PhD proposal

Title:
Software Defined Networking for Quality of Experience-based network optimization and management: application to 5G networks

Supervisors:
Damien Saucez, Chargé de recherche Inria, DIANA team
Lucile Sassatelli[1], Maitre de conferences UNS (I3S), SigNet team

Additional information: Besides Damien Saucez (DIANA, Inria) and Lucile Sassatelli (SigNet, I3S), Chadi Barakat (HDR, DIANA, Inria) and Guillaume Urvoy-Keller (HDR, SigNet, I3S) will take part into the supervision.

Employer:
It is planned to be CNRS, but it can be switched to Inria if further requirements must be met.

Context

Internet traffic is mostly composed of video streaming and multimedia contents in general. The amount of such traffic is growing at a blistering pace: according to [C1] video traffic will constitute 83% of consumer IP traffic in 2018 while it was 73% in 2013. Below we explain two main factors accounting for such a surge.

The new Ultra High Definition, a.k.a. 4K because it packs 4 times as many pixels as 1080p, is the new identified trend in video distribution at the 2014 CES [C2]. Main content provider like Netflix [C3] and Youtube [C4] plan to offer it, and the demand for streaming is expected to increase because such a resolution does not fit onto a DVD or Blu-ray anymore. The required bandwidth for a user to receive 4K videos is 15Mbps [C5], and even though most of the residential access in western countries do not yet receive such a speed, broadband access is expected to quickly grow [C5]. This 4K trend is hence foreseen to put a heavy strain on Internet Service Providers (ISPs). So far, the traditional approach to face traffic increase consists in overprovisioning networks. However, with such dramatic increase, simply upgrading link capacity is not economically viable.

This problem is exacerbated for mobile carriers. According to [C1], “mobile data traffic will grow 11-fold from 2013 to 2018, a compound annual growth rate of 61%.” Indeed, a large fraction of video traffic is and will be generated by mobile devices, such as tablets and smartphones. When a mobile client requests a video, the request is currently directed to a relatively close server that is part of a CDN and that is going to serve the client. The video travels from the server, located in some data-center (within the ISP or in a CDN), all the way down to the client through ISPs, the wireless carrier *Core Network* (CN) and its *Radio Access Network* (RAN). With the above figures (even without 4K) and the current architecture, the RAN is bound to suffer from heavy congestion, thereby degrading the mobile user experience. These shortcomings are getting aggravated with the advent of 5G technologies (LTE, LTE Advanced, small cells,etc.) that are going to greatly increase the last-hop data rates, enticing the users into requesting higher end-to-end rates. It is hence expected that the mobile operators will face a huge backhaul capacity crunch by 2017, unless they heavily invest into the RAN backhaul [L4].

[1] Lucile Sassatelli is currently preparing her HDR and intends to benefit from the free token offered by the doctoral school to be the official supervisor. Please note that Guillaume Urvoy-Keller or Chadi Barakat can alternatively be the official supervisor.

Motivation

On the client and server sides, to address the problem of video streaming towards users with heterogeneous accesses, *HTTP Adaptive Streaming* (HAS) is used to adapt the video encoding rate to the current network speed of the client. The ISO standard of HAS is MPEG-DASH. Multi-layer video descriptions, such as SVC (Scalable Video Coding), allow further adaptability of the video quality to the client conditions, by decomposing the video into multiple layers of incremental quality.

On the network side, the generalization of *Software Defined Networking* (SDN) [C6] and *Network Functions Virtualization* (NFV) [C7] permits to easily and efficiently control the network and its components, e.g., to process videos within the network to adapt them to user terminals, for example using transcoding [C8]. The principle of SDN is to completely decouple the control-plane (i.e., the routing decisions) from the data-plane (i.e., the forwarding actions). The control-plane is implemented by a centralized controller that pushes forwarding decisions on demand to the switches. This approach permits to individually control the exact path for each flow instead of controlling the path at the aggregated level by the means of link weights. NFV leverages virtualization capabilities of modern networks and SDN fine-grained control to move network appliances (e.g., routers, switches, NAT…) in the most suitable positions of the network in real time as a function of the network load and usage.

In addition to that, the storage cost is diminishing so much that nowadays storage units are proposed to be placed in network appliances directly, to push storage capacity as close as possible to the end-users and free bandwidth in the backhaul and in the core network [D3]. These caches store contents that are likely to be consumed by the users connected close to them, whereby reducing the overload of the network.

We believe that the technical solutions proposed so far to transport high quality video streams in networks are only a first step towards high quality video services but that they forget an essential component of the network: the user. More precisely, the *Quality of Experience* (QoE) of users directly dictates their willingness to subscribe to a service or to switch to another provider [D1]. The QoE is a subjective measure of the quality of the service provided by the network. Regrettably, QoE is usually not taken into account while designing and provisioning networks or while taking routing decisions. We claim that taking QoE into account allows improving network efficiency. Indeed, network performance metrics (e.g., delay, bandwidth, packet loss) influence differently the different types of traffic (e.g., video, audio, realtime, games, bulk). Knowing that, it would be relevant to adopt different routing, forwarding, and in-network processing policies based on the specific QoE requirements of each traffic type. To do so, being able to know the impact of a particular network condition on the perception of users would for example allow the operator to deviate delay insensitive traffic to slow paths to reduce the load on links that would be profitable for other types of traffic. This information can also be used to re-allocate virtual network appliance (e.g., caches, transcoders, etc.) on-the-fly to ensure that resources are always used in the most efficient way for both operators and users.

A whole realm of network optimization and management has addressed in the last decades the so-called *Quality-of-Service* (QoS) constraints. QoS denotes a set of metrics directly extracted from the network, such as data rate (bandwidth available in the network, or actual rate experienced by the end application), packet delay, packet loss rate or jitter. Whilst the QoS metrics were sufficient to quantify the user QoE with legacy applications such as file transfer or database interrogation, the advent of sophisticated multimedia applications, such as those mentioned above, render QoS metrics inadequate. For example, it has been shown that video streaming experience is dictated by a number of quantities amongst which the buffering delay the user has to wait before the video starts or resumes to play after a stall event, the number and frequency of stalls, and some non-

linear function of the video coding rate [L1]. These quantities cannot be measured directly in the network anymore, but instead depend on a combination of factors amongst which network parameters, cache locations, client and server behaviors, and human perception [L2, L3].
Driving the network optimization with these objectives is in sharp contrast to legacy methods as basic network functions such as routing, more sophisticated ones such as caching or function chaining, up to client and server software behaviors, but also subjective perceptions are now involved. Such strains call for a network paradigm shift where all the entities can be easily managed jointly, from the network layer up to the application layer, and this is exactly what SDN allows.

Objectives

In this PhD, we will design and evaluate network solutions leveraging SDN to enable QoE-based caching, routing and forwarding decision in 5G networks. To achieve such a goal, we will answer the three fundamental questions below:

1. *How to link subjective Quality of Experience with network metrics directly measurable by the network?*

- Defining the metrics that matter to the user is the first step towards assuring QoE. Some of them, like the buffering delay or the frequency of stall events are mentioned above. Another set of metrics relate to the cache impact. Adding caches close to the user has proven interesting in terms of bandwidth savings or delay and throughput improvements, but it has been underlined recently that they may also be detrimental to the QoE [L3], in the sense that a user may perceive oscillations between a “good” and a “very good” state more negatively than no oscillations in a “good“ state. In [L3] “good” is quantified by a regular network metric, e.g., cache hit ratio. QoE metrics are still quite limited, especially when considering the possibilities ahead for further optimization of content distribution, and require to be more comprehensively investigated.
- We will build on our previous works on QoE modelling [D2] to carefully determine the relevant QoE metrics of interest to us (frequency of stall events, video resolution or cache hit ratio, and functions of their variability).
- We will express these QoE metrics thanks to the combination of low-level measurable (QoS) metrics, so as to relate QoE objectives to the system parameters to optimize (network routing, cache/functions placement).

2. *How to compute and prove QoE-optimal network decisions (e.g., routing, caching, virtualization) taking into account operational requirements?*

So far we have described our concern regarding QoE optimization. However, when operating a network, other costs come at play and the user satisfaction cannot be the sole objective: fairness amongst user (where fairness can be enforced in terms of QoE instead of QoS), resource utilization, maintenance costs or energy expenditure are also critical to operators nowadays. The latter objectives are conflicting with the user QoE metrics, and hence the whole multi-objective optimization must be addressed. In particular in this thesis, we will focus on two main factors that influence QoE and operator costs: cache placement and routing.

- Cache placement is challenging as, to be efficient in offloading core networks while being used by a sufficient number of users to justify its maintenance cost, a content cache must be neither too far away from nor too close to the end users. A number of recent works have looked at better content placement to bear the soaring of video streaming [L5,L6]. However none of them take into account the maintenance costs incurred to the telco by such solutions. We want to address the problem more broadly by considering both maintenance costs and more relevant QoE metrics.

- Routing is the cornerstone of network operation, and performing it by taking into account that most of the carried data is streaming flows is paramount. Regular traffic engineering aims at spreading efficiently the traffic over the possible paths, possibly considering QoS constraints of the flows, not congesting the shortest paths while letting others underutilized. Again, more broadly than what is being done in [L7,L8], we want to address the problem of efficient routing in mobile operator networks encompassing the above mentioned QoE and maintenance objectives. In particular, we want to allow for energy savings in the network by setting caps on the number of routes, or router interfaces, taken by the traffic.

3. How to implement and deploy a QoE-based routing solution in 5G networks?

As soon as each of the multi-objective optimization problems described above is expressed and a solution is devised (either exact or approximated depending on the form of the problem), we intend to implement the solution using SDN. SDN allows for the controller to be aware of all the traffic parameters and network state at each point of time and space. In particular, in the case of video streaming, this allows for this cross-layer joint optimization where the controller can both send rules on how to route each flow to the switches and communicate with the client application to inform the adaptive HTTP streaming algorithm. We state that the assets of SDN can be even more crucial when optimizing wireless carriers CN and RAN, where the current control and data planes are heterogeneous. One particular challenge in mobile network and especially in 5G is the large diversity (e.g., dense cities vs remote countryside spots) and dynamics (e.g., rush hours, punctual events like football matches, new year greetings) of the network components and usage that imposes the solution to be highly reactive, versatile, and particularly efficient.

Complementarity of the partners

By combining the key aspects of todays networks usage, ranging from the primary concern of user QoE in heterogeneous services and access contexts, to the efficiency of the networks serving the user traffic, this PhD proposals highlights the complementarity of the DIANA (at Inria) and SigNet (at I3S) groups.
The DIANA group has expertise in QoS and QoE measurements. For instance, an ongoing LABEX PhD thesis focuses on network measurements to assess the net neutrality and a Ubinet master internship concentrates on the development of a measurement framework to instrument the YouTube video streaming player. In order to enforce user QoE (whose an important component is users' rights and privacy respect), DIANA is developing a notable experience in SDN to allow efficient deployment of in-network functions (such as caching, video transcoders, load balancers, etc.) by the means of collaborative projects focusing on SDN in challenging networks.
The SigNet group has been focused both on wireless networks (going from WiFi to cellular, studying content dissemination and scheduling), and wired networks with an emphasis on data center networking (with the impact of virtualization) and energy-aware routing, for which SDN development is also being done. Lucile Sassatelli and Guillaume Urvoy-Keller are currently supervising a PhD student within a CIFRE contract with Orange Labs, focusing on QoE analytical models for video streaming in mobile networks, and are going to supervise a master intern working on next-generation adaptive streaming techniques, in order to prepare this PhD thesis.
Owing to this complementarity in the network functions, components and tools we address in each research group, we strongly believe that our joint work in the framework of a PhD thesis will be very fruitful.

Moreover, this thesis will foster a new collaboration between two members of Diana and SigNet, namely Damien Saucez and Lucile Sassatelli who are at the initiative of this project. They do not hold HDR yet. This thesis specifically supports their research agenda with the starting streaming activity led by Lucile Sassatelli within the SigNet group and the operationally driven SDN/NFV research ecosystem that Damien Saucez, who joined the Diana group as a CR2 last October, is developing.

Relevance to the LABEX call

Our PhD proposal is related to the "Réseaux orientés contenu" and "Infrastructures: hétérogénéité et efficacité" topics covered by the LABEX.

It lies in "Réseaux orientés contenu" because we will combine the expertise in networking architecture and QoE measurements of the SigNet and DIANA research groups to offer network optimization focusing on users expectations while consuming their contents (e.g., video streams) instead of the usual approach in networking that ignores the user to solely concentrate on technical considerations without totally understanding the impact on the users. In other words, this activity aims at providing contents to users in a better way than today by taking into consideration their consumption expectations.

This project also addresses the efficient management of heterogeneous infrastructures as (i) we will consider various user network accesses, amongst which the challenging mobile access, and (ii) we want to use software-based tools (such as SDN or NVF), oblivious of the physical network appliances or the specific architecture choices of the various providers, to efficiently (for various metrics such as fairness or energy, as aforementioned) manage the provider networks.

References

[C1] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018. Cisco White Paper.
[C2] http://www.wowza.com/blog/five-trends-from-ces-2014-that-will-impact-streaming-media
[C3] http://www.theverge.com/2014/1/10/5291068/how-netflix-won-ces-4k-streaming
[C4] http://www.pcmag.com/article2/0,2817,2428947,00.asp
[C5] http://www.reelseo.com/4k/
[C6] ETSI. “Network Functions Virtualisation” ETSI White Paper. Available at: https://portal.etsi.org/NFV/NFV_White_Paper.pdf
[C7] ONF. "Software-Defined Networking: The New Norm for Networks". White paper. Open Networking Foundation, 2012.
[C8] A. Vetro, C. Christopoulos, H. Sun, "Video transcoding architectures and techniques: an overview," Signal Processing Magazine, IEEE , vol.20, no.2, pp.18,29, Mar 2003
[L1] G. Cermak, M. Pinson, and S. Wolf. The Relationship Among Video Quality, Screen Resolution, and Bit Rate. IEEE Trans. Broadcast., 57:258 –262, 2011.
[L2] A. Mansy, M. Ammar, J. Chandrashekar, and A. Sheth, “Characterizing Client Behavior of Commercial Mobile Video Streaming Services”, in Proceedings of ACM Workshop on Mobile Video Delivery (MoViD'14), 2014..
[L3] D. H. Lee, C. Dovrolis, and A. C. Begen, “Caching in HTTP Adaptive Streaming: Friend or Foe?”, in Proceedings of ACM Network and Operating System Support on Digital Audio and Video Workshop (NOSSDAV '14), 2014.
[L4] Juniper Research, Basingstoke, U.K., “Press release: Mobile network upgrades of up to $840bn required over next five years to meet burgeoning data demand, finds Juniper Research,” Available online: http://www.juniperresearch.com/viewpressrelease.php?id=336&pr=259
[L5] H. Ahlehagh, and S. Dey, "Video-Aware Scheduling and Caching in the Radio Access Network," Networking, IEEE/ACM Transactions on , vol.22, no.5, pp.1444,1462, Oct. 2014
[L6] P. Georgopoulos, M. Broadbent, B. Plattner, and N. Race, "Cache as a service: Leveraging SDN to efficiently and transparently support video-on-demand on the last mile," Computer Communication and Networks (ICCCN), 2014 23rd International Conference on , vol., no., pp.1,9, 4-7 Aug. 2014
[L7] G. Panagiotis, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, “Towards network-wide QoE fairness using openflow-assisted adaptive video streaming”, in Proceedings of the ACM SIGCOMM workshop on Future human-centric multimedia networking (FhMN '13), 2013.
[L8] S. Laga, T. Van Cleemput, F. Van Raemdonck, F. Vanhoutte, N. Bouten, M. Claeys, and F. De Turck, "Optimizing scalable video delivery through OpenFlow layer-based routing," Network Operations and Management Symposium (NOMS), 2014 IEEE , vol., no., pp.1,4, 5-9 May 2014
[D1] Nokia white paper, “Quality of Experience (QoE) of mobile services: Can it be measured and improved?”, 2004.
[D2] S. Afra, D. Saucez, and C. Barakat, “From network-level measurements to expected Quality of Experience: the Skype use case”, Inria technical report, https://hal.Inria.fr/hal-01071373, 2014.
[D3] B. Ahlgren, C. Dannewitz , C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking”, Communications Magazine, IEEE, 50(7):26 –36, July 2012.