Multimedia Content Delivery in SDN & NFV based Towards-5G Networks

Fidel Liberal[1], Akis Kourtis[1,2], Jose Oscar Fajardo[1], Harilaos Koumaras[2],

[1]University of the Basque Country (UPV/EHU), [2]NCSRD Demokritos

Email:{fidel.liberal, joseoscar.fajardo}@ehu.eus, {akis.kourtis, koumaras}@iit.demokritos.gr

1. Introduction

Cellular technologies have quickly evolved from narrowband voice only specific purpose networks to becoming the predominant candidate for accessing multimedia services. Most current analysis [1][2] foresees an imminent explosion of Internet connections over broadband wireless networks, making seamless high speed connectivity finally a reality. Together with raising paradigms that will shape next generation cellular technologies, an equivalent evolution of content delivery networks (CDNs) is taking place to cope with the new challenges associated to myriads of mobile users accessing high quality services.

Among these paradigms, *cloudification* through software defined networking (SDN) and network function virtualization (NFV) entails several opportunities for true convergence between CDNs and 5G networks.

Efficient use of scarce network resources will no doubt become the main motto for these new scenarios. Foreseen solutions will strongly depend on the roles of the significantly different players involved, ranging from technology agnostic e2e service adaptation mechanisms traditionally found in Internet based CDNs, network assisted broadcast-like technologies of mobile telcos and content aware resource scheduling by broadband access equipment manufacturers.

The strategy towards the evolution of cellular networks from 4G to 5G can be summarized by the following trends [3]: 1) evolution of Ratio-Access Technologies (RATs), 2) cell densification, 3) composition of radio access technologies -- particularly 3G/4G and Wi-Fi offloading, 4) heterogeneous networks (HetNets), 5) flexible spectrum management, 6) cloudification, and 7) new scenarios -- device-to-device (D2D), machine-to-machine (M2M), and Internet of Things (IoT).
At first sight trends 1)-5) may appear independent of the service delivered but optimized allocation of radio resources demand a clear input of the content delivery needs in order to carry out network orchestration.

Additionally, in the meantime, hybrid approaches have been already proposed to pave the way for such transitions from 4G specific hardware based architectures to software based 5G platforms.

In the following sections we will focus on how *cloudification* of 5G through different SDN/NFV paradigms may affect content delivery or, on the other hand, how different content delivery mechanisms can be fitted into foreseen 5G network architectures with a particular focus on dynamic adaptive streaming over HTTP (DASH).

Towards 5G Evolution and SDN/NFV

Figure 1 depicts foreseen evolution for cellular networks from LTE (4G) towards 5G.

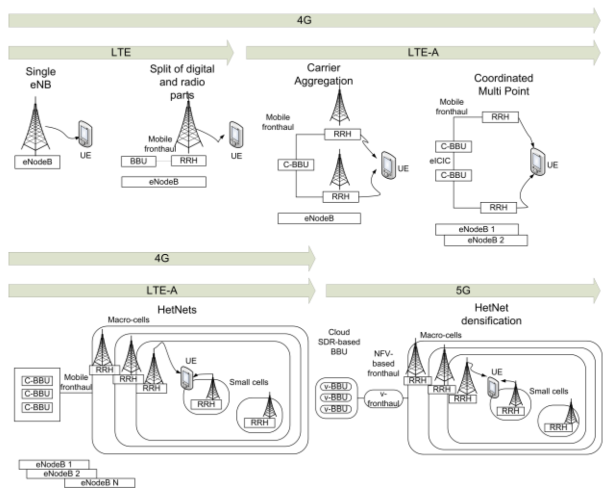

Figure 1. 4G to 5G Evolution

4G LTE technology was designed to cope with the requirements of a flat all-IP architecture. The eNodeB was introduced as the key element within the Radio Access Network (RAN), in charge of managing the allocation of radio resources and of implementing the Adaptive Modulation and Coding (AMC).

As a step forward to foster the deployment of LTE networks, the eNodeB was functionally split into the BaseBand Unit (BBU), capable of performing the digital functions, and the Remote Radio Head (RRH), which actually needs to be located at specific placements in order to perform radio transmissions.

The advent of LTE-Advanced brings new radio capabilities that leverage the achievable performance even in cell edges. First, Carrier Aggregation allows increasing the channel bandwidth by implementing a coordinated scheduling of radio resources. In this case, a Centralized BBU (C-BBU) is able to control several RRHs without further radio synchronization restrictions. Besides, Coordinated MultiPoint (CoMP) addresses the possibility to synchronize the transmissions to/from different eNodeBs.

The logical evolution of this network architecture towards HetNets resides on the co-sited deployment of a number of C-BBUs, each of them controlling a set of RRHs through high speed fronthaul connections and implementing coordinated control with almost negligible delays in the inter-eNodeB exchange of information.

Finally, the forthcoming 5G network is envisaged to be ruled by flexible NFV management, which allows the virtualization of all the RAN elements into virtual appliances. This way, programmable network control will enable a more user- and content-focused allocation of resources. Additionally, some initiatives such as the Mobile Edge Computing [4] are pursuing the flexible usage of general purpose HW/SW platforms at the Cloud RAN (C-RAN). Thus, some lightweight service instances would be able to run at the C-RAN, leveraging user-tailored proximity services.

Caching and CDNs

Various studies have observed that a large portion of the mobile multimedia traffic is generated by duplicate downloads of popular multimedia content [5]. Therefore, research activities in the CDN field have been focused in finding novel caching algorithms operating inside networks, which deliver high-demand content to customers via nearby caching gateways. This concept reduces inter-Internet Service Providers (ISPs) traffic. Additionally edge caching both minimizes content fetching time and, when combined with multicast/broadcast approaches (i.e. eMBMS), it can dramatically reduce radio resource consumption particularly relevant in towards-5G architectures.

The main goal of caching techniques is to attain the optimal trade-off between traffic bandwidth cost and storage cost. Traffic bandwidth cost remains quite expensive for ISPs, whereas storage cost constantly drops. However, available multimedia content is growing rapidly making it impossible to store a great amount of it, thus content demand has to be considered when caching.

Most caching alternatives can be classified as either web caching, which is based on uniform resource locator (URL), or packet redundancy caching, which is protocol-independent and packet oriented. Web caching creates a correlation between the requested content and its URL. The caching server is responsible to deliver the content to the user, when the user requests it from the corresponding URL. Packet redundancy caching offers more complex correlation between the content and the client request. The caching server in this case can store packets, or even chunks of content, which facilitates protocols like Dynamic Streaming over HTTP (DASH). In DASH, video content is encoded in various quality levels, and each quality sequence is divided into small “chunks”. When the client requests a video, depending on the network conditions, it receives the corresponding quality chunk. This enables the service to switch between different quality levels seamlessly, as the content is delivered through a CDN service.

In order to further improve and evolve current CDN technologies, next steps are heading towards the virtualization plane. The current proliferation of 4G networks and future 5G networks creates ample space for CDN technologies to be further extended and exploited regarding virtualized environments. On one hand, SDN is mostly focused on network controllability, which can be applied on a CDN-driven content delivery system, in order to facilitate and enhance network management and content delivery. On the other hand, NFV is related to the network data plane. Virtualized Network Functions (VNFs) which inspect network traffic, e.g. virtualized Deep Packet Inspection, can be used to add content-aware forwarding and caching in CDN environments.

In order to efficiently distribute the highest possible quality content in variable channels in towards-5G networks, SDN and NFV can provide scalable and fine-tuned adaptation loops between network and CDNs. Following Sections will analyse different alternatives and highlight challenges and opportunities.

2. Adaptation Strategies (DASH case study)

Figure 2 illustrates the evolution of adaptation strategies focusing in DASH-based media delivery. Different enhancements have been proposed over recent years to optimize the transmission of multimedia content over the Internet, with special interest in mobile environments.

Figure 2a illustrates the traditional DASH adaptation approach where the adaptation algorithm is deployed at the client device, in order to modify the source of the media content. Figure 2b represents the next evolution, based on the introduction of media-aware network nodes in the provisioning chain. Network-assisted adaptation mechanisms cooperate with the traditional client-driven approach to optimize the media delivery in multi-user scenarios. Such elements have been traditionally focused on network awareness but the interaction with the resource allocation and scheduling mechanisms of underlying radio networks was not typically allowed.

Finally, current 5G trends suggest the standardization and deployment of intelligent network nodes that will enable more powerful adaptation and prioritization frameworks over the whole transmission chain, and especially, at the edge of the mobile access network [6].

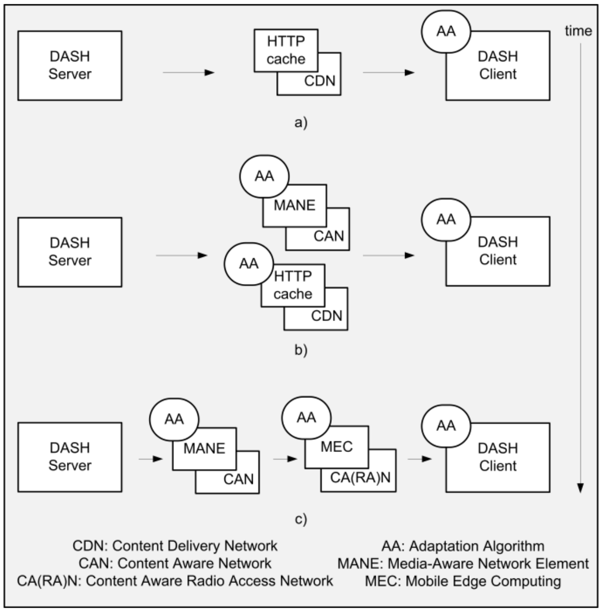

Figure 2. Evolution of mobile Multimedia adaptation strategies a) client-driven a; b) core network-assisted; c) core and access network-assisted.

In this scenario, SDN paradigm may be used to dynamically allocate network resources to different DASH clients [7, 8]. An SDN manager would then be used to dynamically modify video flows and/or network resources in order to achieve a QoE-based fairness between different users.

Figure 2c illustrates such approach aimed at including service adaptation capabilities into the 5G cloud-enabled RAN to integrate user awareness and enhance adaptation responsiveness. This way, the flexible architecture of the Cloud-RAN [9] will enable the convergence of SDN-based mobile networks and MEC-assisted close-to-the-user service instances.
In order to do so, channel and content aware adaptation algorithms would be in charge of making QoE-driven scheduling decisions. This way, the MEC will exploit the benefits of the multi-layer media delivery to incorporate network-assisted adaptations to the client-driven approach. In order to perform the channel aware estimations, the adaptation algorithm needs to be fed with low-level radio channel information (i.e. cell statistics and individual channel quality information). However, in multi-user scenarios, too fine-grain feedback granularity would entail a high load and traffic volume in the interface between MEC server and RAN monitoring elements.

3. Research challenges and opportunities

Cloudification trends in most networking areas and particularly in 5G and CDNs depict an exciting playground for improved multimedia services delivery in the near future. However, a number of challenges and research opportunities must be faced to ensure:

- **Network supported vs. implicit channel awareness**: Although MEC initiative defines specific mechanisms for retrieving channel information from the RAN, such deployment demands collaboration between radically different stakeholders, e.g., CDNs and mobile network operators (MNOs), which might jeopardize resulting architecture. As an alternative, over-the-top (OTT) CDNs and user equipment (UE) manufacturers could incorporate feedback mechanisms in multimedia players and devices and use crowdsourcing/Big Data techniques to infer implicitly channel and cell information to feed their adaptation algorithms and overcome MNO’s resistance.
- **Analysis of the trade-off between improvements vs. quality feedback granularity**. Most research studies confirm that incorporating channel awareness in both radio resource scheduling and multimedia content delivery optimization loops result in enhancements in QoE and resource usage. A proper analysis of the complexity and overload of quality feedback mechanisms versus achieved improvements is of outermost importance in order to guarantee a business case for actual implementations.
- **Relevance of channel modelling in 4G+/5G environments**. Per user channel behaviour modelling in shared wireless channels remains a huge research challenge. A better understanding of 4G+/5G and realistic stochastic models will no doubt help refining channel feedback reporting rates and delay constraints as to relax initial requirements.
- **Impact of new codecs and transport mechanisms**. Recent multilayer video coding mechanisms (e.g., H265/HEVC) and transport mechanisms (e.g., QUIC and HTTP2.0) will affect existing DASH adaptation mechanisms and overall performance. Cross-layer effects with LTE-A and 5G networks need to be also carefully analysed as to better enhance the scheduling/service level adaptation proposals.
- **Role of eMBMS and other caching and broadcasting/multicasting schemes**. Caching mechanisms have been already incorporated into standardised broadcasting mechanism and integrated in 4G network nodes. However, their chances to become an actual alternative to “traditional” TV broadcasting mechanisms (i.e. DVB-T*) is still unclear.
- **Evolution of the ecosystem**. Major OTT Internet players, Cloud providers, Telco manufacturers, MNOs and broadcasters depict a tumultuous but challenging environment. The resulting ecosystem will no doubt determine the real applicability of joint CDN/5G integration initiatives.

Acknowledgments

This work has been partially funded by the Spanish Ministerio de Economia y Competitividad (MINECO) under grant TEC2013-46766-R: QoEverage - QoE-aware optimization mechanisms for next generation networks and services.

References

[1] The Global mobile Suppliers Association (GSA), "Evolution to LTE report", April 9, 2015. http://www.gsacom.com/downloads/pdf/Evolution_to_LTE_report_090415.php4

[2] Ericsson Mobility Report, June 2015. http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf

[3] Demestichas, P.; Georgakopoulos, A.; Karvounas, D.; Tsagkaris, K.; Stavroulaki, V.; Jianmin Lu; Chunshan Xiong; Jing Yao, "5G on the Horizon: Key Challenges for the Radio-Access Network", in *Vehicular Technology Magazine*, IEEE vol.8, no.3, pp.47,53, 2013.

[4] M. Patel et al., "Mobile-Edge Computing Introductory Technical White Paper," Sept. 2014.

[5] Xiaofei Wang; Min Chen; Taleb, T.; Ksentini, A.; Leung, V., "Cache in the air: exploiting content caching and delivery techniques for 5G systems", in *IEEE Communications Magazine*, vol.52, no.2, pp. 131-139, 2014.

[6] Soldani, D.; Manzalini, A., "Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society", in *Vehicular Technology Magazine, IEEE* , vol.10, no.1, pp. 32-42, 2015.

[7] P. Georgopoulos, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, "Towards Network-wide QoE Fairness Using OpenFlow-assisted Adaptive Video Streaming," in *Proceedings of the 2013 ACM SIGCOMM FhMN workshop*, pp. 15-20, 2013.

[8] C.-F. Lai, R.-H. Hwang, H.-C. Chao, M. M. Hassan, and A. Alamri, "A Buffer-Aware HTTP Live Streaming Approach for SDN-Enabled 5G Wireless Networks," in *IEEE Network*, vol.29, no.1, pp. 49-55, 2015.

[9] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud Radio Access Network (C-RAN): A Primer", *IEEE Network*, vol.29, no.1, pp. 35-41, 2015.

Fidel Liberal received his PhD from University of the Basque Country (UPV/EHU) in 2005 for his work in the area of holistic management of quality (both QoS and QoE) in telecommunications services. His research interests include quality management and multicriteria optimization for multimedia services in NGNs and 5G. He has co-authored more than 40 conference and journal papers.

Kourtis Michail-Alexandros, MSc was born in Athens, Greece, in 1989. He received the Diploma in Computer Science, and the Master's Degree in Computer Science from Athens University of Economics and Business, in 2011 and 2014, respectively. From January 2010 he is working at NCSR "Demokritos", in various research projects. Currently he is pursuing his PhD at the University of the Basque Country (UPV/EHU), Bilbao, Spain. His research interests include Video and Image Processing and Quality Assessment, Network Function Virtualization, Software Defined Networks and Quality of Service.

Jose Oscar Fajardo works as research fellow in the the University of the Basque Country (UPV/EHU), at the Faculty of Engineering in Bilbao. He received his M.Sc degree in 2003 and he is now Ph.D. candidate. He has lately worked in the FP7 ICT ADAMANTIUM (2007-2010) and FP7 SEC GERYON (2011-2014), where he has been involved in researching the adaptive management of mobile multimedia services under the framework of IMS. He has co-authored more than 25 journal and conference papers since 2005, mainly in the areas of QoS/PQoS/QoE and service performance assessment, and QoS-aware networking.

Dr. Harilaos Koumaras was born in Athens, Greece in 1980. He received his BSc degree in Physics in 2002 from the University of Athens, Physics Department, his MSc in Electronic Automation and Information Systems in 2004, being scholar of the non-profit organization Alexander S Onassis, from the University of Athens, Computer Science Department and his PhD in 2007 at Computer Science from the University of Athens, Computer Science Department. Dr. Koumaras has been an active research associate since 2003 of Media Net Lab at the National Centre of Scientific Research "Demokritos".