Characterizing Quality of Content Distribution from YouTube like Portals

Anirban Banerjee and Dhiman Barman
University of California,Riverside CA, 92501
{anirban,dhiman}@cs.ucr.edu

1. INTRODUCTION

YouTube like media portals have changed the way users access media content in the Internet. Among many popular media sharing websites, some of the most popular ones are YouTube [7], Google-Video [4], Yahoo-Video [6], Metacafe [5], Flickr [3] and BlipTV [1]. Two factors along with the ubiquity of the Web have bolstered tremendous growth of these sites, namely, **availability** and **media quality**.

Availability: YouTube like portals have enabled users to perceive that they have access to infinite media storage systems. A plethora of media portals allow subscribers to store and share anything from home videos, to popular episodes of TV programs, pictures and more, all absolutely free of charge.

Media Quality: While streaming quality for media from these portal sites perceived by users is a major reason for their popularity [2], video quality might degrade due to different encoding schemes. For example, an uploaded video in YouTube is transcoded into Flash Video format (FLV). The video frame size is scaled to around 320x240 and frame rate to 25-30 frames per second. Audio is transcoded to a lower bit rate and reduced to mono. Other factors which affect streaming quality include network conditions and server load.

While YouTube, Google-Video, Yahoo-Video, Metacafe and BlipTV are all video content sharing website, Flickr [3] is a photo sharing website and web services suite. Recent analysis in [8] shows the characteristics and dynamics of popularity of such portals. Our research quantifies the quality of content delivery (QoCD) supplied by these six large media portals, a key success-factor for these sites. This has implications for understanding and modeling behavior of these entities which contribute a large portion of Internet traffic. Our measurement study spanning nearly 60 days from 140 clients, is divided into two phases: *Spidering* and *Probing*. During the *spidering* phase automated scripts strip off the html links to movie and image files published under various categories by these media portals. During the *probing* phase, the clients access the html links and download contents from different locations in the Internet. This phase ranges from 2-7 hours for different PlanetLab nodes and server/network conditions. In the probing phase we gather all this data from media portals and measure interesting features related to QoCD. As soon as each client completes downloading content using `wget`, it logs information about the probing phase and the spidering phase is re-initiated immediately. We seek to answer the following questions: 1) How does QoCD for a media portal vary with time ? 2) Does usage of commercial CDNs lead to marked improvements in QoCD ?

*CoNEXT'*07, December 10-13, 2007, New York, NY, U.S.A.

2. RESULTS

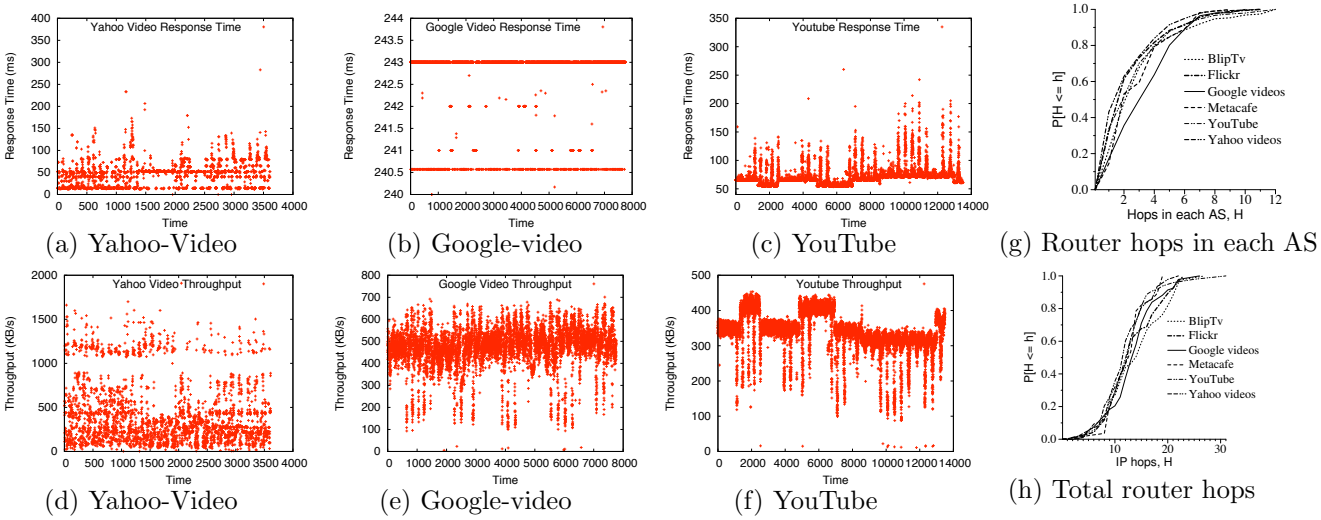
To answer our research questions we conduct extensive measurements and observe that:

(1)**Correlation of throughput and response time**: Throughput and response time to these sites are not strongly correlated, as observed from many clients.

(2)**Network proximity of CDN servers**: Sites which employ CDNs or have large number of edge servers are not closer in terms of IP hops to clients.

(3)**More servers; better service**: We also observe that Yahoo-Video employs Akamai and the number of CDN servers is an order of magnitude more than those employed by competing sites. This results in better user experience from Yahoo-Video.

In Table 1 we present the server-diversity observed from each client (i.e., number of edge servers employed by each portal to serve content). Yahoo-Video employs a much larger number of edge servers, courtesy the Akamai CDN to distribute content, compared to other portals. As we will see this has a noticeable effect on QoCD for this site. Due to space constraints we present a case study from our dataset in Fig. 1(a)-(c), where we observe the response time (ms) time-



(a) Yahoo-Video (b) Google-video (c) YouTube (g) Router hops in each AS

(d) Yahoo-Video (e) Google-video (f) YouTube (h) Total router hops

Figure 1: Response times (ms) measured from kc-sce-plab1.umkc.edu client for (a) Yahoo-Video, (b) Google-Video (c) YouTube. Similarly, throughput (KB/s) measured from the client for (d) Yahoo-Video, (e) Google-Video (f) YouTube. (g) Router Hops in each AS, (h) Total Router hops

series from `kc-sce-plab1.umkc.edu` to Yahoo-Video, Google-Video and YouTube. This client observed predictable response times from YouTube, BlipTv, Flickr and Google-Video with a few outlier observed values.

With no server diversity, response time from Flickr is expected to be around 51.5ms. BlipTv's response time is expected to be around 47ms. The client observed variable response times from Metacafe, Yahoo and YouTube showing diurnal as well as weekly patterns. This effect is prominent in Metacafe and YouTube measurements. Along with diurnal patters, we observe weekly patterns among response time values as well.

In Fig. 1(d)-(f), we observe throughput from these sites to the same PlanetLab client. As before, we observe diurnal and weekly patterns for Flickr, BlipTV, Google-Video and YouTube. In summary: 1) BlipTV and Flickr display base-band behavior with throughput values close to 237 and 550 KB/s respectively; 2) Metacafe and Yahoo-Video display significant variation in throughput values with peak values deviating by 205% and 325% respectively from the average values;3) Google-Video shows much varying behavior compared to stable response time behavior. The average value is around 474 KB/s, with extreme values off by about 74% from the average; 4) Throughputs from Metacafe and Yahoo-Video exceed 1 MB/s. Throughput from BlipTv and Google-Video are stable at values 250 KB/s and 500 KB/s respectively over longer time scale. For other content providers, throughput values lie between 0-700 KB/s.

Usage of a commercial CDN like Akamai enables Yahoo-Video to provide better throughput and lower response times than most competitors. Interestingly, we uncover that portals which provide good throughput and response times, employing more edge servers are not closest in terms of IP-hops to our PlanetLab clients. This is depicted in Fig. 1(g) and (h). This suggests that edge servers are not spread out in the Internet. One possible explanation could be the fact that portals often tie up with a quorum of bandwidth providers and place edge servers in those specific domains only. The high bandwidth links ensure rapid delivery of content but connections incur large number of router-hops in various ASes. Also the number of router hops in each AS for 90% of all connections to each portal is between 5 to 6. However, for Google-Video 90% of connections incur about 7 router hops.

Media Distributor	Max	Avg	Min	Total
Bliptv	2	1.18	1	2
Flickr	1	1	1	1
GoogleVideo	8	2	1	21
Metacafe	1	1	1	1
YouTube	2	1.48	1	2
YahooVideo	21	1.8	1	2218

Table 1: Number of Unique Servers seen by PlanetLab Clients per Media Portal

3. REFERENCES

[1] *BlipTV*. `http://www.bliptv.com`.
[2] *comScore*. `http://www.readwriteweb.com/archives/youtube\_nearly.php`.
[3] *Flickr*. `http://www.flickr.com`.
[4] *Google Videos*. `http://video.google.com`.
[5] *Metacafe*. `http://www.metacafe.com`.
[6] *Yahoo Videos*. `http://video.yahoo.com`.
[7] *YouTube*. `http://www.youtube.com`.
[8] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Proceedings of IMC'07*, 2007.