Trade-Off between QoE and Operational Cost in Edge Resource Supported Video Streaming

Valentin Burger*, George Darzanos†, Ioanna Papafili†, Michael Seufert*

*University of Würzburg
Institute of Computer Science
Würzburg, Germany
{valentin.burger | seufert}@informatik.uni-wuerzburg.de

†Athens University of Economics and Business
Network Economics and Services Lab
Athens, Greece
{ntarzanos | iopapafi}@aueb.gr

***Abstract*—The largest share of today's consumer Internet traffic is video streaming and its demand on content delivery networks is continuously growing. To cope with the increasing demand of video streaming, recent work proposes mitigating end-user equipment to support content delivery at the edge of the network. The throughput of end-user equipment supporting content delivery is limited by the uplink of the users Internet connection. Especially for video streaming insufficient throughput causes the video to stall and affects the Quality of Experience (QoE) of end-users. To prevent video streams from stalling, we consider a tiered caching architecture, which requests higher tier caches to support content delivery, if the uplink throughput drops below a certain threshold. We conduct a simulative performance evaluation of the mechanism to investigate its impact on the QoE of end-users. Our results show that especially if the upload bandwidth of end-user equipment is low the setting of the threshold has a high impact. This can be used by operators to achieve the desired trade-off between QoE and operational cost for cache resources.**

I. Introduction

The Internet has seen a strong move to support overlay applications, which demand a coherent and integrated control in the underlying heterogeneous networks in a scalable, resilient, and energy-efficient manner. Cloud-based applications, i.e. applications which run completely or partly on the cloud or use cloud services, have set a strong need for network efficiency at all layers; therefore, cooperative cross-layer traffic optimization is a major goal to follow and can be achieved with tighter integration of network management and overlay service functionality.

Online Social Networks (OSNs) are cloud-based applications which are increasingly popular, as more and more people are using them to keep in touch with their acquaintances, to entertain themselves, and to share video and other content. Therefore, one can observe that nowadays OSNs have changed the way we use the Internet, since everyday huge volumes of content are being generated, stored initially, and shared at the edges of the network. Moreover, the data exchanged over OSNs represent a significant fraction of the Internet traffic globally. However, only a small percentage of the content shared in OSNs has high popularity, while most of it is created by users, hence is called User Generated Content (UGC) and has a long-tailed popularity distribution with fewer popular objects and a large number of low popularity ones.

Video sharing in OSNs is the main contributor of the traffic created by them. Videos in an OSN are distributed through friends, by users view and share actions. Currently, Facebook has become the second largest video viewing platform after Google websites. In order to assist video distribution most OSN providers either deploy their own CDN following a client/server architecture to distribute videos, or assign the delivery of videos to third-party CDNs. In both cases, such solutions are costly primarily in terms of bandwidth but also in terms of the storage needed to achieve high QoE for the users, due to the long-tailed content being delivered [1]. Moreover, traditional web-caching schemes, CDNs, and P2P assisted video sharing systems cannot deal efficiently with such content, because they dont take social relationships into account.

Consequently, we have identified in [2] the need for a scalable content distribution system for OSNs to efficiently deliver the content, both popular and UGC, achieving high QoE for end-users, while minimizing the operating costs of ISPs and OSN providers, respectively. In particular, we have proposed RBH in [3], [4], [5] and SEConD in [2], two traffic management mechanisms to deal with the delivery of videos over a social network as a case study. The proposed mechanisms are socially-aware as they exploit social relationships, interest similarities with respect to content and AS locality of exchange of OSN content. They address efficient delivery of both popular and long-tailed content in order to reduce the associated costs of ISPs. Hence, our mechanisms are ISP-friendly, while maintaining high QoE of OSN users and requiring only little effort and investment by the OSN provider.

As a variety of mechanisms had been proposed in literature, which are addressing the efficient delivery of OSN content by exploiting information such as: social relationships, interest similarities with respect to content, read patterns of OSN users, time-zone differences and locality of demand for OSN-published content, we developed an evaluation framework to simulate the generation of content in the environment of an OSN, in order to evaluate our mechanisms and compare them with other approaches in the literature.

For instance, initial evaluation results on RBH [4] show that leverage home routers for content delivery highly increases the total cache capacity of the autonomous system, which also saves inter-domain traffic. It further reduces the efficiency of the ISP cache, since the overlay of home routers serves as a filter for requests of popular content, which leaves only requests of rare objects to the ISP cache that are less likely to be hit.

Moreover, evaluation results reported in [2] showed that compared to SEConD, most of the approaches in literature are ineffective in terms of either inter-domain traffic costs, or Quality of Experience (QoE) for the end user, or scalability. The proposed mechanism improves users QoE and simultaneously, reduces traffic in potentially expensive inter-domain links, as well as the contribution of the origin content server. Thus, the SEConD mechanism can lead to benefits for all involved stakeholders, i.e., ISPs, OSNs, CDNs, and end-users.

In this paper, we extend the investigation on RBH and SEConD by further considering the implications of them on users' QoE and the cost of ISPs in terms of ISP cache contribution. To perform these evaluations, the evaluation framework designed and implemented for the purposes of [2] and [4] have been extended. Specifically, a new QoE model has been defined based on discrete-time Markov models for an analytic performance evaluation of video streaming over TCP [6], and a tiered caching architecture has been considered comprising the origin content server residing most probably in data centre(s) controlled by the Content Provider, CDN nodes, i.e., data centre nodes controlled by a CDN Provider which are used to replicate content items for redundancy and performance, and finally, end-host nodes, i.e. edge storage devices running appropriate software so as to participate in the efficient dissemination of the content items.

The remainder of this paper is organized as follows: in Section 2, we provide the necessary background for the investigation conducted in the context of this paper, and we provide an overview of related work in literature summarizing unaddressed issues or limitations of those. In Section 3, we briefly describe the main concept of the two traffic management mechanisms which are important for the analysis of QoE and ISP cache participation, and we describe the tiered content delivery architecture considered. In Section 4, we describe the simulation environment, providing details on the employed QoE model, and the simulation setup. Section 5 presents and discusses evaluation results obtained by simulations, while finally, Section 6 concludes the paper.

II. Background and Related Work

With HTTP video streaming, video data is downloaded to the client, stored in a buffer, and played out concurrently. Because the network conditions are varying, delays can be introduced at the application layer due to the TCP transmission. These delays, in particular, initial delay (i.e., the delay before the play out starts) and stalling (i.e., play-out interruptions because of delayed buffer fill), are the most important influence factors of users' Quality of Experience (QoE) [7], [8]. Therefore, the goal of video service providers must be to ensure smooth streaming of high quality video content to the clients. [9]

The increasing quality of multimedia content raised the quality expectations of users and the demand on content servers. Thus, Content Distribution Networks (CDNs) nowadays form the basic infrastructure and technology for the delivery of video content [10]. In a CDN, data centers are distributed around the globe, such that content is close to end users and can be accessed with low latency. CDNs are connected to important edge points of presence and have peering links to access networks with many end users, which saves transit costs for access network providers. To bring content even closer to users, ISPs can deploy CDN servers inside their own network to serve popular content, e.g., [11].

Additionally, network providers employ caching of popular content to save transit costs and to reduce latency. The performance caches has been extensively studied for different cache replacement strategies [12], [13], [14], [15] as well as static and dynamic request processes [16], [17]. A unified model to analyse the performance of caches for different request processes and replacement strategies is provided in [18]. The caching concept can be further extended to include also caches at the edge or in users' premises. The utilized caches include Nano Datacenters (NaDa) [19], [20], shared WiFi routers [21], [3], or end devices [22]. A two tiered video caching architecture consisting of home routers in access networks and data center caches in the core network is investigated in [23]. However, none of these works considers the upload bandwidth of caches, which is limited, especially if caches are deployed in end users' premises.

In this work, we will evaluate the QoE of video streaming in a three-tiered hierarchical cache architecture, which consists of the CDN data centers hosting the video content, ISP-owned caches, and home routers in the end users' premises. Thereby we consider the upload bandwidth of home routers and we will focus on the influence of tier-3 caches (home routers) on the inter-domain traffic, the ISP-cache contribution, and the resulting ratio of sessions having a good QoE.

III. QoE Aware Multi-tier Caching Concept

In this paper, we introduce a new multi-tier caching architecture that takes advantage of caching capability in different layers of the content delivery network. The main objectives of our approach is the improvement of users' QoE by avoiding stalling events, the reduction of content providers' operating expenses by mitigating its contribution in content delivery, and finally the reduction of transit interconnection costs of ISPs by decreasing the inter-domain traffic generated by content delivery. To this end, our mechanism performs hierarchical caching by taking advantage of equipment owned either by CDNs, ISPs or users and finally enables an ISP/user assisted content delivery. Our approach conducts an in depth evaluation of the impact of caching system that was presented in of the socially aware and ISP-friendly mechanism SEConD [2].

A. *Tiered Architecture*

In our system, we consider a hierarchical system of caches. In the tier-1 of the caching hierarchy are classified the CDN-owned caches, known as datacenters. In the tier-2 of the hierarchy the ISP-owned caches are placed. An ISP may deploy one or multiple caches in each AS and these ISP-caches can exchange content in local level or download content from a tier-1 cache. We assume that tier-2 caches that are located in different ASes cannot exchange content. Finally, in tier-3 of caching hierarchy belongs the user-owned Nano

Data centers (uNaDas). We assume that each user in our system maintains a uNaDa connected at his home network with the functionalities of home routers as described in [3]. The home routers (tier-3) and the ISP-caches (tier-2) located in the same AS form a P2P overlay that enables the content exchange among them. In Fig 1, we show the three tiers of caching hierarchy and their interconnection. This caching architecture and the communication between the different tier are extensively described in [2].

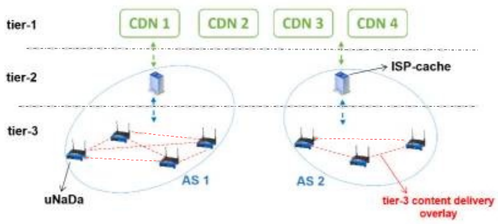

Fig. 1. Architecture of hierarchical caching and caches distribution.

B. QoE Aware Content Delivery

In this paragraph, we describe how the content flows through the hierarchy of caches, we define certain policies for the source selection for content downloading and we set a rule for the ISP-caches participation in the delivery process. Furthermore, we set threshold on the mean available download bandwidth to identify if a user can download a piece of content in acceptable QoE. Next, we present our mechanism in two different approaches for content delivery, the single-source and the multi-source content downloading.

Single-source. In this case, we assume that when a user requests a piece of content, its home router can only download it from a single source, i.e. only one cache regardless of the tier that this cache belongs to. Therefore, when a user requests a piece of content, its home router prioritizes the tier-3 caches in the source selection for fetching the requested content. If more than one tier-3 caches can provide the content in an acceptable QoE, then the home router chooses to download it from the source with the highest available bandwidth in order to ensure the best possible QoE. Otherwise, if there is not a tier-3 cache that can make the content available or even available in an acceptable QoE, then the home router requests it from a tier-2 cache within its AS. In the case where the content is not available, neither by local home routers that are in the same AS nor the ISP-cache, then the home router of the interested user fetches the content for the CDN caches.

Multi-source. In this case, we assume that the home routers and the ISP-cache within each AS form a P2P overlay that can perform a chunk-based Bit-torrent like content dissemination. Note that the role of ISP-cache in this case is diverse, since it operates as Tracker and participate as peer in swarms. Similar to the single-source case, the mechanism uses the QoE threshold in order to decide when and in which swarms the ISP-cache should participate as peer in order to guarantee an acceptable QoE for each requested piece of content.

Consequently, whenever a home router requests a piece of content, the ISP-cache checks if there is already a local P2P swarm for this content id. If there is, the ISP-cache adds the requesting home router in this swarm and stores the video in his cache (if not already there). Otherwise, the ISP-cache first creates a swarm for this video and then stores the video in his cache. Therefore, in this case the new swarm includes only the user requesting the video and the ISP-cache. The ISP-cache assists in content delivery until the total upload bandwidth offered by other peers of the swarm is adequate to achieve the desired QoE. The home routers that are interested in downloading a specific content are added in the swarm with the respective content id in order to simultaneously download and share the content (leechers). On the other hand, the home routers who have stored a content from previous requests, are added in the corresponding swarm in order to contribute in sharing and traffic localization (seeders). Since pre-fetching of content is very inefficient on caches with small capacity, it is not considered in this work.

C. Envisioned benefits

The potential deployment of our mechanism would improve users QoE and at the same time result in savings of inter-domain traffic and the associated transit costs, while it would incur the extra cost for deploying and running the ISP-owned cache. [24] have shown that the energy consumption and cost indeed are reduced. Due to the AS-aware source selection and the significant reduction of inter-AS traffic, we expect similarly important reduction of inter-connection charges. However, a cost benefit analysis is work in progress and concrete results, also on break-even points, will be provided. Finally, takes advantage of users' participation in content delivery. The exploitation of inherent P2P characteristics such as scalability and robustness by the presented mechanism is the reason for increased performance as the system (number of users/peers) scales, and for less demand on centralized servers.

IV. Simulation Model

This section presents the simulation model used to evaluate the multi-tier caching concepts. Therefore, the used QoE model and details of the simulation are presented.

A. QoE Model

The worst quality degradation of video streaming is stalling [7], i.e, the playback interruption because of insufficient downloaded video data. The authors found that users tolerate at most one stalling event of up to three seconds length for good QoE. In our work, a simplified QoE model is used inspired by the work in [6]. The authors used discrete-time Markov models for an analytic performance evaluation of video streaming over TCP. They found that a good streaming performance, which results in a low probability of stalling, can be achieved if the network throughput is roughly twice the video bit rate when allowing a few seconds of initial delay. [25] showed that the impact of initial delays on QoE is not severe, as users are already used to them and tolerate them. Therefore, our simplified QoE model only considers the received throughput of the video streaming connection:

$$QoE = \begin{cases} \text{good}, & \text{if throughput} \geq 2 \cdot \text{video bit rate.} \\ \text{bad}, & \text{otherwise.} \end{cases} \tag{1}$$

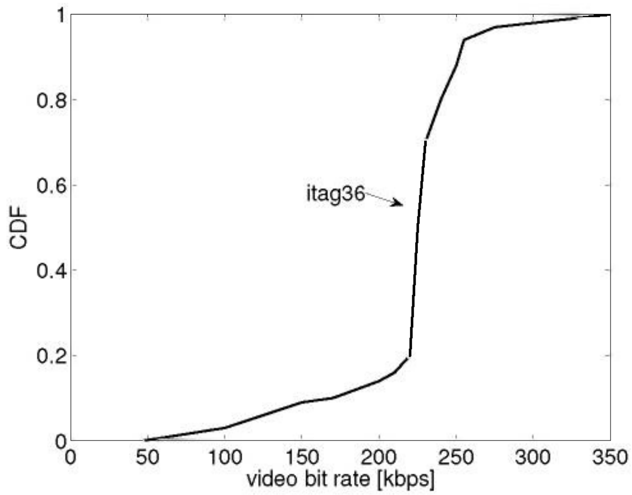

Fig. 2. Bit rate of YouTube videos in *itag*36 format [26].

To derive the bit rate of videos streamed by mobile devices we use the results from [26] where the video formats in mobile networks were characterized by analysing 2000 videos streamed from the video on demand platform YouTube. The authors find that the format *itag*36 is used in 80% of the streams. Figure 2 shows the cumulative distribution of video bit rates for mobile videos in *itag*36. The majority of the videos have a bitrate between 220 and 250 kbps.

B. Simulation Description

To investigate the trade-off between QoE and ISP cache contribution the paper assumes and evaluates a tiered caching architecture with resource locations at three different tiers, including the main data center of the content provider, CDN caches, and end-user equipment. The number of different content items to be downloaded or streamed from the resources is specified by the catalog size N. A Tier-1 resource is the data center of the content provider, where all N content items are stored. Tier-2 resources are edge caches and ISP caches, typically organized in a CDN, which are located close to Internet exchange points or within ISP networks. Requests served by ISPs or edge caches produce less or no inter-domain costs. Thus, these caches are referred to as ISP caches in the following. The capacity of ISP caches is specified by C_{ISP}. The caching strategy of ISP caches is LRU. Within tier-3 the caches are placed on shared HRs that run the RB-HORST mechanism. These caches are referred to in the following as home routers (HRs). The cache capacity of HRs is specified by C_{HR} and their caching strategy is LRU. In this study C_{HR} is set to 4 content items. The number of end-users in the autonomous system is given by n. The probability that an end-user shares its HR for content delivery is given by p_{share}. The probability that a user requests certain content items depends on the content popularity distribution, which is specified by the Zipf exponent α. The shared HRs form an overlay, that means a requested item is looked up in the HR of the user, if it is not found, it is looked up in shared HRs in the same autonomous system using the overlay. If no tier-3 cache in the AS contains the item it is looked up in tier-2 caches and finally in the data center of the content provider. The hierarchic caching strategy is leave-copy-everywhere, with the constraint, that the item is cached in the tier-3 cache only, which was looked up first. To limit the HR upload bandwidth we model the home router upload bandwidth with a normal distribution with mean μ and standard deviation σ. We implement the support threshold θ, which determines the ISP cache participation. If no home router caching the requested item has at least θ available bandwidth left, the ISP cache is requested. In case of ISP cache participation, the request is served by the ISP cache instead of the respective HR. To assess the QoE of end users in video sessions, we implement the QoE model for mobile video streaming described in Sec. IV-A. An overview of the parameters and their default values is given in tab:symbols.

TABLE I. PARAMETERS AND DEFAULT VALUES

Symbol	Description	Default value
α	Zipf exponent of content popularity	0.8
N	Catalogue size	$1e5$
C_{ISP}	ISP cache capacity	$2e4$
C_{HR}	Home router cache capacity	2
n	Number of end-users	$1e5$
p_{share}	Tier-3 cache sharing probability	0.01
μ	Average upload bandwidth of tier-3 caches	500kbps
σ	Standard deviation of upload bandwidth	100kbps

To estimate the amount of inter-domain traffic saved we consider the share of requests served in the different tiers, i.e., by home routers, ISP cache and content provider. We evaluate the load put on the ISP cache by determining the share of requests served by the ISP cache. To assess the QoE of end-users, we determine the fraction of video sessions receiving a good QoE.

V. Numeric Examples

As our focus is on the impact of the support threshold on inter-domain traffic, ISP cache contribution and QoE, we use a fixed sharing probability of $p_{\text{share}} = 10\%$ in the following. Exhaustive studies of the impact of the sharing probabilty on inter-domain traffic and the ISP cache contribution are provided in [4] and [5]. We further use a fixed standard deviation of tier-3 cache upload bandwidth $\sigma = 100$kbps and study only the mean upload bandwidth μ.

In order to investigate the impact of the home router upload bandwidth on the inter-domain traffic we study the share of request served from the different tiers. Figure 3 shows the share of requests served in the different tiers dependent on the mean home router upload bandwidth for support threshold θ=500kbps, home router cache cacacity $C_{HR} = 4$ and $alpha = 0.99$. The share of requests served by the content provider remains constant independent of the tier-3 cache upload bandwidth. If the home router upload bandwidth is zero, all requests that can be served locally are served by the ISP cache. With increasing home router upload bandwidth more requests can be served by tier-3 caches, hence the ISP cache contribution highly depends on the home router upload bandwidth.

We use the default parameters to investiagte the impact of the support threshold θ on the ISP cache contribution and the QoE of video sessions. The ISP cache contribution dependent on upload bandwidth of home routers is shown in figure 4.

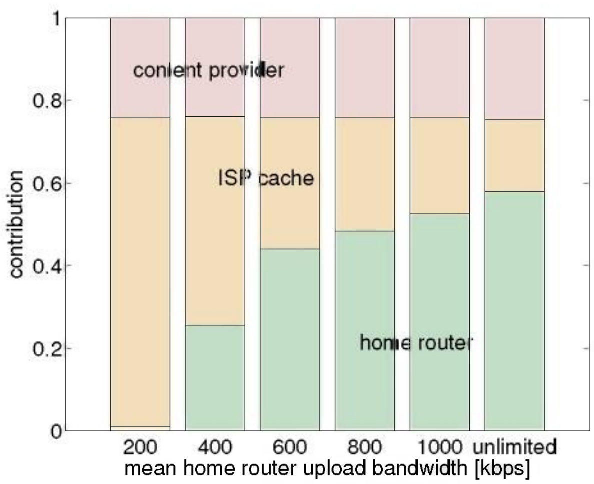

Fig. 3. Share of requests served from different tiers.

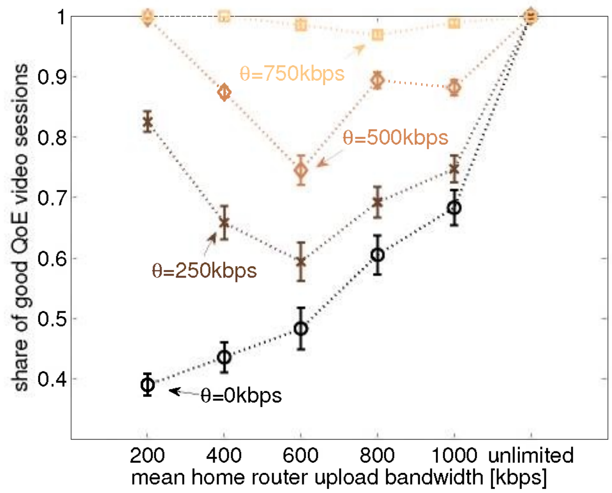

Fig. 5. Share of requests receiving a good QoE.

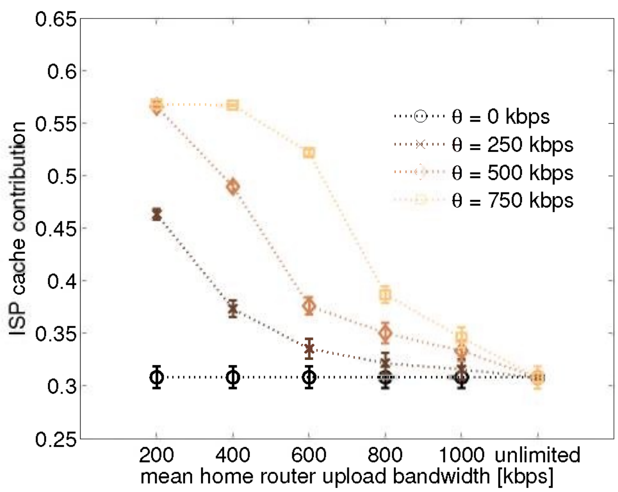

Fig. 4. ISP cache contribution dependent on home router upload bandwidth.

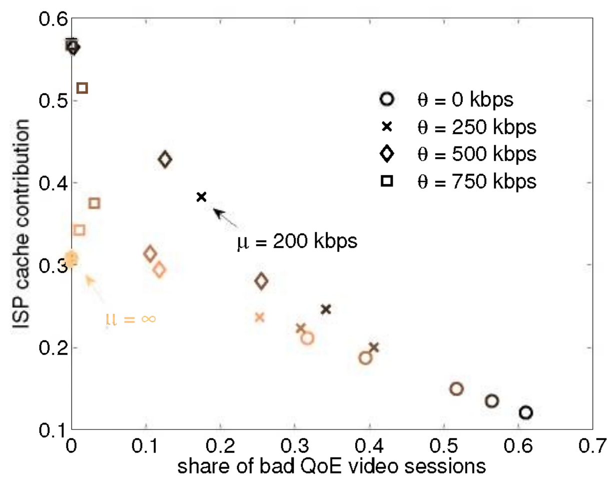

Fig. 6. Trade-off between ISP cache contribution and bad QoE video sessions.

If the support threshold is zero, the ISP cache contribution is minimal independent of the home router upload bandwidth. This depends on the fact that objects that are available on home routers are all served directly from their cache, independent of their upload bandwidth. For higher support thresholds the ISP cache contribution increases, since it is requested if the available bandwidth of home routers drops below the support threshold. As expected, the ISP cache contribution decreases with the home router upload bandwidth.

Figure 5 shows the fraction of requests receiving a good QoE dependent on the home router upload bandwidth for different support thresholds θ. The fraction increases with the support threshold. If the support threshold is 750kbps close to all requests receive a good QoE. Dependent on the support threshold there is a worst-case home router upload bandwidth in terms of QoE. For lower upload bandwidth requests are served by the ISP cache and receive enough bandwidth for seamless playback. For higher upload bandwidth the bandwidth of home routers is enough to ensure good QoE.

The trade-off between the ISP cache contribution and bad QoE video sessions is shown in figure 6 dependent on the mean upload bandwidth of home routers μ and on the support threshold θ. The ISP cache contribution is plotted against the share of bad QoE video sessions. Hence, optimal points that provide good QoE requiring low ISP cache contribution are in the bottom left. For support threshold $\theta = 0$kbps the ISP cache contribution is optimized, since the circles are always bottom most. For support threshold $\theta = 750$kbps the QoE is optimized, since the squares are left most. The mean home router upload bandwidth μ is coded in the color where it is unlimited for the yellow markers. An operator can set the support threshold θ to select the best trade-off between QoE and cache contribution dependend on its requirements and the available upload bandwidth μ. Especially if the upload bandwidth is low, c.f. $\mu = 200$kbps, the setting of the threshold has a high impact.

VI. CONCLUSION

To cope with increasing demand of video on demand services and the increasing number of mobile devices and requests, recent approaches suggest to leverage edge resources such as home routers to assist in content delivery and to get content even closer to users. The bandwidth of home routers is limited and their capability to serve as cache for video requests. Consequently, a low throughput of the video streams served by home routers can lead to stalling and bad QoE for end-users. To avoid quality degradations caused by low bandwidth recent approaches propose participation of ISP caches to support content delivery.

We develop a simulation framework to evaluate the impact of edge resource supported video streaming. We consider a tiered caching architecture and conduct performance studies on the upload bandwidth of home routers. Using a simple QoE model, we investigate the fraction of good QoE video sessions and the ISP cache contribution, to evaluate the impact of resource provisioning on QoE.

Our results show that, especially if the upload bandwidth of home routers is low, the QoE is improved by participation of the ISP cache. E.g., if the mean upload bandwidth is only 200 kbps, increasing the support threshold from 0 to 500 kbps results in almost twice as much video sessions receiving a good QoE. This comes to the cost of almost doubling the ISP cache contribution. Hence, setting the support threshold appropriately, operators may achieve the desired trade-off between QoE and operational costs for cache resources.

ACKNOWLEDGMENT

This work was partly funded in the framework of the EU ICT Project SmartenIT (FP7-2012-ICT-317846). The work of Valentin Burger and Michael Seufert was also funded in the framework of the EU ICT Project INPUT (H2020-2014-ICT-644672) and by Deutsche Forschungsgemeinschaft (DFG) under grants HO 4770/1-2 and TR257/31-2. The authors thank all project partners for their valuable contributions.

REFERENCES

[1] B. Ager, F. Schneider, J. Kim, and A. Feldmann, "Revisiting cacheability in times of user generated content," in *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*. IEEE, 2010, pp. 1–6.

[2] G. Darzanos, I. Papafili, and G. D. Stamoulis, "A socially-aware isp-friendly mechanism for efficient content delivery," in *Teletraffic Congress (ITC), 2014 26th International*. IEEE, 2014, pp. 1–9.

[3] M. Seufert, V. Burger, and T. Hoßfeld, "HORST - Home Router Sharing based on Trust," in *Proceedings of the Workshop on Social-aware Economic Traffic Management for Overlay and Cloud Applications (SETM 2013)*, Zurich, Switzerland, 2013.

[4] M. Seufert, V. Burger, F. Wamser, P. Tran-Gia, C. Moldovan, and T. Hoßfeld, "Utilizing Home Router Caches to Augment CDNs toward Information-Centric Networking," in *Proceedings of the European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.

[5] A. Lareida, G. Petropoulos, V. Burger, M. Seufert, S. Soursos, and B. Stiller, "Augmenting Home Routers for Socially-Aware Traffic Management," in *40th Annual IEEE Conference on Local Computer Networks (LCN)*, Clearwater Beach, FL, USA, Oct. 2015.

[6] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, "Multimedia Streaming via TCP: An Analytic Performance Study," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 4, no. 2, p. 16:116:22, 2008.

[7] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, and P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing," in *Proceedings of the IEEE International Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions (MQoE 2011)*, Dana Point, CA, USA, 2011.

[8] R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chan, "Inferring the QoE of HTTP Video Streaming from User-Viewing Activities," in *Proceedings of the ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, Toronto, ON, Canada, 2011.

[9] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[10] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience," in *Data Traffic Monitoring and Analysis: From Measurement, Classification and Anomaly Detection to Quality of Experience*, E. Biersack, C. Callegari, and M. Matijasevic, Eds. Springer's Computer Communications and Networks Series, Volume 7754, 2013, pp. 264–301.

[11] Google, "Peering & Content Delivery," https://peering.google.com/.

[12] H. Che, Y. Tung, and Z. Wang, "Hierarchical Web Caching Systems: Modeling, Design and Experimental Results," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 7, pp. 1305–1314, 2002.

[13] G. Haßlinger and O. Hohlfeld, "Efficiency of Caches for Content Distribution on the Internet," in *Proceedings of the 22nd International Teletraffic Congress (ITC)*, Amsterdam, The Netherlands, 2010.

[14] C. Fricker, P. Robert, and J. Roberts, "A Versatile and Accurate Approximation for LRU Cache Performance," in *Proceedings of the 24th International Teletraffic Congress (ITC 24)*, Krakow, Poland, 2012.

[15] G. Haßlinger and K. Ntougias, "Evaluation of Caching Strategies Based on Access Statistics of Past Requests," in *Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*. Springer, 2014, pp. 120–135.

[16] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, "Temporal locality in today's content caching: why it matters and how to model it," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 5, pp. 5–12, 2013.

[17] F. Olmos, B. Kauffmann, A. Simonian, and Y. Carlinet, "Catalog dynamics: Impact of content publishing and perishing on the performance of a lru cache," in *Teletraffic Congress (ITC), 2014 26th International*. IEEE, 2014, pp. 1–9.

[18] V. Martina, M. Garetto, and E. Leonardi, "A unified approach to the performance analysis of caching systems," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 2040–2048.

[19] N. Laoutaris, P. Rodriguez, and L. Massoulie, "ECHOS: Edge Capacity Hosting Overlays of Nano Data Centers," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 1, pp. 51–54, 2008.

[20] J. He, A. Chaintreau, and C. Diot, "A performance evaluation of scalable live video streaming with nano data centers," *Computer Networks*, 2009.

[21] A. J. Mashhadi and P. Hui, "Proactive Caching for Hybrid Urban Mobile Networks," Tech. Rep., 2010.

[22] E. Koukoumidis, D. Lymberopoulos, K. Strauss, J. Liu, and D. Burger, "Pocket cloudlets," *ACM SIGARCH Computer Architecture News*, 2011.

[23] C. Fricker, P. Robert, J. Roberts, and N. Sbihi, "Impact of Traffic Mix on Caching Performance in a Content-centric Network," in *Proceedings of the 2012 IEEE Computer Communications Workshop*, Sedona, AZ, USA, 2012.

[24] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the Internet with Nano Data Centers," in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies (Co-NEXT '09)*, Rome, Italy, 2009.

[25] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea," in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX 2012)*, Yarra Valley, Australia, 2012.

[26] J. J. Ramos-Muñoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J. M. López-Soler, "Characteristics of Mobile YouTube Traffic," *IEEE Wireless Communications*, vol. 21, no. 1, pp. 18–25, 2014.