Tracing the Path to YouTube – A Quantification of Path Lengths and Latencies towards Content Caches

Trinh Viet Doan, Ljubica Pajevic, Vaibhav Bajpai, Jörg Ott

***Abstract*—Quantifying the benefits of content cache deployments, both in terms of latency and path lengths, and identifying the disparity of achievable benefits over IPv6 and over IPv4, is essential to identify bottlenecks in content delivery. We approached this matter by collecting and analyzing `traceroute` measurements towards YouTube from ~100 vantage points located in 74 different origin Autonomous Systems (ASes). Using a longitudinal dataset (05/2016 – 05/2018), we show that most of the Google Global Cache (GGC) nodes were reachable within ~6 IP hops and within ~20 ms from users streaming the videos. Further, we observed that in cases where GGC nodes were dual-stacked, the path lengths and latencies were comparable over both address families. However, as generally believed, shorter path lengths did not always correlate with lower latency: when the video was cached by a GGC node over IPv6 only, paths were almost always shorter over IPv6. Yet, latencies were still lower over IPv4, indicating room for improvement over IPv6. GGCs reduced IP path lengths and latencies by up to a third over IPv4 and by up to a half over IPv6, stressing the importance of content cache deployments in ISP networks. To encourage reproducibility of this work, we make the entire dataset available to the community.**

***Index Terms*—Google Global Caches, YouTube, traceroute, SamKnows, IPv6**

I. Introduction

Latency is a critical factor that impacts Quality of Experience (QoE) in networks [1]. To meet the increasing user demands in this respect, content delivery networks (CDNs) make continuous efforts [2], [3] to move popular content as close to the edge as possible by deploying content caches directly in ISP networks [4], [5]. Lately, due to the rapid depletion of the IPv4 address space, prominent ISPs have also rolled out IPv6 both in fixed-line and cellular networks. Furthermore, given that dual-stacked clients prefer requesting content over IPv6 (RFC 8305, RFC 6724) rather than IPv4, the amount of IPv6 traffic on the Internet is also increasing with these rollouts. ISPs such as Comcast and Swisscom estimate IPv6 traffic within their network to be ~25% of the total traffic today [6]. In terms of traffic volume, this amounts to more than 1 Tbps of native IPv6 traffic (as of July 2014) as witnessed by Comcast. Consequently, content providers witness an increasing trend of clients requesting content over IPv6. For instance, Google reports that ~90% of traffic originating from Verizon Wireless clients is over IPv6 [6]. Similarly, according to Facebook and Akamai, more than half of the traffic from 4 major US mobile networks to Facebook and dual-stacked websites hosted on Akamai originates from IPv6 clients alone [6].

All authors are with Technical University of Munich, Germany.

Swisscom reports (as of October 2014) that ~60% of their IPv6 traffic is served by YouTube (with ~5% by Facebook) alone. As such, YouTube is the single largest source of IPv6 traffic in ISP networks. We pose three questions: *How far are GGC nodes in terms of path lengths and latency from vantage points that stream YouTube content, how much benefit do they provide, and how do these metrics compare quantitatively over IPv4 and IPv6?* To investigate these questions, we performed `traceroute` measurements towards YouTube media server locations over both address families (see § II) from ~100 geographically distributed vantage points to ensure the diversity of network origins. These vantage points receive native IPv4 and IPv6 connectivity and belong to different ISPs, covering 74 different origin ASes. This paper presents an analysis of a longitudinal dataset (05/2016 – 05/2018) towards YouTube media server locations (or GGC nodes in situations where the video was cached), collected by these dual-stacked vantage points. Our main **findings** are –

- Path lengths and latencies towards YouTube media server destinations over both address families were similarly distributed, with ~50% of the paths being faster over IPv6, while in ~91% of the cases (see § III for details), path lengths and latencies between address families differed by 5 IP hops and 20 ms of each other.
- GGC nodes were reachable within ~6 IP hops and within ~20 ms (see § IV for details) from vantage points streaming the videos. In situations where GGC nodes were dual-stacked, path lengths and latencies were comparable over both address families. However, contrary to what is generally believed, shorter path lengths did not always correlate with lower latency: when the video was cached by a GGC node over IPv6 only, paths were almost always shorter over IPv6, yet, latencies were still lower over IPv4.
- GGC nodes reduced IP path lengths and latency by up to a third over IPv4 and by up to a half over IPv6 (see § V for details). As such, ISPs need to ensure their caches are dual-stacked to benefit from the reduced latency.

To the best of our knowledge, this is the first study to quantify the latency and path lengths towards YouTube media servers while also considering disparities between address families along with GGC node deployments and the quantification of their benefits. To encourage reproducibility [7], the entire dataset and software (see § IX for details) used in this study will be made available to the community.

 https://doi.org/10.1109/MCOM.2018.1800132

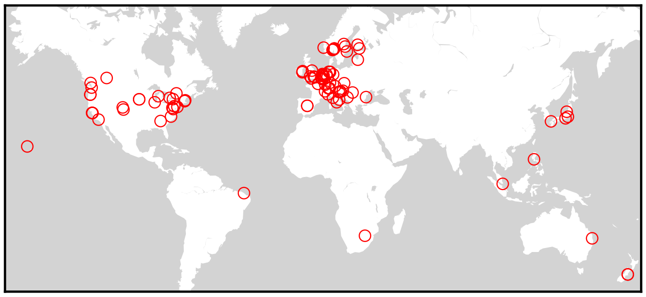


Fig. 1. **Geographic distribution of ~100 dual-stacked SamKnows probes as of June 2018. The metadata for each probe is available online: https://goo.gl/NN81ij [accessed 26-June-2018].**

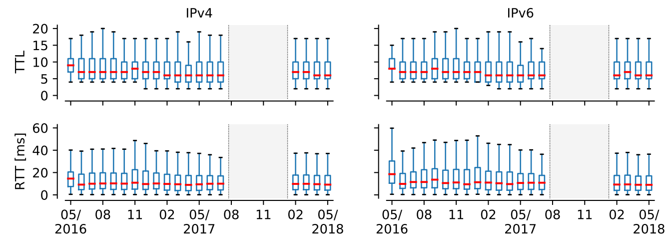


Fig. 2. **Boxplots of TTL (top) and RTT (bottom) values for IPv4 (left) and IPv6 (right) by month over the entire duration of the dataset. During the grayed out area (08/2017 – 01/2018), data was filtered out due to missing and faulty measurements. The path lengths and latencies towards YouTube destinations appear congruent over both address families.**

II. Methodology

We have developed a `youtube` test [8] that downloads and mimics the playing of YouTube videos. The test runs twice, once for IPv4 and subsequently for IPv6, and repeats every hour. We use the YouTube Data API to generate a list of globally popular videos, which is refreshed every 24 hours. This allows us to measure the same video for the entire day, which allows for temporal analysis, while cycling videos on a daily basis allows for larger coverage (~1865 videos as of May 2018) with different characteristics in the longitudinal dataset. We refer the reader to our previous work [8] for a more detailed description of the `youtube` test.

We run `scamper` [9] immediately after the `youtube` test completes. The `scamper` test performs `paris-traceroute` [10] over ICMP towards the media server destinations identified by the `youtube` test (or GGC nodes in situations where the content was cached) both over IPv4 and IPv6. The identification of GGC nodes is performed offline and is not forced during the measurement phase. These measurements are run every hour to keep the load on the volunteers' networks tolerable, while still allowing us to collect a representative dataset of captured latencies and path characteristics for the streamed videos. The measurements are performed from ~100 SamKnows [11] probes, whose locations are shown in Fig. 1. These probes are connected in dual-stacked networks representing 74 different origin ASes. Most of these probes are connected within the RIPE (60 probes) and ARIN region (29 probes), and are mainly hosted in home networks (~75) with native IPv6 connectivity.

The performance dataset collected by the `youtube` test was reported in our previous study [8]. In another recent study [12], we further investigated the stability and load balancing behavior of paths towards YouTube, in particular with regards to interdomain links. In this paper, we take this research forward by analyzing the longitudinal `scamper` dataset (05/2016 – 05/2018) to quantify the path lengths and latency towards YouTube destinations.

III. Temporal and Spatial View

We began the analysis by calculating the rate of successful executions of `traceroute` measurements within our dataset. We observed that `traceroute` measurements towards YouTube destinations had very low failure rates in terms of reachability (0.5-2% failure over both address families). However, the causality analysis revealed different reasons for the failure with the highest contribution. For instance, for IPv4, loops were the most common causes of failures, as opposed to unreachable destinations over ICMP for IPv6. In this way, due to overall low failure rates, the rest of the analysis was performed for situations where `traceroute` was successful.

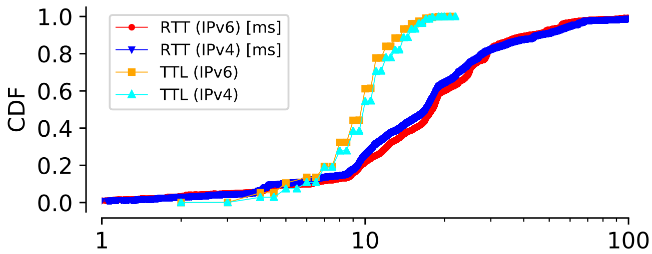
A. Temporal View

Given the longitudinal nature of the dataset, we wanted to know: *have IP path lengths and latencies towards YouTube destinations improved over the observation period?* Towards this pursuit, we studied the distributions of maximum TTL (representing path lengths) and RTT (representing latencies) values for all IPv4 and IPv6 measurements by month, shown in Fig. 2. As can be seen, the interquartile ranges of both metrics appear congruent over both address families across all months. Aggregations by days and hours delivered similar observations, indicating that there have been no substantial changes over the observation period. Additionally, we calculated the deltas as the absolute difference between IPv4 and IPv6 TTL (and RTT) values for a measurement, and observed that the deltas were consistently located around zero for both path lengths and latencies as well. Further analysis of individual vantage points and their measurements added little variance in the deltas over either address family. *Takeaway: The longitudinal view of path lengths and latency towards YouTube media server destinations appeared congruent over both address families over the duration of the dataset.*

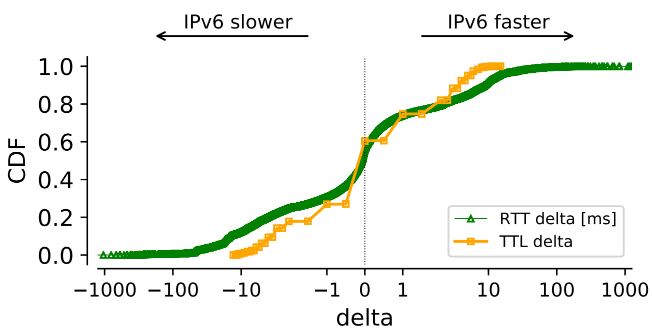
B. Spatial View

For all discovered paths, which we defined as the distinct 3-item tuple of probe, source and destination addresses, we began by analyzing the distribution of the median IP path lengths and latencies. We witnessed a low fraction of outliers and apply median aggregates to avoid bias caused due to limitations (see § VII for details) with `traceroute` measurements. Fig. 3a shows that paths over both address families are very similarly distributed.

We go further and compare IPv4 with IPv6 in terms of path lengths and latencies towards individual YouTube media



(a) **Distribution of IP path metrics.**



(b) **Distribution of destination pair deltas.**

Fig. 3. **(a) Distribution of median TTL and RTT values for paths. The distribution curves of each metric are largely overlapping for both IPv4 and IPv6, suggesting a comparable performance considering the similar number of samples. (b) Distribution of median TTL and RTT deltas of destination pairs, indicating symmetry around zero. Around half of the pairs were faster over IPv4, whereas the other half were faster over IPv6.**

servers to see which address family provides better performance. To do so, we introduced the notion of destination pairs. A destination pair is a tuple of destination endpoints over IPv4 and IPv6 that were streaming a specific video to the same vantage point within an hourly window of each other, yielding ∼43K distinct destination pairs.

Fig. 3b shows the median TTL (and RTT) deltas for such destination pairs. As can be seen, path lengths and latencies towards the destinations follow a very similar distribution. ∼27% of the destination pairs had longer path lengths over IPv6 (with negative deltas) while ∼40% of the destination pairs had shorter path lengths over IPv6 (with positive deltas), meaning that more paths over IPv6 were shorter when compared to IPv4. The remaining ∼33% destination pairs had equal path lengths over both address families (exhibiting median deltas at zero). Similarly, around half of the destination pairs had RTT deltas that were less than −0.05 ms, i.e. less than roughly zero. As such, destination pairs for the latency distribution were split into almost two equally-sized bins, where half of the destinations could be reached faster over IPv6, while the other half were reached more quickly over IPv4. Also, ∼12.5% of the destination pairs were reached more than 10 ms faster over IPv4, while only ∼8% of the destination pairs were more than 10 ms faster over IPv6. Interestingly, even though more paths were shorter over IPv6, more paths had lower latency over IPv4. Looking at the entire distribution, it can be seen that the majority of the path lengths (∼91%) differed by 5 IP hops and the majority of the latencies (∼91%) differed by 20 ms across address families. *Takeaway: The path lengths and latency distributions of YouTube destination pairs over IPv4 and IPv6 followed a very similar distribution.*

IV. How Beneficial are Content Caches?

Content cache deployments bring performance benefits by bringing content closer to the user. However, these benefits have not been quantified in practice. Our longitudinal dataset allowed us to carry out a quantitative comparison of performance (in terms of path lengths and latency) benefits as witnessed when the video was streamed from a GGC deployed directly within the ISP's network versus when the video was streamed from a CDN outside the ISP's network boundary.

Towards this goal, we employed a heuristic for both address families where we mapped the IP endpoint of the YouTube media server destination to the ASes that announce the most specific prefix encompassing the IP endpoints. Given knowledge about the source ASes of the probes and considering that ∼75% of the probes are deployed in home networks, a match in the source and destination ASes indicates that requests did not leave the ISP boundary, revealing the presence of a GGC delivering the YouTube video to the vantage point. For validation, we further performed reverse DNS lookups for each (∼20K unique) destination IP endpoint. These reverse DNS entries (such as `cache.google.com`, `*.cache.googlevideo.com`) were used as additional identifiers for GGC nodes. Non-matching cases were labeled as *no-cache* scenarios. Note that the goal here was not to exhaustively identify GGC deployments, but to measure latency and path lengths for cases in which GGC nodes could be identified with reasonable certainty.

A. Benefits between different address families

Using Fig. 3b as the baseline, we split the path length and latency distribution of destination pairs into four situations based on whether the video was delivered by 1) cache over IPv4 but not over IPv6, 2) cache over IPv6 but not over IPv4, 3) caches over both address families, and 4) caches over neither address family. In situations 1) and 2) where a cache was present over only one address family, one would expect to observe shorter path lengths and latency over the address family that had the video delivered from a deployed cache. Fig. 4 shows the quantification of these benefits in terms of path lengths and latency when a cache delivered the video. It can be seen that the distributions shift away from zero (congruency) in situations when the cache was available on one address family (blue and red) only. The offset of the shift quantifies the benefit the video delivery from a cache brings when compared to the no-cache deployment scenario over the other address family. It must also be noted that the offset (especially with latency) is weaker when the cache was only available over IPv6, potential reasons for this observation are discussed in § IV-B.

In situations where both address families had the video delivered over a cache (purple), the distribution converges towards zero (congruency), indicating that the path lengths and latency were comparable and exhibited less variation in the

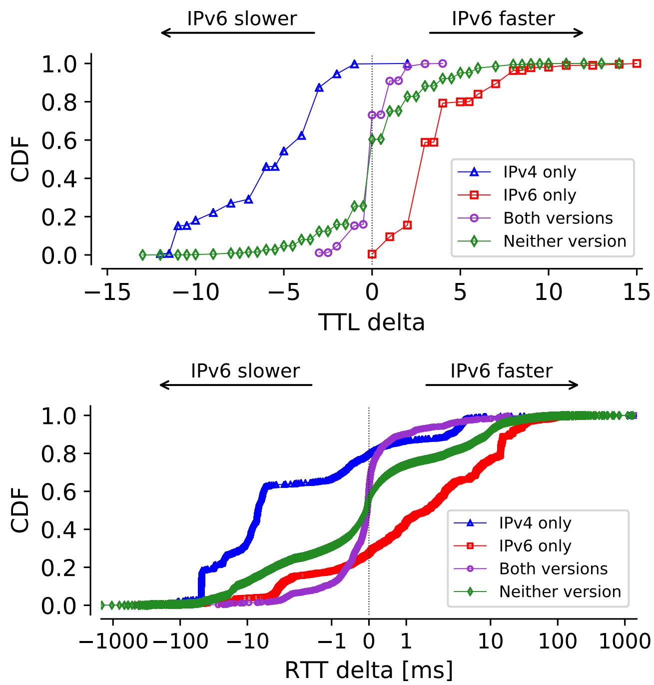


Fig. 4. **Distribution of deltas of TTL (top) and RTT (bottom) for destination pairs, classified by cache presence for IPv4 and IPv6: (1) IPv4 only; (2) IPv6 only; (3) Cache present for both address families; (4) No cache present for either address family. Latency and path length distributions shift away from congruency when a cache was available over one address family only. Latency and path lengths are comparable between address families in situations when caches were dual-stacked.**

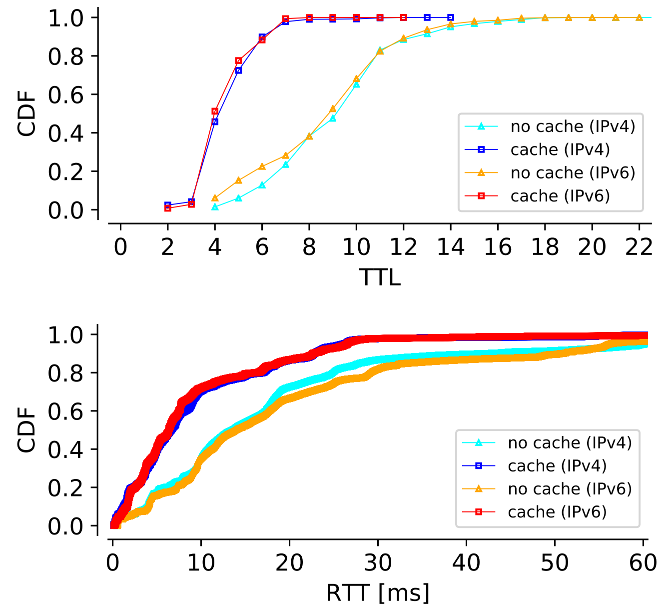


Fig. 5. **Comparison of path lengths (top) and latency (bottom) of individual IPv4 and IPv6 samples, split by cache and no-cache deployments. Caches reduce IP path lengths and latencies by up to ~50%.**

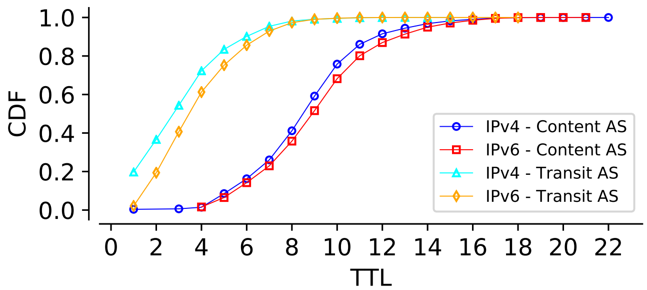
distribution. The green curve represents the situation when the video was not delivered from an identified cache over either address family, with the largest variation in path length and latency distribution between address families. As such, the path lengths and latencies were comparable and predictable when caches were involved in both address families.

Nearly all caches over IPv4 (~99.7%) had path lengths that were at least 1 IP hop shorter than when IPv6 had no cache present. Similarly, ~99.5% of the cases with an IPv6 cache had path lengths that were at least 1 IP hop shorter than when IPv4 had no cache present. In situations where caches were present in both address families, ~60% of the samples were faster over IPv4, while ~40% of the samples were faster over IPv6. However, in these situations, ~79% of the samples were within 1 ms apart from each other, suggesting that the difference in latency when caches were dual-stacked was usually marginally small. When caches were only available over IPv4, we observed that in ~80% of the samples the latency was lower over IPv4, with ~35% of the samples being faster by even more than 10 ms over IPv4.

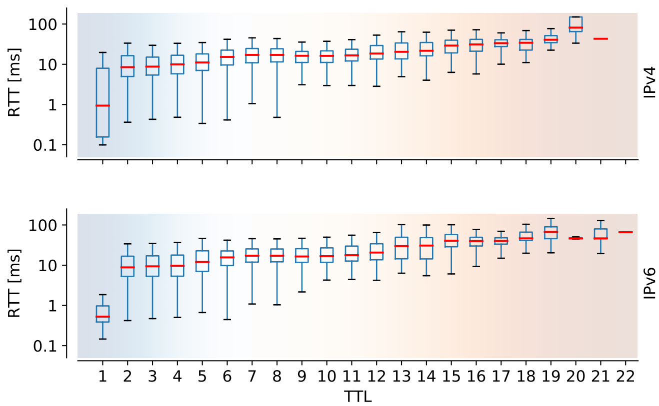
The most interesting finding was for cases where caches were only available over IPv6. Even though paths were shorter towards these IPv6 caches, ~27% of the samples were still slower over IPv6 when compared to IPv4. In these situations, streaming the video from a local cache over IPv6 was still slower than leaving the ISP boundary over IPv4 to stream the video. Overall, we witnessed several cases where content was served from a cache over only one address family, hinting towards areas of improvements in dual-stacking content cache deployments. *Takeaway: In situations where GGC nodes were dual-stacked, IP path lengths and latencies were comparable over both address families. Shorter IP path lengths, however, did not always correlate with lower latency as is generally believed: when the video was cached by a GGC node over IPv6 only, paths were almost always shorter over IPv6, yet, latencies were still lower over IPv4.*

B. Benefits within the same address family

We went further to quantify *by how much* the path lengths and latencies were reduced due to GGC node deployments within the same address family. Fig. 5 shows the distribution of path lengths and latency for all the samples grouped by address family. It can be seen that almost all the samples reached the cache over IPv4 within 7 IP hops. In ~90% of the samples, caches were reached within 6 IP hops, while ~88.5% of the no-cache destinations were reached in 12 hops. As such, caches reduced IPv4 path lengths by up to a half. Regarding IPv6, similar observations were made: nearly all caches were reached in at most 7 IP hops. In ~88% of the samples, caches could be reached within 6 IP hops, while ~89% of no-cache servers were reached in 12 IP hops, leading to the same observation that a cache reduced IP path lengths by up to a half for IPv6. Considering latencies, the majority of samples were able to reach a cache within 20 ms over both IPv4 and IPv6 (~87% for each address family). Interestingly, when quantifying the improvements in latency, we observed that over IPv4 differences between caches and no-cache deployments were smaller than over IPv6. In particular, over IPv6 latencies could be reduced by up to about a half (from ~29 ms to ~16 ms for ~80% of the samples), while over IPv4 latencies could only be reduced by up to roughly



(a) **Distribution of intermediate IP endpoints over path lengths by AS types.**



(b) **Boxplots of latency by IP path lengths.**

Fig. 6. **(a) Distribution of path lengths towards intermediate IP endpoints classified by AS types. Path lengths up to 7 IP hops can serve as a heuristic to delineate the boundary of ISP networks and CDN providers. (b) Boxplots of latency by path lengths over IPv4 (top) and over IPv6 (bottom) as seen from residential vantage points. The blue tone represents caches in ISP networks, while the red tone represents CDNs.**

a third (from ~25 ms to ~17 ms) for a similar amount of samples.

In § IV-A, we observed that for cases in which a video was only streamed from a local IPv6 cache but was streamed from outside the ISP network over IPv4, the IPv6 cache brought less latency benefits than seen for the reverse case. In § III-B, we also showed that even though more destination pairs had shorter path lengths over IPv6, more pairs still had lower latencies over IPv4. When comparing cache benefits within the same address family, we observe that deploying caches over IPv6 brings more benefits in terms of reducing IP path lengths and latency than the cache deployments over IPv4. We speculate the observation could be either due to IPv4 optimization or due to router bufferbloat. However, the dataset does not have the information to provide direct evidence for these assumptions. *Takeaway: Cache deployments over either address family were able to reduce IP path lengths by up to a half (6 IP hops). However, latency was halved over IPv6 but only reduced by a third over IPv4 (up to about 13 ms, respectively 8 ms, of difference), highlighting the importance of IPv6 cache deployments.*

V. How Far are the Content Caches?

We further analyzed the intermediate IP hops on the entire path from source (vantage points) to destination. We mapped each intermediate IP endpoint to its corresponding ASN over both address families. Further, we augmented the ASN by using the CAIDA's AS classification dataset to classify each AS by its network type, such as Transit/Access, Enterprise and Content ASes.

Fig. 6a shows the distribution of path length towards intermediate IP endpoints, further categorized by Transit/Access and Content AS (the number of intermediate IP endpoints assigned to Enterprise ASes was negligible). It can be seen that more than ~93% of the IP endpoints belonging to Transit/Access type ASes had a path length of *at most 7 IP hops* from the probe over both address families. On the contrary, ~85% of the IP endpoints assigned to Content type ASes had a path length of *at least 7 IP hops* from the probe over both IPv4 and IPv6. As such, 7 IP hops can serve as a heuristic to identify the boundary between ISP and CDN providers. Fig. 6b visualizes this trend. It shows the correlation of latency with increasing path lengths and uses the heuristic of 7 IP hops to delineate the boundary of ISP and content providers. The spike from the first to the second IP hop are last-mile latencies within the ISP's network. The shorter path lengths (blue region) represent caches (and their subsequent latency) in ISP networks, while longer path lengths (red region) represent CDN networks. This analysis also supports our previous findings where we demonstrated that most caches were reachable within a path length of 7 IP hops, see Fig. 5. *Takeaway: Path lengths of up to 7 IP hops can serve as a heuristic to delineate the boundary of ISP networks and CDN providers.*

VI. Related Work

Chiu *et al.* [3] (2015) run `traceroute` from Planetlab vantage points to measure the average path of high traffic routes on the Internet. They find that most paths from a large content provider, Google, go directly from Google's network into the client's network, as opposed to average paths that traverse one or two intermediate transit ASes. They also show that, from the same content provider, popular paths serving high-volume client networks tend to be shorter than paths to other networks, observing that some content providers even host servers in other networks, which in effect shortens paths further. Livadariu *et al.* [13] (2016) show that IPv6 exhibits less stability than IPv4 in the control and data plane. The reason for worse performance in the control plane was that most routing dynamics were generated by a small fraction of pathological unstable prefixes, whereas in the data-plane, episodes of unavailability were longer on IPv6 than on IPv4. Basu *et al.* [14] (2018) propose two TTL-based caching algorithms that are able to handle non-stationary request traffic for CDNs. The algorithms are able to adapt to the traffic automatically and can simultaneously achieve target hit rates with high accuracy, evaluated on the basis of a high volume dataset from a production CDN.

VII. Limitations

The observations are biased by the number and location of the vantage points, which largely cover US, EU and JP regions. However, a large fraction of IPv6 deployment is also centered in these regions, but we acknowledge that the state of IPv6 adoption may change in the future.

`traceroute` is also known to have several limitations [15]. For instance, ICMP packets may encounter issues with non-cooperative routers on the path and thus be highly variable in response times, even in unloaded networks. The dataset, however, does not have information about the network load to allow filtering such outliers. It is assumed that in `traceroute` measurements, ICMP responses are generated from the ingress interface of each router on the forwarding path. However, studies have shown that some routers also send ICMP responses using their egress interface. As such, routers that are located on the border of two ASes can be mapped to either of the ASes in such scenarios.

We did not perform region-based analysis of our results for two reasons. First, studies [5] have shown that the AS-based network topology (as revealed by `traceroute`) does not represent the geographical topology of the traversed paths. Second, even with ground-truth information of probe locations, grouping probes by their geographical regions results in a very low sample size per group. As such, it becomes infeasible to reasonably analyze and discuss IP path lengths and latencies taking the regional information into account for this study.

VIII. Conclusion

We presented the benefits of content cache deployments in ISP networks with path length and latency comparisons over both address families. When streaming YouTube, we identified that streaming from a GGC node reduced path lengths and latencies to reach the content within ~6 IP hops and within 20 ms. However, cache deployments were seen to have more benefit for IPv6, as IPv6 caches reduced latencies by up to roughly 50%, whereas IPv4 caches only improved latencies by up to around 33%. As such, ISPs can improve latencies significantly by deploying GGC nodes within their network. Further, we observed that in situations when GGC nodes were dual-stacked, path lengths and latencies were comparable over both address families. Therefore, we recommend ISPs who are either in the process of or are in early stages of IPv6 deployment to ensure that their existing GGC nodes are dual-stacked. We also showed that when the video was cached by a GGC node over IPv6 only, paths were almost always shorter over IPv6. However, latencies were still higher when compared to an IPv4 no-cache deployment, indicating room for improvement regarding routing and processing over IPv6. We hope these insights will help both ISPs and CDN providers to optimize their content delivery over IPv6 to the benefit of user QoE.

IX. Reproducibility Considerations

The dataset collected by running `scamper` from SamKnows probes is stored as a SQLite database (along with the SQL schema) and is made publicly available[1], together with the Jupyter notebooks used in the analysis to generate the plots. Guidance on how to reproduce these results is provided, and reproducers are encouraged to contact the authors for further details and questions.

References

[1] S. Sundaresan, N. Feamster, R. Teixeira, and N. Magharei, "Measuring and Mitigating Web Performance Bottlenecks in Broadband Access Networks," in *Proceedings of the 2013 Internet Measurement Conference, IMC*, 2013. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504741

[2] X. Fan, E. Katz-Bassett, and J. S. Heidemann, "Assessing Affinity Between Users and CDN Sites," in *Traffic Monitoring and Analysis - 7th International Workshop, TMA*, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-17172-2_7

[3] Y. Chiu, B. Schlinker, A. B. Radhakrishnan, E. Katz-Bassett, and R. Govindan, "Are We One Hop Away from a Better Internet?" in *Proceedings of the 2015 ACM Internet Measurement Conference, IMC*, 2015. [Online]. Available: http://doi.acm.org/10.1145/2815675.2815719

[4] F. Chen, R. K. Sitaraman, and M. Torres, "End-User Mapping: Next Generation Request Routing for Content Delivery," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, SIGCOMM*, 2015. [Online]. Available: http://doi.acm.org/10.1145/2785956.2787500

[5] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. S. Heidemann, and R. Govindan, "Mapping the Expansion of Google's Serving Infrastructure," in *Proceedings of the 2013 Internet Measurement Conference, IMC*, 2013. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504754

[6] Internet Society, "State of IPv6 Deployment 2017," [Online; accessed 26-June-2018]. Available: https://goo.gl/u1tiGK.

[7] V. Bajpai, M. Kühlewind, J. Ott, J. Schönwälder, A. Sperotto, and B. Trammell, "Challenges with Reproducibility," in *Proceedings of the Reproducibility Workshop, Reproducibility@SIGCOMM*, 2017. [Online]. Available: http://doi.acm.org/10.1145/3097766.3097767

[8] V. Bajpai, S. Ahsan, J. Schönwälder, and J. Ott, "Measuring YouTube Content Delivery over IPv6," *Computer Communication Review*, 2017. [Online]. Available: http://doi.acm.org/10.1145/3155055.3155057

[9] M. J. Luckie, "Scamper: a Scalable and Extensible Packet Prober for Active Measurement of the Internet," in *Proceedings of the 10th ACM SIGCOMM Internet Measurement Conference, IMC*, 2010. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879171

[10] B. Augustin, T. Friedman, and R. Teixeira, "Measuring Load-balanced Paths in the Internet," in *Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference, IMC*, 2007. [Online]. Available: http://doi.acm.org/10.1145/1298306.1298329

[11] V. Bajpai and J. Schönwälder, "A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts," *IEEE Communications Surveys and Tutorials*, 2015. [Online]. Available: https://doi.org/10.1109/COMST.2015.2418435

[12] R. K. P. Mok, V. Bajpai, A. Dhamdhere, and kc claffy, "Revealing the Load-Balancing Behavior of YouTube Traffic on Interdomain Links," in *Passive and Active Measurement - 19th International Conference, PAM*, 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-76481-8_17

[13] I. Livadariu, A. Elmokashfi, and A. Dhamdhere, "Characterizing IPv6 control and data plane stability," in *35th Annual IEEE International Conference on Computer Communications, INFOCOM*, 2016. [Online]. Available: https://doi.org/10.1109/INFOCOM.2016.7524465

[14] S. Basu, A. Sundarrajan, J. Ghaderi, S. Shakkottai, and R. K. Sitaraman, "Adaptive TTL-Based Caching for Content Delivery," *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 1063–1077, 2018. [Online]. Available: https://doi.org/10.1109/TNET.2018.2818468

[15] B. Donnet and T. Friedman, "Internet Topology Discovery: a Survey," *IEEE Communications Surveys and Tutorials*, 2007. [Online]. Available: https://doi.org/10.1109/COMST.2007.4444750

[1] https://github.com/tv-doan/youtube-traceroutes

Trinh Viet Doan is a research assistant and PhD student at the Chair of Connected Mobility at the Technical University of Munich (TUM), Germany, from which he also received his Master's degree in informatics (2017). His research interests include Internet measurements as well as decentralized Internet infrastructures.

Ljubica Pajevic is a research assistant at the Chair of Connected Mobility at TUM. She holds a Licentiate degree in electrical engineering from KTH Royal Institute of Technology (2013) and an M.Sc. degree in systems engineering and wireless communications from the School of Electrical Engineering, University of Belgrade, Serbia (2009). Her research interests are modeling and performance evaluation of opportunistic mobile networks, and data mining in wireless networks.

Vaibhav Bajpai is a postdoctoral fellow at TUM. He works with Prof. Jörg Ott. He received his PhD (2016) and Master's (2012) degrees in Computer Science from Jacobs University Bremen, Germany, where he was advised by Prof. Jürgen Schönwälder. He is the recipient of the IEEE COMSOC award for the best dissertation in network and service management in 2017. He is interested in Internet measurements, network management and reproducibility of scientific research.

Jörg Ott holds the Chair of Connected Mobility in the Department of Informatics at TUM. He is also Adjunct Professor for Networking Technology at Aalto University. He received his diploma and doctoral (Dr.-Ing.) degree in computer science from TU Berlin (1991 and 1997, respectively), and his diploma in industrial engineering from TFH Berlin (1995). His research interests are in network architecture, (Internet) protocol design, and networked systems, with a focus on (mobile) decentralized services.