

A Framework of Network Connectivity Management in Multi-Clouds Infrastructure

Anwer Al-Dulaimi, *Senior Member, IEEE*, Shahid Mumtaz, *Senior Member, IEEE*, Saba Al-Rubaye, *Senior Member, IEEE*, Siming Zhang, *Member, IEEE*, and Chih-Lin I, *Senior Member, IEEE*

Abstract—The network function virtualization (NFV) transformation is gaining an incredible momentum from mobile operators as one of the significant solutions to improve the resource allocation and system scalability in fifth-generation (5G) networks. However, the ultra-dense deployments in 5G create high volumes of traffic that pushes the physical and virtual resources of cloud-based networks to edge limits. Consider a distributed cloud, replacing the core network with virtual entities in the form of virtual network functions (VNFs) still requires efficient integration with various underlying hardware technologies. Therefore, orchestrating the distribution of load between cloud geo-datacenters starts by instantiating a virtual and physical network typologies that connect involved fronthaul with relevant VNFs. In this article, we provide a framework to manage calls within 5G network clusters for efficient utilization of computational resources and also to prevent unnecessary signaling. We also propose a new scheme to instantiate virtual tunnels for call forwarding between network clusters leading to new logic networks that combine geo-datacenters and fronthaul. To facilitate service chaining in cloud, we propose a new enhanced management and orchestration (E-MANO) architecture that brings network traffic policies from the application layer to the fronthaul for instant monitoring of available resources. We provide analysis and testbed results in support of our proposals.

INTRODUCTION

MIGRATING appliances to the cloud in the form of virtual machines (VMs) enables a centralized dynamic allocation of computational resources compared with fragile uncontrolled schemes in proprietary hardware. The software defined networking (SDN) separates the control functionality from data plane to support interworking between VMs operating in cloud especially those located in different geo-datacenters [1]. The virtualization layers enable new type of pervasive networks that process user data at the edge of the mobile network nearby end-users. To this end, we identify two types of virtual processing units: baseband unit (BBU) that processes the signal of original frequency before it is modulated. Secondly, virtual network core units that provide converged voice and data to operators network. Since the core network is only a software package that can be launched at any datacenter, call processing and data downloading requests can also be maintained near fronthaul to ease the burden of transferring large traffic volumes to core network [2]. Therefore, geo-datacenters need to be deployed within various network domains to provide the computational backup for their cluster topology. This model of local load processing by geo-based servers brings distributed-cloud deeper into radio access segment. Assuming callers located at two different cluster, there still a need to define a controlling scheme that

manages call processing assignment between selected geo-datacenters where involved callers are located. In addition, we need to characterize performance in such separated cloud topology to define the necessary enhancements to the European Telecommunications Standards Institute (ETSI) network function virtualization (NFV) model [3].

The conventional architecture of wireless networks deploys middleboxes in a particular order to filter unwanted traffic and network address translation (NAT) to modify packets source and destination addresses. A manual selection of those middleboxes enable network administrators to steer traffic between sites/routers [4]. The decoupling of control and data planes in SDN based networks transforms the network architecture into centralized programmable model. As the data plane is migrated into VM processing entities, the control plane becomes represented in the form of bare metal controllers that manage data forwarding between network slices. This network formulation provides the necessary components for an autonomous network with SDN that provide OpenFlow protocols to enable communication between the SDN controller and the network nodes on the southbound interface while providing control-application interfaces on the northbound interface [5]. However, this flexibility in managing network interfaces and diverting load flows remains isolated within the SDN domain or geo-datacenter.

Expanding the NFV framework to bundle virtual Evolved Packer Core (vEPC) functions with one or more servers does not necessarily imply a change in the mechanism used for interfacing network nodes or managing network operations [6]. Clearly, this refers to a complete virtualization of all core network components including the IP Multimedia Subsystem (IMS) in the form of VNFs that are connect to an underlying physical network using different capacity ports [7].

At the highest level, a VNF provides various components such as routers, gateways, firewalls, policy control, load balancers, etc. that use open interfaces such as extensible application programming interfaces (APIs) to create a logical network [8]. Assuming that each of these logical networks is managed by a geo-datacenter, the NFV/SDN literature does not provide any framework to process service management between multiple geo-datacenters including resource requirements and complete connectivity information.

In this article, we propose a new scheme to manage calls between users located at two separated network clusters. The goal is to activate one virtual network core for each call to prevent unnecessary consumption of computational resources. Since our model suggests prior knowledge on frontahul incom-

network entities. In this mode, call detail records (CDRs) that define categories of mobile call profiles and classify network usages accordingly, will be maintained at the geo-datacenter and will be made available to other geo-datacenters across the cloud. This also means that a datacenter may maintain many profiles that require more computational resources to preserve a growing data. One solution is to install additional persistent storage volumes that can be accessed by any server within the same cloud. However, it is necessary to allocate certain servers for data storage within each datacenter and allow all authorized cloud datacenters to access and use those storage servers as a common recourse for the operator.

The SDN controllers supports the creation of logic networks and impose relevant policies for traffic forwarding between connected virtual components and physical ports. The network resources and the initialization of virtual machines are defined by orchestrators through the visualization layer [8]. In our model, the basic NFV elements represent core network functionalities that may run for long times and need to have the necessary resources to scale in response to traffic changes. Therefore, preactive and proactive service monitoring will help to allocate the necessary computational resources to enable elasticity. This will scale the size of VMs or create new similar VMs based on information obtained from the hypervisor indicators or any extended telemetry services on the RAN. Generally, automating elasticity should be defined through certain policies and business models during the service initiation supported by fully automated platform features. These APIs are the key interfacing technology allow connecting various applications/modules to orchestrates for highly flexible control across datacenter. However, creating virtual resources and instantiating interfaces between cloud and the RAN raises the question whether orchestrators are the suitable technologies for provisioning network architecture and determine service fitness in tight NFV/RAN framework.

NETWORK PROVISIONING USING SOFTWARE-DEFINED DATACENTER

The binding of SDN and NFV continue to emerge as a new approach to facilitate higher levels of service management in cloud based networks. The SDN provides switching abstraction support to handle traffic across network sites without eliminating the network infrastructure. However, traffic forwarding is provided on an end-to-end rating policy and controllers cannot change hop-by-hop interactions. This means that a SDN controller layer lacks the necessary policies to actively conducting mid-route changes when starting traffic forwarding. Considering NFV, the hypervisor provides the foundation for cloud computing with features that enable creation and deletion of VMs/VNFs, virtualized resources, multi-tenant environments, and different types of high volume virtual NICs [5]. To this point, policies defines the life cycle transitions for VMs and do not influence the underlying SDN layer to configure interworking between separated networks. Therefore, there is a need to define the hierarchy between various SDNs to cascade basic policies from platform SDN to logic networks controllers.

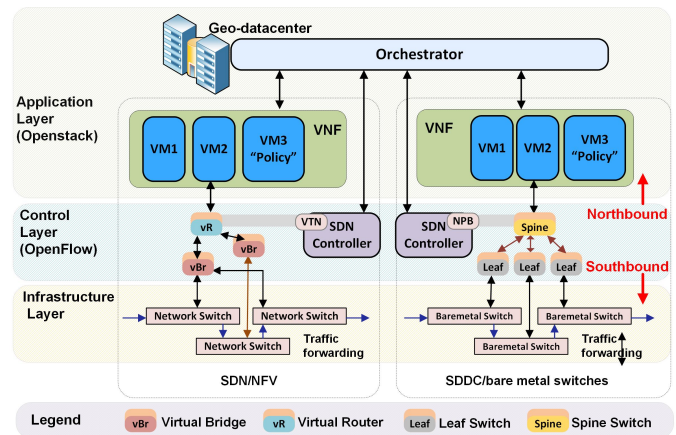


Fig. 2. The SDN functionalities in typical hypervisor vs. SDDC platform.

The SDN provides the ability of creating logic networks that interface various physical and virtual ports without impacting the underlying physical network. The SDN provides such functionality using virtual tenant controller (VTN) coordinator that allows to define networking for layer 2 and layer 3, regardless of physical network topology. The controller functionalities are enabled through API interfaces that also allow to set the flow entries for various Openflow switches. The SDN deploys a set of virtual bridges and routers that connect VMs to switches in compliance with ETSI model [3]. In such design, the network architecture will be entirely controlled by the NFV layer, allowing the execution of dynamic flow forwarding using a priority queue and/or capacity changes in the underlying network [13]. In our model, we assume that policies may also be deployed within VNFs to provision data forwarding to infrastructure especially when virtual networking to other VNFs/ports may be dynamically configured. Therefore, VMs can instantiate direct interfaces through network ports/gateways to access bare metal switches at different sites as long as they all interfaced to the same logic network, as shown in Fig. 2.

A datacenter that exploits SDN features to enable advent of controlling through virtualization layer is known as *Software-defined Datacenter (SDDC)*. The SDDC can be defined as a hardware-accelerated datacenter that can access all sorts of infrastructure appliances using SDN enhanced APIs. In this case, the SDN employs network packet broker (NPB) functionality that allows to create an overlaid network of leaf and spine virtual switches that connect physical ports and VMs, respectively. This indicates a new platform that provides sophisticated management of traffic forwarding between VMs and underlying virtual/physical network architecture. In a logic network, the datacenter is the host for vEPC/vIMS VNFs that provide the core network support to the network cluster in its local geographical location. Therefore, efficient traffic management may not be able to avoid the lack of hardware resources making SDDC a complementary additive value rather than a complete solution. Since the main concern for this article is the ability of individual vEPC/vIMS entities that are installed in separate VMs to handle operations with

very limited resources within a geo-datacenter, the following section explains a new service chaining model that leans more towards interconnected distributed clouds. It is also necessary to highlight that scaling resources using elasticity is only successful for clouds preserving a significant amount of resources to process arrival traffic and there still a need to incorporate clouds as one resources rather than segregated model [14]

CHAINING SERVICES BETWEEN GEO-DATACENTERS

In a cloud geo-datacenters model, calls between local UEs within the same cluster are handled by local geo-datacenter VNFs. Theoretically, each cluster/geo-datacenter combination is considered as a logic network since it consists of RAN and cloud backbone. The connectivity between various geo-datacenters allows the call detail records (CDRs) to be available through cloud to all datacenters. To setup a call between two UEs located at two different clusters, two geo-datacenters need to be involved in such operation. This can be unrealistic as those two clusters might belong to the same operator causing unnecessary duplicate resource and power consumption at different sites within the same operator [15].

We propose a new framework to chain vEPC/vIMS VNFs in geo-datacenters to avoid duplicated processes when connecting callers located at different network clusters. Consider Fig. 3, a call originating from UE#1 to UE#2 will be forwarded from geo-datacenter#1 to geo-datacenter#2. The procedure includes:

- 1) All access points are connected through Ethernet ports (eth ports) to the cloud. The capacity of those ports are variant and it depends on the capacity of each access point. All call initiated from UE#1 in Cluster#1 is sent to the geo-datacenter for further processing.
- 2) When the UE#2 location is announced through the control plane (C-plane) LTE positioning protocol (LPP) and secure user plane location (SUPL), the call request is sent from core network in geo-datacenter#1 to core network in geo-datacenter#2.
- 3) Using orchestrator predefined policies, core network in geo-datacenter#2 announces call forwarding from UE#2 to geo-datacenter#1. This operation is performed through SDN controller that updates the the Open vSwitch Database Management Protocol (OVSDB) server and editing the flows designations of the involved Open vSwitches (OVSS). The OVS are normally attached to the Ethernet ports that connect each access point to local datacenter.
- 4) The core network in geo-datacenter#1 will access the CDRs of UE#1 and UE#2 to confirm credits and start the call. There are three VNFs that perform connectivity to Wi-Fi, EPC, and IMS.
- 5) During the call, the core network services in geo-datacenter#2 will be blinded and wont process or access any CDRs for either UE#1 or EU#2.
- 6) This framework can also support group calls between multiple users who may be located at different clusters.

To enable highly configurable cloud platform that monitors traffic volumes prior to arrival to the geo-SDDC, the ETSI NFV standard [3] may need to include KPI manager at the RAN side. This assumes creating a new KPI monitoring APIs that can access the access points hardware managers and collect real-time data on resource availability for better assisting cloud platforms to know what type of traffic on the way before requests arrival. The new extended MANO that supports a KPI manager may be renamed to Enhanced-MANO (E-MANO) that expose all RAN obtained information to the orchestrator to help higher traffic monitoring and maybe control from hop-to-hop, as shown in Fig. 3.

PERFORMANCE EVALUATION

System Setup

A simulation testbed was created using two endpoint groups (EPGs) that emulate the RAN segment. Each EPG has a 1GPE Ethernet port that interfaces to a server operating single node Openstack setup. Those servers provide the vEPC/vIMS core functionalities using software installed VMs similar to the model shown in Fig. 3. The VNFs are instantiated using Heat Orchestration Templates (HOTs) using MANO VNF manager. Generally, the capacity of any VM to process a certain number of requests is impacted by cardinality between functional software installed inside the VM and the virtual resources allocated to that VM. As we use VoIP to examine performance, we identified the upper limit for call processing in call session control function (CSCF) VMs using aggression testing to determine the threshold for the maximum number of SIP requests before starting to reject calls. We identified four cases to examine the performance differences when two UEs call each other from different clusters. In case#1, we test a logic network model where each EPG is managed by a core VNF and calls between slices are managed by CSCFs to provide media streaming. In case#2, we repeat the same test for case#1 with elasticity enabled to bypass the restrictions imposed by CSCF software cardinality. Similarly, we repeat test cases #1 and #2 in test cases#3 and #4, respectively. In test case#3, we test chaining between cluster#1 and cluster#2 to facilitate a call between UE#1 and UE#2 using the model shown in Fig. 3. Finally in test case#4, we test case#3 while enabling elasticity for chaining model. We increase the number of VoIP requests between the two clusters by simultaneously increasing the number of UEs attached to each access point. Therefore, the number of requests, in our results, reflects the number of UEs attached to EPGs. The parameters of simulation are given in Table 1.

The main goal of the simulated scenarios is to analyze performance when processing calls in a clustered network model governed by geo-datacenters. Results are envisioned to help in developing a policy for future 5G deployments involving cloud architectures.

Test Results

In the following, we report performance results for the four test cases that we identified in the previous subsection using

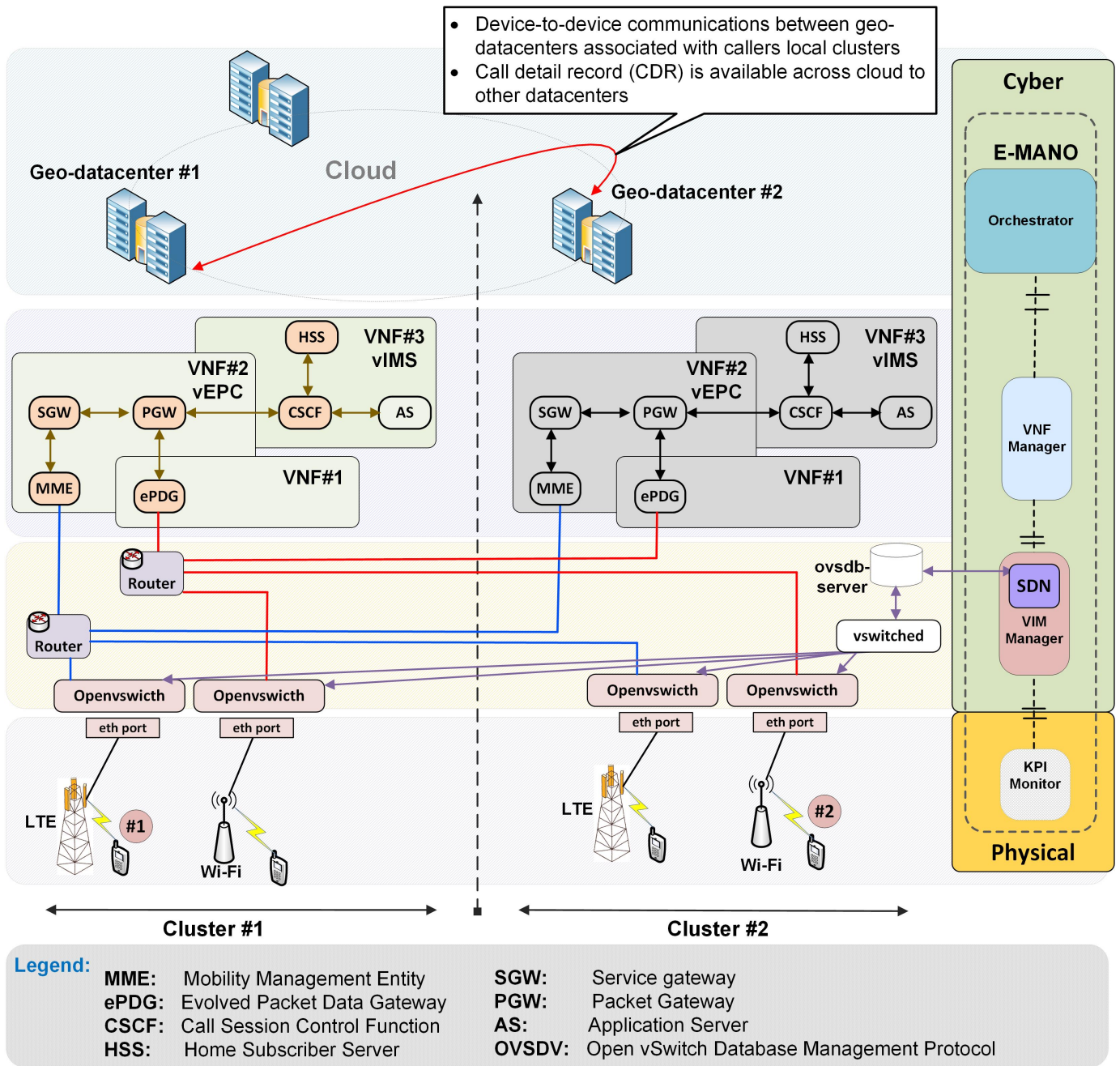


Fig. 3. Chaining services between different VNFs to perform multi-slice calls.

the number of VoIP requests. The overall trend for end-to-end latency performance figures, in Fig. 4, is increasing as the number of arrival VoIP requests increases. This increase in time delay happens because of the time consumed for processing the requests by individual vEPC/vIMS VMs in the geodatacenters. However, we notice major differences between test cases subject to the scheme used for call processing. For example, logic networks experience higher latency compared with chained networks scheme. The reason is that calls are processed individually in each geo-datacenter without considering the fact that both slices belong to the same mobile operator.

Therefore, higher latency is resulted from system processing model rather than actual call processing functions. Oppositely, a chained model will process requests using a single set of vEPC/vIMS VNFs leading to more efficient system design. Both logic and chained networks are hit by a significant increase in latency once VoIP requests reach 70 per second. This performance degradation is a result of limited resources available to CSCF when VM virtual CPU reaches 98% of its processing capability. When enabled, elasticity provides an automated features to scale resources allocated to VMs subject to the number of received VoIP requests. This means that

TABLE I
SIMULATION PARAMETER VALUES

Parameter	Value
RAN layout	2 EPGs, variable number of UEs
Packet inter-arrival time	10ms
Voice packet length	80 bytes
Application Layer	
Encoder Scheme	G.711 (silence)
Voice Frames Per Packet	1
Type of Service	Best Effort (0)
Signaling	SIP
Max. ACK Delay	0.2 msec
Max. ACK Segments	2
Fast Recovery	Reno
NFV Platform	
Openstack	RDO Mitaka release
Servers	2 Dell

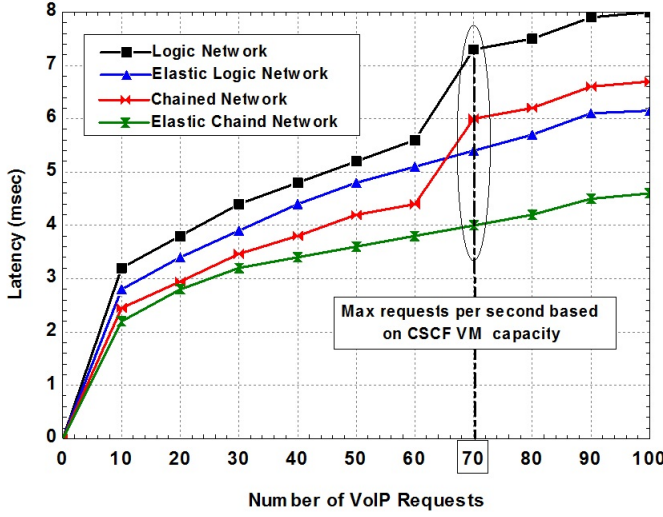


Fig. 4. End-to-end latency vs. number of VoIP requests.

CSCF VM flavor may change at any time during the scanning cycle of load received from the RAN.

The same performance can be also noticed in Fig. 5, which shows the total number of dropped calls versus the number of VoIP requests. The logic network calling scheme shows the highest percentage of dropped calls compared with all other test cases. The increase in the number of dropped calls results in higher number of retransmission attempts causing more disturbance to the whole system. Therefore, it is necessary to maintain lower number of dropped calls through using the NFV features such as elasticity that restabilize the whole system through steady performance and avoiding the bottleneck imposed by VMs resources. The chained network model shows the same behavior but with lower numbers of dropped calls. The VM resources shows a similar impact on performance characteristics as reflected with higher number of dropped calls. However, elasticity provides the necessary component to make chained network model the best probable solutions for such cloud based networks.

In summary, VNF environment provides the necessary features to instantiate interfaces between different entities and allocate resources on fly to meet user requirements. For net-

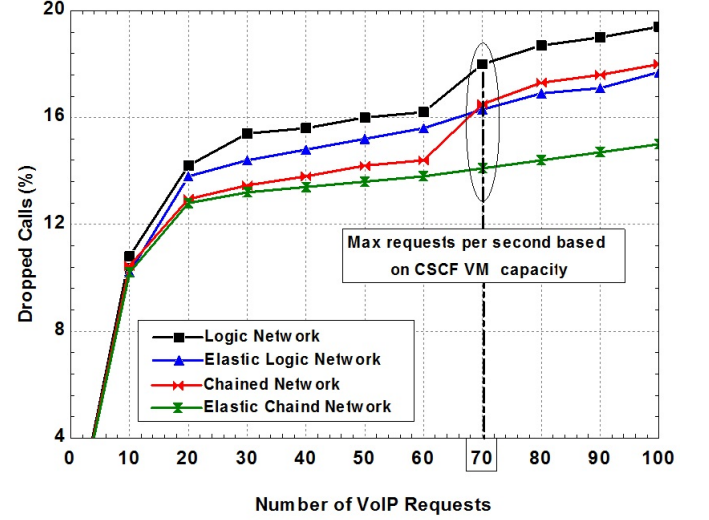


Fig. 5. Number of dropped calls vs. number of VoIP requests.

work planning engineers, NFV represents the tool to achieve resilient network that can meet the ultra-dense deployments of 5G era.

CONCLUSION

In this article, we presented various advantages that can be obtained from employing geo-datacenters to manage local network clusters as independent logic networks. We have shown how vEPC/vIMS VNFs created at geo-datacenters support interoperability between various wireless technologies such as VoIP calls between LTE and Wi-Fi. We also provided a new framework to manage calls between various slices using single virtual core network on the originating geo-datacenter. Simulation results show performance improvements when employing chained model compared with logic calling model. Tailoring this chaining process to NFV, we shown the advantages of enabling elasticity to avoid degradation in VM performance by removing limitations on using virtual resources. This article provided a long term vision for connecting different network clouds using dynamically instantiated logic networks using SDN enhanced features.

AUTHOR INFORMATION

Anwer Al-Dulaimi (anwer.al-dulaimi@exfo.com) received the Ph.D. degree in electronic and computer engineering from Brunel University, London, UK in 2012. Currently, he is a Technical Product Owner in Center of Excellence at EXFO, Toronto, Canada. His research interests include 5G networks, cloud networks, V2X and Internet of Things. He is the chair of IEEE 1932.1 Working Group "Standard for Licensed/Unlicensed Spectrum Interoperability in Wireless Mobile Network and IEEE Distinguished Lecturer.

Shahid Mumtaz has more than 10 years of wireless industry experience and is currently working as Senior Research Scientist at Instituto de Telecomunicaes Aveiro, Portugal. He

received his MSc and Ph.D. degrees in Electrical & Electronic Engineering from Blekinge Institute of Technology (BTH) Karlskrona, Sweden and University of Aveiro, Portugal. Dr. Shahid has more than 150 publications in international conferences, journal papers, and book chapters. He is IEEE senior member and ACM distinguished speaker.

Saba Al-Rubaye received her Ph.D. degree in Electrical and Electronic Engineering from Brunel University London, United Kingdom. Currently, she is an associate professor in the Manufacturing Informatics Centre at Cranfield University. She led and managed several research/industrial projects from conception to completion in the field of secure manufacturing, vehicle to everything, grid modernization & digitization, industrial Internet of things, SDN control systems, cyber-physical system, and networking & communications. She is registered as a Chartered Engineer (CEng) by the British Engineering Council. She is recognized as Associate Fellow of the British Higher Education Academy (AFHEA) in the UK, and a Senior Member of IEEE.

Siming Zhang received her Ph.D. degree in wireless communications from the University of Bristol (UK) in 2016. She currently works at China Mobile Research Institute (Beijing). She is one of the co-leads on NGMN Trial and Testing Initiative. She is also co-leading WG1 of the Wireless Artificial Intelligence Alliance. She is the Associate Managing Editor of the IEEE Future Networks Tech Focus. Her current research interests include PHY layer design on 5G/6G mobile networks, especially on channel modeling and prototyping, wireless big data analysis, and ML/AI application in the RAN domain.

Chih-Lin I is CMCC Chief Scientist of Wireless Technologies. She received Ph.D. EE from Stanford University. She has won IEEE ComSoc Stephen Rice Prize, Fred W. Ellersick Prize, and 2015 Industrial Innovation Award. She is the Chair of FuTURE 5G SIG, the Chair of WAIA Executive Committee, the Chair of O-RAN TSC, an Executive Board Member of GreenTouch, a Network Operator Council Founding Member of ETSI NFV, a Steering Board Member and Vice Chair of WWRF, a Steering Committee member and the Publication Chair of IEEE 5G Initiative, a member of IEEE ComSoc SDB, SPC, and CSCN-SC, and a Scientific Advisory Board Member of Singapore NRF. Her current research interests center around "From Green & Soft to Open & Smart".

- [4] J. G. Herrera and J. F. Botero, "Resource Allocation in NFV: A Comprehensive Survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, Sept 2016.
- [5] Q. Duan, N. Ansari, and M. Toy, "Software-defined Network Virtualization: an Architectural Framework for Integrating SDN and NFV for Service Provisioning in Future Networks," *IEEE Network*, vol. 30, no. 5, pp. 10–18, September 2016.
- [6] J. Liu, Z. Jiang, N. Kato, O. Akashi, and A. Takahara, "Reliability Evaluation for NFV Deployment of Future Mobile Broadband Networks," *IEEE Wireless Communications*, vol. 23, no. 3, pp. 90–96, June 2016.
- [7] N. Kumar, S. Zeadally, and S. C. Misra, "Mobile Cloud Networking for Efficient Energy Management in Smart Grid Cyber-physical Systems," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 100–108, October 2016.
- [8] B. Naudts, W. Tavernier, S. Verbrugge, D. Colle, and M. Pickavet, "Deploying SDN and NFV at the Speed of Innovation: Toward a New Bond between Standards Development Organizations, Industry Fora, and Open-source Software Projects," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 46–53, March 2016.
- [9] Z. Zhou, S. Mumtaz, K. M. S. Huq, A. Al-Dulaimi, K. Chandra, and J. Rodriguez, "Cloud Miracles: Heterogeneous Cloud RAN for Fair Coexistence of LTE-U and Wi-Fi in Ultra Dense 5G Networks," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 64–71, June 2018.
- [10] A. Al-Dulaimi, S. Al-Rubaye, J. Cosmas, and A. Anpalagan, "Planning of Ultra-Dense Wireless Networks," *IEEE Network*, pp. 12–18, March/April 2017.
- [11] "3GPP TS 36.300 v14.1.0 (2016-12), Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2."
- [12] S. Cash, V. Jain, L. Jiang, A. Karve, J. Kidambi, M. Lyons, T. Mathews, S. Mullen, M. Mulsow, and N. Patel, "Managed Infrastructure with IBM Cloud Openstack Services," *IBM Journal of Research and Development*, vol. 60, no. 2-3, pp. 6:1–6:12, March 2016.
- [13] W. Li, D. Guo, A. X. Liu, K. Li, H. Qi, S. Guo, A. Munir, and X. Tao, "Coman: Managing Bandwidth Across Computing Frameworks in Multiplexed Datacenters," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 5, pp. 1013–1029, May 2018.
- [14] G. Premsankar, K. Ahokas, and S. Luukkainen, "Design and Implementation of a Distributed Mobility Management Entity on Openstack," in *2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom)*, Nov 2015, pp. 487–490.
- [15] A. Al-Dulaimi, S. Al-Rubaye, and Q. Ni, "Energy Efficiency using Cloud Management of LTE Networks Employing Fronthaul and Virtualized Baseband Processing Pool," *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1–1, 2016.

REFERENCES

- [1] Y. Nam, S. Song, and J. M. Chung, "Clustered NFV Service Chaining Optimization in Mobile Edge Clouds," *IEEE Communications Letters*, vol. 21, no. 2, pp. 350–353, Feb 2017.
- [2] Y. Jia, C. Wu, Z. Li, F. Le, and A. Liu, "Online Scaling of NFV Service Chains Across Geo-Distributed Datacenters," *IEEE/ACM Transactions on Networking*, vol. 26, no. 2, pp. 699–710, April 2018.
- [3] "ETSI GS NFV 002 v1.1.1 (2013-10), Network Functions Virtualisation (NFV); Architectural Framework."

2019-02-21

A framework of network connectivity management in multi-clouds infrastructure

Al-Dulaimi, Anwer

IEEE

Al-Dulaimi A, Mumtaz S, Al-Rubaye S, Zhang S & Chin-Lin I (2019) A framework of network connectivity management in multi-clouds infrastructure, IEEE Wireless Communications, Volume 26, Issue 3, 2019, pp. 104-110

<https://doi.org/10.1109/MWC.2019.1800166>

Downloaded from Cranfield Library Services E-Repository