Adaptive video streaming and multipath: can less be more?

Vitalii Poliakov, Lucile Sassatelli, Damien Saucez

Adaptive video streaming and multipath: can less be more?

Vitalii Poliakov, Lucile Sassatelli
Université Côte d'Azur, CNRS, I3S, France
Email: {last}@i3s.unice.fr

Damien Saucez
Université Côte d'Azur, Inria, France
Email: {first.last}@inria.fr

***Abstract*—A prominent portion of the traffic carried by Internet Service Providers (ISP) is delivered by Content Delivery Networks (CDNs). CDNs usually deploy their own caches at ISPs in addition to the servers deployed in their own premises. With the adoption of multipath transport protocols such as MPTCP, content consumers are able to retrieve their data from either caches deployed in their ISPs or server in the CDN premises using multiple paths in parallel. In this paper we present a study on the interplay between actors of a multipath-enabled video delivery system. For that we design a realistic numerical evaluation experiment; results obtained suggest that multipath can be both beneficial and harmful for ISP and CDN in different representative scenarios.**

I. Introduction

The answer to increased bandwidth demand has always been to over provision Internet Serivice Provider (ISP) networks. However, the demand becomes so important nowadays that this approach starts to show its limits. To tackle this issue, bandwidth aggregation is a candidate that gains momentum. The idea behind bandwidth aggregation is to group multiple Internet connections (e.g., wifi and LTE) together to sum their bandwidth ([1]). Aggregation is commonly used in servers and Ethernet networks to group links and ports to increase their bandwidth and resiliency [2]. However it is restricted to one hop only and thus cannot be deployed on user equipments such as smartphones in the Internet especially since using multiple radio networks simultaneously on a device was still impossible a few years ago. Without this feature, only one network can be used at a time. This means that the entire traffic of the device must cross the one network selected. It is possible to change it over time, but it requires to move all individual users' traffic from one network to the other.

Modern devices can simultaneously use multiple radio interfaces to transmit and receive data. This offers new perspectives in terms of bandwidth aggregation at the Internet level. In fact, if all network interfaces can be used simultaneously, it is possible to put a fraction of the flows on one network and the rest on the other or even to split a flow on two networks. It results that applications can balance the load of their traffic over multiple networks and potentially get higher quality of experience without ISPs to invest in re-provisioning their network as the traffic peaks can be smoothened over multiple networks. Initial deployments of this solution show its feasibility [3] and call for studying the case of Internet-wide bandwidth aggregation more carefully.

In this paper we seek an answer to the following question: "Can multipath be harmful for adaptive video delivery over Internet?". To do this, we model the case of a video delivery system incolving Contend Delivery Networks (CDN) that leverages the Internet wide bandwidth aggregation to use multiple ISPs simultaneously by the mean of a multipath transport. In order not to study the effects of implementation details, we instead focus on conceptual, higher level view of the problem. We discover that while multipath can increase the average quality of experience, it comes at an expense for ISPs that see their congestion disproportionally increasing under high load because multipath is able to scrounge the last bits of available bandwidth on every ISP reducing then the number of served requests. The situation improves when caches are used but without completely solving the contention problem.

The paper is structured as follows. We start with a short related work overview in Sec. II. Sec. III explains the system we study, as well as implications of using multipath in it. Evaluation results are presented in Sec. IV, which is followed by concluding remarks in Sec. V.

II. Related work

MultiPath TCP ([4]) gave a second life to the well-researched concept multipath data delivery ([5], [6]). Chen et al. [7] performed an in-depth study of real world MPTCP performance and concluded that in basic scenarios (like arbitrary data download from a remote server) MPTCP proves to be a resilient and efficient transport protocol. On the other hand, Arzani et al. ([8]) studied scheduling policies in MPTCP and provided a slightly opposite point of view on its performance. According to them, current MPTCP implementation can perform worse compared to usual TCP in certain cases. In order to avoid the impact of MPTCP implementation design, we decided instead to focus on high-level evaluation to study how does the concept of multipath behave in our scenario.

As for the concept of CDN and its server selection, Nikravesh et al. ([9]) identify, among other, a technical aspect of selecting CDN servers for MPTCP-enabled environments. Apostolopoulos et al. ([10]) have provided an overview of conceptual challenges for CDNs when employing MPTCP. Our study is more narrow; we focused on a particular yet representative scenario involving mobile networks and limitations related to them in terms of cache placement.

Finally, Jiang et al. in [11] provide a mathematical framework for cooperation between CDN and ISP in a singlepath environment, which we base our study on.

III. Video delivery and multipath

Video delivery typically involves a CDN hosting videos, one or multiple ISPs, and a group of clients accessing the CDN via their ISPs. With the generalization of Internet connectivity it is common for a client to be multihomed, i.e., connected to the Internet directly via multiple ISPs. In multihomed environments, clients can use each connection independently or aggregate them as one single logical connection by the means of *multipath* protocols such as MPTCP [12].

To study how does such multipath delivery work in a scenario of video distribution, we designed a representative numerical evaluation experiment. This experiment demonstrates the interplay between four actors, namely: one CDN, two independent ISPs and a large group of clients, as depicted in Fig. 1. Let us consider their roles and behavior in detail.

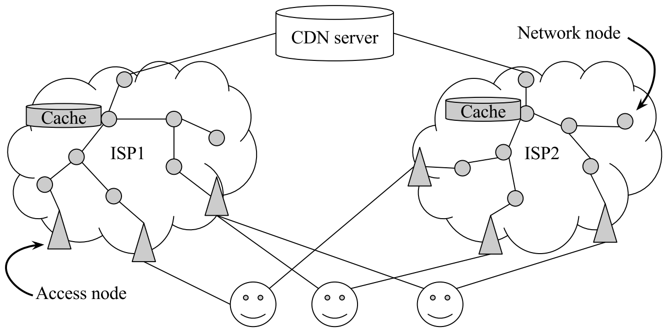


Figure 1: General view of a multipath video delivery. A group T of clients is connected to two ISPs (*ISP1* and *ISP2*) that are connected to a CDN.

A. Content Distribution Network

CDNs datacenters are connected to virtually all their client's ISPs either by private circuits (e.g., leased lines), via Internet eXchange Points (IXPs), or via transit providers [13] and can thus leverage the possibility for multipath video delivery to their clients. Video streaming servers are located in these datacenters but CDNs can optionally deploy caches directly in ISPs premises to reduce latency and path length.

We model a scenario where CDN has one datacenter with a pool of servers deployed behind a multipath capable load balancer (i.e., clients see the datacenter as one single server reachable from both ISPs and hosting the entire video catalog). The datacenter is directly connected to both ISPs with a private circuit. When caches are used, each cache is a replica of the entire CDN video catalog and does not support multipath. To avoid cross-ISP traffic, caches can only serve clients in the customer cone of their ISP. The video catalog is constant and all videos are pre-loaded on every server and caches.

CDNs aim at delivering videos to as many subscribers as possible to maximize their incomes so we assume that the CDN is provisioned enough to achieve that and it believes that so is the ISP. In this paper the CDN can take advantage of video quality adaptation in order to select a bitrate the request can sustain while being served [1]. Using this capability, CDN hence aims at maximizing the request bitrate. We model it with the following integer linear optimization problem.

Let our CDN hosts a server, a set of caches deployed at two ISPs, and a set of clients (each of them is connected to both ISPs). This can be represented as a graph $O = (V_o, E_o)$ where V_o are network nodes and E_o are directed links connecting them. Let us denote the root server and caches as $s \in V_o$, and clients as $t \in V_o$. In this way, a flow between a cache or server and a client is x^{st}, while the path it follows has a capacity of C_{st}. We assume that ISPs' network topologies are not exposed to the CDN, hence any t is only one hop away from any accessible s.

Considering that clients make requests (which we denote as i) to the CDN, we can represent server selection for it as $\max \sum_i x_i^{st}$, where x_i^{st} is the served request bandwidth that we use as an indicator of client's perceived video quality. Since this problem is hard to solve for big amount of requests, and considering that it is not viable to know the future requests, we can transform our problem into an online one as presented in Problem 1, which makes decision only based on one request.

$$
\begin{aligned}
\max \quad & x^{s_1 t} + x^{s_2 t} + x^{s_3 t}_{f_1} + x^{s_3 t}_{f_2} \\
\text{s.t.:} \quad & \\
\textstyle\sum_s p^{st} &= 1, && \forall t \\
x^{st} &\leq M * p^{st}, && \forall (s,t) \\
x^{s_3 t} &= x^{s_3 t}_{f_1} + x^{s_3 t}_{f_2} \\
x^{s_1 t} &\leq C_{1t} \\
x^{s_2 t} &\leq C_{2t} \\
x^{s_3 t}_{f_1} &\leq \min(C_{1t}, C_{31}) \\
x^{s_3 t}_{f_2} &\leq \min(C_{2t}, C_{32}) \\
x^{s_1 t} + x^{s_2 t} + x^{s_3 t}_{f_1} + x^{s_3 t}_{f_2} &\leq max\_rate
\end{aligned}
\tag{1}
$$

Here M is a big number; max_rate is the bitrate of the highest quality representation of the requested video; p^{st} is a binary variable to ensure that only one server can serve a request at the same time (we do not allow multiple sources). The last constraint limits bitrate selection by the bitrate of a maximum available representation for a given video class. Fulfilling the limitation of its minimum bitrate is performed in the experiment environment itself. Note that solving this problem not only tells CDN which source to select, but also a bitrate the request can achieve while being served from it.

B. Internet Service Providers

Clients connect with the CDN through two ISPs having identical partially-meshed tree topology (see Fig 2). Each node of the topology can be a switch, a router, or a user access node (e.g., an eNodeB); it can also be a CDN cache at the same time. In this paper access nodes are located at the bottom of

[1] Note that the actual video quality at the client depends on how efficient is the quality selection in the player; this is beyond the scope of the paper.

the topology figure, and caches are located at the second from top level. Such design is representative of common mobile backhaul architectures where caches are deployed only at PGWs because of the limitations of GTP [14].

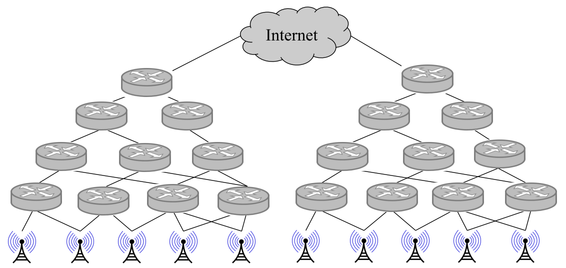


Figure 2: ISP network with a partially-meshed tree topology

Each ISP is interested in keeping its links load as low as possible while still carrying all its client traffic. While not being able to influence the traffic it has to carry, it optimizes its routing in the network by solving the single path multi-commodity network flow problem. This problem is well presented by Jiang et al. ([11], Eq. 1) and fits perfectly in our environment, so we reuse it in this work.

The cost function in the ISP problem has to be a decreasing convex function; in our experiments, we reuse one from *Fortz and Thorup* [15]. An example of such a function for our experiment settings is presented at Fig. 3. As its would be impractical to solve this problem each time a flow enters the system, this optimization is performed periodically.

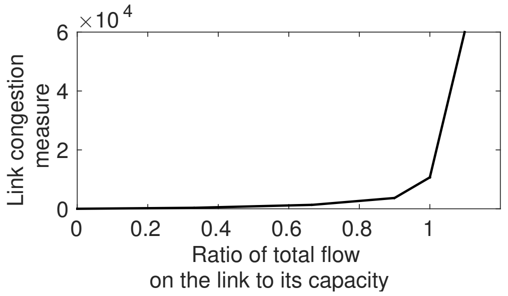


Figure 3: ISP cost functions for link capacity of 1Gbps

C. Clients

The interest of a client is to maintain the highest possible quality of experience given achievable bitrate decided by CDN for its request. A client can only stream one video at a time. Clients do not have any means of influencing the decisions of the CDN and ISPs. Therefore, the role of the client is solely to issue a request to the CDN and, by this, bring the load onto ISPs' networks.

D. Application of multipath, and disagreement between ISP and CDN

Multipath is useful in multiple applications and in our video delivery system it is used to increase the client's download bitrate. Video representations, however, have limited encoding bitrates, which can be less than client's connectivity to the network. A request in this case might be well served through single path, so multipath becomes indispensable only when aggregated available bandwidth allows to provide greater video download bitrate than that of a single connectivity.

When the CDN does not have caches, both multi- and single path deliveries have to cross the entire ISP network on the way to the CDN server – an ISP should carry the same amount of data regardless of delivery mode. On the contrary, caches at the ISPs allows the CDN to serve requests from these caches, and the traffic will not have to cross the entire ISP backhaul.

On the other hand, the CDN might want to use their multipath capabilities fairly often to guarantee high QoE. For instance, the average LTE downlink bandwidth is less than 15 Mbps in many developed countries (e.g., 13.4 Mbps in France and 12.3 Mbps in USA [16]), which is barely enough to stream Full-HD 60fps videos.

Different and potentially contradictory objectives of ISPs and CDNs cause tension between them and Sec. IV shows how multipath can take part in the tussle.

IV. Evaluation results

This section presents the results of a numerical evaluation of how does multipath video delivery affect the system described in Sec. III.

To quantify the effect, we assess the total congestion on all links of both ISPs, acceptance rate of the requests by CDN and the average achieved bitrate of those requests, as well as its distribution. We look at these quantities when CDN has no caches at ISPs and when it does; two demand scenarios are evaluated: *low load* – when ISP networks are over provisioned for the arrival of video requests, and a *high load* – when the demand is higher than the ISP networks can handle.

The video catalog is made of $10,000$ videos following a $Zipf(0.8)$ popularity distribution. Clients request videos according to a Poisson process with two arrival rate modes as explained earlier: $\lambda = 0.02$ (requests per second per each ISP access node) for *low load*, and $\lambda = 0.16$ for *high load*.

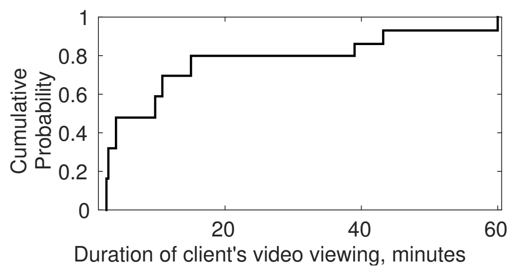


Figure 4: Users' duration of viewing in the incoming workload

Videos that are requested by clients can be divided into three classes, depending on the maximum bitrate they are encoded into. These classes can correspond to different amount of action for each one: *Static* (maximum bitrate is $8\ Mbps$), *Medium* ($max\_rate = 16\ Mbps$), and *Action* ($max\_rate = 30\ Mbps$). Distribution of these classes in the incoming

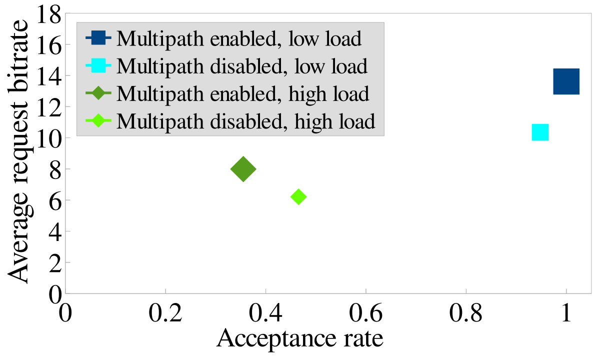
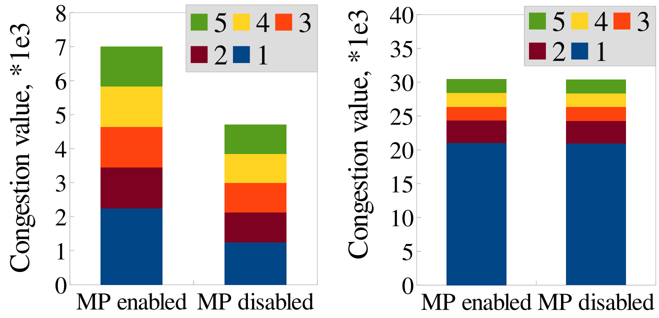


Figure 5: Average request bitrate vs request acceptance rate, without caches.



(a) Low load scenario

(b) High load scenario

Figure 6: ISP congestion, without caches.

workload into the system is uniform. Note thSimilarly to *Li et al.* [17], video duration distribution includes considerations regarding users who abandon videos before it ends as well as video popularity, and is presented on Fig. 4. The bitrate of the lowest quality representation (min_rate) is set to be $0.1 \times max\_rate$ for all three video classes; no service for a request is possible if the CDN estimates that the available bandwidth is inferior to the minimum possible rate.

Capacity of all the links in ISP networks are set to be 1 Gbps. To simulate fluctuation of clients' bandwidth until ISP's access node (e.g., eNodeB and the effect of distance with the client) the access capacity is randomly picked in the set $\{2, 5, 10, 15, 20, 25\}$ Mbps [16].

A. A case when CDN does not have caches

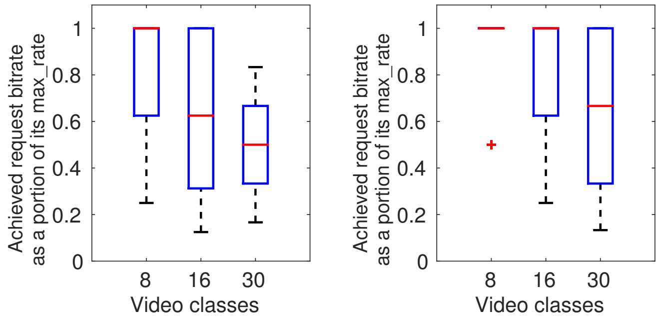
Fig. 5 shows the average request bitrate and the acceptance ratio for low and high load scenarios. The closer a scenario is to the top right corner of the figure, the better. The best possible case is obtained at low load and with multipath enabled. Without multipath, this ideal cannot be achieved even at low load as for some requests the access capacity is lower than the minimum allowed bitrate for the requested video. In this case, even though the acceptance ratio is nearly 1, the average bitrate significantly decreases, which highlights the importance of using an efficient video bitrate adaptation algorithm that manages to adapt the video quality to the actual network performance. While at low load using multipath is beneficial as it improves both acceptance ratio and the average bitrate, it is not as evident in case of high load. Under these circumstances using multipath increases the average bitrate but reduces the acceptance ratio.

While average bitrates and acceptance ratio are important metrics from the CDN perspective, congestion is another important metric to consider from an ISP standpoint. Fig. 6 shows the evolution of congestion with one of the ISPs (very similar between both due to symmetry) and its breakdown between different levels of the ISP. On these plots, level 1 is the closest to the CDN while level 5 is the level connecting the access node to the ISP network (see Sec. III). At low load, multipath can increase the congestion in the ISP network as both bitrate and acceptance increase at the same time. In contrast, using multipath at high load does not influence the overall congestion as the first level of the network is already completely congested so increasing average request bitrate can only come at the expense of reducing the number of videos streamed in the network, i.e., reducing the acceptance ratio.

Using multipath permits to increase the average bitrate of streamed video requests, however, the quality improvement is not uniform over the classes of video with a clear bias against the class Action videos, as shown by Fig. 7 and Fig. 8. These plots show the distribution of achieved download bitrates for each accepted video request, normalized by their max_rate. Fig. 7 and Fig. 8 show the quartiles and outliers with their bounds, for each video class, with and without multipath activated. Regardless of multipath we can see that videos with high max_rate are proportionally more degraded than the videos with low ones, however the reasons are different. At low load and with no multipath, not all requests can be accepted because the access link limits the maximum achievable bitrate. The bias against the video classes with more action can then be explained by the fact that it is more likely to be able to fit a lower-quality representation of a video on a limited access link bandwidth. At low load when multipath is enabled most of requests can be accepted at their full rate, except a few outliers for which the sum of access bandwidth is smaller than max_rate. This case becomes more common when the max_rate of video increases which explains that not all requests can be satisfied even at low load. Nevertheless, using multipath improves the video bitrate for most of videos. In case of high load the trend against higher-action videos classes holds regardless of using multipath. Similarly, using multipath increases the video bitrate of most videos, but now at the expense of videos from the Action class.

B. A case when CDN comprises caches

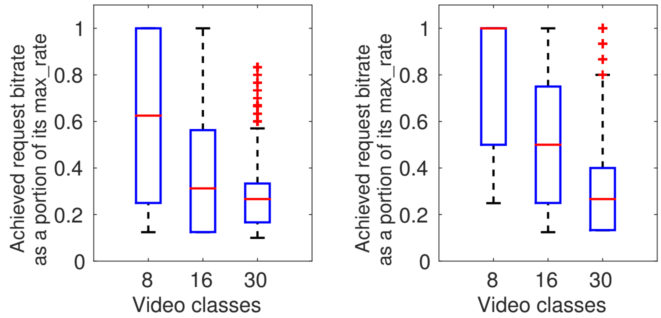
To improve their services and reduce the burden on the ISPs most CDNs replicate their contents in caches located within the ISPs of their clients [18]. In this situation, caches are used instead of the servers in the CDN unless it is impossible to use them. In this section, we study the impact of using multipath in such a case.



(a) Without multipath.

(b) With multipath.

Figure 7: Normalized achieved request bitrate at low load, without caches, for three video classes: 8 – "Static", 16 – "Medium", 30 – "Action"



(a) Without multipath.

(b) With multipath.

Figure 8: Normalized achieved request bitrate at high load, without caches, for three video classes: 8 – "Static", 16 – "Medium", 30 – "Action"

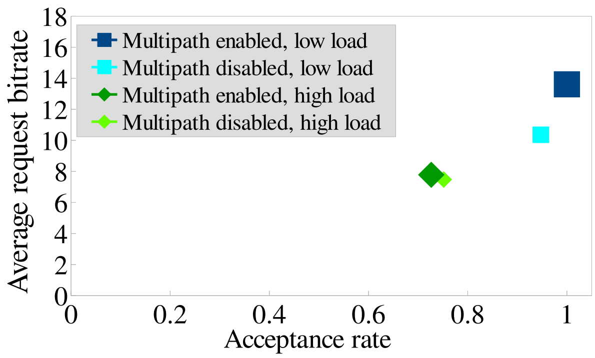
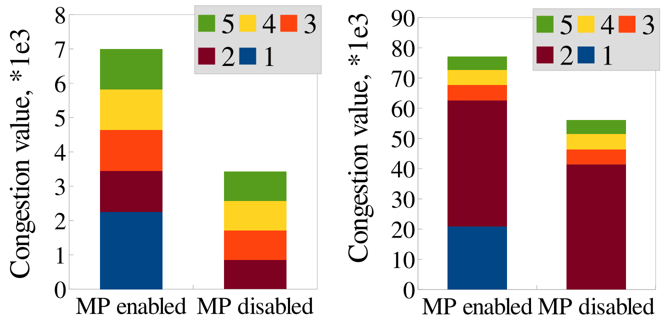


Figure 9: Average request bitrate vs request acceptance rate, with caches.



(a) Low load scenario

(b) High load scenario

Figure 10: ISP congestion, with caches.

Fig. 9 shows the average request bitrate and the acceptance ratio for low and high load scenarios. Compared to Fig. 5, at high load the acceptance ratio is improved with caches as they allow to avoid that all the traffic crosses the entire network to reach the CDN. Nevertheless, adding multipath still improves the average request bitrate but degrades the acceptance ratio under high load.

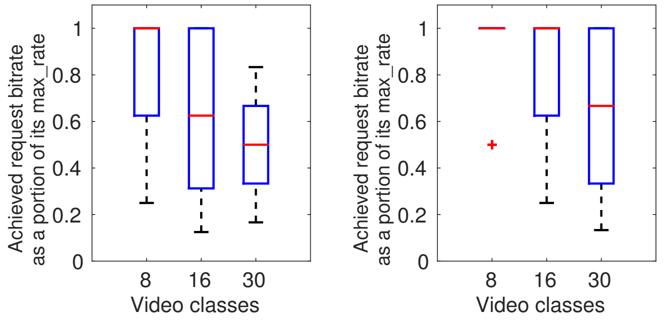
Similarly, at low load the congestion follows the same pattern than without caches. However, using caches at high load causes high congestion as the acceptance ratio is high. Without multipath, we can see that caches free the link at the ISP provider-edge but when multipath is used, the total congestion is increased and congestion re-appears at the provider-edge link.

Finally, Fig. 11 and Fig. 12 show a general bias against the Medium and Action classes and shows that the bitrate improvement observed while using multipath benefits more the Static videos. Interestingly, we can see that while it is beneficial for most of Static videos to use multipath at high load, the 25% most degraded observations gets more degraded with multipath than without. On the contrary, the 25% least degraded Action videos get better quality when multipath is used. This can be explained by the fact that in our setup the maximum access bandwidth is lower than the requested rate for the class Action videos (25 Mbps vs 30 Mbps).

V. Discussion and Future Work

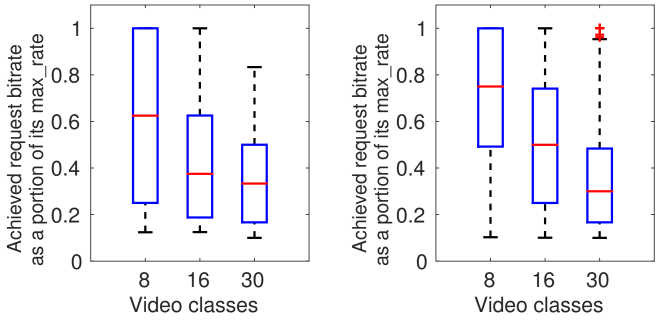
Previous sections show that using multipath may either be beneficial or detrimental. This section aims at discussing the conditions when it is interesting to use multipath and when it is preferable not to use it, and why.

When it comes to multipath, one must distinguish two cases: the case where ISP networks are over provisioned and suffer no contention, and the case where ISPs experience contention that cause rejection of requests. In the first case in our settings, all requests can potentially be fulfilled but the access link itself becomes the limiting factor – for instance, bandwidth that an LTE user equipment can obtain depends on its distance from the base station. In such circumstances, the user equipment can use its multipath capability to increase its available bandwidth without harming the traffic in the networks. It results in better Quality of Experience as the client can stream videos with a higher bitrate than over one single link, the drawback being that videos have to be streamed from deeper in the Internet core to benefit from using multiple access nodes, causing more overall load in the network. It is worth noting that in terms of bandwidth using caches at the edge is of no benefit for



(a) Without multipath. (b) With multipath.

Figure 11: Normalized achieved request bitrate at low load, with caches, for three video classes: 8 – "Static", 16 – "Medium", 30 – "Action"



(a) Without multipath. (b) With multipath.

Figure 12: Normalized achieved request bitrate at high load, with caches, for three video classes: 8 – "Static", 16 – "Medium", 30 – "Action"

the CDN as the network is able to handle all the load; it is, however, beneficial for ISPs as the traffic in this case is concentrated near its customer edge.

For what concerns bandwidth, when the network infrastructure is over provisioned using multipath can only be beneficial and can be used to implement bandwidth aggregation. Regrettably, such conclusion doesn't hold when contention occurs. In this case, the capacity of the network has to be shared between all traffic flows and the higher the load is, the lower the individual bandwidth is. One may expect that in this case flows could benefit from using multipath as they would be able to get every single bit of available bandwidth over the multiple access networks. Unfortunately, when the network is congested multipath circumvents the effect of congestion by using the last bits of available bandwidth in every network, causing even more congestion. As a result, requests that can be accepted slightly improve their quality (i.e., higher bitrate), but prevent other ones to be accepted leading to less fulfilled demands overall.

In congested networks, blindly using multipath results in marginal quality increase but dramatic contention increase and less served requests. To avoid the negative effects of contention, the solution is to add resources, either by increasing link capacity, which is not practical, or by re-deploying the degraded services away of the congestion, for example by using caches at the edge.

All the more so, it is important to seek for a solution that can help multipath transport to showcase its potential and to be more beneficial when employed in a video delivery systems such as we studied in this paper. We believe that the key for a solution lies in cooperation between the actors of the system, which we plan to study as a future work.

References

[1] B. Peirens, G. Detal, S. Barre, and O. Bonaventure, "Link bonding with transparent Multipath TCP," Internet RFC Draft, RFC Draft, July 2016.

[2] M. Sato, S. Nakajima, and K. Suzuki, "Ethernet link aggregation," Sep. 25 2012, uS Patent 8,274,980. [Online]. Available: https://www.google.com/patents/US8274980

[3] S. Seo, "Kt's giga lte," *Presentation at IETF*, vol. 93, 2015.

[4] S. Barré, C. Paasch, and O. Bonaventure, "Multipath tcp: from theory to practice," *NETWORKING 2011*, pp. 444–457, 2011.

[5] H. Han, S. Shakkottai, C. V. Hollot, R. Srikant, and D. Towsley, "Multi-path tcp: a joint congestion control and routing scheme to exploit path diversity in the internet," *IEEE/ACM Transactions on networking*, vol. 14, no. 6, pp. 1260–1271, 2006.

[6] W.-H. Wang, M. Palaniswami, and S. H. Low, "Optimal flow control and routing in multi-path networks," *Performance Evaluation*, vol. 52, no. 2, pp. 119–132, 2003.

[7] Y.-C. Chen, Y.-s. Lim, R. J. Gibbens, E. M. Nahum, R. Khalili, and D. Towsley, "A measurement-based study of multipath tcp performance over wireless networks," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 455–468.

[8] B. Arzani, A. Gurney, S. Cheng, R. Guerin, and B. T. Loo, "Impact of path characteristics and scheduling policies on mptcp performance," in *Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on*. IEEE, 2014, pp. 743–748.

[9] A. Nikravesh, Y. Guo, F. Qian, Z. M. Mao, and S. Sen, "An in-depth understanding of multipath tcp on mobile devices: Measurement and system design," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 189–201.

[10] J. G. Apostolopoulos and M. D. Trott, "Path diversity for enhanced media streaming," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 80–87, 2004.

[11] W. Jiang, R. Zhang-Shen, J. Rexford, and M. Chiang, "Cooperative content distribution and traffic engineering in an isp network," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 37, no. 1. ACM, 2009, pp. 239–250.

[12] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses," Internet Requests for Comments, RFC 6824, July 2013.

[13] B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H. V. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, and H. Zeng, "Engineering egress with edge fabric," in *Proceedings of the ACM SIGCOMM 2017 Conference (SIGCOMM'17). ACM, New York, NY, USA*, 2017.

[14] J. Zhu, J. He, H. Zhou, and B. Zhao, "Epcache: In-network video caching for lte core networks," in *Wireless Communications & Signal Processing (WCSP), 2013 International Conference on*. IEEE, 2013, pp. 1–6.

[15] B. Fortz and M. Thorup, "Internet traffic engineering by optimizing ospf weights," in *INFOCOM 2000. Nineteenth annual joint conference of the IEEE computer and communications societies. Proceedings. IEEE*, vol. 2. IEEE, 2000, pp. 519–528.

[16] "Global state of mobile networks (08/2016)," www.opensignal.com/reports/2016/08/global-state-of-the-mobile-network, accessed: 10/2017.

[17] J. Li, A. Aurelius, M. Du, H. Wang, A. Arvidsson, and M. Kihl, "Youtube traffic content analysis in the perspective of clip category and duration," in *IEEE NoF*, Oct 2013.

[18] A. Vakali and G. Pallis, "Content delivery networks: Status and trends," *IEEE Internet Computing*, vol. 7, no. 6, pp. 68–74, 2003.