Radio Resource Provisioning for Network Slicing with Coverage Constraints

Quang-Trung Luu, Sylvaine Kerboeuf, Alexandre Mouradian, Michel Kieffer

▶ To cite this version:

Quang-Trung Luu, Sylvaine Kerboeuf, Alexandre Mouradian, Michel Kieffer. Radio Resource Provisioning for Network Slicing with Coverage Constraints. 2020 IEEE International Conference on Communications (ICC 2020), Jun 2020, Dublin (virtual), Ireland. pp.1-7, 10.1109/ICC40277.2020.9148897 . hal-03097210

HAL Id: hal-03097210

https://centralesupelec.hal.science/hal-03097210v1

Submitted on 5 Jan 2021

Radio Resource Provisioning for Network Slicing with Coverage Constraints

Quang-Trung Luu*†, Sylvaine Kerboeuf*, Alexandre Mouradian†, and Michel Kieffer†
*Nokia Bell Labs, Nozay, France
†Laboratoire des Signaux et Systèmes (L2S),
Univ. Paris-Sud – CNRS – CentraleSupélec – Univ. Paris-Saclay, France
E-mails: {quangtrung.luu, michel.kieffer}@centralesupelec.fr, sylvaine.kerboeuf@nokia-bell-labs.com

***Abstract*—With network slicing, Mobile Network Operators can accommodate on a common network infrastructure various customized services from Service Providers (SPs). Usually, the Service Function Chains belonging to a slice are deployed on a best-effort basis. Nothing ensures that enough infrastructure resources can be allocated to satisfy the demands of SPs. This paper introduces a radio resources provisioning approach to satisfy the demands of slices with radio coverage constraints. By provisioning, we ensure that enough resources are reserved for further SFC deployment. Numerical results show the effectiveness of the proposed provisioning framework for a slice deployment on a mobile network infrastructure satisfying a minimum data rate for users in the geographical areas where services have to be made available.**

***Index Terms*—Network slicing, resource provisioning, coverage constraints, 5G.**

I. INTRODUCTION

Network slicing has emerged as a new paradigm for 5G networks to meet various service needs from diverse Service Providers (SPs) such as mobile virtual network operators, vertical industries and OTT service providers [1–3].

A network slice consists of collection of Service Function Chains (SFCs) involving physical network resources, which are dynamically allocated to build a customized logically isolated virtual network. Each SFC consists of several interconnected Virtual Network Functions (VNFs) describing the processing applied to a data flow related to a given service.

Several entities are involved in network slicing [4]. The Mobile Network Operator (MNO) deploys network slices onto its own infrastructure resources (Remote Radio Heads (RRHs)), fronthaul and backhaul networks, data centers) and/or infrastructure resources leased to third-party Infrastructure Provider (InP). The SP exploits the slices supplied by the MNO, and provides to his customers the related services. Service needs are forwarded by the SP to the MNO within a Service Level Agreement (SLA). The SLA describes, at a high level of abstraction, characteristics of the service with the desired QoS, the number of devices (or the device/user density), the geographical region where the service has to be made available for the end-users, *etc*. The MNO translates the SP high-level demands into SFCs able to fulfill the service requirements. SFCs are then deployed on the network infrastructure so that QoS requirements are satisfied. With virtualization, SFCs and VNFs can be easily and flexibly initialized, launched, chained, and scaled to meet changeable workload requests [4–6]. Nevertheless, many research challenges remain when network slicing incorporates the wireless part of legacy or 5G networks [7, 8], where the radio access has to be considered.

This work extends the provisioning approach introduced in [9] to situations where radio coverage constraints have to be taken into account. In [9], we adopted the point of view of the InP, who aims at identifying the infrastructure nodes on which the VNF within a slice are deployed and the links able to transmit data between these nodes. In the present work, we additionally account for rate and coverage constraints for mobile end-users of the slice services. These constraints have to be satisfied by the InP, leading to a problem of appropriate RRH node selection and provisioning of their radio resources.

In the rest of the paper, Section II introduces some related work. Section III presents the model of the infrastructure network and of the slice resource demands. The slice resource provisioning problem is then formulated in Section IV as a Mixed Integer Linear Programming (MILP) problem, accounting for radio resource constraints for the deployment of multiple slices. Numerical results are presented in Section V. Finally, Section VI draws some conclusions and perspectives.

II. RELATED WORK

Early results on assigning infrastructure resources to virtual network components may be found, *e.g.*, in [10, 11]. Due to its capability of sharing efficiently resources in 5G networks, the concept of network virtualization has gained renewed attention in the literature [3, 12–14], via the concept of network slicing. Network slice resource allocation is a complex problem. When a slice instance is seen as a collection of SFCs, slice embedding needs to deploy the SFCs on a shared infrastructure while satisfying various constraints. Most of prior work related to SFC and VNF deployment do not account for coverage constraints [15, 16]. The design of efficient allocation mechanisms for virtualized radio resources has been recently addressed in [17]. This paper aims at minimizing the leasing cost of Base Stations (BSs) so as to meet SP demands, while providing, with a given probability, a minimum data rate for any user located in the BS coverage area. Users are assumed to be served by their nearest BS, which facilitates the evaluation of the rate constraint and reduces the dimension of the problem. This assumption reduces somehow the potentiality of achieving the optimal sharing of radio resources.

A heterogeneous spatial user density is considered in [18]. Joint BS selection and adaptive slicing are formulated as a two-stage stochastic optimization problem. A reduced-complexity deterministic optimization problem is obtained by generating several random realizations of user locations. A genetic algorithm is then used for the optimization.

In [19], a network slicing framework for multi-tenant heterogeneous cloud radio access network is introduced. Slicing is formulated as a weighted throughput maximization problem, which aims at maximizing the total rate obtained by users connected to given RRHs on given sub-channels. Nevertheless, the proposed framework does not consider computing and memory resource associated to the processing within the Baseband Units (BBUs). Such resource is assumed to be properly scaled so as to support the required service rate. Moreover, the proposed framework addresses only downlink data services.

A game theory-based distributed algorithm is proposed in [20] to address the wireless network slicing problem. This algorithm accounts for the limited availability of wireless resources and considers different aspects such as congestion, deployment costs and the RRH-user distance. This work considers the coverage area of RRH, but ignores the possible coverage constraints required by the slices.

III. System Model

Consider a set of SPs whose aim is to provide different services, indexed by $\sigma = 1, \ldots, |\mathcal{S}|$, to mobile users over some geographical area. The geographical area under study is denoted by $\mathcal{A}$ and the subarea over which service σ has to be made available is denoted by $\mathcal{A}^\sigma$. Figure 1 illustrates three geographical subareas over which three different services have to be deployed. For that purpose, each SP forwards its service requirements to an MNO. The MNO sends to the InP a Slice Resource Demand (SRD) consisting of (*i*) an SRD graph accounting for the structure and SLA of the slice and (*ii*) SRD coverage information related to the area $\mathcal{A}^\sigma \subset \mathcal{A}$ over which the service will have to be made available. The InP is then in charge of provisioning enough infrastructure resources to deploy the SFCs, whose resource demands are described by the SRD graph. This work focuses on the specific aspect of the SRD related to the slice radio coverage constraints, which impose a minimum data rate for users in the geographical areas where services have to be made available.

Our aim, with resource provisioning is to reserve, somewhat in advance, enough infrastructure resources to ensure that the MNO will have access to properly located radio resources with service characteristics as stated in the SLA.The time scale at which provisioning is performed is much larger than that at which SFCs are deployed and adapted to meet actual time-varying user demands. One focuses on a time interval over which resources will be provisioned so as to be compliant with the variations of user demands within a slice. The duration of this time interval results from a compromise between the need to update the provisioning and the level of conservatism in the amount of provisioned resources required to satisfy fast fluctuating user demands.

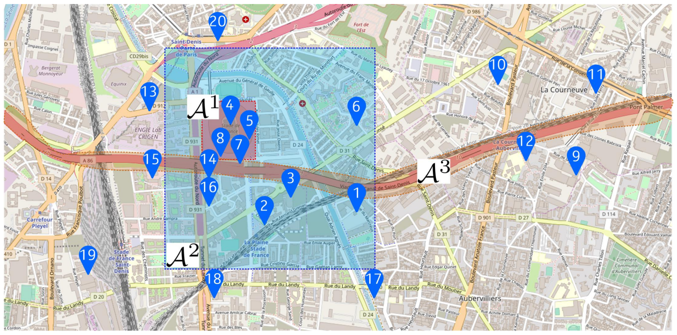


Fig. 1. The considered metropolitan area including the Stade de France (covered by the red rectangle representing $\mathcal{A}^1$), its surrounding (blue rectangle representing $\mathcal{A}^2$), and part of the A86 highway (orange band representing $\mathcal{A}^3$); blue markers show the locations of RRH nodes.

Table I summarizes all notations used in this paper.

TABLE I
Table of Notations

Symbol	*Description*
$\mathcal{G}_\mathrm{I}$	Infrastructure network graph, $\mathcal{G}_\mathrm{I} = (\mathcal{N}_\mathrm{I}, \mathcal{E}_\mathrm{I})$
$\mathcal{N}_\mathrm{I}$	Set of infrastructure nodes
$\mathcal{N}_\mathrm{Ir}$	Set of RRHs, $\mathcal{N}_\mathrm{Ir} \subset \mathcal{N}_\mathrm{I}$
$\mathcal{E}_\mathrm{I}$	Set of infrastructure links
$a_\mathrm{r}(i)$	Available RBs at RRH $i \in \mathcal{N}_\mathrm{Ir}$
$c_\mathrm{f}(i)$	RRH node disposal cost
$c_\mathrm{r}(i)$	Per-unit cost of RB for node $i \in \mathcal{N}_\mathrm{Ir}$
$\mathcal{S}$	Set of all slices indexed σ
$\mathcal{A}$	The geographical area under study
$\mathcal{A}^\sigma$	Coverage area of slice σ
$\mathcal{Q}^\sigma$	Set of all divided subareas q in $\mathcal{A}^\sigma$
$\mathcal{A}_q^\sigma$	Subarea q, $q \in \mathcal{Q}^\sigma$
$\rho^\sigma(x)$	User density function of slice σ, $x \in \mathcal{A}$
$\underline{R}_\mathrm{u/d}^\sigma$	Min data rate for UL/DL traffic for each user in σ
$r_\mathrm{u/d}(v_\mathrm{r})$	Aggregated UL/DL data rate requirement at v_r
$\eta_\mathrm{u/d}^\sigma(i, q)$	RB proportion provisioned by RRH i to users in $\mathcal{A}_q^\sigma$ for UL/DL traffic
$\widetilde{\eta}^\sigma(i)$	Binary variable indicating whether RRH $i \in \mathcal{N}_\mathrm{Ir}$ provisions RBs to slice σ
$b_\mathrm{u/d}(x_i^\mathrm{r}, \mathcal{A}_q^\sigma)$	UL/DL data carried by a RB for a user located in $\mathcal{A}_q^\sigma$
λ	Discount factor
c_rr	Cost function for the radio resources

A. Infrastructure model

Consider an infrastructure network managed by some InPs. We devise the special case of the cloud mobile network architecture with RRH nodes connected to datacenter nodes at edge and central locations as depicted on Figure 2. This network is represented by a directed graph $\mathcal{G}_\mathrm{I} = (\mathcal{N}_\mathrm{I}, \mathcal{E}_\mathrm{I})$, where $\mathcal{N}_\mathrm{I}$ is the set of infrastructure nodes and $\mathcal{E}_\mathrm{I}$ is the set of infrastructure links, which correspond to the wired connections between nodes of the infrastructure network.

Each infrastructure node and link in $\mathcal{G}_\mathrm{I}$ is characterized by a given amount of supported resources (*e.g.*, computing and storage for nodes as well as bandwidth for links), which may

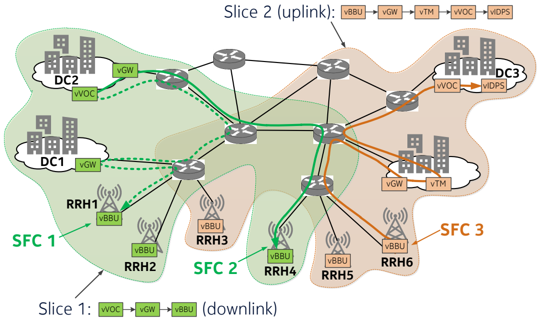


Fig. 2. Example of deployment of two slices in a cloud mobile network. Slice 1 is dedicated to a video streaming service and Slice 2 aims at providing video surveillance and traffic monitoring. The SFCs of Slice 1 consists of a virtual Video Optimizer Controller (vVOC), a virtual Gateway (vGW), and a virtual BBU (vBBU). The SFCs of Slice 2 consists of a vBBU, a vGW, a virtual Traffic Monitor (vTM), a vVOC, and a virtual Intrusion Detection Prevention System (vIDPS).

be allocated to network slices. Radio resources are exclusively provided by a subset $\mathcal{N}_{\text{Ir}} \subset \mathcal{N}_{\text{I}}$ of RRH nodes, whose location in some Cartesian frame attached to $\mathcal{A}$ is denoted x_i^{r}. The cost associated to the use of a RRH i consists of a fixed part $c_{\text{f}}\left(i\right)$ for node disposal, and a variable part $c_{\text{r}}(i)$, which depends linearly on the radio resource amount provided by that RRH.

B. SRD Model

A SRD is defined on the basis of the SLA between the SP and the MNO. The SLA may consider several time intervals over each of which the service characteristics and constraints are assumed constant, but may vary from one interval to the next one. These time intervals translate, *e.g.*, day and night variations of user demands. They last between tens of minutes and hours. It is of the responsibility of the SP and MNO to properly scale the requirements expressed in the SLA, by considering, for example, similar services deployed in the past.

In this work, one considers a given time interval specified in the SLA. The SLA is expressed in terms of supported service type and targeted QoS such as a minimum average data rate $\underline{R}_{\text{u}}^{\sigma}$ and $\underline{R}_{\text{d}}^{\sigma}$ for the wireless uplink (UL) and downlink (DL) traffic of each client. The geographical distribution function $\rho^{\sigma}(x)$, with $x \in \mathcal{A}$, describes the *maximum* user/device density to be served around x within the considered time interval.

A slice is also characterized by the amount of resources required for running all VNFs and links supporting all the SFCs to achieve the SLA with the SP. The SRD represents the aggregated resource requirements of these SFCs and its graph mimics the graph of SFCs, with SRD nodes corresponding to the VNFs of the SFC. Each SRD node is characterized by a given amount of *required* resources, *e.g.*, computing and/or storage, to sustain the aggregated demand for all instances of a given VNF in the slice. Similarly, each SRD link can be characterized by the bandwidth required to sustain the aggregated traffic demand between the VNFs at two ends.

One assumes that the UL and DL resource demands are aggregated within a single node v_{r} of the SRD graph. The resulting aggregated UL and DL data rates $r_{\text{u}}\left(v_{\text{r}}\right)$ and $r_{\text{d}}\left(v_{\text{r}}\right)$ resulting from the coverage constraint of slice σ are

$$r_{\text{u}}\left(v_{\text{r}}\right) = \underline{R}_{\text{u}}^{\sigma} \int_{\mathcal{A}^{\sigma}} \rho^{\sigma}\left(x\right) \mathrm{d}x, \tag{1}$$

$$r_{\text{d}}\left(v_{\text{r}}\right) = \underline{R}_{\text{d}}^{\sigma} \int_{\mathcal{A}^{\sigma}} \rho^{\sigma}\left(x\right) \mathrm{d}x. \tag{2}$$

IV. COVERAGE-AWARE RESOURCE PROVISIONING FOR NETWORK SLICES

This section describes a formulation of the slice resource provisioning problem. In Section IV-A, we propose a framework to provision radio resources for multiple slices. The problem of joint radio and other network resource provisioning is discussed in Section IV-B.

A. Radio Provisioning

Our aim is to provision radio resources from the RRHs for the slices managed by an MNO with minimum deployment cost. For slice σ, the InP has to provide a minimum average data rate ($\underline{R}_{\text{u}}^{\sigma}$ for UL and $\underline{R}_{\text{d}}^{\sigma}$ for DL) to each mobile user spread over $\mathcal{A}^{\sigma}$ with a density $\rho^{\sigma}\left(x\right)$. For that purpose, the InP will have to provision resources from the RRH nodes in $\mathcal{N}_{\text{Ir}}$. One assumes that every RRH node is able to provide a fixed amount $a_{\text{r}}\left(i\right)$ of Resource Blocks (RB) per time unit to exchange data (UL and DL) with users. The amount of data transmitted using a single RB depends on the characteristics of the RRH, of the User Equipment (UE), and on the transmission channel between the RRH and the user.

During the resource provisioning phase, the locations of users are unknown. To address this problem, an approach inspired by the subarea partitioning technique introduced in [21] is considered. $\mathcal{A}^{\sigma}$ is partitioned into Q^{σ} convex subareas $\mathcal{A}_q^{\sigma}, q \in \mathcal{Q}^{\sigma} = \{1, \dots, Q^{\sigma}\}$. Instead of allocating RBs to users, RRH nodes allocate RBs to subareas. The way the partitioning is performed is not detailed here. One may consider, *e.g.*, a partitioning into rectangles of equal surfaces or a partitioning based on ρ^{σ} that provides an equal average number of users per subarea.

To formulate the radio resource provisioning problem, we introduce the sets of variables $\boldsymbol{\eta}$ and $\widetilde{\boldsymbol{\eta}}$, where $\boldsymbol{\eta} = \left\{\eta_{\text{u}}^{\sigma}\left(i, q\right), \eta_{\text{d}}^{\sigma}\left(i, q\right)\right\}_{i \in \mathcal{N}_{\text{Ir}}, q \in \mathcal{Q}^{\sigma}, \sigma \in \mathcal{S}}$, represents the proportion of RBs provisioned by RRH i to the users in $\mathcal{A}_q^{\sigma}$ for UL and DL traffic, respectively. The elements of $\boldsymbol{\eta}$ take real values in $[0, 1]$. The set of binary variables $\widetilde{\boldsymbol{\eta}} = \left\{\widetilde{\eta}^{\sigma}\left(i\right)\right\}_{i \in \mathcal{N}_{\text{Ir}}}$ identifies whether a RRH $i \in \mathcal{N}_{\text{Ir}}$ has provisioned some RBs to any subarea for slice σ, with $\widetilde{\eta}^{\sigma}\left(i\right) = 1$ if $\sum_{q \in \mathcal{Q}^{\sigma}} \eta^{\sigma}\left(i, q\right) > 0$, and $\widetilde{\eta}^{\sigma}\left(i\right) = 0$ otherwise.

1) Single Slice : The cost related to the radio resource provisioning for slice $\sigma \in \mathcal{S}$ consists of a fixed part $c_{\text{f}}\left(i\right) \widetilde{\eta}^{\sigma}\left(i\right)$ related to the use of a RRH and a variable part $c_{\text{r}} a_{\text{r}}\left(i\right) \eta^{\sigma}\left(i, q\right)$ related to the amount of RBs provided by each RRH to the slice. A bias towards RB allocation by RRHs providing high spectral efficiency is obtained by the introduction of a rate-related discount $\lambda b\left(x_i^{\text{r}}, \mathcal{A}_q^{\sigma}\right) a_{\text{r}}\left(i\right) \eta^{\sigma}\left(i, q\right)$, where λ is a positive discount factor. The resulting cost function for the radio resources is

$$
\begin{aligned}
c_{\mathrm{rr}}\left(\boldsymbol{\eta},\widetilde{\boldsymbol{\eta}}\right) &= \sum_{i\in\mathcal{N}_{\mathrm{Ir}}} c_{\mathrm{f}}\left(i\right)\widetilde{\eta}^{\sigma}\left(i\right) \\
&+ \sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\sum_{q\in\mathcal{Q}^{\sigma}}\left[c_{\mathrm{r}}-\lambda b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\right]a_{\mathrm{r}}\left(i\right)\eta_{\mathrm{u}}^{\sigma}\left(i,q\right) \qquad (3)\\
&+ \sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\sum_{q\in\mathcal{Q}^{\sigma}}\left[c_{\mathrm{r}}-\lambda b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\right]a_{\mathrm{r}}\left(i\right)\eta_{\mathrm{d}}^{\sigma}\left(i,q\right).
\end{aligned}
$$

The minimization of $c_{\mathrm{rr}}\left(\boldsymbol{\eta},\widetilde{\boldsymbol{\eta}}\right)$ has to be such that several constraints are satisfied.

The summed proportions of RBs provided by a given RRH i must be less than one

$$
\sum_{q\in\mathcal{Q}^{\sigma}}\left(\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)+\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)\right)\leqslant 1,\ \forall i\in\mathcal{N}_{\mathrm{Ir}}. \qquad (4)
$$

For each subarea $\mathcal{A}_q^{\sigma}$, the total data rate provided by the allocated resource blocks should satisfy the minimum average user demand. Thus, $\forall q\in\mathcal{Q}^{\sigma}$, one should have

$$
\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\geqslant\underline{R}_{\mathrm{u}}^{\sigma}\int_{\mathcal{A}_q^{\sigma}}\rho^{\sigma}\left(x\right)\mathrm{d}x, \qquad (5)
$$

$$
\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\geqslant\underline{R}_{\mathrm{d}}^{\sigma}\int_{\mathcal{A}_q^{\sigma}}\rho^{\sigma}\left(x\right)\mathrm{d}x, \qquad (6)
$$

which corresponds to the satisfaction of the geographical coverage constraints for UL and DL. Here, $b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$ and $b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$ denote the amount of data carried by a RB for a user located in $\mathcal{A}_q^{\sigma}$ for UL and DL. Depending on the level of conservatism, $b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$ and $b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$ may represent the minimum or the average amount of data evaluated over the possible locations of users in $\mathcal{A}_q^{\sigma}$. The terms $b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$, $b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)$, and $\int_{\mathcal{A}_q^{\sigma}}\rho^{\sigma}\left(x\right)\mathrm{d}x$ are fixed quantities that only depend on the RRH location x_i^{r}, on the user density ρ^{σ}, and on the way the partitioning of $\mathcal{A}^{\sigma}$ has been performed. These terms may thus be evaluated in advance, see Section V-C. Summing (6) over all $q\in\mathcal{Q}^{\sigma}$ and using (2), one gets

$$
\sum_{q\in\mathcal{Q}^{\sigma}}\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\geqslant r_{\mathrm{u}}\left(v_{\mathrm{r}}\right), \qquad (7)
$$

$$
\sum_{q\in\mathcal{Q}^{\sigma}}\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\geqslant r_{\mathrm{d}}\left(v_{\mathrm{r}}\right), \qquad (8)
$$

which ensures, for slice σ, the satisfaction of the part of the SRD graph related to the UL and DL radio resource demands.

For each RRH, the amount of provisioned UL and DL resources should be proportional to the demand expressed in the SRD graph through $r_{\mathrm{u}}\left(v_{\mathrm{r}}\right)$ and $r_{\mathrm{d}}\left(v_{\mathrm{r}}\right)$, see (1) and (2). This avoids provisioning RRH resources taking care only of the UL or only of the DL traffic. This has to be ensured for all subareas $q\in\mathcal{Q}^{\sigma}$

$$
\frac{\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)}{r_{\mathrm{u}}\left(v_{\mathrm{r}}\right)}=\frac{\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)a_{\mathrm{r}}\left(i\right)b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)}{r_{\mathrm{d}}\left(v_{\mathrm{r}}\right)}. \qquad (9)
$$

The nonlinear relation between $\eta^{\sigma}\left(i,q\right)=\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)+\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)$ and $\widetilde{\eta}^{\sigma}\left(i\right)$ can be linearly expressed as

$$
0\leq\widetilde{\eta}^{\sigma}\left(i\right)-\sum_{q\in\mathcal{Q}^{\sigma}}\eta^{\sigma}\left(i,q\right)<1,\ \forall i\in\mathcal{N}_{\mathrm{Ir}},\forall\sigma\in\mathcal{S}. \qquad (10)
$$

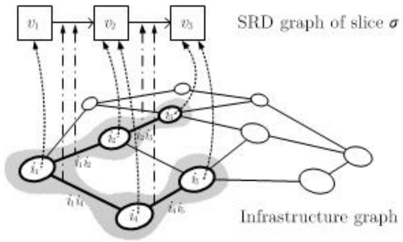


Fig. 3. Provisioning of infrastructure resource to an SRD graph: Resources from the infrastructure node i_1 is provisioned for SRD node v_1; Resources from i_2 and i_4 are provisioned for SRD node v_2; and resources from i_3 and i_5 are provisioned for SRD node v_3. Correspondingly, the infrastructure links i_1i_2 and i_1i_4 are provisioned for SRD link v_1v_2 and resources from links i_2i_3 and i_4i_5 are provisioned for SRD link v_2v_3.

Finally, the radio provisioning problem consists in finding

$$
\left(\widehat{\boldsymbol{\eta}},\widehat{\widetilde{\boldsymbol{\eta}}}\right)=\arg\min_{\boldsymbol{\eta},\widetilde{\boldsymbol{\eta}}}c_{\mathrm{rr}}\left(\boldsymbol{\eta},\widetilde{\boldsymbol{\eta}}\right),\ \text{subject to: (4) -- (10),}
$$

which is an MILP problem.

2) Multiple Slices: When several SRD graphs for slices indexed $\sigma\in\mathcal{S}$ have to be considered, (3) becomes

$$
\begin{aligned}
c_{\mathrm{rr}}\left(\boldsymbol{\eta},\widetilde{\boldsymbol{\eta}}\right) &= \sum_{\sigma\in\mathcal{S}}\sum_{i\in\mathcal{N}_{\mathrm{Ir}}} c_{\mathrm{f}}\left(i\right)\widetilde{\eta}^{\sigma}\left(i\right) \\
&+ \sum_{\sigma\in\mathcal{S}}\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\sum_{q\in\mathcal{Q}^{\sigma}}\left[c_{\mathrm{r}}-\lambda b_{\mathrm{u}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\right]a_{\mathrm{r}}\left(i\right)\eta_{\mathrm{u}}^{\sigma}\left(i,q\right) \qquad (11)\\
&+ \sum_{\sigma\in\mathcal{S}}\sum_{i\in\mathcal{N}_{\mathrm{Ir}}}\sum_{q\in\mathcal{Q}^{\sigma}}\left[c_{\mathrm{r}}-\lambda b_{\mathrm{d}}\left(x_i^{\mathrm{r}},\mathcal{A}_q^{\sigma}\right)\right]a_{\mathrm{r}}\left(i\right)\eta_{\mathrm{d}}^{\sigma}\left(i,q\right),
\end{aligned}
$$

with the constraints (5)-(10) and the constraint (4) replaced by

$$
\sum_{\sigma\in\mathcal{S}}\sum_{q\in\mathcal{Q}^{\sigma}}\left(\eta_{\mathrm{u}}^{\sigma}\left(i,q\right)+\eta_{\mathrm{d}}^{\sigma}\left(i,q\right)\right)\leqslant 1,\ \forall i\in\mathcal{N}_{\mathrm{Ir}}. \qquad (12)
$$

B. Full resource provisioning for slices

To complete the slice resource provisioning, one has to account for other resources in terms of computing, storage, and networking requirements to run the interconnected VNFs that are parts of the slice. A joint provisioning of radio and other network resources should be considered.

This provisioning is represented by a mapping between the infrastructure graph $\mathcal{G}_{\mathrm{I}}$ and the SRD graph, as depicted in Figure 3. In this example, the slice σ is described by an SRD graph aggregating the demands of several linear SFCs.

An additional set of constraints for these other resources needs to be introduced. This set of constraints, which expresses the other resource demands for each $\sigma\in\mathcal{S}$, has to be consistent with the coverage constraints presented in Sections IV-A1 and IV-A2. For instance, the other types of network resources (*e.g.*, link bandwidth or node storage capacity) provisioned by an infrastructure node should be commensurate with the provisioned radio resources. When considering demands on both UL and DL traffic, it is relevant to ensure the balance between radio and other network resources provisioned by infrastructure nodes and links. Details of those constraints can be found in [22], where a two-step algorithm to solve the joint radio and other network resources provisioning problem is introduced. The radio provisioning proposed in Section IV-A2 is first achieved. Then, only the subset of infrastructure nodes that contains RRHs involved in the slice coverage and identified from the first step are considered in the resource provisioning for the remaining functions and links of the slice.

V. EVALUATION

In this section, one compares via simulations the performance of three different radio resource provisioning schemes for network slicing: a Sequential approach (`SR`), a Joint approach (`JR`) and a Baseline approach. In `SR`, resources are provisioned slice by slice by solving the single slice MILP problem (3)-(10). In `JR`, provisioning is performed taking into account all slices simultaneously and solving the multiple slices MILP problem (11)-(12). All simulations are performed with the CPLEX MILP solver interfaced with MATLAB. The Baseline approach works slice-by-slice, as the `SR` approach. For each slice, the RRHs provisioning resources to users located in subarea q are selected starting from those providing the highest Signal-to-Noise (SNR) ratio, whatever the cost for using those RRHs, contrary to what is done by the `JR` and `SR` approaches.

Heterogeneous cellular networks will be considered. Due to the large difference between the transmit power of macrocells and of smaller cells with lower power, *e.g.*, microcells or picocells, in the Baseline approach, an offset is introduced to favor the utilization of smaller stations when they are closer to the users. This association strategy is known as *Cell Range Expansion* (CRE) [23], which allows a UE to associate to lower power stations.

A. Infrastructure Network

We consider a $1.4\,\mathrm{km}\times 5\,\mathrm{km}$ area around the Stade de France near Paris, shown in Figure 1. The map includes real coordinates of RRH nodes (indicated by blue markers) taken from the open database provided by the French National Agency of Frequencies[1].

B. Slice Resource Demand (SRD)

Three types of slices are considered:

- Slices of type 1 aim to provide video streaming services at 3 Mbps for at most 250 VIP users within the stadium (DL traffic);
- Slices of type 2 are dedicated to a video streaming services at 0.5 Mbps for at most 1000 users, and cover the blue-highlighted area in Figure 1 (DL traffic);
- Slices of type 3 aim to provide video surveillance and traffic monitoring service at 1 Mbps for 50 cameras installed on the A86 highway (UL traffic).

Slices of type 1 and 2 have the same components in their SRD graphs, illustrated by Slice 1 in Figure 2. Slices of type 3 have an SRD graph similar to that of Slice 2 in Figure 2. In what follows, different scenarios are considered with an increasing number of slices whose distribution among each type is given in Table II. This represents situations where slices of the same type are provided by different SPs. The coverage area $\mathcal{A}^\sigma$ associated to each slice is partitioned into subareas $\mathcal{A}_q^\sigma$ of $90\,\mathrm{m}\times 103\,\mathrm{m}$.

TABLE II
TYPE OF EACH SLICE

Slice	1	2	3	4	5	6	7	8
Type	3	1	2	3	2	3	3	3

C. Rate Function

The amount of data carried by a RB for a user located in $\mathcal{A}_q^\sigma$ and served by a RRH in x_i^r is $b_\mathrm{d}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ and $b_\mathrm{u}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$, respectively for DL and UL. Models for $b_\mathrm{d}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ and $b_\mathrm{u}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ are now considered.

Let $d\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ be the distance between x_i^r and the center of each rectangle $\mathcal{A}_q^\sigma$. Focusing on DL traffic, one considers the following model taken from [24]

$$b_\mathrm{d}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right) = W_i \log_2\left(1 + SNR_\mathrm{d}\left(i, q\right)\right), \qquad (13)$$

where W_i is the bandwidth (in Hz) of a RB provided by RRH i. $SNR_\mathrm{d}\left(i, q\right)$ is the DL Signal-to-Noise (SNR) ratio from RRH i to users located in subarea q, given by $SNR_\mathrm{d}\left(i, q\right) = P_\mathrm{rx,d}\left(d\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)\right)/P_\mathrm{n}$, where $P_\mathrm{n} = W_i N_0$ is the noise power, N_0 is the noise power spectral density. $P_\mathrm{rx}(d) = P_\mathrm{tx,d} + G_\mathrm{tx,d} + G_\mathrm{rx,d} - PL(d)$ is the signal power at the receiver, where P_tx is the transmission power of the transmitter, G_tx and G_rx are the antenna gains of the transmitter and the receiver. $PL(d) = 10\alpha\log_{10}(d) + \beta + 10\gamma\log_{10}(f_i)$ is the path loss given by the adapted $\alpha\beta\gamma$-model introduced in [25] for 5G mobile networks, where α and γ are respectively coefficients accounting for the dependency of the path loss with distance and frequency f_i, β is an optimized offset value for path loss (in dB). The model (13) is also considered for $b_\mathrm{u}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$.

A heterogeneous cellular network is considered. The RRHs 4, 5, 7, and 8 are microcells, whereas the rest are macrocells. The parameters for the models $b_\mathrm{d}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ and $b_\mathrm{u}\left(x_i^\mathrm{r}, \mathcal{A}_q^\sigma\right)$ are: $a_\mathrm{r}(i) = 100$, $f_i = 2.6$ GHz, $W_i = 0.2$ MHz, $(P_\mathrm{tx,d}, P_\mathrm{tx,u}) = (43, 12)$ dBm for macrocells and $(P_\mathrm{tx,d}, P_\mathrm{tx,u}) = (35, 8)$ dBm for microcells, $(G_\mathrm{tx,d}, G_\mathrm{tx,u}) = (15, 3)$ dBi, $N_0 = -174$ dBm/Hz, and $(\alpha, \beta, \gamma) = (3.6, 7.6, 2)$, see [26]. Moreover, c_f, c_r, λ, are set to 100, 1, and 0.1 in (3) and (11). The offset considered in the Baseline algorithm is set to 0 dB for a macrocell and to 4 dB for a microcell.

D. Results

Figure 4 illustrates the utilization of RBs per RRH for each slice, with $|\mathcal{N}_\mathrm{lr}| = 8$ and $|\mathcal{S}| = 8$. Thanks to the rate-related discount introduced in the objective function of the MILP problem, RRHs that are close to the coverage area of each slice and provide strong signal are chosen in priority in all the considered approaches. For instance, with the `JR` approach, Slice 2 of type 1, which covers the stadium, has its resource demand provisioned by RRH 4 and RRH 7. The advantage of

---

[1] L'Agence nationale des fréquences (ANFR): https://data.anfr.fr/

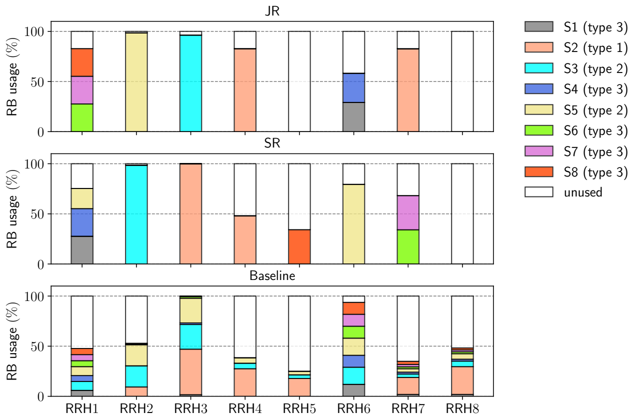
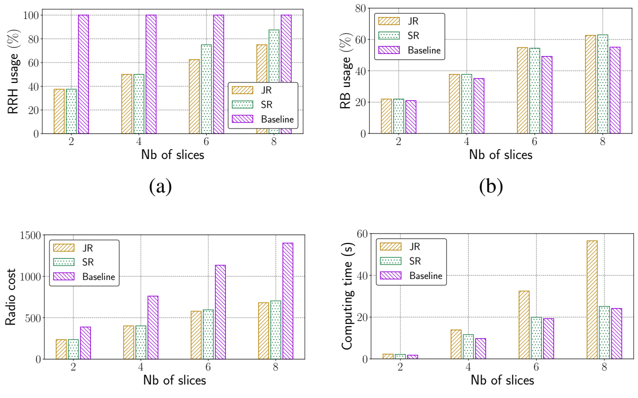


Fig. 4. Distribution of RBs per RRH for each slice: with the JR (top), SR (middle), and Baseline (bottom) approach.



(a) (b)

(c) (d)

Fig. 5. Results of (a) utilization of RBs, (b) utilization of RRHs, (c) radio cost, and (d) computing time as a function of $|\mathcal{S}|$, with $|\mathcal{N}_{\text{Ir}}| = 8$.

the JR approach over the two other approaches can also be observed: with the JR approach, only six RRHs are required to provision resources, whereas with the SR and Baseline approach, seven and eight RRHs are needed, respectively.

Figures 5a, 5b, 5c, and 5d depict respectively the percentage of provisioned RRHs and RBs, the radio cost, and the computing time for different $|\mathcal{S}|$, see Table II, with $|\mathcal{N}_{\text{Ir}}| = 8$. Figures 6a, 6b, 6c, and 6d consider the same metrics with different values of $|\mathcal{N}_{\text{Ir}}|$, when $|\mathcal{S}| = 6$.

The JR approach outperforms the SR and Baseline approaches in terms of radio cost and utilization of RRH nodes: the JR approach aims at finding the optimal solution for the whole problem, *i.e.*, provisioning for all the slices, while the SR and Baseline approaches only account for the constraints of each slice sequentially. The difference in performance of these three methods increases when four slices and more are simultaneously considered. For instance, with six slices, a cost reduction of 2.9% compared to the SR approach and 96% compared to the baseline approach is observed in favor of the JR approach, see Figure 5b. With eight slices, the provisioning cost reductions are 3.3% and 105%.

Nevertheless, the Baseline approach slightly outperforms the other approaches in terms of RB utilization, since it is not constrained by the node disposal cost when selecting RRHs, see Figures 5b and 6b.

Consider now one slice of each type. Figure 7 shows the maximum supported data rate with the three considered approaches as a function of the aggregated data rate demand from users, *i.e.*, $\sum_{\sigma\in\mathcal{S}} u^{\sigma}\underline{R}^{\sigma}$, where u^{σ} is the number of users in σ, when $|\mathcal{N}_{\text{Ir}}| = 8$. $\underline{R}^{\sigma}$ remains constant for each slice σ. The total number of users u^{σ} associated to each slice varies, but their relative proportions among slices remain constant. With the JR approach, a larger aggregated data rate is supported: provisioning of slices with more users is then possible.

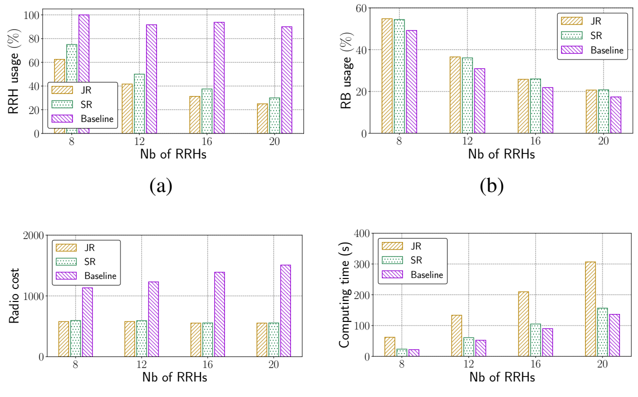
Nevertheless, as expected, the price to be paid is a larger computing time for the JR approach compared to the SR and Baseline approaches, as shown in Figures 5d and 6d. Increasing the number of slices leads to an increase of the cardinality of the set of variables $\left(\boldsymbol{\eta}^{\sigma}_{iq}, \tilde{\boldsymbol{\eta}}^{\sigma}_{i}\right)$.

VI. CONCLUSIONS

This paper considers the problem of radio resource provisioning for network slicing in future mobile networks. Among the SLAs between MNOs and SPs, one focuses in this paper on those involved in service coverage constraints, *i.e.*, related to the geographical distribution of the end-users to which the service has to be provided.

Adopting the point of view of the InP, one tries to minimize the cost related to the usage of the network infrastructure, in particular the radio access network, while satisfying radio coverage constraints, to ensure a minimum data rate for users in the geographical areas where services have to be made available. This problem is cast in the framework of MILP problem.

Two provisioning approaches are considered and compared to a Baseline approach. The sequential approach provisions resources slice by slice and the joint approach considers



(a) (b)

(c) (d)

Fig. 6. Results of (a) utilization of RBs, (b) utilization of RRHs, (c) radio cost, and (d) computing time as a function of $|\mathcal{N}_{\text{Ir}}|$, with $|\mathcal{S}| = 6$.

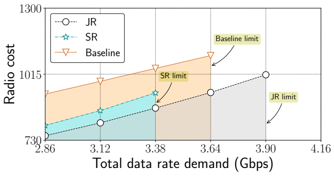


Fig. 7. Maximum supported data rate associated to the `JR`, `SR`, and baseline provisioning approaches when 3 slices have to be deployed.

the constraints of all slices simultaneously. The sequential approach scales better to network topologies of realistic size, due to the exponential complexity in the number of variables of the MILP problem. The price to be paid is a somewhat degraded RRH and RB utilization and a higher provisioning cost compared to the joint approach. When compared to a Baseline approach, where only radio efficiency is considered, the two proposed approaches provide a lower slice provisioning cost.

Several issues will be addressed in future work. For instance, our optimization model allows UEs to be served by far RRHs when the closer RRHs are saturated. This can generate undesired inter-cell interference. In future work, this problem should be carefully addressed. On the other hand, adaptive provisioning techniques will be proposed to cope with time-varying constraints using, *e.g.*, iterative solutions.

ACKNOWLEDGMENT

This work is partly supported by the French Agence Nationale de la Recherche (ANR), project MAESTRO-5G ANR-18-CE25-0012.

REFERENCES

[1] 5G America, “Network Slicing for 5G and Beyond,” *White Paper*, 2016.

[2] IETF, “Network Slicing Architecture,” *Internet-Draft*, pp. 1–27, 2017.

[3] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega, D. Aziz, and H. Bakker, “Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks,” in *IEEE Commun. Mag.*, vol. 55, no. 5, 2017, pp. 72–79.

[4] C. Liang and F. R. Yu, “Wireless Network Virtualization: A Survey, Some Research Issues and Challenges,” *IEEE Commun. Surveys Tuts.*, pp. 1–24, 2014.

[5] A. Basta, W. Kellerer, M. Hoffmann, H. J. Morper, and K. Hoffmann, “Applying NFV and SDN to LTE Mobile Core Gateways; The Functions Placement Problem,” in *ACM AllThingsCellular*, 2014, pp. 33–38.

[6] D. H. Luong, H. T. Thieu, A. Outtagarts, and Y. Ghamri-Doudane, “Cloudification and Autoscaling Orchestration for Container-Based Mobile Networks toward 5G: Experimentation, Challenges and Perspectives,” in *Proc. IEEE VTC*, 2018, pp. 1–7.

[7] X. Li, M. Samaka, A. H. Chan, D. Bhamare, L. Gupta, C. Guo, and R. Jain, “Network Slicing for 5G: Challenges and Opportunities,” *IEEE Internet Comput.*, vol. 21, no. 5, pp. 20–27, 2017.

[8] A. Kaloxylos, “A Survey and an Analysis of Network Slicing in 5G Networks,” *IEEE Commun. Std. Mag.*, vol. 2, no. 1, pp. 60–65, 2018.

[9] Q.-T. Luu, M. Kieffer, A. Mouradian, and S. Kerboeuf, “Aggregated Resource Provisioning for Network Slices,” in *Proc. IEEE GLOBECOM*, Abu Dhabi, 2018.

[10] Y. Zhu and M. Ammar, “Algorithms for Assigning Substrate Network Resources to Virtual Network Components,” in *Proc. IEEE INFOCOM*, 2006.

[11] M. Chowdhury, M. R. Rahman, and R. Boutaba, “ViNEYard: Virtual Network Embedding Algorithms,” *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 206–219, 2012.

[12] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, “Network Slicing in 5G: Survey and Challenges,” *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 94–100, 2017.

[13] A. Nakao, P. Du, Y. Kiriha, F. Granelli, A. A. Gebremariam, T. Taleb, and M. Bagaa, “End-to-end Network Slicing for 5G Mobile Networks,” *J. Inf. Process.*, vol. 25, pp. 153–163, 2017.

[14] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions,” *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 2018.

[15] R. Riggio, A. Bradai, D. Harutyunyan, T. Rasheed, and T. Ahmed, “Scheduling Wireless Virtual Networks Functions,” *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 2, pp. 240–252, 2016.

[16] P. Vizarreta, M. Condoluci, C. M. Machuca, T. Mahmoodi, and W. Kellerer, “QoS-driven Function Placement Reducing Expenditures in NFV Deployments,” in *Proc. IEEE ICC*, 2017.

[17] S. Chatterjee, M. J. Abdel-rahman, and A. B. Mackenzie, “Virtualization Framework for Cellular Networks with Downlink Rate Coverage Probability Constraints,” in *Proc. IEEE GLOBECOM*, 2018.

[18] K. Teague, M. J. Abdel-rahman, and A. B. Mackenzie, “Joint Base Station Selection and Adaptive Slicing in Virtualized Wireless Networks: A Stochastic Optimization Framework,” in *Proc. ICNC*, 2018.

[19] Y. L. Lee, J. Loo, and T. C. Chuah, “A New Network Slicing Framework for Multi-Tenant Heterogeneous Cloud Radio Access Networks,” in *Proc. ICAEES*, 2016, pp. 414–420.

[20] S. D’Oro, F. Restuccia, T. Melodia, S. Member, S. Palazzo, and S. Member, “Low-Complexity Distributed Radio Access Network Slicing: Algorithms and Experimental Results,” *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2815–2828, 2018.

[21] Y. Shi and Y. T. Hou, “Approximation Algorithm for Base Station Placement in Wireless Sensor Networks,” in *Proc. IEEE SeCON*, 2007, pp. 512–519.

[22] Q.-T. Luu, S. Kerboeuf, A. Mouradian, and M. Kieffer, “A Coverage-Aware Resource Provisioning Method for Network Slicing,” *arXiv:1907.09211 [cs.NI]*, pp. 1–14, 2019.

[23] S. S. Sun, W. Liao, and W. T. Chen, “Traffic Offloading with Rate-Based Cell Range Expansion Offsets in Heterogeneous Networks,” in *IEEE WCNC*, no. Icic, 2014, pp. 2833–2838.

[24] D. Tse and V. Pramod, *Fundamentals of Wireless Communication*, 2004.

[25] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, “Propagation Path Loss Models for 5G Urban Micro- and Macro-Cellular Scenarios,” in *Proc. IEEE VTC*, 2016, pp. 1–6.

[26] ETSI, “Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception,” *Tecnical Specification - ETSI TS 136 101 V10.21.0 (2016-04)*, 2016.