IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 1, 2018, accepted March 12, 2018, date of publication March 23, 2018, date of current version April 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2818751

Service-Aware Multi-Resource Allocation in Software-Defined Next Generation Cellular Networks

OMER NARMANLIOGLU[1], (Student Member, IEEE), ENGIN ZEYDAN[2], (Member, IEEE), AND SUAYB S. ARSLAN[3], (Member, IEEE)

[1]P.I. Works, 34912 Istanbul, Turkey
[2]Türk Telekom Labs, 34889 Istanbul, Turkey
[3]MEF University, 34396 Istanbul, Turkey

Corresponding author: Omer Narmanlioglu (omer.narmanlioglu@piworks.net)

ABSTRACT Network slicing is one of the major solutions needed to meet the requirements of next generation cellular networks, under one common network infrastructure, in supporting multiple vertical services provided by mobile network operators. Network slicing makes one shared physical network infrastructure appear as multiple logically isolated virtual networks dedicated to different service types where each Network Slice (NS) benefits from on-demand allocated resources. Typically, the available resources distributed among NSs are correlated and one needs to allocate them judiciously in order to guarantee the service, MNO, and overall system qualities. In this paper, we consider a joint resource allocation strategy that weights the significance of the resources per a given NS by leveraging the correlation structure of different quality-of-service (QoS) requirements of the services. After defining the joint resource allocation problem including the correlation structure, we propose three novel scheduling mechanisms that allocate available network resources to the generated NSs based on different type of services with different QoS requirements. Performance of the proposed schedulers are then investigated through Monte-Carlo simulations and compared with each other as well as against a traditional max-min fairness algorithm benchmark. The results reveal that our schedulers, which have different complexities, outperform the benchmark traditional method in terms of service-based and overall satisfaction ratios, while achieving different fairness index levels.

INDEX TERMS Resource allocation, SDN, vertical industries, network slicing, MNOs.

I. INTRODUCTION

Telecommunication technologies have been transformational in providing new and advanced ways of enabling end-to-end communication services. 5G technology is envisioned to provide enhanced services, connecting new industries and empowering new user experiences for the decades to come. In addition to providing high capacity and increased data rates, the 5G era is expected to revolutionize the ways applications communicate based on a flexible and ubiquitous 5G infrastructure. This infrastructure is required to promote: Optimization of scarce resources (e.g. frequency spectrum, Radio Access Network (RAN)/transmission/core equipment); Reduce capital expenditure (CapEx) and operating expenditure (OpEx); support the Machine Type Communications (MTC) services and the Internet of things (IoT), while seamlessly integrating both IT and telco domains. 5G requirements (i.e. low latency, high capacity, performance and spectrum access) lead to a novel mobile network architecture which benefits from Software Defined Networking (SDN), Network Function Virtualization (NFV), Cloud RAN (C-RAN), and Multi-Access Edge Computing (MEC) technologies and provides the means to support the expected higher service diversity, performance, flexible deployments, and network slicing for Mobile Network Operators (MNOs).

The requirements of 5G architecture necessitate efficient resource management, enhanced smart connectivity, and delivery of a flexible architecture. Tools to enable these solutions include open platforms (that can provide integrated network services provided by third parties), intelligent Operation, Administration and Management (OAM) for autonomous fault management, and modular services for flexibe on-demand deployment. The means to optimize the network operation, and thus facilitate dependable service delivery for the vertical stakeholders. This has got vital

importance which creates novel-value-added services for telecommunication providers. 5G networks are driven by the flexibility and programmability of all network domains including RANs, fronthaul & backhaul transmission networks, and core networks located at cloud environments and/or network edges. Hence, network and compute visualization, slicing, NFV, programmability and softwarization technologies are required across all domains.

Application of 5G technology into the vertical markets is beginning to develop solutions based on each vertical industry's specific needs. In particular, use cases from automotive, transport and logistics, finance, health and wellness, smart cities, agriculture industries should be targeted jointly with the relevant telecommunication capabilities [1]–[4]. For example, the water industry requires ubiquitous connectivity, network reliability and cost reduction delivered by also heterogeneous Massive MTC (mMTC) support, industrial applications need to satisfy stringent latency, throughput, and reliability requirements, whilst other application types may live with longer delays and intermittent connectivity, but require minimal energy consumption.

A fundamental enabling technology of 5G is network slicing which makes one shared physical network infrastructure appear as multiple logically isolated virtual networks dedicated to different service types, and where each Network Slice (NS) benefits from guaranteed and on-demand allocated resources. The aim of network slicing is to provide end-to-end level partitioning of the physical network while allowing traffic grouping, tenant isolation and resource configuration at the macro level. Network slicing and virtualization technologies can offer harder guarantees in terms of availability of telecommunications systems. Thanks to network slicing, an MNO can split its physical resources into multiple logical slices and lease each to interested vertical industries. For instance, an energy utility company producing electricity may lease a long term NS for the reliable connectivity of its smart grid infrastructure comprising meters, sensors, controllers, and so on. Conversely, a big event organizer may lease a short term lease NS with streaming ultra-high definition (UHD) video and voice-over-IP (VoIP) connectivity to support concerts, sports events, etc. Network slicing and virtualization can also lead to CapEx/OpEx reduction due to easier management and re-utilization of resources.

A. RELATED WORK

Due to the excitement of business opportunities foreseen in network slicing, Standards Developing Organisations (SDOs) such as the 3rd Generation Partnership Project (3GPP) [5]–[7], European Telecommunications Standards Institute (ETSI) NFV [8], Institute of Electrical and Electronics Engineers (IEEE) [9], and other industry and open source communities have generated lots of related technical definition materials. 3GPP has initiated an overall OAM framework for Infrastructure Providers (InPs) in order to manage the slices as part of its virtualized NFV network. Moreover, signalling strategies and procedures between user equipment (UE) and network components is also being discussed under SA-2 specifications [5]–[7]. ETSI NFV has issued a white paper on network slicing, prioritizing NFV for 5G systems [8]. The IEEE's perspective on network slicing has been published in a recent 5G Roadmap whitepaper [9].

There are various works on network slicing in the context of mobile network infrastructure [10]–[20]. Algorithmic aspects of network slicing that describes the challenges of introducing slicing in future wireless networks is described in [10]. End-to-end network slicing for 5G mobile networks, with three representative use case scenarios of Ultra Reliability and Low Latency Communications (URLLC), mMTC, and enhanced Mobile Broadband (eMBB), have been studied in [11]. A NS architecture towards 5G communications, and its demonstration using state-of-the art techniques, have been shown in [12]. Orchestration and activation of NSs that cover the entire life cycle in a cloud-native environment is demonstrated in [13]. For multi-tenancy support, 5G crosshaul network slicing and the corresponding architecture that focuses on transport network enhancements are proposed in [14]. An important contribution that considers network slicing deployment by integrating both SDN and NFV technologies into the referenced architecture is studied in [15]. In the past, joint resource allocations involving interference and power consumption while providing different quality-of-service (QoS) requirements have been discussed [16]. In [17], a sub-channel allocation is considered in addition to efficient power utilization. Finally, network virtualization in the context of mobile cellular networks for multi-MNOs were investigated in our previous works [18], [19].

In this paper our focus is concentrated on solving slice optimization problems in order to maximize the satisfaction ratios of MNOs while leveraging the benefits of the SDN-based network slicing architecture. Resource allocation and management have been further investigated within [21]–[28]. In [21], resource allocation and isolation in virtualized network environments are investigated and challenges are presented. The authors in [22] propose a joint base station (BS) assignment and resource block (RB) allocation in conjunction with power allocation mechanism for UEs associated with different slices that have minimum data rate constraint in a virtualized wireless network. Similarly, the authors in [23] propose a dynamic resource allocation method for UEs associated with different slices in a C-RAN environment in order to maximize system data rate under the minimum data rate constraint assumption for each slice. In [24] and [25] a novel cloud-based radio over fiber network architecture including optimization of radio frequency, optical spectrum, and baseband unit (BBU) processing resources in order to maximize radio coverage under the consideration of QoS requirement is proposed. LeAnh *et al.* [26] propose a resource allocation method for NSs in conjunction with transmit power optimization under the consideration of backhaul network limitations. In [27], resource orchestration posed as a multi-objective optimization problem in terms of load balance, energy cost and resource consumption is investigated. In [28], a novel

two-stage resource allocation algorithm for C-RAN architecture is proposed with the limitations of fronthaul network and cloud computation limits in order to maximize the profits of both MNOs and infrastructure owner. However, none of the above works consider the correlation and joint optimization of multi-resource allocations in a sliced network architecture between different MNOs/slices while exploiting the benefits of SDN and network virtualization. In this paper, we propose different scheduling algorithms along with an SDN-based network slicing framework that can allow immediate resource sharing and slicing based on demands of services, while taking into account the guarantees or Service Level Agreements (SLAs) with multiple MNOs and vertical industries. The contributions of this paper are three fold: (i) We propose a 5G system architecture design targeted for vertical sectors that can improve the performance of communication/application solutions within the existing 5G framework. (ii) We showcase the performance improvements compared to traditional methods (such as Max-Min Fairness (MMF)) in terms of service-satisfied ratio and ratio of allocated resource to demand levels, by introducing carefully designed cost functions and a weighted joint multi-resource allocation process. (iii) Service awareness of resources are modeled into our joint optimization methodology using weights which are calculated by the Analytical Hierarchical Process (AHP) method.

The rest of the paper is organized as follows. In Section II, we describe an SDN-based network slicing framework for 5G networks covering various 5G use cases for MNOs/vertical sectors. In Section III, we describe the joint resource allocation problem as a result of multiple MNOs and services with different QoS requirements. In Section IV, we study the optimization problem and propose three scheduling solutions in 5G-enabled cellular networks together with the AHP tool to effectively characterize service awareness of resources. We provide numerical comparison results of the studied joint optimization solutions in Section V, and we finally conclude the paper in Section VI.

Notation: $[.]^{\mathrm{T}}$ and $[.]^{\mathrm{H}}$ denote transpose and Hermitian operations. tanh(.) is Hyperbolic Tangent function. $\lceil . \rceil$ is the *ceiling* function. The sets are denoted by upper case calligraphic symbols. The scalars are represented by regular symbols and vectors are denoted by bold face regular letters, e.g., $\mathbf{x}$ where $\boldsymbol{x}(k)$ denotes the k-th element of $\mathbf{x}$.

II. SYSTEM ARCHITECTURE

In order to reliably handle data traffic and build a robust architecture, virtualization techniques utilizing principles of SDN and NFV can be considered. In this paper, we propose a multi-MNO resource sharing framework for heterogeneous network (HetNet) deployments using an SDN-based network slicing concept. The overall proposed architecture for the SDN-based network slicing, covering vertical use cases and hierarchical SDN controllers is depicted in Fig. 1 where two MNOs, each providing three different services, are sharing the same RAN, transmission network and core network infrastructures via a virtualization controller and several C-RAN controllers owned by an InP. In the proposed architecture, the virtualization controller, which is managed by Virtual Infrastructure Manager (VIM), stands between the MNO's applications/controllers and the 5G network infrastructure, whereas *virtual SDN controllers (vSDN-C)* of each MNO, communicate with this virtualization controller via VIM and the *Slice Resource Orchestrator and Manager*. Each MNO can control their own NSs via their own virtual Mobility Management Entity (vMME), virtual Home Subscriber Station (vHSS) and virtual Policy and Charging Rules Function (vPCRF) instantiated as Virtual Network Functions (VNFs). Therefore, the capabilities in NS are not restricted to data plane, i.e., they can also provide control plane relevant capabilities.

The architecture given in Fig. 1 shows the interactions between an InP and multiple MNOs (e.g. Mobile Virtual Network Operators (MVNOs), with over-the-top (OTT) services) as an extension of the system architecture defined in [18]. The main role of the InP is to provide the resources while ensuring co-existence of different NS. The key elements, and their corresponding architectural descriptions are as follows:

Virtualization Controller and C-RAN Controllers enable end-to-end network control while supplying abstraction towards higher layers (e.g. service layer). The virtualization controller and MNO specific SDN controllers enable control operation between each MNO's applications and the 5G softwarized network infrastructure underneath. The slice management of each MNO can be performed similar to physical network resource management, once a slice is assigned to an MNO. For this reason, the InP exposes some control functionalities to MNOs in order to allow them to control their own dedicated slices. Various applications and services of each MNO are served by different slices based on the requirements of the use cases such as factory automation, connected vehicles, tactile internet, media and entertainment, etc.

Slice Resource Orchestrator and Manager provides life-cycle management for each individual slice (creation, activation, deletion). Slice resource orchestrator and manager is responsible for 5G infrastructure management. This resource orchestrator and manager ensures that network resource allocations are efficient, flexible and adaptive among MNOs. The allocation is performed based on various requirements of vertical sectors and services of MNOs. The NS requests arriving from each vertical to MNO, as well as from MNO services into the InP, need to create virtual networks over the physical network infrastructure.

Infrastructure hosts the physical and virtual resources. The infrastructure comprises communication, storage and other vertical specific (e.g. sensing, actuating) equipment provided by sensor networks, access/core networks and cloud infrastructures. All those resources are programmable and can be utilized by controller elements.

For the slice management life-cycle that is controlled by a slice resource orchestrator and manager, (i) as the first

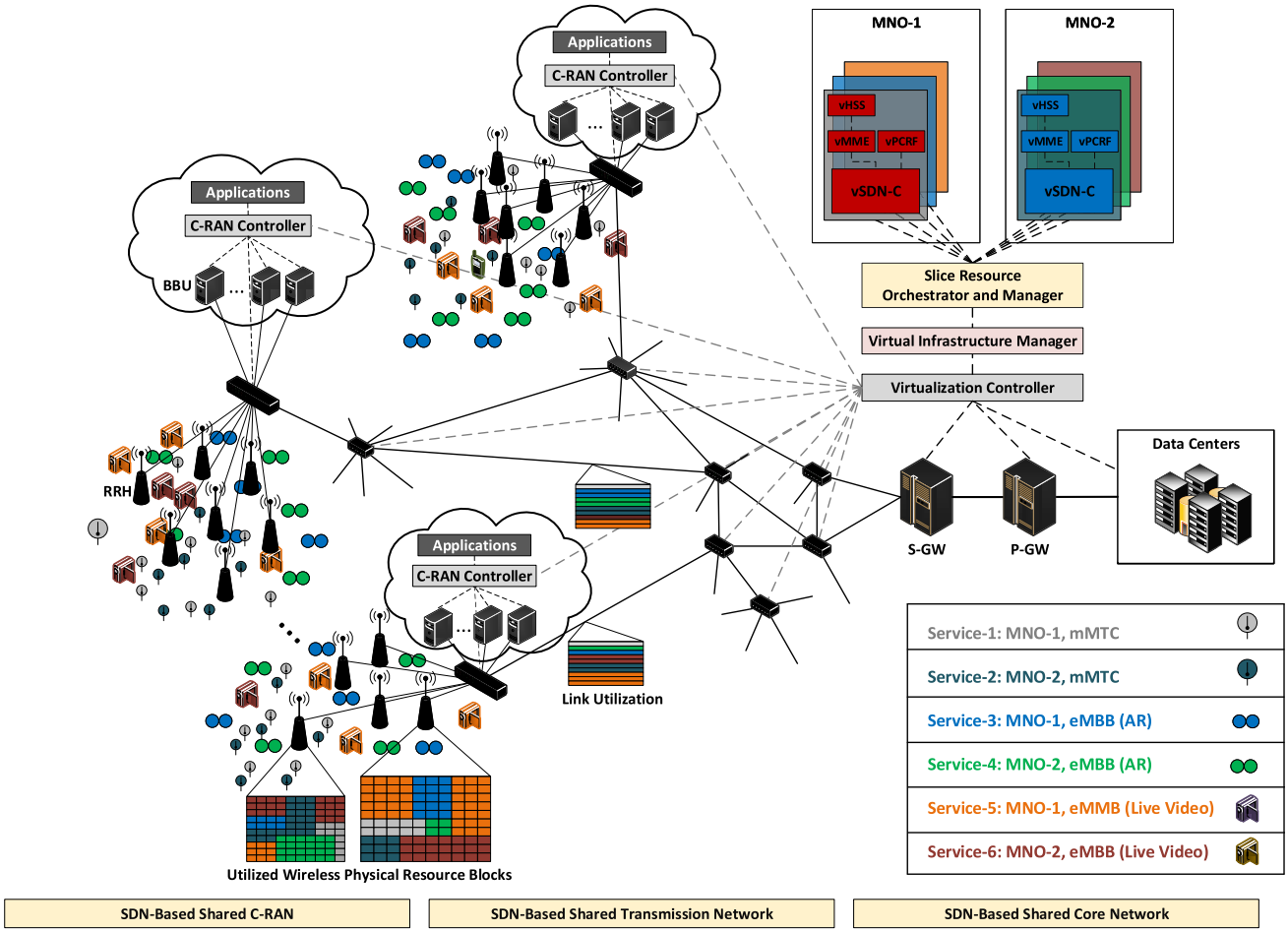

FIGURE 1. SDN-based network slicing framework for 5G networks covering various 5G use cases for MNOs/vertical sectors.

step, MNOs perform a slice requirement plan, taking into account the demands of their vertical coordinators as well as the demands of their subscribers. In this step, MNOs reserve compute, storage and network resources from an InP similar to cloud reservations of resources. (ii) In the second step, an InP manages slices based on service demands. (iii) In the final step, MNOs perform intra-slice network management and an InP performs inter-slice orchestration between MNOs. For optimal network capacity planning, dynamic scheduling of resources can help improve both costs and coverage.

A. 5G APPLICATION USE CASES

The architecture is designed to provide the technical means to enable and support competing traffic types of MNOs servicing for different vertical use cases. On top of this architecture, MNOs can obtain differentiated SLA requirements based on different dimensions such as bandwidth, latency, reliability, etc. The use cases defined in this architecture base their definitions on 5G Infrastructure Public Private Partnership (5G-PPP)'s and different SDOs' three main use cases for 5G NSs [1]:

i) URLLC applications: URLLC is one generic mode of MTC which can be used for mission-critical applications such as reliable remote robotic actions, remote surgery or coordination among vehicles. With the new applications that URLLC relies on, ultra-reliable wireless connectivity can be enabled with 99.999% availability for products and systems used in, for example Industry 4.0. This use case investigates challenges for enabling high reliability with guaranteed latency. For these types of applications, end-to-end latency should also be secured, including the communication setup. For some applications such as alarms, robotic motion control, the feedback loop of the system should be less than one milliseconds which cannot be met with the current 3GPP Long Term Evolution (LTE) infrastructure. On the other hand, ensuring high reliability can only be obtained at the expense of latency and/or capacity cost. For example, connected and autonomous vehicles are beginning to become one of the most significant use case of the increasingly connected world. However, due to nature of mobility, it is not easy to support real time reliability, safety and stability. For vehicular network applications, network slicing in combination with MEC can decrease the latency and increase capacity, especially

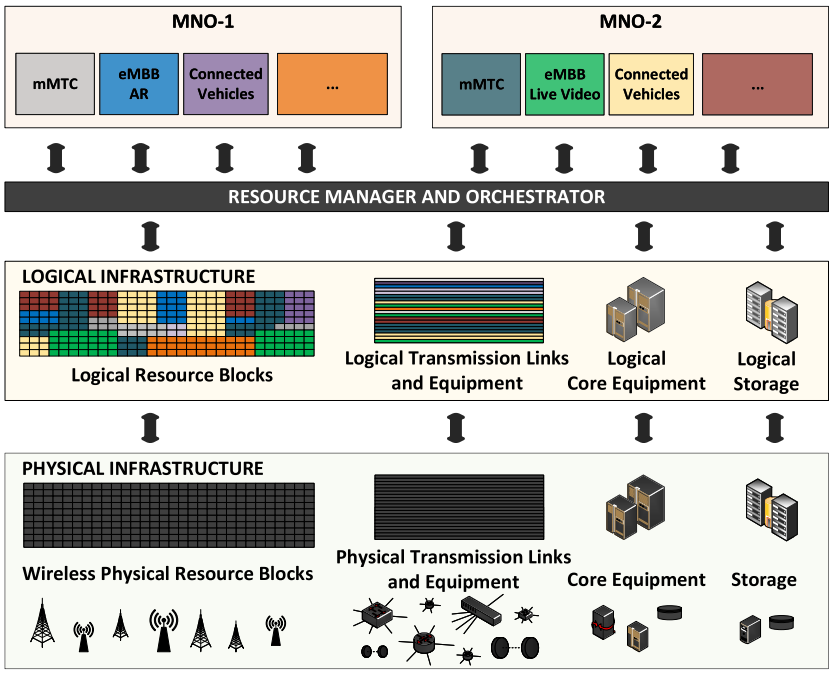

FIGURE 2. Network sharing and allocation of resources to multiple MNOs each having different services.

when VNFs are deployed to the edge and multi-level priority management is imposed in the network. MEC can also provide higher battery lifetime of devices due to offloading, and reliability due to potential redundancy.

ii) eMBB applications: Another important use case for 5G is the eMBB scenario. 5G new radio Phase 1 (Release 15, 2018) is mostly focusing on eMBB. This standard slice type is expected to provide high data rate and high traffic densities. For example, streaming delivery of 4K UHD live video and other content-driven applications (including other bandwidth demanding applications such as Augmented Reality (AR)/Virtual Reality (VR)) with the assured user quality-of-experience (QoE), or fast large file downloads are the application areas of this use case. More advanced media delivery services including 8K resolution, 6-degree of freedom (6DoF) video are also on the horizon which require on the order of gigabits per second data transfer rates [29].

iii) mMTC applications: This use case investigates challenges for ubiquitous coverage and massive connection support for delay-tolerant traffic. In 3GPP standardization, massive deployment of connected devices is left for 3GPP Release 16 (2020). Typically, deployed mMTC devices can be located in a remote area with no cellular connection. Therefore, ubiquitous coverage should include remote and rural areas. For instances, smart meter networks deployed throughout the infrastructure can save physical resources and reduce labour costs. Most of the mMTC devices are expected to transmit only a few bytes of data for long periods of time. Hence, capacity scaling as well as spectrum utilization are important for mMTC applications. LoRa, Sigfox, Narrowband IoT (NB-IoT) and LTE Cat-M1 (LTE-M) are some of the prominent technologies in the context of low-power wide area network (LPWAN) [30]. SigFox [31] and LoRa [32] are more suited to long distance, low bit rate short bursty traffic used in unlicensed bands (e.g. 868 MHz in Europe). However, high loads of traffic can saturate the system performance resulting in increased latency, high power consumption and eventually potential loss of data which in turn affects reliability. On the other hand, 3GPP LTE CAT-M1 (or enhanced MTC (eMTC)) and LTE Cat NB1 (NB-IoT) have been designed so as to provide better coverage, energy savings and longer battery life, lower device cost and a higher node density for massive connectivity. LTE CAT-M1 has reduced bandwidth (1.4 MHz), single antenna UE, resulting in decreased UE cost, and extended coverage. It also supports the ability to handover from cell to cell, which is beneficial for mobile services. NB-IoT, has bandwidth further reduced to 180 kHz, with half-duplex mode option, and single tone (15 kHz) transmission. This results in an additional extended coverage as compared with LTE CAT-M1, providing further reduction of UE cost and a battery lifetime up to 10 years. However, it is limited by cell re-selection mechanism and therefore applicable only for fixed applications.

B. MULTI-SLICE RESOURCE MANAGEMENT

We consider a multi-slice resource sharing mechanism, as shown in Fig. 2. Each slice is assumed to serve multiple UEs which are spatially distributed. Furthermore, each slice connects to the distributed virtualization controller which is responsible for assigning resources from a common pool

to the slices. Based on the SLAs, the virtualization controller exchanges messages with NFV Slice Management and Orchestration (MANO) and Operations Support System (OSS)/Business Support System (BSS) components to define a mutual agreement for spectrum sharing policy. We study a resource allocation problem where the virtualization controller allocates several resources (e.g., RBs, virtual baseband units (vBBUs), transmission link capacities, core network equipment (NE), storage) to a slice based on the demands and guarantees of the slice itself.

TABLE 1. The QoS requirements for different use cases [34].

Use case	Throughput	Latency	Reliability	Storage
mMTC	250 kbps	<10 sec	-	No
URLLC (CCAS)	<5 Mbps	10 msec	10^{-5}	No
eMBB (4K Live Video)	40 Mbps	500 msec	-	Yes
eMBB (AR)	100 Mbps	100 msec	10^{-2}	No
SDS (HPC)	2 Gbps	$1-2$ msec	10^{-7}	No
SDS (Amazon Glacier)	150 Mbps	$1-5$ min	10^{-9}	Yes

We assume that there can be many NSs of each MNO. Each slice represents a corresponding SLA signed with the MNO. Therefore, each slice corresponds to a different use case with different QoS requirements (e.g. as presented in Table 1). For example, the first NS can represent a URLLC use case for cooperative collision avoidance system (CCAS) in a Vehicular ad-hoc network (VANET) for a given UE set of MNO-k, whereas another NS with eMBB use case of AR should give another guaranteed SLA for a given UE set of MNO-k [33]. CCAS is a life-saving use case of URLLC and therefore it needs ultra-high reliability with 10^{-5} with lower delay tolerance [34]. On the other hand, AR and 4K live video of eMBB use case for entertainment purposes have higher delay tolerance than CCAS and require higher throughput.

III. SYSTEM MODEL AND PROBLEM DESCRIPTION

We assume a network region including multiple-MNOs that serve their UEs using different applications with different QoS and SLA requirements. Fig. 1 illustrates the considered particular network region where hardware resources and UEs are spatially distributed. We assume orthogonal frequency division multiple access (OFDMA) scheme where each BS's carrier (also called as *cell* with a pre-defined bandwidth amount such as 5 MHz, 10 MHz, 20 MHz, etc.) can only serve to one UE with a set of sub-carriers in a specific time slot (known as RB in LTE domain). It should be noted that a given BS can contain multiple cells operating at different carrier frequencies (with the same or different amount of bandwidth) in the same sector, and it has more than one sector in which same carrier frequencies are utilized, typically known as *unity frequency reuse*.

A. SYSTEM MODEL

Consider a set of multiple MNOs given by $\mathcal{K} = \{1, 2, \ldots, K\}$ with K MNOs. Let the set of NSs in the available infrastructure is given by $\mathcal{S} = \{1, 2, \ldots, S\}$ with S NSs. Let $\mathcal{S}_k \subset \mathcal{S}$ be the set of NSs assigned to MNO-k with S_k NSs. Also let the set of NEs used by a MNO-k for a given NS-i be given by $\mathcal{F}_k^i = \{1, 2, \ldots, F_k^i\}$ with F_k^i NEs where NEs can be cells, storage elements, gateways, core network elements, etc. It should be noted that each MNO can be allocated with different number of NEs for a given NS. Moreover, let $\mathcal{F} = \cup_{k\in\mathcal{K}} \cup_{i\in\mathcal{S}_k} \mathcal{F}_k^i$ be the set of all NEs. The set of UEs utilizing NE-$f \in \mathcal{F}_k^i$ are given by $\mathcal{U}_k^{i,f}$ with $U_k^{i,f}$ UEs. For a given NE-f in $f \in \mathcal{F}_k^i$, we denote UE-$u \in \mathcal{U}_k^{i,f}$ as its associated UE. Let $\mathcal{B}_{k,j}^{i,f}$ be the set of resources available at NE-$f \in \mathcal{F}_k^i$ with $B_{k,j}^{i,f}$ resources where $j \in \{0, 1, 2, \ldots, R-1\}$ denotes the type of resource and R is number of different types of resources (e.g. storage available at data centers, bandwidth available at cells, link capacities available at transmission network equipment) and denote resource-$b_j \in \mathcal{B}_{k,j}^{i,f}$ as the assigned resource. In order to meet the requirements of UEs, the NE-$f \in \mathcal{F}_k^i$ can select a subset of the resources in $\mathcal{B}_{k,j}^{i,f}$.

It should be noted that after resource allocation to the service-based slices is done, conventional scheduling algorithm (e.g., proportional fairness, round robin etc.) is performed in order to share the allocated resources to their relevant UEs with relatively smaller periods with respect to service-based allocations.

B. PROBLEM DESCRIPTION

In this paper, we consider an example use case scenario that focuses on providing resource allocation capabilities for distributing NSs created dynamically based on service demands inside the given network infrastructure. As of today, network infrastructures consist of heterogeneous and proprietary sets of network equipment. Therefore, it is expected that major paradigms such as SDN and NFV will shift the mobile infrastructure from configurable networks into programmable networks, which will facilitate dynamic resource allocation capabilities. Our scenario addresses the need for dynamic allocation of resources and demands for MNOs on the network infrastructure. One of the main innovations of our scenario is the increased service satisfaction ratios that allow reactive actions to be taken to allocate resources to services based on the demands of MNOs and their corresponding vertical industries. One of the technical challenges is to provide an intelligent allocation of resources in order to achieve better service satisfaction rates while considering the available and guaranteed resources as well as demands. Additionally, the solution should provide a quick and dynamic response in order to guarantee NS SLAs of MNOs for their customers (e.g. vertical industries, subscribers, etc) in a cost-efficient way. This paper assumes not only a single resource demand and the satisfaction of a service, but also considers simultaneous demands for multiple resources (such as bandwidth and storage) that can be correlated with each other. The general flowchart of the described use case scenario is summarized in Fig. 3.

In this subsection, we shall define our multi-resource allocation problem in which resource capacities are not randomly described or determined. Individual resource allocation

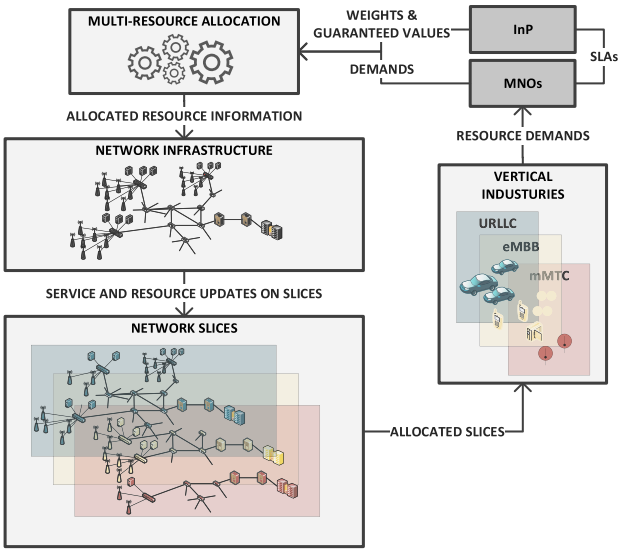

FIGURE 3. The flow of the proposed scenario.

problem is defined based on a set of constraints and a generic objective function to be minimized. Conventionally, a fixed amount of scalar resource, usually called *capacity*, must be shared among $M = S_k$ users (users represent the number of NSs in our context) based on their respective resource demands. Also, in a number of applications, users are guaranteed a particular amount of resource due to their SLA. There are two approaches to the resource allocation problem: (i) a reasonably defined objective/cost function is optimized based on a set of constraints or (ii) a solution to the optimization problem is seeked using an iterative algorithm that satisfies a qualitative measure such as a fairness criterion.

In a myriad of real applications, more than one resource is shared among M NSs simultaneously and the allocation is performed independent of each other. Plus, some of the resources such as bandwidth can be reused whereas other resources cannot. On the other hand, it is observed that the requested capacities of these resources are related as seen in MNOs' SLA offerings. In other words, neither the resources that are offered nor the users sharing the resources are determined independently. For instance, a high-rate link connection usually comes with low or medium latency guarantees. Additionally, users with higher bandwidth assignment are allowed to make more connections and can cache more data (for high speed access) than the users with lower bandwidth requirements whereby MNO subscriber tariff plans are also in accordance with this observation [35]. This is often due to maximizing the user satisfaction with the aim of meeting the pre-agreed SLA requirements.

Let us assume R resources with capacities $\mathbf{C} = \{C_j \geq 0, 0 \leq j < R\}$ where C_j represents $B_{k,j}^{s,f}$ where we suppressed subscripts k, s and f for simplicity. We denote by $\mathbf{d}_i$ (e.g. $= W_i$ for throughput) the individual demand vector of NS-$i \in \mathcal{S}_k$, where $\mathbf{d}_i(j)$ is the demand of NS $i \in \{0, 1, \ldots, M-1\}$ for the resource j. Similarly, we use $\mathbf{c}_i(j)$ to denote the resource amount allocated to NS-$i \in \mathcal{S}_k$ for the resource j. Moreover, NS-$i \in \mathcal{S}_k$ is guaranteed $\mathbf{g}_i(j)$ from resource j and uses weight $w_{j,i}$ to quantify how important a given resource is in order to be able to meet the SLA. We assume that there is at least one resource j' such that $\sum_{i=0}^{M-1} \mathbf{d}_i(j') > C_{j'}$ i.e., it is not possible to fully satisfy all NSs for all the resources they demand and we do not let guaranteed resources to be beyond the total capacity available for that resource i.e., $\sum_{i=0}^{M-1} \mathbf{g}_i(j) < C_j$. In addition, we follow the general logical rule that any NS cannot be assigned more capacity than what is demanded.

Under the light of these assumptions, we have the following set of constraints for the multi-resource allocation problem for $0 \leq j < R$ and $0 \leq i < M$,

$$\min\left(\mathbf{g}_i(j), \mathbf{d}_i(j)\right) \leq \mathbf{c}_i(j) \leq \mathbf{d}_i(j), \quad \sum_{i=0}^{M-1} \mathbf{c}_i(j) \leq C_j \tag{1}$$

where *min* function is used to cover the general case that if the demand is smaller, then the guaranteed amount is not strictly allocated.

Our final step is to determine the objective function. Due to the guaranteed resource allocation, we cannot claim pure fairness among the NSs as it is not something intended. On the other hand, we are interested in the overall average satisfaction ratio of all M NSs. Therefore, for a given resource j, we can quantify the discrepancy for NS-$i \in \mathcal{S}_k$ as $\mathbf{d}_i(j) - \mathbf{c}_i(j)$ that we treat it as the soft error term that needs to be minimized. To be able to normalize the soft error terms coming from different resources (potentially with different domains and support) we divide it by $\mathbf{d}_i(j)$.

However from marginal utility point of view, it is more important whether the user is completely satisfied or not relative to soft satisfaction ratio i.e., a binary decision have to be made in order that resource to be any meaningful use to the assigned party. In order to encode this information into our problem, it is favorable to scale the normalized error to a range between zero and unity. The mathematical function that provides this mapping is the *ceiling* function. This finally gives us the following optimization problem.

Optimization Problem 1 (Original Joint Multi-Resource Allocation Based on SLA Requirements):

$$\min_{\mathbf{C}} \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i} \left\lceil \frac{\mathbf{d}_i(j) - \mathbf{c}_i(j)}{\mathbf{d}_i(j)} \right\rceil, \tag{2}$$

subject to

$$\min\left(\mathbf{g}_i(j), \mathbf{d}_i(j)\right) \leq \mathbf{c}_i(j) \leq \mathbf{d}_i(j), \quad 0 \leq j < R, \; 0 \leq i < M, \tag{3}$$

$$\sum_{i=0}^{M-1} \mathbf{c}_i(j) \leq C_j, \quad 0 \leq j < R, \; \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i}^2 = 1 \tag{4}$$

in which ceiling function is used to map soft information to hard decisions "0" for satisfied and "1" for unsatisfied states. However with this small change, the optimization problem with the given non-smooth i.e., non-differentiable

cost function would become quite hard to solve.[1] One of the useful approximations to ceiling function is tanh with positive support. Thus, our weighted cost (error) function can be expressed as

$$\sum_{j=0}^{R-1}\sum_{i=0}^{M-1} w_{j,i}\tanh\left(\frac{\mathbf{d}_i(j)-\mathbf{c}_i(j)}{\mathbf{d}_i(j)}\right) \tag{5}$$

where the weights $w_{j,i}$ can be adjusted based on how the NS-$i \in \mathcal{S}_k$-th satisfaction is affected by competing for resource type $j \in \{0, 1, 2, \ldots, R-1\}$. Alternative functions could have been used though the concavity of tanh is deemed useful from optimization point of view i.e., instead of minimizing the equation (2), we shall minimize the equation (5) which provides a convex (concave) cost and convex (concave) constraint set whose local solution is guaranteed to be global optimum. Note that the equation (5) is a concave function of $\{\mathbf{c}_i(j)\}$, i.e., the sum of non-negative weighted concave functions of a linear function of $\{\mathbf{c}_i(j)\}$ is a concave function of $\{\mathbf{c}_i(j)\}$. To see this, let $\mathbf{c}_i^{(1)}(j), \mathbf{c}_i^{(2)}(j) \in [0, \mathbf{d}_i(j)]$, then for a constant $\theta \in (0, 1)$ we have

$$\tanh\left(1-\frac{\theta\mathbf{c}_i^{(1)}(j)+(1-\theta)\mathbf{c}_i^{(2)}(j)}{\mathbf{d}_i(j)}\right) \tag{6}$$

$$=\tanh\left(\theta-\frac{\theta\mathbf{c}_i^{(1)}(j)}{\mathbf{d}_i(j)}+1-\theta+\frac{(1-\theta)\mathbf{c}_i^{(2)}(j)}{\mathbf{d}_i(j)}\right) \tag{7}$$

$$\geq \theta\tanh\left(1-\frac{\mathbf{c}_i^{(1)}(j)}{\mathbf{d}_i(j)}\right)+(1-\theta)\tanh\left(1-\frac{\mathbf{c}_i^{(2)}(j)}{\mathbf{d}_i(j)}\right) \tag{8}$$

where the last inequality is due to concavity of tanh for the range of values of interest. Finally, since non-negative weighted sum of concave functions is also concave, the result follows.

One of the things we realize that the hyperbolic function that we use i.e., tanh maps $[0, \infty)$ to $[0, 1)$. However, the function argument $1-\mathbf{c}_i(j)/\mathbf{d}_i(j)$ has the range $[0, 1-\mathbf{g}_i(j)/\mathbf{d}_i(j)]$ for $\mathbf{g}_i(j) < \mathbf{d}_i(j)$ and zero otherwise. To be able to express it without the condition, we can further express the range as $[0, 1-\min(\mathbf{g}_i(j), \mathbf{d}_i(j))/\mathbf{d}_i(j)]$. Note that this may lead to the use of different ranges of the function by different user and resource combinations. In order to address this issue, we introduce a scaling coefficient $\alpha_{j,i}$ that satisfies the following inequality

$$1-\tanh\left(\alpha_{j,i}\left(1-\frac{\min(\mathbf{g}_i(j),\mathbf{d}_i(j))}{\mathbf{d}_i(j)}\right)\right) \leq \eta \tag{9}$$

where $\eta \in (0, 1)$ is a small fixed tolerance value (such as 10^{-4}) to be able to constrain the input range to $[0, 1-\eta]$. Thus, depending on the guaranteed as well as the demanded resource quantities, the domain of tanh function is properly adjusted to ascertain that each term in the sum expression (5) uses a similar functional transformation. Using the equation (9) along with the known identity $\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$ for $x \in \mathbb{R}$, we can simply find the minimum $\alpha_{j,i}$ that can be used in the following set of optimization problems.

$$\begin{aligned}\alpha_{j,i} &= \frac{\tanh^{-1}(1-\eta)}{1-\min(\mathbf{g}_i(j),\mathbf{d}_i(j))/\mathbf{d}_i(j)}\\ &= \frac{\ln(2/\eta-1)}{2(1-\min(\mathbf{g}_i(j),\mathbf{d}_i(j))/\mathbf{d}_i(j))}\\ &= \frac{\mathbf{d}_i(j)\ln(\sqrt{2/\eta-1})}{\mathbf{d}_i(j)-\min(\mathbf{g}_i(j),\mathbf{d}_i(j))}\end{aligned} \tag{10}$$

Finally, $\alpha_{j,i}$ gets multiplied by the argument of tanh function in equation (5) the result of which can be expressed as

$$\frac{\ln(\sqrt{2/\eta-1})(\mathbf{d}_i(j)-\mathbf{c}_i(j))}{\mathbf{d}_i(j)-\min(\mathbf{g}_i(j),\mathbf{d}_i(j))} \tag{11}$$

which will appear as the new argument of tanh.

Unfortunately, the challenge with multi-resource allocation problem is that an allocation strategy that meets a certain fairness criterion for a given resource $j \in \{0, 1, 2, \ldots, R-1\}$, might not satisfy the same fairness criterion for the allocation of resource $j' \in \{0, 1, \ldots, R-1\}, j' \neq j$. It might even be impossible to satisfy fairness at the same time for all the resources. So the objective of this study is to find a trade-off point that maximizes the overall utility prescribed by weights, while minimizing the gap between the unconstrained solution and the fair allocation in a max-min sense.

We can identify two distinct directions to allocation strategy. The initial direction is to construct the appropriate optimization problem and address the weighted joint multi-resource allocation. The joint allocation strategy shall utilize the correlation structure between different SLAs through judiciously chosen weights and will provide better performance relative to individual resource allocations. The second direction will be towards finding a low complexity iterative scheme that approximates the solution of the original optimization problem while satisfying a desirable qualitative property such as weighted max-min fairness as much as possible (addressing the trade-off mentioned earlier). In the latter direction, we shall adapt *water filling* type of algorithms for our joint multi-resource allocation problem in a multi-user scenario.

IV. MULTI-RESOURCE ALLOCATION OPTIMIZATION PROBLEMS AND SOLUTIONS

Using the discussion of previous section, let us provide two different optimization problems that attempt to solve the trade-off mentioned earlier with different complexities. Then, we provide a low complexity iterative approach for solving these optimization problems as well as explain how we model SLA correlations into our optimization problems. We begin with a single-phase joint multi-resource allocation (abbreviated as SPATIAL) problem using equation (11) as follows.

[1] In the context of non-linear programming, this usually refers to the set of non-convex optimization problems. In addition, due to non-differentiability, iterative methods also suffer to converge to the global optimum.

*Optimization Problem 2 (**S**ingle-**Ph**ase Joi**n**t Multi-Resource **A**llocation (SPATIAL) Based on SLA Requirements):*

$$\min_{\mathbf{C}} \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i} \tanh\left(\frac{\ln\left(\sqrt{\frac{2}{\eta}-1}\right)(\mathbf{d}_i(j)-\mathbf{c}_i(j))}{\mathbf{d}_i(j) - \min(\mathbf{g}_i(j), \mathbf{d}_i(j))} \right), \tag{12}$$

subject to

$$\min(\mathbf{g}_i(j), \mathbf{d}_i(j)) \leq \mathbf{c}_i(j) \leq \mathbf{d}_i(j), \quad 0 \leq j < R,\ 0 \leq i < M, \tag{13}$$

$$\sum_{i=0}^{M-1} \mathbf{c}_i(j) \leq C_j, \quad 0 \leq j < R, \quad \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i}^2 = 1 \tag{14}$$

Next, we realize that SPATIAL is derived from the following relatively more fair problem that also takes into account the guaranteed resource management. Contrary to above case, the latter problem consists of two phases.

In the initial phase (phase 0), each slice is allocated with its guaranteed resources under the constraint of the demanded values (as before to cover up the condition we use minimum function to find the allocated capacity at phase 0), e.g.,

$$\mathbf{c}_{i,0}(j) = \min(\mathbf{g}_i(j), \mathbf{d}_i(j)), \quad 0 \leq j < R,\ 0 \leq i < M, \tag{15}$$

where $\mathbf{c}_{i,0}(j)$ denotes the initial capacity allocation for $i-$th slice and $j-$th resource. Based on equation (15), a new demand vector and remaining capacities can be derived as follows,

$$\mathbf{d}_{i,F}(j) = \mathbf{d}_i(j) - \mathbf{c}_{i,0}(j), \quad 0 \leq j < R,\ 0 \leq i < M, \tag{16}$$

$$C_{j,F} = C_j - \sum_{i=0}^{M-1} \mathbf{c}_{i,0}(j), \quad 0 \leq j < R, \tag{17}$$

where notation F denotes *Final* and $C_{j,F}$ are the remaining capacities after the initial allocation.

In the first phase (phase 1), we solve the following new optimization problem. Note that the scaling coefficient for this optimization problem should not depend on $\mathbf{g}_i(j)$ as its effect is stripped off. Since the range of the new argument is $[0, 1]$, it would not depend on the user or the resource indexes, so we name the new coefficient as β which satisfies

$$1 - \tanh(\beta) \leq \eta \tag{18}$$

from which we can deduce $\beta = \tanh^{-1}(1-\eta) = \ln\left(\sqrt{2/\eta - 1}\right)$ similar to our previous arguments. Note that the coefficient β does not depend on the resource type or user, due to the fact that we subtracted the effect of guaranteed resource from the optimization context. This double-phase joint multi-resource allocation (abbreviated as DORSAL) strategy is characterized by the following optimization problem.

*Optimization Problem 3 (**D**ouble-Phase **Jo**int Multi-**R**esource **A**llocation (**DORSAL**) Based on SLA Requirements):*

$$\min_{\mathbf{C}} \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} \times w_{j,i} \tanh\left(\frac{\left(\ln\left(\sqrt{\frac{2}{\eta}-1}\right)\right)(\mathbf{d}_{i,F}(j)-\mathbf{c}_{i,F}(j))}{\mathbf{d}_{i,F}(j)} \right), \tag{19}$$

subject to

$$0 \leq \mathbf{c}_{i,F}(j) \leq \mathbf{d}_{i,F}(j), \quad 0 \leq j < R,\ 0 \leq i < M, \tag{20}$$

$$\sum_{i=0}^{M-1} \mathbf{c}_{i,F}(j) \leq C_{j,F}, \quad 0 \leq j < R, \quad \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i}^2 = 1 \tag{21}$$

where $\mathbf{C}_F = \{C_{j,F} \geq 0, 0 \leq j < R\}$. It is not hard to realize that this optimization problem is quite similar to optimization problem 1 except the constant multiplying term $\ln\left(\sqrt{2/\eta - 1}\right)$ and the same concavity arguments apply for the cost function. This shall increase the chances of applying convex optimization tools and be able to quickly obtain the optimization solution.

Suppose the result of the optimization problem yields the minimum values $\mathbf{c}^*_{i,F}(j)$. Then, the final allocation will be given by

$$\mathbf{c}^*_i(j) = \mathbf{c}^*_{i,F}(j) + \mathbf{c}_{i,0}(j), \quad 0 \leq j < R,\ 0 \leq i < M. \tag{22}$$

We note that the tolerance value should also be judiciously adjusted. This is because we observe that if $\eta \to 0$ the result of the minimization is $\sum_j \sum_i w_{j,i}$ which is actually the maximum value possible for cost function to attain. Similarly, if we let $\eta \to 1$, the result of the minimization is zero which is the trivial solution. However, in this case we are far from our approximation of the *ceiling* function since as the argument of tanh approaches zero, tanh return values close to zero whereas *ceiling* function always returns 1 as long as the argument never equals zero. As we shall see in numerical results section, non-trivial best performing η can be found using heuristics.

A. ITERATIVE METHOD

Although the second optimization has broken the problem into two separate phases, the phase 1 can still be computationally infeasible for power and energy-hungry systems such as small ubiquitous IoT devices which may have only limited resources. Additionally, solving optimization problem in a dynamically changing environment is too much work to be handled by 5G devices. Thus, it is desirable to provide a low complexity iterative program, named *JENNER* (see Algorithm 1), that approximates the result of the first optimization problem.

It is important to identify that for the kth iteration of *JENNER* for $k > 0$, the leftover amount for the resource j

is given as

$$\sum_{i=0}^{M-1} (\mathbf{d}_i(j) - \mathbf{c}_{i,k-1}(j)) \tag{23}$$

Note that if the NS is satisfied, then the leftover resource is zero and does not contribute to the final error summation. On the other hand, since we cannot provide capacity more than we have, C_j minus the capacity given at the $(t-1)$-th iteration (denoted by $\mathbf{c}_{i,t-1}$) puts a constraint on the leftover resource which we model into our expression using minimum function. Finally, we use weights to differentiate between the UEs. So the resource allocated to the NS-$i \in \mathcal{S}_k$ for resource j shall be given by

$$\mathbf{c}_{i,t}(j) = \mathbf{c}_{i,t-1}(j) + \min\left(\mathbf{d}_{i,t}(j), \frac{w_{j,i}^2}{\sum\limits_i w_{j,i}^2} \times \min\left(C_{j,t}, \sum_{l=0}^{M-1} \mathbf{d}_{l,t}(j)\right)\right) \tag{24}$$

where $\mathbf{d}_{i,t}(j)$ and $C_{j,t}$ are new demand values for NS-$i \in \mathcal{S}_k$ and resource $j \in \{0, 1, 2, \ldots R-1\}$ and the remaining j^{th} resource at the beginning of the t^{th} step, respectively. They can be calculated by

$$\mathbf{d}_{i,t}(j) = \mathbf{d}_i(j) - \mathbf{c}_{i,t-1}(j) \tag{25}$$

$$C_{j,t} = C_j - \sum_{l=0}^{M-1} \mathbf{c}_{l,t-1}(j) \tag{26}$$

for all $t > 0$. As you may note that we also need the first minimum operation in equation (24) which ensures that no NS gets resources more than needed.

The iterations cease either when the equalities in equation (21) are satisfied i.e., all the capacity is distributed, or all the demands are satisfied. The initial condition ($t = 0$) for the iterations is given in equation (15). We note that the correlation structure of various SLAs of multiple resources are captured by the weights $w_{j,i}$ which makes the calculations of such weights an important task to carry out for the accuracy of the optimization solution. The complete algorithm is provided in **Algorithm 1**. One choice for the calculation of weights is based on AHP that we provide the details in the next section.

Algorithm 1 **J**oint **R**esource Allocatio**n** Usi**n**g It**e**rative Weighted Max-Min Fai**r**ness (*JENNER*)

Input: $\{\mathbf{g}_i(j)\}$, $\{C_j\}$, $\{\mathbf{d}_i(j)\}$, R, M for $0 \le j < R$, $0 \le i < M$

Output: $\mathbf{c}_i(j)$: Allocated share for resource $j \in \{0, 1, 2, \ldots R-1\}$ and NS-$i \in \mathcal{S}_k$.

Define $\delta_{M\times R} = \{\delta_{i,j} = 1\}$ for $0 \le j < R$, $0 \le i < M$
for $0 \le j < R$ **do**
  for $0 \le i < M$ **do**
    $\mathbf{c}_{i,0}(j) = \min(\mathbf{g}_i(j), \mathbf{d}_i(j))$
    if $\mathbf{d}_i(j) < \mathbf{g}_i(j)$ **then**
      $\delta_{i,j} \longleftarrow 0$
    end
    $\mathbf{d}_{i,1}(j) = \mathbf{d}_i(j) - \mathbf{c}_{i,0}(j)$
  end
  $C_{j,1} = C_j - \sum_{l=0}^{M-1} \mathbf{c}_{l,0}(j)$
end
$t \longleftarrow 1$
while $\sum_j C_{j,t} > 0$ ***and*** $\sum_i \sum_j \delta_{i,j} \neq 0$ **do**
  for $0 \le j < R$ **do**
    for $0 \le i < M$ **do**
      if $\delta_{i,j} = 1$ **then**
        $\mathbf{c}_{i,t}(j) = \mathbf{c}_{i,t-1}(j) + \min\left(\mathbf{d}_{i,t}(j), \frac{w_{j,i}^2}{\sum_i w_{j,i}^2} \min\left(C_{j,t}, \sum_{l=0}^{M-1} \mathbf{d}_{l,t}(j)\right)\right)$
        $\mathbf{d}_{i,t+1}(j) = \mathbf{d}_i(j) - \mathbf{c}_{i,t}(j)$
        if $\mathbf{d}_{i,t+1}(j) \le 0$ **then**
          $\mathbf{c}_{i,t}(j) = \mathbf{d}_i(j)$
          $C_j = C_j + \mathbf{c}_{i,t}(j) - \mathbf{d}_i(j)$
          $\gamma_{i,j} \longleftarrow 0$
        end
      end
    end
    $C_{j,t+1} = C_j - \sum_{l=0}^{M-1} \mathbf{c}_{l,k-1}(j)$
  end
  $t = t + 1$
end
return $\mathbf{c}_i(j)$

B. WEIGHTS THROUGH AHP

AHP is a structured technique for organizing and analyzing complex decisions [36]. Since resources are provided based on different SLA offerings, priorities can change with respect to the significance of each resource for the selected application.

Let us suppose that we consider video streaming application with demanded resources throughput, storage space, latency and reliability. Note that this application typically require relatively large throughput and limited latency. In order to use AHP, we need to either get pairwise significance of each resource for this application from web-pages or customer surveys that are periodically carried out to measure and respond to NS dis/satisfaction.

Suppose we are given Table 2, also known as the significance map, to assess the relative significance of each resource with respect to video streaming application. Note that the (i, j) entry of the significance map, represented by a matrix $\mathbf{S}$, is the reciprocal of the (j, i) entry i.e., the entries are not independent of each other. The way we interpret this table is for instance as follows: For application A, latency is three times more important than the resource storage.

TABLE 2. An example significance map of different network resources for application A.

	Throughput	Latency	Storage	Reliability
Throughput	1	3/2	5/2	5
Latency	2/3	1	3	2
Storage	2/5	1/3	1	1/2
Reliability	1/5	1/2	2	1

One of the key findings of AHP analysis is that the ranking of the priorities of these resources is given by the eigenvector of the maximum eigenvalue of the significance matrix $\mathbf{S}$. Fortunately, this eigenvector (also known as the priority column vector $\mathbf{p_A}$ for application A) can be closely approximated by normalizing the columns of $\mathbf{S}$ and and taking the average of the rows of $\mathbf{S}$. For instance, the priority vector for the data in Table 2 is given by $[0.797 \quad 0.512 \quad 0.195 \quad 0.254]^{\mathrm{T}}$ for application A. Alternatively, we could have used geometric mean for the rows to calculate the non-normalized priority vector.

Suppose that we have $N = S_k$ slices (or we have M users) called $A_1, \ldots, A_N$ with an application priority vector p_N i.e., relative significance of each application compared to others. For each application A_m, we can use AHP analysis to find its own priority vector $\mathbf{p}_{\mathbf{A}_m}$. Then, we can simply express weights $w_{n,m} = p_{A_m}(n)$. If we would like to find a single weight factor that does not depend on the application but the resources, our weights can be computed given by the following simple matrix multiplication operation

$$\begin{bmatrix} w_0 & w_1 & \ldots & w_{R-1} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} p_{A_1} | p_{A_2} | \ldots | p_{A_N} \end{bmatrix} p_N$$

We finally note that although AHP analysis provides a subjective measure of calculating weights, it may be replaced with future objective methods and this replacement shall not change our previous arguments about the proposed optimization problems.

V. PERFORMANCE EVALUATION

In this section, we will demonstrate that based on network virtualization, which allows the sharing of underlying communication system resources, the proposed solution can provide better QoS and, where applicable, yields better QoE levels that meet the requirements of different vertical domains. We start with the description of our simulation setup and continue with some of the numerical results that support the arguments and claims of this paper.

A. SIMULATION SETUP

We consider a particular network region where one BS including 4 cells is serving UEs at different locations within a circle with a radius of 1 km. Each UE is utilizing a unique and different vertical application, one of *MNO*−1 *mMTC*, *MNO*−2 *mMTC*, *MNO*−1 *eMBB (AR)*, *MNO*−2 *eMBB (AR)*, *MNO*−1 *eMBB (Live Video, L.V.)*, and *MNO*−2 *eMBB (Live Video, L.V.)*. The cells within the BS under consideration operates at 900 MHz band. Each cell uses a different carrier with 20 MHz bandwidth. It is assumed that the number of UEs related to different vertical domains have uniform distribution with different mean values. The mean values, throughput and storage demands of each UE related to different vertical services are given in Table 3. The randomly generated UEs share the same bandwidth resource and storage[2] with the total capacity of 4 TB. The guaranteed values for storage resource are set to 1.5 TB and 1 TB for *MNO*−1 *eMBB (Live Video, L.V.)*, and *MNO*−2 *eMBB (Live Video, L.V.)*, respectively and the remaining guaranteed values are set to zero.

TABLE 3. Mean values of uniformly distributed number of UEs and their throughput/storage requirements.

		mMTC	eMBB (AR)	eMBB (L.V.)
MNO-1	**Mean Values**	600	30	60
	Throughput	0.25 Mbps	100 Mbps	40 Mbps
	Storage	–	–	30 GB
MNO-2	**Mean Values**	600	30	45
	Throughput	0.25 Mbps	10 Mbps	40 Mbps
	Storage	–	–	30 GB

TABLE 4. Downlink channel simulation parameters.

Parameter	Value
HS-DSCH power ($P_{\mathrm{HS-DSCH}}$)	46 dBm
RRH transmitter antenna gain (G_{Antenna})	18 dBi
Cable loss (L_{Cable})	2 dB
Noise Power Spectral Density	−179 dBm/Hz
Height of RRH antenna (h_B)	80 m
Height of UE antenna (h_M)	1.5 m

The downlink channel simulation parameters utilized throughout our simulations are presented in Table 4. We consider urban environment Okumura − Hata path loss model [37] between each UE and BS which can be written as

$$H_{\mathrm{PL}}\,[\mathrm{dB}] = 69.55 + 26.16\log(f) - 13.82\log(h_B) - C_H + (44.9 - 6.55\log(h_B))\log(d), \quad (27)$$

where d is the UE distance to BS in km and C_H is antenna height correction factor and for small and medium-sized cities, it is given by

$$C_H = 0.8 + (1.1\log(f) - 0.7)h_M - 1.56\log(f), \quad (28)$$

where f is the operating frequency of each cell under BSs (set to 900 MHz). The received signal power at the i-th UE side, denoted by P_{RX}, is equal to

$$P_{\mathrm{RX}}\,[\mathrm{dBm}] = P_{\mathrm{HS-DSCH}}\,[\mathrm{dBm}] + G_{\mathrm{Antenna}}\,[\mathrm{dB}] - L_{\mathrm{Cable}}\,[\mathrm{dB}] - H_{\mathrm{PL}}\,[\mathrm{dB}], \quad (29)$$

without fast fading and shadowing effects. We further consider single-input single-output (SISO) transmission model for mMTC services and 8×8 multiple-input multiple-output (MIMO) transmission for the remaining services under the assumption that downlink channel state information is available at the centralized NEs. The Shannon Capacity for mMTC

[2]Two different resources (or QoS requirement dimensions) are shared among six different services (or slices under the assumption that each service contains one slice).

TABLE 5. Significance matrix S used in simulations.

			MNO-1 mMTC		MNO-2 mMTC		MNO-1 eMBB (AR)		MNO-2 eMBB (AR)		MNO-1 eMBB (L.V.)		MNO-2 eMBB (L.V.)	
			Throughput	Storage	Throughput	Storage	Throughput	Storage	Throughput	Storage	Throughput	Storage	Throughput	Storage
MNO-1	mMTC	Throughput	1	1	0.2	1	0.2	1	0.2	1	0.2	0.2	0.2	0.2
		Storage	1	1	1	1	1	1	1	1	1	1	1	1
MNO-2	mMTC	Throughput	5	1	1	1	0.2	1	0.2	1	0.2	0.2	0.2	0.2
		Storage	1	1	1	1	1	1	1	1	1	1	1	1
MNO-1	eMBB (AR)	Throughput	5	1	5	1	1	1	0.2	1	10	10	5	5
		Storage	1	1	1	1	1	1	1	1	1	1	1	1
MNO-2	eMBB (AR)	Throughput	5	1	5	1	5	1	1	1	10	10	5	5
		Storage	1	1	1	1	1	1	1	1	1	1	1	1
MNO-1	eMBB (L.V.)	Throughput	5	1	5	1	0.1	1	0.1	1	1	0.2	0.2	0.2
		Storage	5	1	5	1	0.1	1	0.1	1	5	1	0.2	0.2
MNO-2	eMBB (L.V.)	Throughput	5	1	5	1	0.2	1	0.2	1	5	5	1	0.2
		Storage	5	1	5	1	0.2	1	0.2	1	5	5	5	1

UEs can be written as

$$C_{\mathrm{mMTC}}(u) = B(u)\log_2\left(1+\frac{P_{\mathrm{RX}}}{\sigma_N^2}\right) \tag{30}$$

where σ_N^2 is the noise power and $B(u)$ is equal to allocated bandwidth for $u^{\mathrm{th}} \in \mathcal{U}_k^{i,f}$ UE. Under the assumption that UEs of eMBB services have the capability for multiple-antenna (up to 8) transmission, let $\mathbf{H} \in \mathbb{C}^{8\times 8}$ denote the channel matrix between cells and eMBB UEs includes normalized Rayleigh fading effect. Every $\mathbf{H}$ can be decomposed accordingly to its singular values such as

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{H}} \tag{31}$$

where the matrices $\mathbf{U}$ and $\mathbf{V}$ are unitaries and $\mathbf{D}$ is non-negative diagonal matrix where diagonal elements $\sigma_i \in \{0, 1..7\}$ denote the singular values of the channel matrix. When the transmitted signal vector with the dimension of 8 is multiplied with $\mathbf{V}$ and received signal vector with the same dimension is multiplied with $\mathbf{U}^{\mathrm{H}}$, MIMO channel model is transformed to independent SISO sub-channels with the channel gain of σ_i^2. The Shannon Capacity with respect to singular values can be found as

$$C_{\mathrm{eMBB}}(u) = B(u)\sum_{i=0}^{7}\log_2\left(1+\frac{P_{\mathrm{RX}}\times\sigma_i^2(u)}{\sigma_N^2}\right) \tag{32}$$

Regarding to service type of $u^{\mathrm{th}} \in \mathcal{U}_k^{i,f}$ UE, first required throughput (approximated by Shannon Capacity, $C(u)$) is determined and then, required $B(u)$ is calculated using known $C(u)$, P_{RX}, $\sigma_i^2(u)$, and σ_N^2.

B. NUMERICAL RESULTS

The significance matrix $\mathbf{S}$ that is used in our simulations for AHP is given in Table 5. Since mMTC and eMBB (AR) do not require any storage, their associated values are set to one with respect to each other and remaining services. The priority of eMBB (AR) and eMBB (Live Video) services with respect to mMTC is increased through setting relevant values in $\mathbf{S}$ to higher than one, specifically five. Between mMTC services, priority of MNO-2 is assumed as higher so that relevant value is set to five as well. We further prioritize eMBB (AR) services with regard to eMBB (Live Video) services for both MNO-1 and MNO-2. In the same matrix, it can also be observed that MNO-2 has higher priority for eMBB (AR) and eMBB (Live Video) services compared to MNO-1's.

The eigenvector corresponding to maximum eigenvalue of $\mathbf{S}$ is $\mathbf{p}_A = [\,0.08\ 0.18\ 0.10\ 0.18\ 0.49\ 0.18\ 0.65\ 0.18\ 0.13\ 0.17\ 0.23\ 0.30\,]^{\mathrm{T}}$. The weight coefficient vectors for resource-1 (bandwidth) and resource-2 (storage) can be found as $w_{1,i} = p_A(2i)$ and $w_{2,i} = p_A(2i+1)$ for $i \in \{0, 1, \ldots, M-1\}$ and they become $\mathbf{w}_1 = [0.08\ 0.10\ 0.49\ 0.65\ 0.13\ 0.23]^{\mathrm{T}}$, $\mathbf{w}_2 = [0.18\ 0.18\ 0.18\ 0.18\ 0.17\ 0.30]^{\mathrm{T}}$ respectively. We utilized MATLAB's optimization toolbox for nonlinear programming for the first and second proposed methods and the results are obtained through Monte Carlo simulations. We conduct 4000 independent simulations where the location of UEs are randomly selected under the consideration of uniform distribution within a radius of 1 km and the number of UEs related to different services is determined with respect to uniform distribution with the mean values defined in Table 3 starting from zero. In each simulation, it is determined whether the slice is fully satisfied based on which the resource-specific satisfaction index is set to one (if slice is fully satisfied) or zero (if slice is not fully satisfied). Finally, service satisfied-ratio is calculated by averaging the satisfaction indexes in each simulation which are depicted in Fig. 4a and Fig. 4b where MMF[3] allocation is considered as our benchmark. We further calculate the ratio of allocated resource to demand and the results are summarized in Fig. 5a and Fig. 5b.

Regarding $w_{1,i}$, it is expected that the weighted coefficient increase the service satisfied-ratio and the ratio of allocated resource to demand probabilities of the 4-th service slice, i.e., MNO-2 eMBB (AR) during bandwidth resource allocation. It is followed by 3-rd, 6-th, 5-th, 2-nd, and 1-st slices, respectively. However, these coefficients is not guaranteed the certain ratio values due to non-linearly of tanh and different amount of service demands. For $\eta = 0.2384$, which basically sets $\ln(2/\eta - 1)$ term inside (12) and (20) to one, Fig. 4a shows that the service satisfied-ratio of the 4-th service is the highest with respect to *SPATIAL* where it is equal to 99.3%. It is followed

[3]In MMF allocation [38], resources are orderly allocated to slices with respect to their increasing demands and unsatisfied slices are given equal share of the remaining resource.

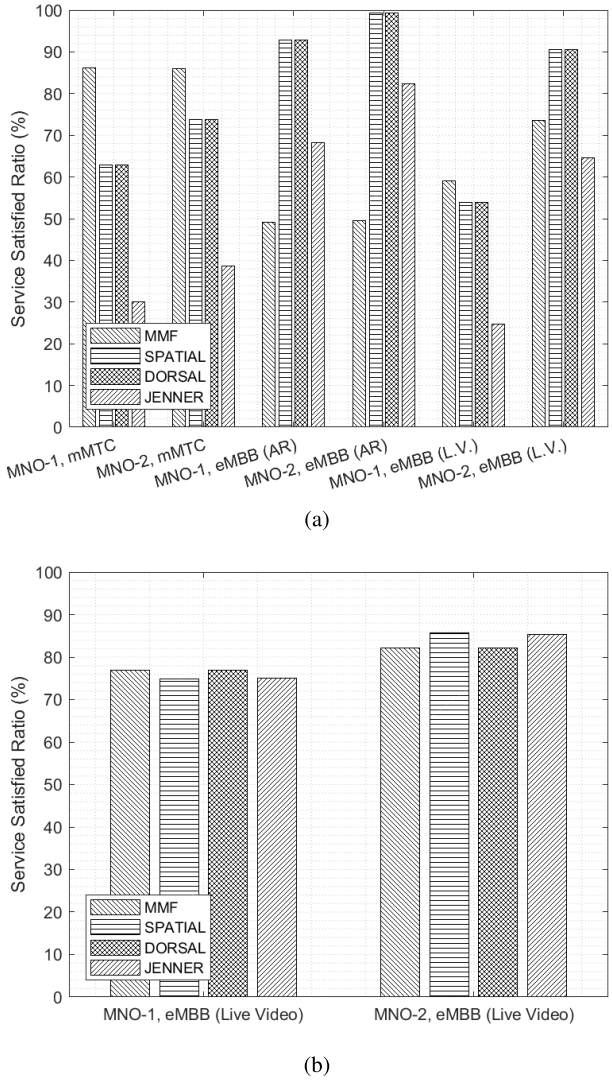

FIGURE 4. Service satisfied ratio of each slice on (a) bandwidth and (b) storage utilization using MMF and our proposed schedulers assuming $\eta = 0.2384$.

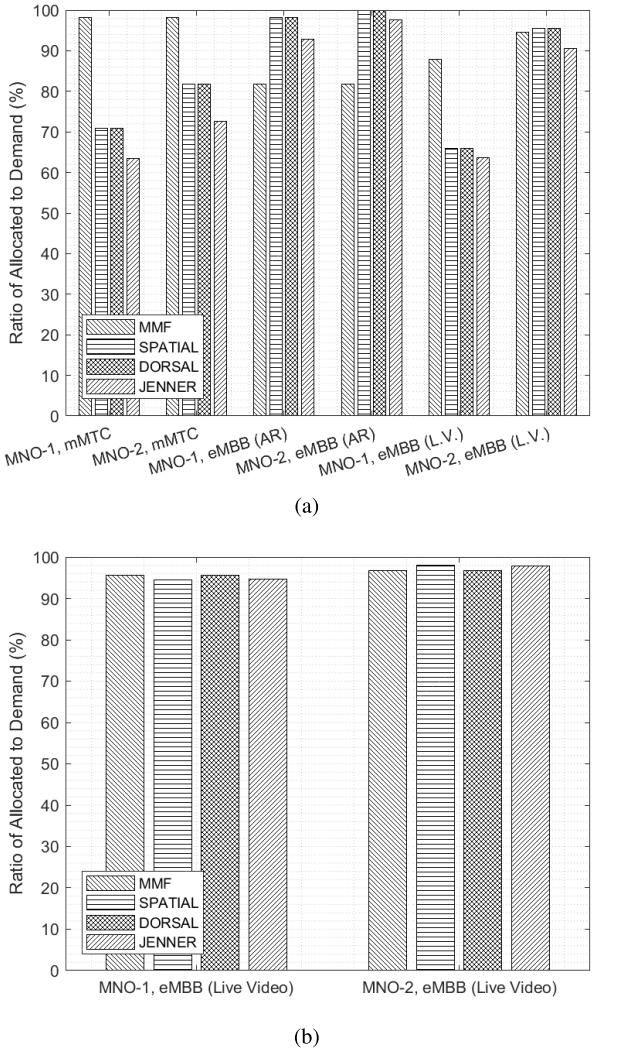

FIGURE 5. Ratio of allocated to demand of each service on (a) bandwidth and (b) storage utilization using MMF and our proposed schedulers assuming $\eta = 0.2384$.

by 3-rd and 6-th services with the ratios of 92.83% and 90.47% in proportion to their priority levels as a consequence of different weight coefficients. On the other hand, service satisfied-ratio levels of 1-st and 2-nd services are higher than 5-th service due to the non-linearity of tanh that suppresses the error term of the highly demanded 5-th service. It should also be noted that since there is no guaranteed amount for the bandwidth resource, *DORSAL* allocates the same amount of resources with *SPATIAL* as is seen from Fig. 4a. *JENNER* shows similar behaviour while allocating the resources to services. However, the ratios are relatively lower than the proposed first two methods, similar to allocation result of MMF. *JENNER* outperforms MMF for 3-rd and 4-th services as a consequence of weighted allocation mechanism. In Fig. 4b, we provide the service satisfied-ratio of 5-th and 6-th services on storage utilization. It can be observed that *MNO-1, eMBB (Live Video)* achieves the service satisfied-ratios of 74.78%, 76.97%, and 75.05% with *SPATIAL*, *DORSAL*, and *JENNER*, respectively, whereas MMF provides 74.78% service satisfied-ratio level for this service. For the other service whose weighted coefficient is approximately twice of the first one, these amounts are increased to 85.67%, 82.10%, and 85.25% using *SPATIAL*, *DORSAL*, and *JENNER*, respectively, whereas MMF increases the ratio to 82.10%. The reason of higher satisfied ratio obtained through MMF is less average UE count leading to less demand requirement, whereas to include weighted coefficient results in a increment on the ratio with our proposed schedulers. The similar behaviour can be observed in Figs. 5a and 5b for all of the schedulers. The ratio of allocated resources to demand levels are relatively higher than service satisfied-ratio levels due to hard classification performed to determine whether the service is satisfied or not.

In order to compare the performance of proposed schedulers with each other and MMF, we present our results in terms of both average service satisfied-ratio and average ratio

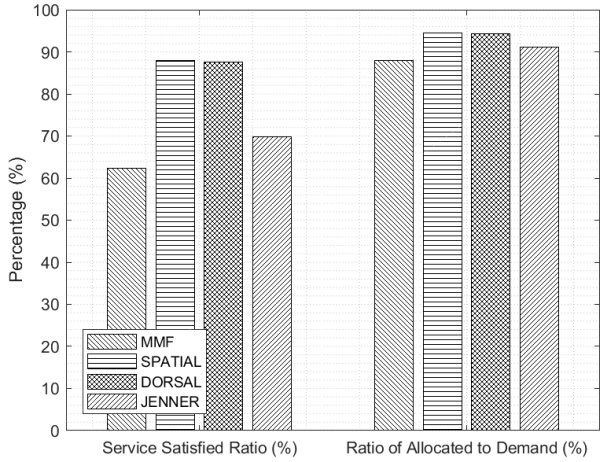

FIGURE 6. Overall satisfaction ratio and overall ratio of allocated resource amounts to demands assuming $\eta = 0.2384$.

of allocated resource to demand (see Fig. 6), which can be calculated by

$$\bar{m} = 100 \times \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i} m_i(j) \Big/ \sum_{j=0}^{R-1} \sum_{i=0}^{M-1} w_{j,i}, \qquad (33)$$

where $m_i(j)$ either denotes the satisfied-ratio or ratio of allocated resource to demand metric for service i for resource j. The first proposed method, *SPATIAL*, achieves the highest service satisfied-ratio of 87.87% while the *DORSAL* and *JENNER* achieve 87.54% and 69.74%, respectively, whereas MMF allocation provides a ratio of 62.28%. For the second metric at the same η value, while MMF allocation secures the value of 88.01%, our proposed methods, *DORSAL*, *SPATIAL*, and *JENNER* achieve 94.44%, 94.33%, and 91.19%, respectively. On the other hand, when Jain's Fairness index is considered, MMF achieves the highest metric, numerically 0.971 and 0.998 for bandwidth and storage resources, respectively. It is then followed by *JENNER*, *SPATIAL*, and *DORSAL* with the values of 0.912, 0.874, and 0.874 for bandwidth resource and 0.997, 0.9965, and 0.9961 for storage resource, respectively. The results reveal that when the resources are not equally important to services, the overall system satisfaction ratio and fairness can jointly be optimized to provide better resilience compared to individual sub-optimal resource allocations.

The performance results in terms of different η are further demonstrated in Fig. 7a and Fig. 7b for *SPATIAL* and *DORSAL*. In Fig. 7a, the results reveal that *SPATIAL* achieves the highest service satisfied-ratio value, 88.49% with η of 0.1. This value is decreased to 87.87%, 85.32%, and 82.78%, respectively, with the η of 0.2384, 0.01, and 0.001. Similar behaviour is observed for the ratio of allocated resource to demand metric depicted in Fig. 7b. Numerically, the ratio becomes 94.44%, 94.60%, 92.08%, and 90.61%, respectively, for the η equaling 0.2384, 0.1, 0.01, and 0.001. *DORSAL* has the same behaviour for the same metrics. It achieves the service satisfied-ratio and ratio of

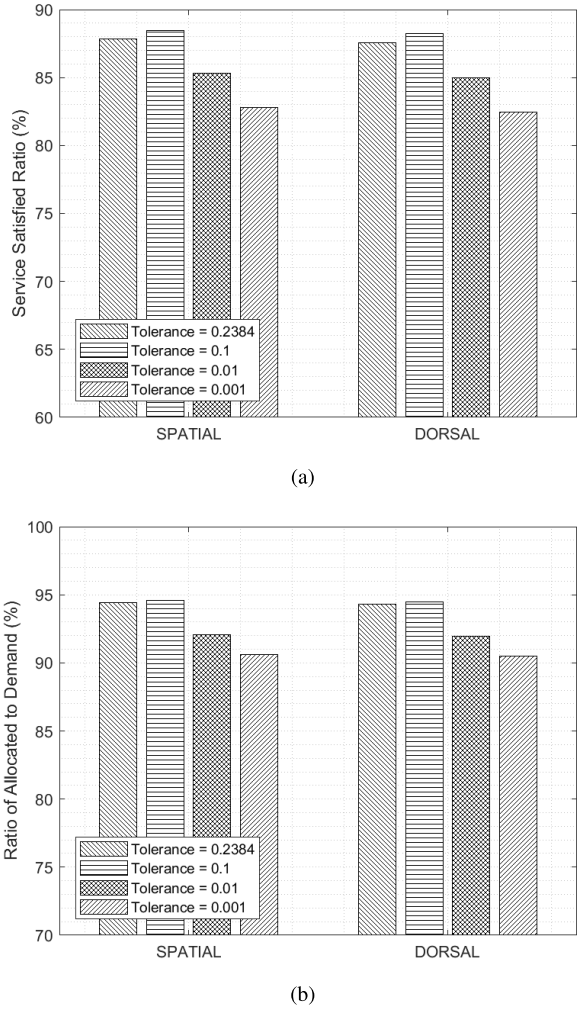

FIGURE 7. (a) Overall service satisfied-ratio and (b) ratio of allocated resource to demand of our proposed schedulers with respect to different η values.

allocated resource to demand of 88.21% and 94.50% with $\eta = 0.1$. It is then decreased to 87.54%, 84.99%, and 82.46%, respectively, for service satisfied-ratio metric and 94.33%, 91.97%, and 90.49%, respectively, for ratio of allocated resource to demand metric with the values of 0.2384, 0.01, and 0.001. On one hand, fairness index of *SPATIAL* and *DORSAL* becomes 0.889 and 0.916 using $\eta = 0.1$ and 0.01 for bandwidth resource. It is further increased to 0.929 from 0.874 using $\eta = 0.001$. When the second resource is considered, there is an improvement with a value between 0.001 and 0.002 i.e., using lower η values. The results reveal that different η values can be optimal depending on demands and guaranteed values. Thus, optimal tolerance value can be found using a simulation-based heuristics.

VI. CONCLUSION

In this paper, we have introduced an SDN-based network slicing framework within a 5G network infrastructure. Additionally, by utilizing the benefits of this architecture, we have investigated three different solutions for joint optimization

of multiple resource allocations over the dimensions of a given set of QoS requirements. An iterative version of the solution to the optimization problem is also provided, and its performance is evaluated with respect to optimum solutions through Monte-Carlo simulations. The results are compared using a traditional MMF scheduler as the benchmark in order to demonstrate the improvements over satisfaction ratios.

REFERENCES

[1] 5GPP. (2017). *View on 5G Architecture (Version 2.0) White Paper*. Accessed: Dec. 14, 2017. [Online]. Available: https://5g-ppp.eu/white-papers/
[2] NGMN Alliance. (2016). *Perspectives on Vertical Industries and Implications for 5G (White Paper)*. Accessed: Dec. 14, 2017. [Online]. Available: https://goo.gl/pd3y23
[3] C. Xu, S. Jia, M. Wang, L. Zhong, H. Zhang, and G.-M. Muntean, "Performance-aware mobile community-based VoD streaming over vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1201–1217, Mar. 2015.
[4] C. A. Detweiler and K. V. Hindriks, "A survey of values, technologies and contexts in pervasive healthcare," *Pervasive Mobile Comput.*, vol. 27, pp. 1–13, Apr. 2016.
[5] *Study on Network Slicing*, document 3GPP TR 28.801, 3GPP, 2017.
[6] *Provisioning of Network Slicing for 5G Networks and Services*, document 3GPP TS 28.531, 3GPP, 2017.
[7] *System Architecture for the 5G System*, document 3GPP TS 23.501, 3GPP, 2017.
[8] ETSI. (2017). *Network Operator Perspectives on NFV Priorities for 5G (White Paper)*. Accessed: Dec. 14, 2017. [Online]. Available: https://goo.gl/ojvWo4
[9] IEEE 5G Initiative Technology Roadmap Working Group. (2017). *IEEE 5G and Beyond Technology Roadmap (White Paper)*. Accessed: Dec. 14, 2017. [Online]. Available: https://goo.gl/KUYeoe
[10] S. Vassilaras *et al.*, "The algorithmic aspects of network slicing," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 112–119, Aug. 2017.
[11] A. Nakao *et al.*, "End-to-end network slicing for 5G mobile networks," *J. Inf. Process.*, vol. 25, pp. 153–163, Feb. 2017.
[12] K. Katsalis, N. Nikaein, E. Schiller, A. Ksentini, and T. Braun, "Network slices toward 5G communications: Slicing the LTE network," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 146–154, Aug. 2017.
[13] S. Sharma, R. Miller, and A. Francini, "A cloud-native approach to 5G network slicing," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 120–127, Aug. 2017.
[14] X. Li *et al.*, "5G-crosshaul network slicing: Enabling multi-tenancy in mobile transport networks," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 128–137, Aug. 2017.
[15] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 80–87, May 2017.
[16] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.
[17] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wen, and M. Tao, "Resource allocation in spectrum-sharing OFDMA femtocells with heterogeneous services," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2366–2377, Jul. 2014.
[18] O. Narmanlioglu and E. Zeydan, "New Era in shared cellular networks: Moving into open and virtualized platform," *Int. J. Netw. Manage.*, vol. 27, no. 6, p. e1986, 2017. [Online]. Available: http://dx.doi.org/10.1002/nem.1986
[19] O. Narmanlioglu and E. Zeydan, "New era in shared C-RAN and core network: A case study for efficient RRH usage," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[20] M. R. Palattella *et al.*, "Internet of Things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.
[21] M. Richart, J. Baliosian, J. Serrat, and J.-L. Gorricho, "Resource slicing in virtual wireless networks: A survey," *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 3, pp. 462–476, Sep. 2016.
[22] S. Parsaeefard, R. Dawadi, M. Derakhshani, and T. Le-Ngoc, "Joint user-association and resource-allocation in virtualized wireless networks," *IEEE Access*, vol. 4, pp. 2738–2750, 2016.
[23] S. Parsaeefard, R. Dawadi, M. Derakhshani, T. Le-Ngoc, and M. Baghani, "Dynamic resource allocation for virtualized wireless networks in massive-MIMO-aided and fronthaul-limited C-RAN," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9512–9520, Oct. 2017.
[24] H. Yang, J. Zhang, Y. Ji, and Y. Lee, "C-RoFN: Multi-stratum resources optimization for cloud-based radio over optical fiber networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 118–125, Aug. 2016.
[25] H. Yang, Y. He, J. Zhang, Y. Ji, W. Bai, and Y. Lee, "Performance evaluation of multi-stratum resources optimization with network functions virtualization for cloud-based radio over optical fiber networks," *Opt. Exp.*, vol. 24, no. 8, pp. 8666–8678, Apr. 2016. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-8-8666
[26] T. LeAnh, N. H. Tran, D. T. Ngo, and C. S. Hong, "Resource allocation for virtualized wireless networks with backhaul constraints," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 148–151, Jan. 2017.
[27] B. Yu, W. Zheng, X. Wen, Z. Lu, L. Wang, and L. Ma, "Dynamic resource orchestration of service function chaining in network function virtualizations," in *5G for Future Wireless Networks*, K. Long, V. C. Leung, H. Zhang, Z. Feng, Y. Li, and Z. Zhang, Eds. Cham, Switzerland: Springer, 2018, pp. 132–145.
[28] V. N. Ha and L. B. Le, "End-to-end network slicing in virtualized OFDMA-based cloud radio access networks," *IEEE Access*, vol. 5, pp. 18675–18691, 2017.
[29] Qualcomm. (Feb. 2017). *Augmented and Virtual Reality: The First Wave of 5G Killer Apps (White Paper*. Accessed: Dec. 14, 2017. [Online]. Available: https://goo.gl/3TB7Gd
[30] D. Flore, *3GPP Standards for the Internet of Things*, document, GSMA MIoT, 2016.
[31] (2017). *SigFox*. Accessed: Dec. 14, 2017. [Online]. Available: http://www.sigfox.com
[32] (2017). *LoRa*. Accessed: Dec. 14, 2017. [Online]. Available: https://www.lora-alliance.org/
[33] 5GPP. (2016). *5G Automative Vision (White Paper)*. Accessed: Dec. 14, 2017. [Online]. Available: https://5g-ppp.eu/white-papers/
[34] J. Liu, J. Wan, B. Zeng, Q. Wang, H. Song, and M. Qiu, "A scalable and quick-response software defined vehicular network assisted by mobile edge computing," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 94–100, Jul. 2017.
[35] comScore Inc. (2016). *Turning Big Data Into Mobile Subscriber Insights: How Mobile Operators Can Capture, Retain and Grow Their Subscriber Base (White Paper)*. Accessed: Dec. 14, 2017. [Online]. Available: https://goo.gl/HHStMr
[36] T. L. Saaty, "How to make a decision: The analytic hierarchy process," *Eur. J. Oper. Res.*, vol. 48, no. 1, pp. 9–26, Sep. 1990.
[37] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*. Hoboken, NJ, USA: Wiley, 2010.
[38] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 4, pp. 5–17, 4th Quart., 2008.

OMER NARMANLIOGLU received the B.Sc. degree from the Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey, in 2014, and the M.Sc. degree from Ozyegin University, Istanbul, Turkey, in 2016, where he is currently pursuing the Ph.D. degree. He is currently with P.I. Works. His research interests include the physical and link layer aspects of communication systems and software-defined networking paradigm for radio access, transmission, and packet core networks.

ENGIN ZEYDAN received the B.Sc. and M.Sc. degrees from the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, in 2004 and 2006, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, in 2011. He was a Research and Development Engineer with Avea, a mobile operator in Turkey, from 2011 to 2016. He is currently with Türk Telekom Labs as a Senior Research and Development Engineer. He has also been a part-time instructor with the Electrical and Electronics Engineering Department, Ozyegin University, since 2015. His research interests include the area of telecommunications and big data networking.

SUAYB S. ARSLAN received the B.Sc. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, San Diego, CA, USA, in 2009 and 2012, respectively. He was with Mitsubishi Electric Research Laboratory, Boston, MA, USA, in 2009, where he was involved in research and development of image and video processing algorithms for biomedical applications. In 2011, he joined Quantum Corporation, Irvine, CA, USA, where he conducted research on advanced detection and coding algorithms for increased capacity storage and cloud systems. He is currently with MEF University as an associate professor. His research interests include digital communication and storage, joint source-channel coding, information and reliability theory, image/video processing, and cross layer design optimizations.