A Software-Defined Architecture for Next-Generation Cellular Networks

Vassilios G. Vassilakis*, Ioannis D. Moscholios†, Bander A. Alzahrani‡, Michael D. Logothetis§
* School of Computing, Engineering & Mathematics, University of Brighton, Brighton, United Kingdom
† Dept. of Informatics & Telecommunications, University of Peloponnese, Tripolis, Greece
‡ Information Systems Department, King Abdulaziz University, Jeddah, Saudi Arabia
§ Dept. of Electrical & Computer Engineering, University of Patras, Patras, Greece

***Abstract*—In the recent years, mobile cellular networks are undergoing fundamental changes and many established concepts are being revisited. New emerging paradigms, such as Software-Defined Networking (SDN), Mobile Cloud Computing (MCC), Network Function Virtualization (NFV), Internet of Things (IoT), and Mobile Social Networking (MSN), bring challenges in the design of cellular networks architectures. Current Long-Term Evolution (LTE) networks are not able to accommodate these new trends in a scalable and efficient way. In this paper, first we discuss the limitations of the current LTE architecture. Second, driven by the new communication needs and by the advances in aforementioned areas, we propose a new architecture for next-generation cellular networks. Some of its characteristics include support for distributed content routing, Heterogeneous Networks (HetNets) and multiple Radio Access Technologies (RATs). Finally, we present simulation results which show that significant backhaul traffic savings can be achieved by implementing caching and routing functions at the network edge.**

Keywords*—Software-defined networking; network function virtualization; cellular network architecture.

I. Introduction

In the recent years we are witnessing a widespread use of end user devices with advanced capabilities, such as smartphones and tablet computers, and the emergence of new services and communication technologies. This new evolved ecosystem, however, imposes very strict requirements on the network architecture and its functionality. Enabling small end-to-end latency and supporting a large number of connections at the appropriate level, is not possible to be achieved in current Long-Term Evolution (LTE) networks. The fundamental limitations of current approaches lie in their centralized mobility management and data forwarding, as well as in insufficient support for multiple co-existing Radio Access Technologies (RATs) [1].

Today, a large variety of RATs and heterogeneous wireless networks have been successfully deployed and used. However, under the current architectural framework, it is not easy to integrate or to enable adequate coordination of these technologies. Despite the fact that the coverage of such wireless and cellular networks has increased by deploying more Base Stations (BSs) and Access Points (APs), the Quality-of-Experience (QoE) of Mobile Users (MUs) does not increase accordingly. For example, the current architectural approach does not enable an MU selecting the best available network in a dynamic and efficient way. It also does not enable simultaneous and coordinated use of radio resources from different RATs. This leads to highly inefficient use of hardware resources (wireless infrastructure) and spectrum, which is worsened even more with almost uncontrollable inter-RAT interference [2].

In this paper, we propose a new architectural framework for next-generation cellular networks. We benefit from the recent advances in Software Defined Networking (SDN) [3] and Network Function Virtualization (NFV) [4], which are natively integrated into the new architecture. Traditionally, SDN and NFV, although not dependent on each other, are seen as closely related and complementary concepts [5]. This integration enables good scalability in terms of supporting a large number of connections and heavy mobility scenarios. Also, the introduction of new services and applications becomes much easier. Decoupling control and data planes, and abstracting network functions from the underlying physical infrastructure, brings much greater flexibility to efficiently utilize radio and computing resources both in the Radio Access Network (RAN) [6] as well as in the Mobile Core Network (MCN) [7]. Furthermore, our proposed approach enables the incorporation of Mobile Edge Computing (MEC) services in an easy and straightforward way. As our experiments show, by bringing the content and the decision functions to the network edge (RAN level instead of MCN level), significant network capacity savings can be achieved.

This paper is organized as follows. In Section II, we present the related work on SDN-based cellular architectures. In Section III, we briefly discuss the main limitations of the current LTE networks. In Section IV, we propose a new architectural framework for next-generation cellular networks, which exploits the benefits of SDN and NFV, and enables the incorporation of MEC services. In Section V, we present our simulation results. We compare the proposed approach with the traditional one in terms of backhaul capacity savings. We conclude in Section VI. Also, in Tables I and II, we present the list of common abbreviations used in the literature and the list of newly introduced abbreviations, respectively.

II. Related Work

In this section, we present the most important, recent works that apply SDN concepts in cellular networks.

The OpenRAN architecture [8] adopts a software-defined RAN approach, achieved via radio network virtualization. OpenRAN comprises the following parts: a) wireless spectrum resource pool, b) cloud computing resource pool, and c) SDN

controller. The controller determines the strategies for allocating appropriate resources to virtual radio access elements. The communication between the controller and the virtual elements is performed via an SDN agent though some SDN protocol. It is shown that this solution can efficiently address the challenges in interconnecting heterogeneous networks. Although this is a very interesting approach, for the moment it does not go beyond the RAN.

The SoftRAN, proposed in [9], is a software-defined control plane for LTE RAN. According to this approach, a virtual macro BS acts as an abstracted centralized SDN controller. Its role is to control a number of distributed data plane radio elements to achieve load balancing and user QoE improvements. The radio elements perform local control decisions within a cell, whereas the abstracted macro BS performs the inter-cell control tasks. Some of the limitations of SoftRAN is that it does not consider HetNet scenarios and SDN abstractions in the MCN.

In [10], the SoftNet, a decentralized SDN based architecture is proposed as an alternative to LTE. The adopted design principles include high network availability, efficiency, and scalability. SoftNet supports multi-RAT coordination, achieved via the NFV infrastructure. Also, the centralized SDN controller is responsible for managing virtual computing and storage resources. Evaluation of the proposed approach shows significant reduction in signalling overhead, compared to LTE networks. This is achieved by moving a part of mobility and data forwarding functionalities closer to the network edge. Also, the proposed approach shows enhanced cellular system capacity and improved data forwarding efficiency.

Hence, most of the approaches proposed in the literature are focusing on radio resources virtualization, rather than on NFV. Also, they do exploit the interactions between RAN and MEC towards bringing the network intelligence to the edge (e.g., via MEC), which is essential to provide adequate QoE in cases of fast user mobility and in densely-populated areas, such as shopping malls and stadiums.

TABLE I. LIST OF COMMON ABBREVIATIONS

ANDSF	Access Network Discovery and Selection Function
AP	Access Point
API	Application Programming Interface
BS	Base Station
D2D	Device-to-Device
EPC	Evolved Packet Core
GTP	GPRS Tunnelling Protocol
HetNet	Heterogeneous Network
IMS	IP Multimedia Subsystem
IoT	Internet of Things
LIPA	Local IP Access
LTE	Long-Term Evolution
MCC	Mobile Cloud Computing
M-CDN	Mobile Content Delivery Network
MCN	Mobile Core Network
MEC	Mobile Edge Computing
MSN	Mobile Social Networking
MU	Mobile User
NFV	Network Function Virtualization
PDN	Packet Data Network
P-GW	Packet Data Network Gateway
QoE	Quality-of-Experience
RAN	Radio Access Network
RAT	Radio Access Technology
SDN	Software-Defined Networking
SIPTO	Selected IP Traffic Offload
S-GW	Serving Gateway
TA	Tracking Area
VM	Virtual Machine

TABLE II. LIST OF NEW ABBREVIATIONS

CCRF	Core Content Resolution Function
CMMF	Core Mobility Management Function
CSC	Core SDN Controller
DCF	Device Control Function
LCCF	Local Content Caching Function
LCRF	Local Content Routing Function
LMMF	Local Mobility Management Function
LO-GW	Local Offload Gateway
LRRF	Local Request Resolution Function
LSA	Local SDN Agent
LSC	Local SDN Controller
MBS	Macro-cell BS
MRCF	Multi-RAT Coordination Function
RMF	Resource Management Function
SBS	Small-cell BS

III. LIMITATIONS OF CURRENT LTE NETWORKS

The main driver behind the design of the current LTE networks was the requirement for supporting all-IP communication paradigm. In particular, the main focus was on supporting IP-based multimedia services via the introduction of the IP Multimedia Subsystem (IMS) [11]. However, the emergence of more advanced mobile services, such as Mobile Social Networking (MSN) [12], Mobile Cloud Computing (MCC) [13], and Internet of Things (IoT) [14], imposes very tight constraints on end-to-end and handover latency. Below, we briefly discuss the main limitations of current LTE networks.

A. Inefficiency of content routing

The current approach demands that all IP traffic passes via the MCN (Evolved Packet Core (EPC) in LTE terminology). This is highly inefficient, since today we have many localised services, where users of social networks or file sharing systems are located in close geographical proximity to each other [15]. The demand for all this traffic to traverse MCN introduces unnecessary delays and consumes scarce network resources. This problem is worsened even more by the demand to establish connections via Packet Data Network (PDN) Gateways (P-GWs), which are usually very sparsely deployed within the mobile network, and the use of tunnelling, which hinders the routing optimization.

B. Significant signalling overhead

Requirement for persistent PDN connections, to ensure efficient support of IP multimedia services in LTE networks, imposes significant signalling overhead for IoT and mobile Internet services. This is especially true for MSN and MCC applications, which typically send small packets with long periods of inactivity within the same connection. Also, sensor networks or monitoring services, which typically rely on data from a large number of devices in different locations, greatly suffer from huge signalling overhead [16].

C. Poor support for multiple radio access technologies

Although the introduction of the Access Network Discovery and Selection Function (ANDSF) [17] enables WiFi network selection by an MU, there is no mechanism to support

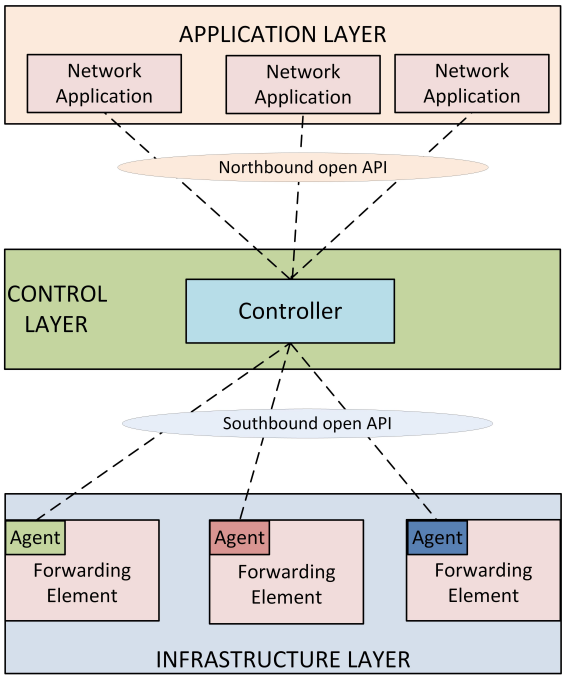

Fig. 1. Layering Concept in Software-Defined Networking.

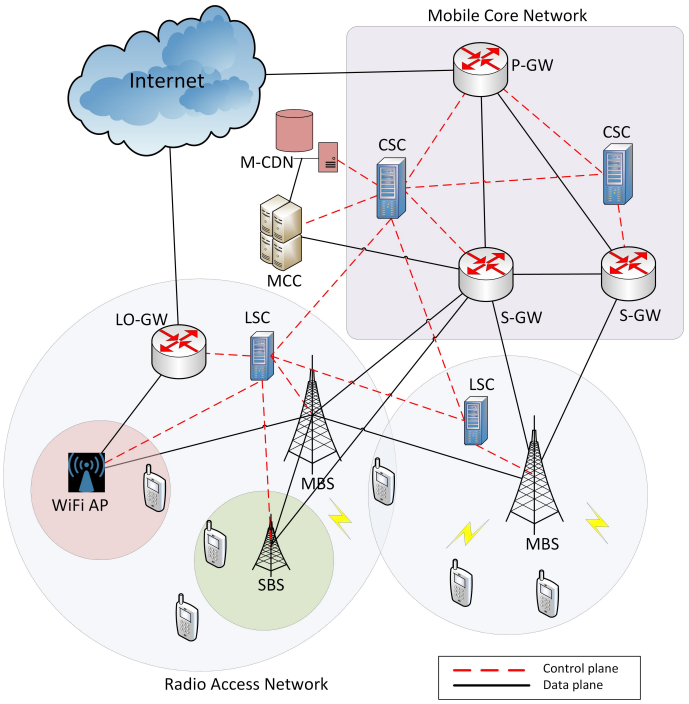

Fig. 2. A Software-Defined Cellular Network Architecture.

efficient use of radio resources when multiple RATs, using both licenced and licence-exempt spectrum, are present. Also, no coordination among different RATs can be easily supported, due to LTE requirement to forward data via the P-GW in case of vertical handovers [18]. This may be too slow in case, e.g., of MCC applications, which typically require low latency and demand considerable amount of radio resources to guarantee good QoE to MUs.

D. Scalability limitations for IoT services

Current LTE architecture is not scalable for supporting IoT services, which typically are designed for a large number of resource-constrained devices with intermittent connectivity. However, the current LTE attachment strategy, that demands each device maintaining at least one connection to the P-GW, is inefficient and not scalable in case of IoT and smart grid devices, such as sensors, actuators, and smart meters [19].

IV. A Software-Defined Architecture for Cellular Networks

Motivated by the limitations of current cellular architectures, as discussed in the previous section, and facilitated by the advances in SDN and NFV technologies, we propose a new software-defined cellular network architecture. The layering concept of SDN is shown in Fig. 1. The SDN controller provides a global view of available resources to the network applications via the *northbound open API*. At the same time, the SDN controller configures flow tables at the forwarding elements via the *southbound open API*.

The design objectives for the new architecture are the following. First, it must efficiently support a wide range of services and applications, from those that require high bandwidth and stable connectivity to those that send small chunks of data with long periods of inactivity. Second, it must provide content routing and resolution functions that can efficiently and quickly adapt to locality characteristics of the communication, supporting at the same time multiple RATs. Below, we describe the RAN and the MCN of the proposed architecture (see also Fig. 2).

A. Radio Access Network

RAN consists of Small-cell BSs (SBSs), Macro-cell BSs (MBSs), WiFi APs, Local Offload Gateways (LO-GWs), and MUs. These are coordinated by the Local SDN Controller (LSC). RAN is divided into clusters, with each cluster covering one or more macro cells and being controlled by a dedicated LSC. The notion of clusters is similar to the notion of Tracking Areas (TAs) in LTE networks. For simplicity, in Fig. 2 each one of the two LSCs controls a cluster covering only a single marco cell.

LSC is responsible for receiving connection requests from its cluster and for allocating appropriate destination address. This is performed by the Local Request Resolution Function (LRRF). In particular, LRRF will facilitate the connection establishment either with an in-cluster entity (MU, SBS, etc.) or will forward the request to MCN. Hence, LRRF is aware of the network topology within the cluster and of egress nodes connections towards MCN and other clusters.

To enable efficient resource utilization and according to the demand from new emerging services, we distinguish two types of connections: synchronous and asynchronous. A synchronous connection is used to support the traditional multimedia applications, such as real-time video streaming, teleconferences, voice, etc. On the other hand, an asynchronous connection (also referred to as virtual connection), is used to support

IoT-type applications, push notification services [20], and in general, services with intermittent connectivity behaviour or requirements. In that case, there is no need to keep MUs attached to MCN all the time and to reserve resources for tunnelling protocols (such as GPRS Tunnelling Protocol (GTP) in LTE).

Other functions of LSC include the Multi-RAT Coordination Function (MRCF) and the Local Content Caching Function (LCCF). MRCF is responsible for allocating radio resources in geographical areas where more than one RAT is available. It can be seen as ANDSF, enhanced with traffic offloading capabilities, using schemes such as Selected IP Traffic Offload (SIPTO) [21] or Local IP Access (LIPA) [22].

LCCF observes content requests from in-cluster MUs and keeps track of the localized content popularity. Based on that and on the knowledge of available storage resources in the cluster, LCCF is responsible for caching decisions within the cluster. The caching decision logic could be based on a number of algorithms proposed in the literature (e.g., [23], [24], [25]). Most of caching solutions exploit the fact that the popularity distribution of content objects follows the Zipf-Mandelbrot distribution [26]. This means that even by allocating relatively small storage space, high cache hit ratio can be achieved [27]. This is expected to greatly reduce the content access delay from MUs and the traffic going via MCN.

Local routing decisions are performed by the Local Content Routing Function (LCRF). LCRF receives requests from LRRF to construct the content delivery path to a local source. Then it configures the flow tables at the data plane forwarding elements. LSC is also responsible for steering the Device-to-Device (D2D) communication via the Device Control Function (DCF). This can be achieved using technologies such as LTE Direct [28], WiFi, or Bluetooth for data transfer between MUs, while the control channels to/from BS may use licenced spectrum [29]. Furthermore, by co-designing LCRF with LCCF, joint optimization of caching and routing logic can be achieved by exploiting schemes such as [30].

Finally, LCS is handling the in-cluster mobility via the Local Mobility Management Function (LMMF). Hence, this information is not passed to the MCN, which greatly reduces the processing and signalling overhead, due to reduced paging messages [31]. This also enables native and elegant incorporation of distributed mobility management schemes [32]. Furthermore, SDN-assisted mobility management can efficiently support even fast moving users/vehicles assuring acceptable QoE [33].

B. Mobile Core Network

MCN consists of a distributed set of Core SDN Controllers (CSCs), MCC infrastructure, Mobile Content Delivery Network (M-CDN) servers, P-GWs, and Serving Gateways (S-GWs). CSC is responsible for receiving and handling connection requests from a set of dedicated clusters, performed by the Core Content Resolution Function (CCRF), and for carrying out the mobility management, via the Core Mobility Management Function (CMMF). Management of storage (i.e., CDN or in-network caches), computing (i.e., MCC), spectrum, and energy resources, as well as QoE support, is performed by the Resource Management Function (RMF). RMF's decisions

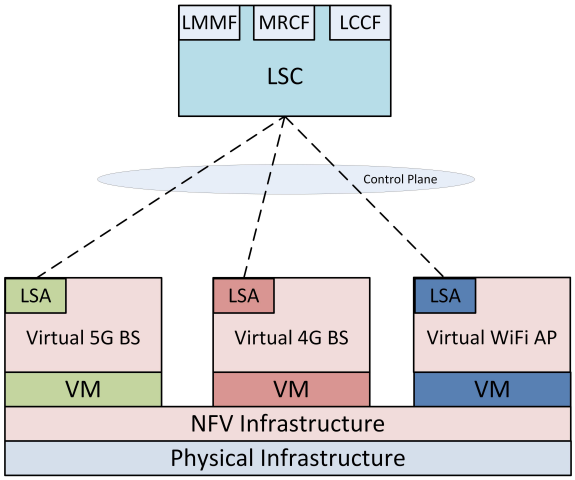

Fig. 3. Virtualized Radio Access Network.

on the allocation of (both physical and virtualized) resources may be based on a number of factors, such as current demand and consumption, monitored radio network conditions, MU density and mobility patterns. This can be greatly benefited by the studies on predictive human mobility (e.g., [34], [35]). To support energy-efficient operation, RMF is responsible for moving the virtualized resources away from heavily underutilized clusters and for switching some of the equipment off. This would enable energy savings during off-peak hours.

The role of P-GW and S-GW is similar to role of homonymous entities in LTE networks, but is restricted to data plane. The corresponding control plane functionality is performed by the CSC (in accordance with the SDN concept). In particular, P-GW is used to access external IP networks (such as the Internet), whereas S-GW is used to access the RAN.

C. Virtualizing Network Functions

In Fig. 3, we present a model for virtualized RAN using SDN and NFV. A number of Virtual Machines (VMs), running on the same physical infrastructure, may enable virtualized implementation of various BSs and APs via any available NFV infrastructure [36]. As it is shown, LSC controls each virtual BS or AP via a dedicated Local SDN Agent (LSA).

Realization of a particular virtual function, e.g., such as content routing from 5G BS to WiFi AP, can be performed as shown in Fig. 4. In this example, Bob wants to receive a *content object* from Alice. Let us assume that the request resolution has already taken place (via LRRF) and LSC knows that the content source is Alice. The virtual LCRF will need to construct the delivery path and to configure the flow tables along the path by sending *route configuration* messages. After that, when the data plane forwarding entities (i.e., virtual 5G BS and virtual WiFi AP) receive the requested object, they forward it to next hop according to the flow table.

To enable efficient and smooth collaboration of CSCs within the MCN, we adopt the solutions developed in the area of NFV. NFV decouples the network functions from the underlying hardware, thus making these functions virtualised by allowing them to be migrated and instantiated on demand.

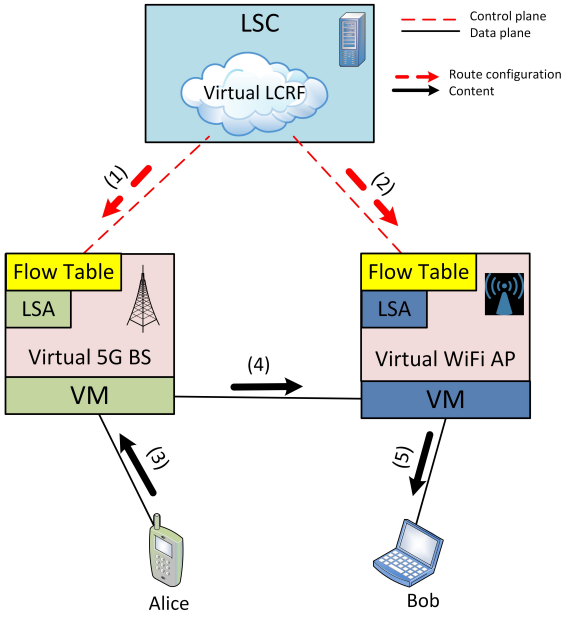

Fig. 4. Realization of Local Content Routing Function via NFV and SDN.

This can be relatively easy realized via VMs, using tools like VMware [37] or VirtualBox [38]. Furthermore, the rise of low-cost and low-demand virtualization technologies such as unikernels [39] opens up avenues for migrating network functions and services even to MUs. A recent example is the implementation of caching functionality at D2D level [40].

V. Evaluation

In this section, we evaluate the performance of our proposed solution and compare it with the traditional LTE approach. To this end, we have built an ns3-based simulator [41] and determine the savings in the backhaul capacity that are achieved by our approach. The backhaul is considered one of major problems in the path towards advanced 5G services and applications. It also seems that various wireless technologies will be considered for backhauling future 5G networks and, hence, achieving significant bandwidth savings is of major importance [42], [43].

In the first scenario, we consider 50 MUs that are accommodated within a cluster of macro cells. Hence, they are controlled by a single LSC. MUs request content objects of different sizes which consist of one or more 20KB-sized chunks. For each chunk there is a separate request message (although alternative approaches, such as requesting multiple chunks or the whole content with a single message, can be easily supported). Initially we generate on average 10 chunk requests per second per MU. Next, we increase the average request rate to 50, 250, 1250, and 62500 requests/sec. Recall that LSC supports local caching though the LCCF. Hence, if the requested chunk is found in the local cache, it can be returned immediately to the requesting MU without the need to deliver the chunk from the remote server. This can result in backhaul traffic reduction, as shown in Fig. 5. Traffic savings greatly depend on the availability of the requested content in the local cache, which in turn affects the cache hit ratio. In Fig. 5, in addition to results for the traditional approach which does

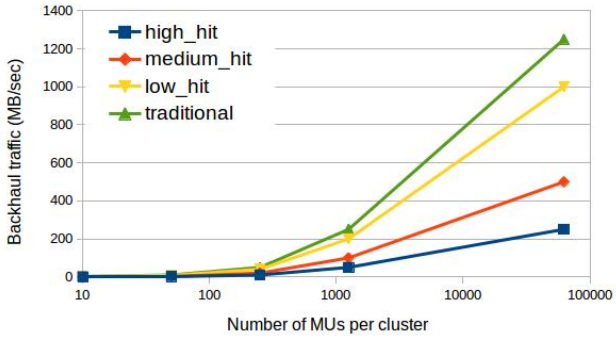

Fig. 5. Backhaul traffic vs MU number.

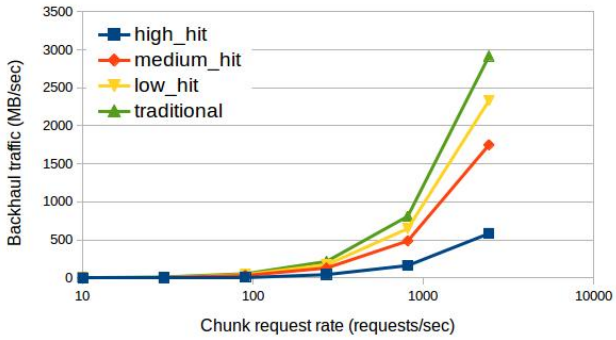

Fig. 6. Backhaul traffic vs content request rate.

not utilise edge caching, we present the results of our proposed scheme for three cases of cache hit ratio: low (0.2), medium (0.4), and high (0.8). We observe that significant bandwidth savings can be achieved in the backhaul, especially when the number of MUs in the cluster is high. Traffic reduction also greatly depends on the cache hit ratio, which can be achieved by preloading local caches with popular content.

In the second scenario, we increase the request rate from 10 to 30, 90, 270, 810, and 2430 chunk requests per second. At the same time we increase the number of MUs in the cluster from 10 to 60 in steps of 10. In Fig. 6 we present the resultant backhaul traffic in the traditional approach as well as in the proposed scheme for low, medium, and high cache hit ratios. We observe that significant bandwidth savings can be achieved in the backhaul, especially when the chunk request rate is high (i.e., cases when large content objects are requested).

VI. Conclusion

In this paper, we first discuss the limitations of the current LTE networks. Their poor support for IoT and cloud computing applications, results in inefficient routing, high signalling overhead, overloaded backhaul, and scalability issues. Second, we propose a novel cellular network architecture based on SDN and NFV concepts. We introduce virtualized network functions for both RAN and MCN, to address the problems of mobility management, multi-RAT coordination, and efficient content routing. Finally, we evaluate the proposed approach my means of computer simulations. In particular, we show that by enabling intelligent routing and caching at the network edge, significant savings of the backhaul capacity can be achieved.

REFERENCES

[1] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5G on the horizon: Key challenges for the radio-access network," IEEE Vehicular Technology Magazine, vol. 8, no. 3, 2013, pp. 47-53.

[2] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," IEEE Communications Magazine, vol. 51, no. 3, 2013, pp. 136-144.

[3] B. Nunes, M. Mendonca, X. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," IEEE Communications Surveys & Tutorials, vol. 16, no. 3, 2014, pp. 1-18.

[4] N. Chowdhury, M. Kabir, and R. Boutaba, "A survey of network virtualization," Computer Networks, vol. 54, no. 5, 2010, pp. 862-876.

[5] E. Haleplidis, J. H. Salim, S. Denazis, and O. Koufopavlou, "Towards a network abstraction model for SDN," Journal of Network and Systems Management, vol. 23, no. 2, April 2015, pp. 309-327.

[6] R. Shrivastava, S. Costanzo, K. Samdanis, D. Xenakis, D. Grace, and L. Merakos, "An SDN-based framework for elastic resource sharing in integrated FDD/TDD LTE-A HetNets," Proc. 3rd IEEE International Conference on Cloud Networking (CloudNet), 2014, pp. 126-131.

[7] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "Softcell: Scalable and flexible cellular core network architecture," Proc. 9th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT), 2013, pp. 163-174.

[8] M. Yang, Y. Li, D. Jin, L. Su, S. Ma, and L. Zeng, "OpenRAN: a software-defined ran architecture via virtualization," Proc. ACM SIGCOMM, 2013, pp. 549-550.

[9] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software defined radio access network," Proc. 2nd ACM SIGCOMM workshop on Hot Topics in Software Defined Networking, 2013, pp. 25-30.

[10] H. Wang, S. Chen, H. Xu, M. Ai, and Y. Shi, "SoftNet: A software defined decentralized mobile network architecture toward 5G," IEEE Network, vol. 29, no. 2, 2015, pp. 16-22.

[11] M. Poikselkä and G. Mayer, "The IMS: IP multimedia concepts and services," John Wiley & Sons, 2013.

[12] N. Vastardis and K. Yang, "Mobile social networks: architectures, social properties, and key research challenges," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, 2013, pp. 1355-1371.

[13] C. Magurawalage, M. Sarathchandra, K. Yang, L. Hu, and J. Zhang, "Energy-efficient and network-aware offloading algorithm for mobile cloud computing," Computer Networks, vol. 74, 2014, pp. 22-33.

[14] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, 2010, pp. 2787-2805.

[15] T. Chung, J. Han, H. Lee, J. Kangasharju, T. Kwon, and Y. Choi, "Spatial and temporal locality of content in BitTorrent: A measurement study," IFIP Networking Conference, 2013, pp. 1-9.

[16] J. Zhang, L. Shan, H. Hu, and Y. Yang, "Mobile cellular networks and wireless sensor networks: Toward convergence," IEEE Communications Magazine, vol. 50, no. 3, 2012, pp. 164-169.

[17] D. Triantafyllopoulou, T. Guo, and K. Moessner, "Energy efficient ANDSF-assisted network discovery for non-3GPP access networks," Proc. 17th IEEE International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2012, pp. 297-301.

[18] A. Ahmed, L. M. Boulahia, and D. Gaïti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," IEEE Communications Surveys & Tutorials, vol. 16, no. 2, 2014, pp. 776-811.

[19] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "Performance evaluation of power demand scheduling scenarios in a smart grid environment," Applied Energy, vol. 142, 2015, pp. 164-178.

[20] I. Sato, A. Bouabdallah, and X. Lagrange, "A new LTE/EPC control-plane based transmission procedure to cope with short data push services," Wireless Pers. Commun., vol. 72, no. 3, 2013, pp. 1723-1735.

[21] T. Taleb, K. Samdanis, and S. Schmid, "DNS-based solution for operator control of selected IP traffic offload," Proc. IEEE International Conference on Communications (ICC), 2011, pp. 1-5.

[22] K. Samdanis, T. Taleb, and S. Schmid, "Traffic offload enhancements for eUTRAN," IEEE Communications Surveys & Tutorials, vol. 14, no. 3, 2012, pp. 884-896.

[23] V. A. Siris, X. Vasilakos, and G. C. Polyzos, "Efficient proactive caching for supporting seamless mobility," Proc. 15th IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2014, pp. 1-6.

[24] S. Vural, P. Navaratnam, N. Wang, C. Wang, L. Dong, and R. Tafazolli, "In-network caching of internet-of-things data," Proc. IEEE International Conference on Communications (ICC), 2014, pp. 3185-3190.

[25] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Multicast-aware caching for small cell networks," Proc. IEEE Wireless Communications and Networking Conference (WCNC), 2014, pp. 2300-2305.

[26] Z. Silagadze, "Citations and the Zipf-Mandelbrot's law," Complex Systems, vol. 11, 1997, pp. 487-499.

[27] X. Zhang, N. Wang, V. G. Vassilakis, and M. P. Howarth, "A distributed in-network caching scheme for P2P-like content chunk delivery," Computer Networks, vol. 91, 2015, pp. 577-592.

[28] B. Raghothaman, E. Deng, R. Pragada, G. Sternberg, T. Deng, and K. Vanganuru, "Architecture and protocols for LTE-based device to device communication," Proc. IEEE International Conference on Computing, Networking and Communications (ICNC), 2013, pp. 895-899.

[29] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," IEEE Comm. Mag., vol. 50, no. 3, 2012, pp. 170-177.

[30] V. G.Vassilakis, M. F. Al-Naday, M. J. Reed, B. Alzahrani, K. Yang, I. D. Moscholios, and M. D. Logothetis, "A cache-aware routing scheme for information-centric networks," Proc. IEEE/IET 9th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), pp. 721-726, 2014.

[31] D. Xenakis, N. Passas, L. Merakos, and C. Verikoukis, "Mobility management for femtocells in LTE-Advanced: Key aspects and survey of handover decision algorithms," IEEE Comm. Surveys & Tutorials, vol. 16, no. 1, pp. 64-91, 2014.

[32] F. Giust, L. Cominardi, and C. Bernardos, "Distributed mobility management for future 5G networks: overview and analysis of existing approaches," IEEE Commun. Mag., vol. 53, no. 1, 2015, pp. 142-149.

[33] V. G. Vassilakis, I. D. Moscholios, A. Bontozoglou, and M. D. Logothetis, "Mobility-aware QoS assurance in software-defined radio access networks: An analytical study," Proc. IEEE Workshop on Software-Defined 5G Networks (Soft5G), London, U.K., April 2015.

[34] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, no. 7196, 2008, pp. 779-782.

[35] A. Mohamed, O. Onireti, S. A. Hoseinitabatabaei, M. Imran, A. Imran, and R. Tafazolli, "Mobility prediction for handover management in cellular networks with control/data separation," Proc. IEEE International Conference on Communications (ICC), 2015, pp. 3939-3944.

[36] J. Batallé, J. F. Riera, E. Escalona, and J. A. Garcia-Espin, "On the implementation of NFV over an OpenFlow infrastructure: Routing Function Virtualization," Proc. IEEE Software Defined Networks for Future Networks and Services (SDN4FNS), 2013, pp. 1-6.

[37] VMware, http://www.vmware.com/ [Feb. 2016].

[38] VirtualBox, https://www.virtualbox.org/ [Feb. 2016].

[39] A. Madhavapeddy, D. J. Scott, J. Lango, M. Cavage, P. Helland, and D. Owens, "Unikernels: Rise of the virtual library operating system," Communications of the ACM, vol. 11, no. 11, January 2014.

[40] G. Chandrasekaran, N. Wang and R. Tafazolli, "Caching on the move: towards D2D-based information centric networking for mobile content distribution," Proc. 40th IEEE Conference on Local Computer Networks (LCN), Florida, U.S.A., 2015.

[41] Network Simulator NS-3, http://www.nsnam.com [Feb. 2016].

[42] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: challenges and research advances," IEEE Network, vol. 28, no. 6, 2014, pp. 6-11.

[43] C. Dehos, J. L. Gonzlez, A. De Domenico, D. Ktenas, and L. Dussopt, "Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?," IEEE Communications Magazine, vol. 52, no. 9, 2014, pp. 88-95.