Measuring YouTube from Dual-Stacked Hosts

Saba Ahsan[1(✉)], Vaibhav Bajpai[2], Jörg Ott[1], and Jürgen Schönwälder[2]

[1] School of Electrical Engineering, Aalto University, Espoo, Finland
{saba.ahsan,jorg.ott}@aalto.fi

[2] Computer Science, Jacobs University Bremen, Bremen, Germany
{v.bajpai,j.schoenwaelder}@jacobs-university.de

Abstract. There is rapid growth in the number of IPv6 users and IPv6 compliant services on the Internet. However, few measurement studies exist about the quality of user experience on IPv6 in comparison to IPv4 for dual-stacked hosts. We present results from a measurement trial consisting of 21 active measurement probes deployed across Europe and Japan connected behind dual-stacked networks, representing 19 different Autonomous System (AS)s. The trial ran for 20 days in September, 2014 and conducted two types of measurements: (a) YouTube performance tests and (b) Speed tests to nearest dual-stacked Measurement Lab (M-Lab) server, both over IPv4 and IPv6. Our results show that a disparity exists in the achievable throughput as indicated by speed tests. We also witness disparity in content delivery servers used for YouTube media for some networks, resulting in degradation of experience over a specific address family.

1 Introduction

The World IPv6 Launch[1] that began in June, 2012 marked its second anniversary this year, reporting an increase in IPv6 usage by 500 % in the past two years. Google reports that as of 2014, over 4 % of their users access Google services over IPv6 in contrast to less than 0.5 % in 2011[2]. With more and more ISPs offering native IPv6 to their customers, there is a need for more measurement studies that can quantify the Internet performance aspects for early adopters of this technology. According to Sandvine Global Internet Phenomena report of 2014, audio and video streaming is the largest traffic category on fixed and mobile networks with YouTube as the largest single source of video streaming around the world[3]. Hence performance of Internet video in general, and YouTube in particular can impact Internet user experience to a great extent.

This paper presents a measurement study carried out in September 2014 that shows a comparison of YouTube performance over IPv4 and IPv6 actively measured over 21 probes distributed over Europe and Japan. To the best of our knowledge, this is the first study to compare YouTube performance over IPv4 and IPv6 from different dual-stacked networks. The probes receive native IPv6

[1] http://www.worldipv6launch.org.

[2] http://www.google.com/intl/en/ipv6/statistics.html.

[3] http://www.sandvine.com/trends/global-internet-phenomena.

J. Mirkovic and Y. Liu (Eds.): PAM 2015, LNCS 8995, pp. 249–261, 2015.
DOI: 10.1007/978-3-319-15509-8_19

connectivity and belong to different ISPs, covering 19 different IPv4 and IPv6 ASes. They run two kinds of measurements; speed tests and YouTube tests. Each test is run over IPv4 and then IPv6, giving us a comparison of performance over each. In this paper we make three contributions: (1) We find that there is disparity in the availability of YouTube content caches over IPv4 and IPv6, whereby the content-caches over IPv6 are largely absent, which can affect YouTube performance, (2) The measured YouTube throughput over IPv4 and IPv6 shows significant difference for some probes, resulting in support for better bit rates and thus higher resolution videos over one address family and not the other and (3) We find that Transmission Control Protocol (TCP) connect times over IPv6 are just not high enough for the happy eyeballs algorithm [13] to prefer a connection over IPv4, potentially choosing an IPv6 connection over IPv4, even when the observed throughput over IPv6 is lower. We release[4] the entire dataset to the measurement community.

The paper is organized as follows. We present related work in Sect. 2. Our metric, measurement test, and the methodology describing the measurement setup, trials and decision process is presented in Sect. 3. Insights derived from data analysis are presented in Sect. 4 with conclusions in Sect. 5.

2 Related Work

A number of early studies have focussed on characterization of YouTube videos. For instance, Phillipa Gill *et al.* in [6] (2007) study YouTube workload patterns by measuring local traffic in a campus setting and observing trends of popular videos. Features such as access patterns, file properties, video popularity, reference behaviors, and transfer characteristics are compared against traditional web workloads. Meeyoung Cha *et al.* in [3] (2007) show how YouTube content popularity is driven by truncated power-law distributions. They also study the prevalence and impact of content duplication and illegal uploads on system characteristics. They show how peer-assisted content delivery and caching schemes can offload server-side traffic by as much as 50 %.

These studies have been followed by a number of passive measurement efforts. Vijay Kumar Adhikari *et al.* in [1] (2010) study YouTube traffic dynamics from the perspective of a large tier-1 ISP. Using flow-level data collected at multiple Point of Presence (PoP)s, they show how the employed load-balancing strategy is location-agnostic. They also compare load-balancing strategies employed by YouTube against routing policies used by the ISP and study relationships between them. Alessandro Finamore *et al.* in [5] (2011) compare YouTube experience from mobile and PC-based devices. Using a week-long passively monitored dataset collected from 5 vantage points, they show how user access patterns are device and location agnostic. They also show how YouTube is heavily optimized for PC-based devices and leverages excessive buffering policies. This often leads to more data being fetched than is used for playback. Georgios Dimopoulos *et al.* in [4] (2013) study user-experience from YouTube video sessions. Using a

[4] http://www.netlab.tkk.fi/tutkimus/rtc/PAM2015/.

week-long passively collected dataset from within a campus network, they show how redirections to the destination media server is the primary contributor to initial delays. They show how statistical information sent back by the client can be used to identify stall events. They also measured the impact of advertisements on playback abandonment rates.

In recent years, we have witnessed a shift towards actively measuring the playback quality of a YouTube video. For instance, Parikshit Juluri *et al.* in [7] (2011) introduce the python based `Pytomo`, a tool that models a YouTube client to measure download statistics and estimate playback interruptions. Our YouTube test is inspired but improves upon this tool in three ways: (a) It is written in C, which has allowed us to deploy it on Customer Premises Equipment (CPE)-like devices such as SamKnows, (b) It supports multiple container formats such as MP4, WebM and FLV (unlike `Pytomo` which supports FLV only), and (c) Our test is more aware of available bit rates and resolutions. Vijay Kumar Adhikari *et al.* in [2] (2012) use PlanetLab vantage points to crawl a finite subset of YouTube videos. They use this dataset to show how: (a) the video ID space is flat, (b) multiple (anycasted) DNS namespaces are used to logically organize media servers and (c) a 3-tier physical cache hierarchy is used to deliver content. Parikshit Juluri *et al.* in [8] (2013) go further and use `Pytomo` to measure YouTube experience from within three ISP networks. They witnessed noticeable difference in experienced quality across ISPs. They reason that latency is not the primary factor when choosing a video server, but the selection mechanism is largely based on delivery policies and individual agreements with ISPs. Hyunwoo Nam, *et al.* in [9] (2014) introduce YouSlow, a browser-based plugin that can detect and report live buffer stalling events when watching YouTube videos that are delivered using Adaptive Bitrate Streaming (ABR) technology.

3 Methodology

We utilize two metrics in this study. A Youtube test that measures performance against dual-stacked YouTube media servers, and a SamKnows speed test that measures line rates against dual-stacked Google M-Lab servers. A detailed description of the implementation is given below:

3.1 Metrics

YouTube Performance Test: We have designed a test that can download and mimic playout of YouTube videos. It measures TCP connection establishment times, achievable throughput, and number of stall events as indicators of performance when streaming a YouTube video. The measures are taken over both audio and video streams separately. The test takes a YouTube URL as input, and scrapes the fetched HTML page to extract the list of container formats, available resolutions and URL locations of media servers hosting the streams. The test then locally resolves Domain Name System (DNS) names and establishes two concurrent HTTP sessions to fetch audio and video streams in the desired

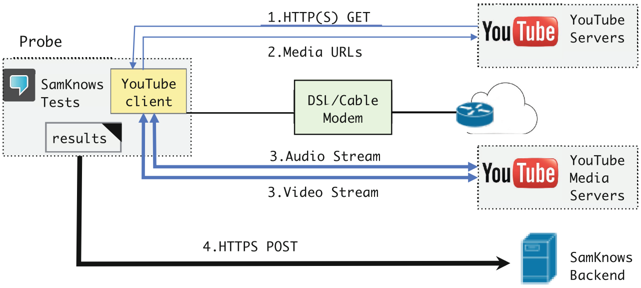


Fig. 1. A measurement setup on top of the SamKnows platform. A dual-stacked probe in addition to the standard SamKnows tests, executes the YouTube test. The YouTube test runs every hour and measures a set of performance indicators to endpoints delivering YouTube audio and video both over IPv4 and IPv6. The locally collected measurement results are pushed every hour to the SamKnows data collection server using HTTP.

format and resolution. The client ensures temporal synchronization between the streams, which means that playout only occurs if both audio and video frames have arrived.

In this process, the test records the time it takes for the `connect(...)` system call to complete as a measure of TCP connect times to both audio and video streams. The DNS resolution time is not taken into account in this measure. The test then measures throughput over the single TCP connection separately (and combined) over both audio and video streams. During playback, a stall event is declared when a frame is not received before its playout time. We use a 2-second prebuffering time, which means that 2 s of audio and video content is downloaded before starting the playout timer. In case a stall occurs, 1 s of media rebuffering is done before resuming the playout timer. The test does not at any time render content, but it only reads the format container to extract frame timestamps. The payload is eventually discarded.

Speed Test: The measurement test is part of the SamKnows' test suite [11] and is used to measure achievable throughput over the line. It uses three simultaneous TCP connections that fetch a portion of a 1 GB, non-zero, randomly generated binary file. Each TCP connection initiates a HTTP GET request to the nearest M-Lab[5] server and the recorded result is an aggregate of the observed values during the measurement. The test was modified to enable throughput measurements over IPv6. We use results from the SamKnows speed test as a baseline to compare the throughput measured from the YouTube test.

3.2 Measurement Setup

We cross-compiled the YouTube test for the OpenWrt platform and deployed it on SamKnows probes. The probes in addition to the YouTube test also run

[5] http://www.measurementlab.net.

standard SamKnows tests (which also includes the modified speed test). The YouTube test runs twice, once for IPv4 and subsequently for IPv6 and repeats every hour. For the speed test, each probe selects its nearest dual-stacked M-Lab server based on latency results. The same dual-stacked server is used to measure line rates both over IPv4 and IPv6. The test runs hourly during peak evening hours, and once every six hours after midnight. The data collected is stored on the SamKnows backend as shown in Fig. 1.

Selection of YouTube Videos: We use the YouTube v3 API[6] to generate a list of globally popular videos. We make use of globally popular charts to ensure our measurements become comparable across geographically located vantage points. We also prune out videos from the list that meet any of the three criteria: (a) Video duration is less than 60 s, (b) Video has regional restrictions, or (c) Video is unavailable in Full HD format. The list is generated on the SamKnows backend and is refreshed every 12 h. Each probe pulls this list on a daily basis. This allows us to measure against the same video for the entire day, which enables temporal analysis. On the other hand, cycling videos on a daily basis allows larger coverage of videos with different characteristics.

MA	IPv4 AS	IPv6 AS	LOCATION	PROVIDER	TYPE
#01	AS31334	AS31334	BREMEN	KABELDEUTSCHLAND	RESIDENTIAL
#02	AS3320	AS3320	BREMEN	DEUTSCHE TELEKOM	RESIDENTIAL
#03	AS50989	AS1257	STOCKHOLM	SITAB	RESIDENTIAL
#04	AS4685	AS4718	FUKUOKA	ASAHI NET	RESIDENTIAL
#05	AS12715	AS12715	MADRID	JAZZ TELECOM	RESIDENTIAL
#06	AS9031	AS9031	ALLEUR	EDPNET	RESIDENTIAL
#07	AS3320	AS3320	BREMEN	DEUTSCHE TELEKOM	RESIDENTIAL
#08	AS2518	AS2516	SHIZUOKA	BIGLOBE NEC	RESIDENTIAL
#09	AS513	AS513	CERN	CERN	RESEARCH
#10	AS680	AS680	BREMEN	DFN	NREN
#11	AS2614	AS2614	TIMISOARA	ROEDUNET	NREN
#12	AS2611	AS2611	LOUVAIN	BELNET	NREN
#13	AS680	AS680	BREMEN	DFN	NREN
#14	AS1741	AS1741	HELSINKI	FUNET	NREN
#15	AS5607	AS5607	LONDON	BSKYB-BROADBAND	LAB
#16	AS3269	AS3269	TORINO	TELECOM ITALIA	LAB
#17	AS8903	AS8903	MADRID	BT ESPANA	LAB
#18	AS2856	AS5400	IPSWICH	BT UK	LAB
#19	AS18070	AS18070	NIIGATA	NDAC	IXP
#20	AS24956	AS24956	BRAUN-SCHWEIG	GAERTNER DATENSYSTEME	BUSINESS
#21	AS13030	AS13030	OLTEN	INIT SEVEN	BUSINESS

Fig. 2. Deployment status of our measurement trial as of August 2014. Each vantage point is a SamKnows probe which is part of a larger SamKnows measurement platform. Most of these probes are deployed behind residential networks and receive native IPv6 connectivity from their ISP. A part of these probes are also connected within NREN.

[6] https://developers.google.com/youtube/v3/docs/videos/list.

Selection of Video Bitrate: YouTube servers provides a list of available resolutions and required bit rates for the requested video. The YouTube test currently does not support Dynamic Adaptive Streaming over HTTP (DASH) [10] during playout, however, it has two modes of operation for dealing with throughput constraints: (a) A *non-adaptive* mode where the test downloads the same video resolution despite video stalls and (b) A *step-down* mode where we step down to a lower resolution if a stall occurs. The test then chooses the next highest bit rate and begins the download from the beginning. The *non-adaptive* mode does not portray the behavior of most YouTube players but is useful in comparing characteristics between IPv4 and IPv6 while keeping conditions identical. The *step-down* mode on the other hand, shows a more user-oriented result in the form of the highest resolution that the client can playout without disruptions over a particular connection. To avoid unnecessary stalling we use results from speed tests to limit the maximum bit rate that the client will attempt to download.

3.3 Measurement Trials

The trial was conducted for 20 days (05^{th}–25^{th} September, 2014) using 21 SamKnows probes deployed behind 19 different ASes across Europe and Japan. These probes are also deployed inside different flavors of networks such as residential, NREN, business, and ISP test labs. Figure 2 provides a list of all probes along with their location, IPv4 and IPv6 AS, ISPs and network types.

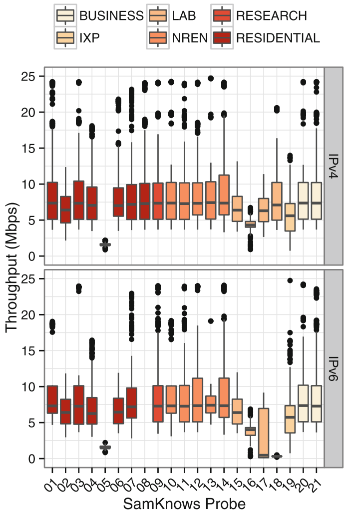
4 Data Analysis

A summary of all results is given in Fig. 3. A number of YouTube tests failed over IPv6 due to the unavailability of dual-stacked media servers or connectivity issues. Probe #08 was behind a Google blacklisted resolver[7], and consistently reported 100 % failure for YouTube IPv6 tests. The table shows the *Success Rate* of YouTube tests indicating the number of tests that successfully connected to media servers to download a YouTube video. The throughput graph shows disparity between IPv4 and IPv6 throughput. A detailed analysis, exploring the other aspects shown in the table follows.

4.1 Google Global Caches

YouTube videos are served to users through the Google's content delivery platform. Operators with a qualifying level of traffic can deploy servers as content caches within their networks in order to serve content closer to the users. These caches form Google Global Caches (GGC) and help increase performance and minimize transit bandwidth. Google estimates that 70–90% of their cacheable traffic is served from GGC[8].

[7] http://cnds.eecs.jacobs-university.de/users/vbajpai/googleipv6.
[8] https://peering.google.com/about/ggc.html.



	Success Rate		Stall Rate		Speedtest (Mbps)		GGC
	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6	
01	100%	55%	0%	0%	92.56	72.35	–
02	100%	100%	7%	1%	11.55	11.37	–
03	100%	60%	0%	0%	61.82	57.99	IPv4
04	100%	92%	0%	4%	10.68	7.55	IPv4
05	100%	100%	29%	39%	1.49	1.47	IPv4
06	100%	100%	0%	1%	27.83	6.16	IPv4
07	100%	100%	0%	2%	44.24	43.45	IPv4
08	100%	0%	0%	0%	13.14	9.80	IPv4
09	100%	100%	0%	0%	83.20	25.06	–
10	100%	55%	0%	0%	92.29	88.54	–
11	100%	100%	0%	0%	37.87	39.10	Both
12	100%	91%	0%	0%	92.15	77.40	–
13	100%	61%	0%	0%	217.99	170.46	–
14	100%	99%	0%	0%	87.09	86.34	Both
15	96%	100%	0%	0%	10.99	10.82	Both
16	100%	100%	5%	30%	4.35	4.31	IPv4
17	100%	100%	1%	57%	9.17	3.49	–
18	100%	100%	0%	100%	20.80	0.29	–
19	100%	99%	7%	5%	11.83	24.14	–
20	100%	100%	0%	0%	93.37	91.83	Both
21	100%	100%	0%	0%	88.08	64.04	–

Fig. 3. A summary of all test results. Box plots of the throughputs observed during YouTube tests (left) during the trial. Note that the graph is only used as a show of disparity and not the cause; throughput depends on the selected video, the selected resolution and throttling due to the limited length of the playout buffer, in addition to connection failure and connection bandwidth constraints. The table (right) shows for each probe (i) Success rate, a percentage of YouTube tests that successfully resolved and connected to media servers, (ii) Stall rate, percentage of successful YouTube tests that experienced one or more stall events, (iii) Speedtest (Mbps), the average throughput observed during the entire duration of the trial, (iv) GGC, the availability of GGC over an address family. The table represents results for the data collected in September 2014.

In our analysis, we identified GGC by looking up the Autonomous System Number (ASN) information for the contacted media servers. As expected, many of the GGC served content only over IPv4 and the probes used Google centralized content servers for IPv6. Among residential networks, 6 (out of 8) probes used GGC when using IPv4, but all used central content servers over IPv6. Within lab networks 2 probes used GGC, of which only 1 (#15) also used a GGC over IPv6. NREN and business probes were different in respect that all their IPv4 media servers belonged to a single ASN and this was the same for IPv6 media servers. Subsequently, we observed a degree of stability exhibited in the TCP connection establishment times of these two categories (see Sect. 4.2). Table 1 gives a list of the ASes we observed during our tests along with their categorization and the number of probes they served while the availability of GGC over each address family is shown under GGC in Fig. 3.

Table 1. Categorization of YouTube content (audio and video) delivery by AS as observed over all probes both over IPv4 and IPv6. It can be seen how content-caches and delivery from YouTube CDN is largely absent over IPv6.

CATEGORY	IPV4	n(PROBES)	IPV6	n(PROBES)
CONTENT CACHES	COMHEM (AS39651)	01	-	-
	ASAHI (AS4685)	01	-	-
	JAZZNET (AS12715)	01	-	-
	EDPNET (AS9031)	01	-	-
	DTAG (AS3320)	02	DTAG (AS3320)	02
	BIGLOBE (AS2518)	01	-	-
	ROEDUNET (AS2614)	01	ROEDUNET (AS2614)	01
	NORDUNET (AS2603)	01	NORDUNET (AS2603)	01
	BSKYB (AS5607)	01	BSKYB (AS5607)	01
	SEABONE (AS6762)	01	-	-
	QSC (AS20676)	01	QSC (AS20676)	01
	NG (AS48161)	01	-	-
CDN	GOOGLE (AS15169)	20	GOOGLE (AS15169)	19
	YOUTUBE (AS43515)	03	-	-
	YOUTUBE (AS36040)	02	-	-
	LEVEL3 (AS3356)	01	-	-
IXP	-	-	INTERLAN (AS39107)	01

4.2 TCP Connect Times and Happy Eyeballs

Figure 4 shows the distribution of raw TCP connection establishment times to YouTube media servers both over IPv4 and IPv6 as seen from each probe. These are the TCP connections that are later used to fetch YouTube video and audio streams separately. It can be seen how TCP connect times tend to show more variation for residential (#01–08) and lab (#15–18) probes. Probes deployed behind NREN networks (#09–14) and business lines (#20–21), on the contrary appear to be more stable.

TCP connect times are largely comparable over both address families. This is important to measure because applications (on top of TCP) running on dual-stacked hosts will prefer connections made over IPv6. This is mandated by the destination address selection policy [12]. As such, `getaddrinfo(...)` tends to resolve DNS names in an order that prefers an IPv6 upgrade path. However, the happy eyeballs algorithm [13] allows these applications to switch to IPv4 in situations where IPv6 connectivity is bad. The connectivity is considered bad when connections made over IPv4 can tolerate the 300 ms advantage imparted to IPv6 and still complete the TCP connection establishment in less time. Figure 5 shows the distribution of TCP connect times across all probes and the values for IPv6 are generally lower than 300 ms. As such, the happy eyeballs algorithm would prefer connections over IPv6.

4.3 Stall Events

Stall events occur due to throughput constraints, which are caused by a bottleneck at any point on the path between the media server and the probe. We observed stall events on 9 probes, 3 of which belonged to lab networks, 1 was in IXP while the remaining were all residential. Some of these cases are discussed below.

In 3 probes #02, #07 and #16, stalling events occurred only during peak hours, however, speed tests showed sufficient throughput values with no degradation during these hours. All 3 probes reported media servers in more than one AS, and the stalling events were specific to a particular AS only. In probe #02, the stall events are specific to servers in AS43515, which is only seen over IPv4 during peak hours and the stalls are also limited to IPv4 only. In case of probe #16, the stall events are seen for servers in AS15169 and the AS appears for both address families, causing stalling events in both cases as well. Figure 6 show hourly trend of YouTube and speed tests for probe #02 and 16. Stall events for probe #07 all occurred for the same video that was downloaded in Ultra HD, with a bit rate of 13 Mbps, which is 4 times the bit rate required during other tests that ran on the probe. While the ASs of media servers used for the video download varied for different hours during the day, all stall events were observed for servers in AS15169. Graphs for probe #07 were not included due to space limitations.

In case of 6 probes (#04–06,17,18,19) the measured throughputs during speed tests indicated insufficient bandwidth and YouTube tests also exhibited stall

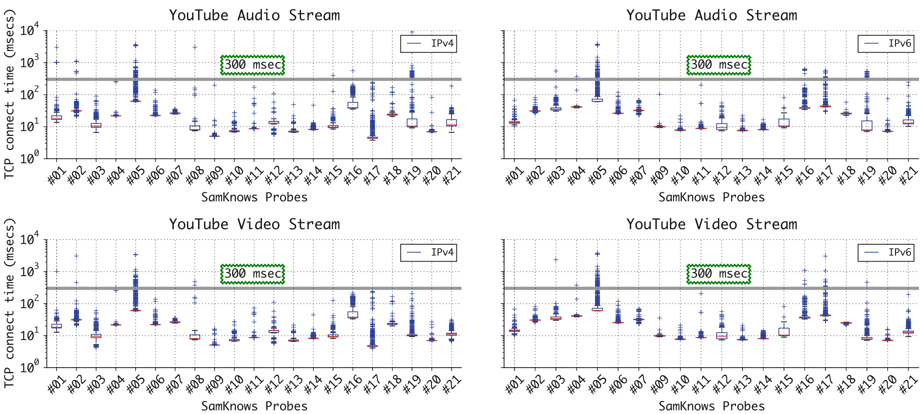


Fig. 4. Boxplots of TCP connection establishment times (in log-scale) to YouTube audio (above) and video (below) streams from each vantage point both over IPv4 (left) and IPv6 (right). The raw TCP connect times to YouTube media servers are comparable over both address family.

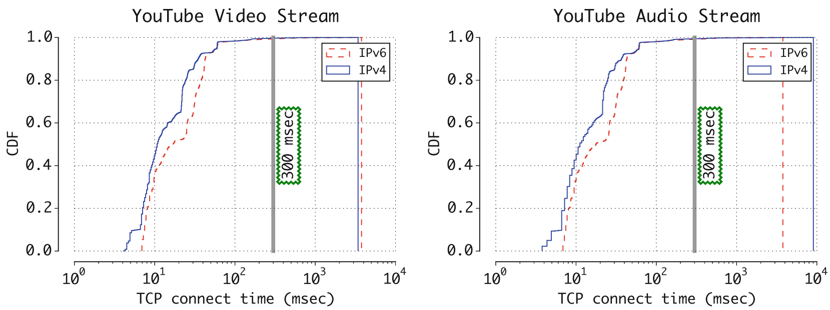


Fig. 5. Distribution of TCP connection establishment times (in log-scale) to YouTube video (left) and audio (right) streams both over IPv4 and IPv6 combined over all probes.

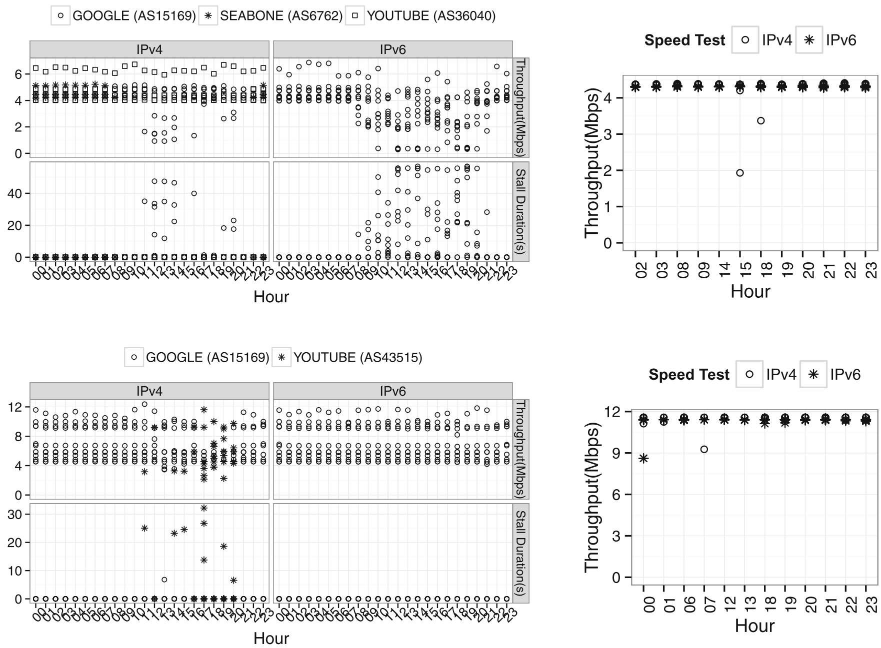


Fig. 6. Hourly trend of stall events, YouTube throughput and speed tests as observed on probe #16 (top) and #02 (bottom) during Trial Phase 1. For both probes, stall events are specific to media servers in a particular AS. We note that the disparity in media servers for each address family leads to stalling only in IPv4 in the bottom graph, while in the top graph it results in more stall events in IPv6 than IPv4. Speed tests, which are run only for specific hours during the day are shown on the right.

events. Figure 7, shows the speed test results for all residential probes, and also the lab probes that exhibit stalling. Note that all 6 of these probes contain some very low throughput measurements. In case of probes #05, #16 and #17, sometimes the competing audio stream consumed too much bandwidth resulting in an insufficient share for the heavier video stream. We identified this as a flaw in our test and noted that pacing audio traffic can help avoid stalls in some cases where the required and available throughput are very close.

4.4 Summary

Among our trial probes, 16 were deployed in home, office or university/research networks and represented real end users with dual-stacked hosts. Disparity in throughput measurements over IPv4 and IPv6 was observed in 10 of them. From the remaining 6, 4 probes showed inconsistent results for YouTube in terms of content delivery, IPv6 connectivity to media servers and/or TCP connect times.

Speed tests revealed a range of achievable throughput for residential networks. 5 out of the 8 residential probes showed disparity in measured throughputs over IPv6 and IPv4, all of them having lower values over IPv6. All these probes used centralized servers for fetching media over IPv6, whereas 6 of them used content caches over IPv4. Half of the probes suffered from connectivity issues to YouTube media servers over IPv6.

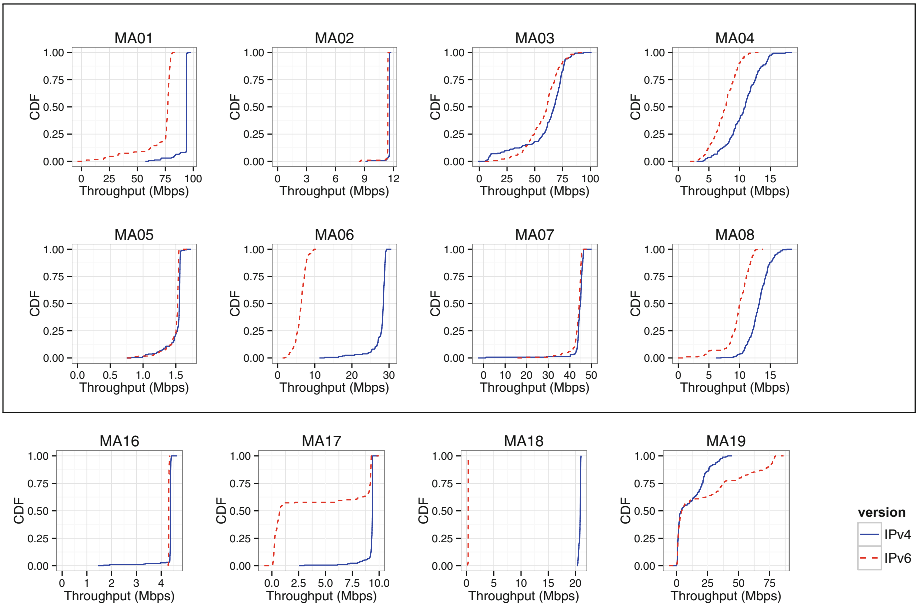


Fig. 7. Distribution of line rates observed by probes wired in behind a residential gateway (boxed) and operator's lab network (unboxed). Line rates are measured using speed test against dual-stacked M-Lab servers.

Office, research and NREN used in the trial were all high-speed networks with even the slowest one reporting an average throughput of over 25 Mbps. YouTube HD content has a typical range of 3–5 Mbps and about 4 times that for 4 K UltraHD, and hence from a required throughput perspective, these networks can easily support YouTube. This was exhibited in the form of 100 % stall-free YouTube tests for these networks. However, networks are typically used by more than one user and even single users run simultaneous tasks. The speed test results in some of these networks show lower throughput values in case of IPv6, which can result in performance degradation for users.

The trial included 4 probes that were deployed in testbeds for ISPs that have not launched IPv6 to customers yet, in order to ascertain network performance before actual IPv6 rollouts. We found erratic results or performance issues on 3 of them, while one showed smooth performance that was consistent over IPv4 and IPv6.

5 Conclusion

We measured YouTube performance from 21 dual-stacked probes deployed in Europe and some parts of Japan and observed two causes of degraded YouTube performance over IPv6 in comparison to IPv4 or vice versa: (i) a disparity in available bandwidth leading to insufficient throughput for a particular address family and (ii) different media content servers for each address family, of which, servers from a certain prefix exhibited lower throughput connections with the probe.

Overall, we observed that network performance over IPv4 and IPv6 is dissimilar in a majority of the networks we studied. From a set of 16 probes deployed in residential, official/educational networks, we observed only 3 probes (MA#11,14,20) with similar network conditions and performance for both IPv6 and IPv4 in terms of speed tests and YouTube delivery. This extent of disparity shows the significance of performance measurements at end points to better understand and improve the quality of services.

Acknowledgements. This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) grant no. 317647 (Leone). We would like to thank all the volunteers who hosted a SamKnows probe for us. We would also like to thank Sam Crawford, Jamie Mason and Cristian Morales Vega (SamKnows) for providing us technical support on the SamKnows infrastructure. We also would like to thank Steffie Jacob Eravuchira (Jacobs University Bremen) for reviewing the manuscripts.

References

1. Adhikari, V.K., Jain, S., Zhang, Z.L.: Youtube traffic dynamics and its interplay with a tier-1 ISP: an ISP perspective. In: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, IMC '10. ACM, New York (2010)

2. Adhikari, V., Jain, S., Chen, Y., Zhang, Z.L.: Vivisecting youtube: an active measurement study. In: 2012 Proceedings IEEE, INFOCOM (March 2012)
3. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.Y., Moon, S.: I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, pp. 1–14. ACM, New York (2007)
4. Dimopoulos, G., Barlet-Ros, P., Sanjuas-Cuxart, J.: Analysis of youtube user experience from passive measurements. In: 2013 9th International Conference on Network and Service Management (CNSM), pp. 260–267 (October 2013)
5. Finamore, A., Mellia, M., Munafò, M.M., Torres, R., Rao, S.G.: Youtube everywhere: impact of device and infrastructure synergies on user experience. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11, pp. 345–360. ACM, New York (2011)
6. Gill, P., Arlitt, M., Li, Z., Mahanti, A.: Youtube traffic characterization: a view from the edge. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, pp. 15–28. ACM, New York (2007)
7. Juluri, P., Plissonneau, L., Medhi, D.: Pytomo: A tool for analyzing playback quality of youtube videos. In: 2011 Teletraffic Congress (ITC) (September 2011)
8. Juluri, P., Plissonneau, L., Zeng, Y., Medhi, D.: Viewing youtube from a metropolitan area: what do users accessing from residential isps experience? In: IFIP/IEEE International Symposium on Integrated Network Management (IM) (May 2013)
9. Nam, H., Kim, K.H., Calin, D., Schulzrinne, H.: Youslow: a performance analysis tool for adaptive bitrate video streaming. In: Proceedings of the 2014 ACM conference on SIGCOMM, pp. 111–112. ACM (2014)
10. Stockhammer, T.: Dynamic adaptive streaming over http: – standards and design principles. In: Proceedings of the Second Annual ACM Conference on Multimedia Systems, MMSys '11, pp. 133–144. ACM, New York (2011)
11. Sundaresan, S., de Donato, W., Feamster, N., Teixeira, R., Crawford, S., Pescapè, A.: Broadband internet performance: a view from the gateway. In: SIGCOMM '11. ACM (2011)
12. Thaler, D., Draves, R., Matsumoto, A., Chown, T.: Default address selection for internet protocol version 6 (IPv6). RFC 6724 (Proposed Standard), Internet Engineering Task Force (September 2012)
13. Wing, D., Yourtchenko, A.: Happy eyeballs: success with dual-stack hosts. RFC 6555 (Proposed Standard), Internet Engineering Task Force (April 2012)