This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2017.2760826, IEEE Communications Letters

On the Need of Joint Bandwidth and NFV Resource Orchestration: a Realistic 5G Access Network Use Case

Jose-Juan Pedreno-Manresa, Pouria Sayyad Khodashenas, Muhammad Shuaib Siddiqui, Pablo Pavon-Marino

***Abstract*—5G envisages a "hyper-connected society" where an enormous number of devices are inter-connected anywhere and at any time. Cloud-enabled radio access networks (RAN) where intelligence is placed in conjunction with the radio heads at the proximity of end users is a promising solution to fulfil the 5G expectations of sub-millisecond latency, huge traffic volumes and higher data rates. Network Functions Virtualization (NFV) and Software Defined Networking (SDN) developments enable end users to access advanced features such as configurability, automation, scalability, improved resource utilization and multi tenancy over the cloud-enabled RANs. Management and orchestration techniques are the ultimate factor that determine the effectiveness of the novel SDN/NFV features being introduced. Our focus in this study is the resource allocation in a realistic cloud-enabled RAN, taking into account the dynamics of ~100,000 persons movement in a crowded event, i.e. a football match. The proposed solution jointly orchestrates NFV and bandwidth resources, as one resource affects the other. Simulation results clearly verify the benefits of the proposed solution over traditional disjoint schemes.**

***Index Terms*—5G, access networks, Net2plan, network orchestration, NFV, SDN, service chaining**

I. Introduction

5G promises a ubiquitous solution featuring aspects like extraordinarily high speeds and capacity, multi-tenancy, fixed and wireless access network convergence, unconventional resource virtualization, on-demand service-oriented resource allocation and automated management [1]. It calls for a fundamental change on the telecommunication infrastructures, shifting them from only data transport media to intelligence entities equipped with IT assets, i.e. compute and storage resources.

European 5G PPP [2] has set a direction, aiming to significantly improve the technology level towards higher throughputs and flexibilities. Efforts done at the 5G PPP so far have led to solutions able to preliminary demonstrate some 5G Key Performance Indicators (KPIs) such as higher radio coverage, more bandwidth per mobile device, increased number of connected devices, logical network slicing, lower latency and cost/energy savings.

Network Functions Virtualization (NFV) and Software Defined Networking (SDN) technologies [3][4] are two important enablers in realization of 5G PPP vision. With the help of NFV and SDN, the embedded IT resources at the network are employed to offer added value services, e.g. innovative media services, aiming to improve the end user's Quality of Experience (QoE) and creating new business opportunities for service providers. With a closer look to the added value services, it is possible to infer that they are complex operations composed by one or a set of software applications, e.g. Virtual Network Functions (VNF), cooperating together towards a common goal. For example, an end to end innovative media service is composed of a series of software each taking a specific responsibility. A virtual machine imitating the functionalities of a specialized "hard-wired" device like a firewall. This virtual device helps pre-filtering the incoming video files from end users to the mobile network edge. Non-blocked contents are directed to a video production application designed for sport events. The processed video by the application is then broadcasted locally to the end users who are interested to have a 360 vision of the goal. To guarantee the Quality of Service (QoS) at any traffic status the whole media service is constantly monitored and optimized by a Self-Optimizing Network (SON/Self-X) functionality [5] against the agreed QoS metrics on the Service Level Agreement (SLA).

Depending on factors like: expected performance, the required resources to run properly, maximum number of concurrent hits, etc. in an end to end service offering scenario, applications forming an added value service might be placed at different networking domain, e.g. the network edge, backhaul or in a remote data center far from the end user. With the help of resource orchestration techniques [6] it is possible to coordinate interaction among each pieces and deliver the desired added value service.

Manuscript received XXXXXX, 2017.

The research leading to these results has been supported by the European Commission H2020 5G-PPP project SESAME (G.A. 671596), H2020-ICT-2016-2 METRO-HAUL (G.A. 761727), the FPU Spanish fellowship program (ref. no. FU14/04227), and the Spanish project grants TEC2014-53071-C3-1-P (ONOFRE).

J.-J. Pedreno-Manresa (e-mail: josej.pedreno@upct.es) and P. Pavon-Marino are with the Department of Information and Communication Technologies, Universidad Politécnica de Cartagena, Plaza del Hospital, 1, 30202 Cartagena, Spain.

P.S. Khodashenas and M.S. Siddiqui are with the Fundació i2Cat, C/ Gran Capità 2-4, Nexus I Building, 2nd floor, Office 203, Barcelona 08034, Spain

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2017.2760826, IEEE Communications Letters

There are several studies available in the literature focusing on the placement and application coordination problem from a static perspective [7-9]. In particular, [7] modeled the resource allocation problem with a mixed-integer program and proposed a solution to incorporate limited physical resources into the NFV resource allocation. Authors in [8] studied the resource allocation problem, taking into account the scalability issue, which might be caused by hosting multiple VNFs over the same hardware. Last but not least, in [9] authors in a static scenario studied the impact of latency requirements on the placement of VNFs constituting an added value service.

Despite the importance of the previous studies, given the targeted 5G use cases, there is no surprise to see that there is a great need to investigate dynamic user behavior. For instance, 5GPPP suggests use cases where 5G capabilities are offered in sport stadiums or shopping malls. In this case, the traffic profile changes drastically over time as crowded events occurs sporadically. This definitely demands an effort to propose solutions able to cope with the traffic changes on the fly. Actually, this topic has been covered separately at some extent in the cloud and the radio world. NFV management related actions are listed under solutions for scaling up/down resources, application migrate from one point of presence to another, etc. On the network management side, allocating more/less frequency bands and/or wavelengths, alternative path allocation, etc. are typical solutions.

In this paper we will show how a holistic and joint decision making process is able to significantly improve the network performance compared to a case where only NFV or network related actions are employed. To this end, we selected a real life two tier network topology made of so called "micro nodes" – close to the end user – and "macro nodes" – far from the end user. Then with the help of simulations results generated by our Net2Plan [10][11] tool, we will prove that the proposed joint NFV-network management solution will significantly outperform the disjoint alternatives.

II. Selected Network and Assumptions

Our simulation scenario involves an event, i.e. football match, during which traffic and performance requirements increase drastically due to the amount of people (~100,000) present in a relative small area.

We consider an area of 1.5km radius around the Camp Nou stadium in Barcelona with a real radio access network topology [12] populated with two types of nodes: traditional macro-cells around the map, and we assume the presence of 200 micro-cells distributed in the stadium premises, i.e. the seating area, so coverage is equally distributed for all persons.. Each macro-cell is directly connected to a Central Office (CO) via a 10 Gbps optical link, while all micro-cells inside the stadium are connected via a 1 Gbps optical link to a central gateway, which is connected to the CO with a 10 Gbps optical link.

Distribution of the macro-cell antennas is depicted in Fig 1. We assume all links are bidirectional and macro-cells, micro-cells and CO are NFV-enabled with capacity to instantiate VNFs composing the added value services. CO also hosts the Evolved Packet Core (EPC) and management logics.

Simulations are performed modeling 99,354 individual persons (stadium capacity) originating from specific source points around the edge of the scenario, represented by metro/bus stations, parking lots, etc. Each person travels at a random walking speed to the stadium, and once everyone is inside, the football match begins. For simplicity, we assume the match duration as 5 minutes and during this time, all people remain seated and stationary inside the stadium. Once the match has ended, each person travels back to its origin source point, and the simulation ends when everyone has reached its destination.

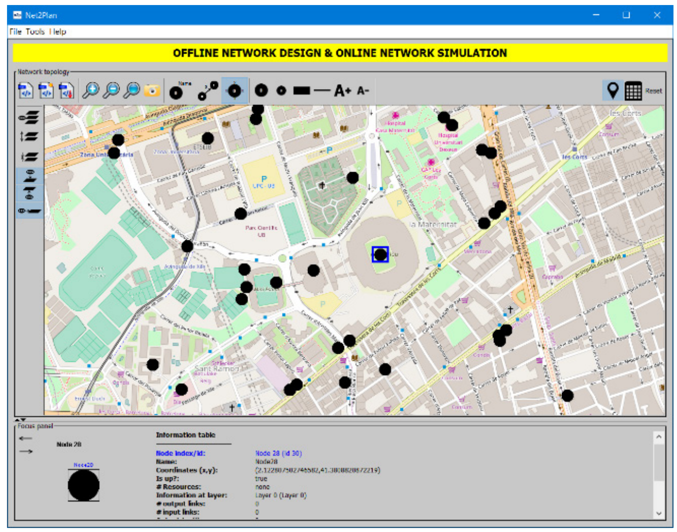

Fig. 1 Radio Access Network topology in Barcelona.

Throughout all the simulation time frame, at each time slot every person has a probability to request a service (Binomial distribution based on the total offered traffic, number of persons and service types). The different services types are 5G Service Bundle (Web, 4K video streaming, etc.), Video Conference and VoIP. Each service has a minimum bandwidth requirements and chained VNFs associated [8], as seen in Table I.

We contemplate two different scenarios of management logic, "non-orchestrated" and "orchestrated". In the first case (non-orchestrated), each service request is handled by the nearest cell [13]. First we check if the handing cell has enough link bandwidth with the CO to carry the service offered traffic, blocking the request otherwise. Then, the allocation algorithm tries to find available VNFs to traverse the chain dictated by Table I (in order). If a particular VNF from the chain is not instantiated or lacks enough capacity, a new instance will be allocated. In case the cell has not enough hardware resources, the algorithm tries to create the remaining VNFs from the chain in the CO. We remark that in order to process a request, the VNFs have to be traverse in strict order from the handling cell to the CO (no going backwards or loops). If the CO also lacks available hardware resources, the service request is blocked.

In the second case (orchestrated), a different approach is used to reflect an orchestrated management, were the serving cell is decided jointly with the service chain allocation. If the nearest cell lacks enough bandwidth to carry a service, the request is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2017.2760826, IEEE Communications Letters

not automatically blocked (as the previous case). Instead, all nearby antennas in coverage (and with enough link bandwidth) are ordered (closest first). Then the allocation algorithm will try to allocate the request on the first available cell in the list that has enough IT resources (following the same procedure explained in the non-orchestrated scenario). Blocking occurs when the service chain cannot be allocated in any cell in coverage. In both scenarios we assume cells have enough radio capabilities to serve a request [14].

Lastly, once all requests at a given time slot have been processed (either in non-orchestrated or orchestrated case), idle VNFs will be de-instantiated to free unused resources.

TABLE I
Service type requirements and associated service chains

Service	*Chained VNFs **	*Bandwidth req.*
VoIP	NAT-FW-TM-FW-NAT	250 Kbps
Video Conference	NAT-FW-TM-VOC-IDPS	2 Mbps
5G Service Bundle	NAT-FW-TM-WOC-IDPS	4 Mbps

* IDPS: Intrusion Detection Prevention, FW: Firewall, NAT: Network Address Translation, TM: Traffic Monitor, VOC: Video Optimization Controller, WOC: WAN Optimization Controller

III. SIMULATION TOOL AND RESULTS

Using Net2Plan, two sets of simulations were performed. For both sets, we establish the following input parameters: 15 Gbps average total offered traffic, service type probabilities of 50% for 5G Service Bundle, 20% for Video Conference and 30% for VoIP. All service types have a holding time of 100 seconds. Each macro-cell site is assumed to have a server with 16 CPU cores, 64 GB of RAM and 10 TB of HDD. Micro-cell sites have a server with 8 CPU cores, 32GB of RAM and 7 TB of HDD. The CO contains 100 CPU cores, 480 GB of RAM and 27 TB of HDD. The hardware requirements to instantiate each VNF and their concurrent number of operations is shown in Table II [15].

TABLE II
Number of concurrent operations and hardware requirements per VNF instance

VNF	*# of concurrent operations*	*Hardware req.*
IDPS	2500	CPU: 2 cores, RAM: 2GB, HDD: 10GB
FW	2500	CPU: 2 cores, RAM: 3GB, HDD: 5GB
NAT	3000	CPU: 1 core, RAM: 1GB, HDD: 2GB
TM	2500	CPU: 1 core, RAM: 3GB, HDD: 2GB
VOC	1000	CPU: 2 cores, RAM: 2GB, HDD: 20GB
WOC	1500	CPU: 1 core, RAM: 2GB, HDD: 10GB

In this first simulation set, no orchestration between radio resources and NFV management occurs. Each service request is assigned to the nearest cell and a service chain allocation is attempted, preferring always VNF instantiation closer to the users, as described in Section II. Using the proposed allocation algorithm, the average blocking in this case is 40%, and the instant blocking is depicted in Fig. 2. We appreciate a peak occurring during the football match. This can be attributed to the fact that during this time, the aggregation links between micro-cells and the central gateway as well as the link between the gateway and the CO saturate. This bottleneck and its consequent high blocking rate is unacceptable in terms of QoS/QoE.

The second set of simulations were performed using an orchestrated management (as explained in Section II), were the serving cell and the allocation VNFs is decided jointly.

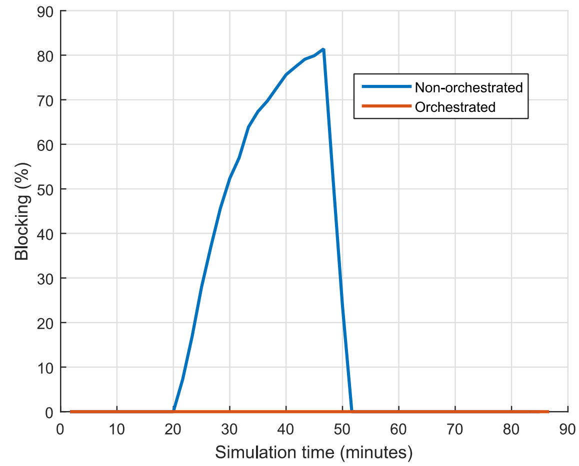

Fig. 2 Blocking in each instant of the simulation

Fig. 3 and 4 show the average resource utilization time pattern in both set of simulations for each type of node: macro-cells, micro-cells and CO. We can observe micro-cells have the highest utilization during most of the simulation for all three resources. This is to be expected, since the server on micro-cells have less hardware capabilities than their macro-cells counterparts; and as people approach and sit in the stadium the number the utilization grows until the match ends. At that time we also note a sudden increase of resource utilization in macro-cells, from people leaving the stadium, slowly decreasing as they abandon the simulation area.

It is important to note that both cases present similar patterns, with the crucial difference that in the orchestrated case, all requests that were previously blocked in micro-cells due to the lack of bandwidth are now rerouted through nearby macro-cells.

Fig 5 illustrates this by showing the average link utilization of micro and macro-cells in both cases (orchestrated and non-orchestrated). In both situations, the micro-cell utilization is the same, blocking services that cannot be allocated in the non-orchestrated case (and with no utilization of the nearby macro-cells). However, the orchestrated case reflects that all the user requests that cannot be allocated by micro-antennas, are rerouted through nearby macro-antenna raising their link utilization while achieving 0% blocking. We can also observe a peak of utilization around minute 50 (the end of the football match) in the non-orchestrated case. Since all people leaving the stadium are trying to request service from the nearest macro-cell antennas this causes a peak in link utilization. Meanwhile, in the orchestrated case the utilization increase is less pronounced since macro-cells are already used during the football match, and the service requests are distributed among several antennas further from the stadium.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2017.2760826, IEEE Communications Letters

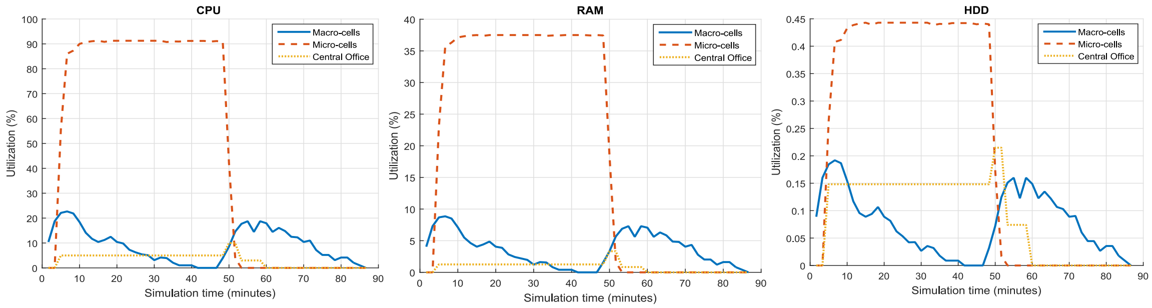

Fig. 3 Hardware utilization, non-orchestrated case.

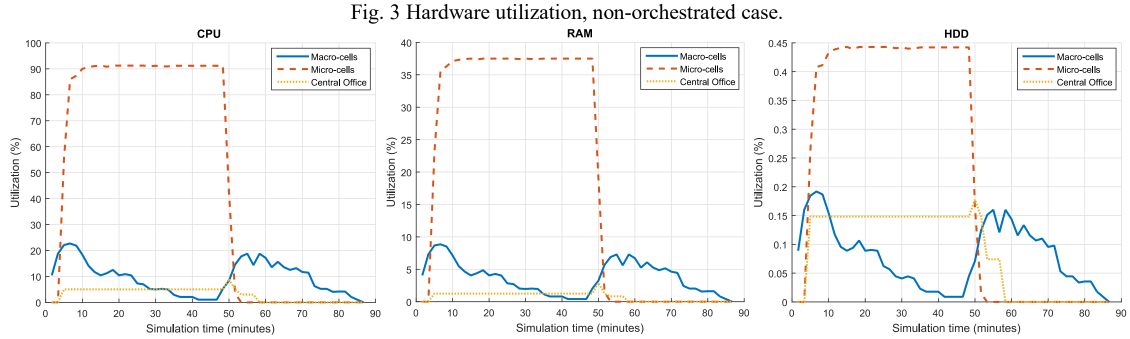

Fig. 4 Hardware utilization, orchestrated case.

IV. Conclusions

The emerging 5G networks and mobile edge computing, leveraging on SDN and NFV technologies, are promising solutions able to fulfil the QoS requirements of diverse IoT verticals (e.g. low latency and huge transmission capacity). Effective orchestration and radio status aware resource allocation is an essential element to form an efficient ecosystem able to address 5G requirements. Such an orchestration needs to be able to coordinate all heterogeneous elements, resources and services in such a complex and sensitive multi-vendor environment. In this paper we show how careful resource orchestration of 5G NFV-enable radio access networks, taking into account not only radio but also bandwidth and NFV requirements, can improve the overall 5G network performance. Such joint resource coordination allows for a better allocation of resources, improving performance and QoE on 5G networks, while reduce blocking and connection failures.

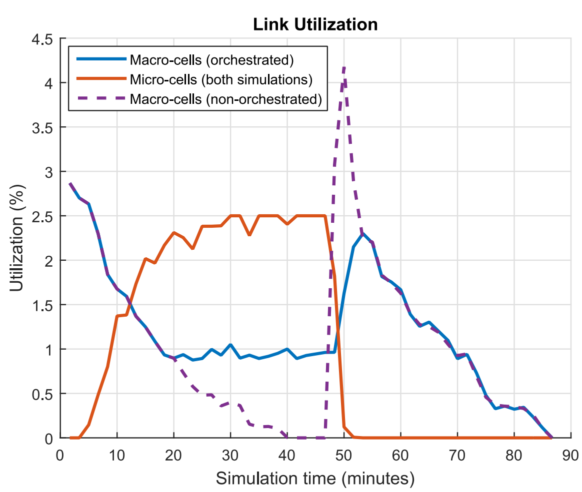

Fig 5. Average link utilization in both simulations.

References

[1] P. S. Khodashenas, J. Aznar, A. Legarrea, C. Ruiz, M. S. Siddiqui, et al., "5G Network Challenges and Realization Insights," Proc. of 18th International Conference on Transparent Optical Networks (ICTON) 2016

[2] The 5G Infrastructure Public Private Partnership. [Online] Available: https://5g-ppp.eu/ [Last Accessed: October 2017]

[3] ETSI NFV white paper, "Network Function Virtualisation: An Introduction, Benefits, Enablers, Challenges and Call for Action," (2012).

[4] R. Jain and S. Paul, "Network Virtualization and Software Defined Networking for Cloud Computing – A Survey," IEEE Comm. Mag., vol. 51, no. 11, pp 24-31, (2013).

[5] J. P. Romero, O. Sallent, C. Ruiz, A. Betzler, P. S. Khodashenas, et al., "Self-X in SESAME," Proc. EuCNC, Athens, (2016).

[6] ETSI, "NFV Management and Orchestration - An Overview," GS NFV-MAN001 V1.1.1, (2014).

[7] T. Lin, Z. Zhou, M. Tornatore and B. Mukherjee, "Optimal Network Function Virtualization Realizing End-to-End Requests," Proc. IEEE GLOBECOM, San Diego, (2015).

[8] M. Savi, M. Tornatore and G. Verticale, "Impact of Processing Costs on Service Chain Placement in Network Functions Virtualization," Proc. IEEE NFV-SDN, San Francisco, (2016).

[9] M. Savi, A. Hmaity, G. Verticale, S. Höst and M. Tornatore, "To distribute or not to distribute? Impact of latency on Virtual Network Function Distribution at the edge of FMC Networks," Proc. ICTON, Trento, (2016).

[10] Net2Plan – The open-source network planner. [Online] Available: http://www.net2plan.com [Last accessed: October 2017]

[11] P. Pavon and J.-L. Izquierdo-Zaragoza "Net2Plan: An Open Source Network Planning Tool for Bridging the Gap between Academia and Industry," IEEE Network, vol. 29, no. 5, pp. 90-96, Sep. 2015.

[12] Spanish Ministry of Energy, Tourism and Digital Agenda – Information about Radioelectric Installations and Radiation Exposure Levels. [Online] Available: https://geoportal.minetur.gob.es/VCTEL/vcne.do [Last accessed: October 2017]

[13] C.-C. Lin, "Handover Mechanisms in 3GPP Long Term Evolution (LTE)", Ph.D. dissertation, University of Technology, Sydney 2013

[14] 3GPP TR25.814 version 7.1.0, "Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA) (Release 7)," 3GPP, Tech. Rep. September 2006

[15] SESAME D4.3 "Techniques for efficient VNF Deployment with relevant VIM extensions, Evaluation framework" Jun 2017. [Online] Available: http://www.sesame-h2020-5g-ppp.eu/Deliverables.aspx [Last accessed: October 2017]