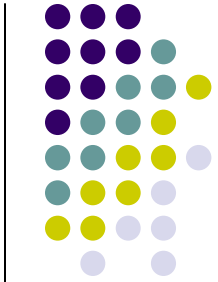
Distribution of Service Delay for different videos.

Median service delay for YouTube is 6.5s.
For Dailymotion and Metacafe, it is less than 1.25s.

YouTube appears to be 5 times slower as compared to Dailymotion and Metacafe!

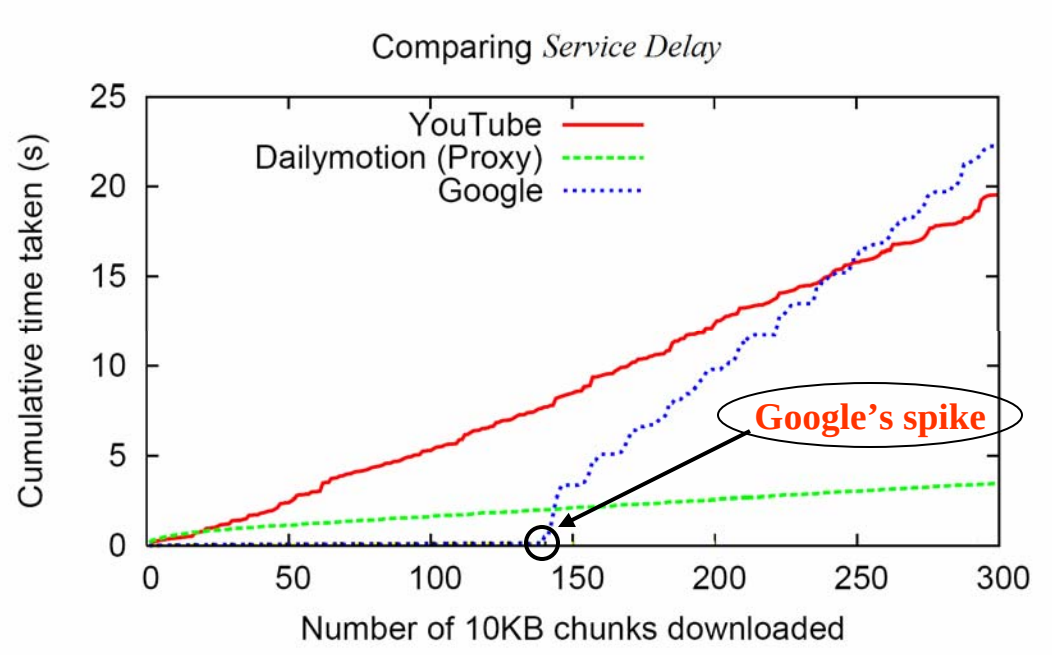
What is the end-user's perception of this difference?

End-user's Perception



- Flash video bit rate estimates
  - Video's file size (HTTP header)/ Run length (meta-information).
  - Typical run-length of 1 Mbyte Flash video is ~ 20-30 s.
  - Difference of 5 seconds – not perceived by end-user!
- VCR playback functionality
  - When user jumps forward/backward in the stream.
  - Need to investigate service delay at finer granularity.

Optimized Content Delivery



YouTube servers deliver at a constant rate – slower than Dailymotion and Metacafe.

Google serves at a high rate initially and later stabilizes to a lower rate. Allows for faster buffering initially – better strategy for popular content!

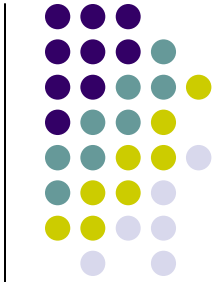
High Definition: Future of Video Traffic

- HD bit rates - orders of magnitude higher than Flash videos.
- Today’s pseudo-streaming tricks may not work!
- How can we still promise a good end-user experience?

Conclusions

- Analyzed three video services in Web 2.0
  - Multiple vantage points over PlanetLab.
- Inferred storage and content delivery frameworks.
  - Selective pushing models are mostly based on content popularity.
- Compared service delay variations globally.
- Insights into optimized content delivery methods.
- Where to next?
  - Investigate patterns due to local popularity of content.
  - Study the impact on other performance metrics.

Thanks



Questions?