RESEARCH   **Open Access**

On the automation of RAN slicing provisioning: solution framework and applicability examples

R. Ferrús[*], O. Sallent, J. Pérez-Romero and R. Agustí

Abstract

Network slicing is a fundamental feature of 5G systems that allows the partitioning of a single network into a number of segregated logical networks, each optimized for a particular type of service, or dedicated to a particular customer or application. While support for network slicing (e.g. identifiers, functions, signalling) is already defined in the latest 3GPP Release 15 specifications, solutions for efficient automated management of network slicing (e.g. automatic provisioning of slices) are still at a much more incipient stage, especially for what concerns the next-generation Radio Access Network (NG-RAN). In this context, and consistently with the new service-based management architecture defined by 3GPP for 5G systems, this paper presents a functional framework for the management of network slicing in a NG-RAN infrastructure, delineating the interfaces and information models necessary to support the dynamic and automatic deployment of RAN slices. A discussion on the complexity of such automation follows together with an illustrative description of the applicability of the overall framework and information models in the context of a neutral host provider scenario that offers RAN slices to third party service providers.

Keywords: Radio slicing, Network slicing, 5G New Radio, RAN slice, Network slicing management

1 Introduction

5G systems target the simultaneous support of a wide range of application scenarios and business models (e.g. automotive, utilities, smart cities, high-tech manufacturing) [1]. This expected versatility comes with a high variety of requirements on network functionalities (e.g. security, mobility, policy control features) and expected performance (e.g. peak rates above 10 Gbps, latencies below 1 ms with 10^{-5} reliability, 500 km/h mobility target) that cannot always be met through a common network setting. In this respect, the support for network slicing in 5G has become a foundational requirement to allow operators to compose and manage dedicated logical networks (i.e. *network slices*) with specific functionality, without losing the economies of scale of a common infrastructure [2]. This is especially relevant for the Radio Access Network (RAN), which is the most resource-demanding (and costliest) part of the mobile

network and the most challenged by the support of network slicing [3].

A network slice can be tailored to provide a particular system behaviour (i.e. slice type) through the use of a specific control plane and/or user plane functions to best support specific service/applications domains. For instance, a user equipment (UE) for smart metering applications can be served through a network slice with radio access tailored to very small, infrequent messages and with no need to implement unnecessary functions (e.g. no mobility support). Similarly, a network slice can also be used to provide a particular tenant (i.e. an organization or business entity) with a given level of guaranteed network resources and isolation with regard to the operation of other concurrent slices. For instance, UEs/subscribers of a public safety (PS) agency can be served through a network slice that guarantees a minimum capacity during network congestion periods. The necessary capabilities for network slicing support in 5G systems have been already defined in the latest 3GPP Release 15 specifications approved in June 2018, which

* Correspondence: ferrus@tsc.upc.edu
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

include the definition of the network slice identifiers, denoted as Single Network Slice Selection Assistance Information (S-NSSAI) as well as the signalling procedures and functions necessary for network slice selection between the UEs, the next-generation RAN (NG-RAN) and the new 5G Core Network (5GC) [4, 5]. From a service perspective, 3GPP defines a network slice, identified by a S-NSSAI, as a particular behaviour delivered by a 5G network. Hence, for service delivery, UEs must register first to the 5G mobile network, which is identified based on a Public Land Mobile Network Identifier (PLMNId), and then establish a 5G connectivity service, denoted as a PDU session, associated with a given S-NSSAI offered within that network. It is worth noting that, from a QoS management perspective, one or multiple QoS forwarding treatments, denoted as 5G QoS flows, could be activated within each PDU session associated with the given S-NSSAI as per the 5G QoS model described in [4].

In the recent years, attention has been also paid to how the NG-RAN resources (e.g. *spectrum*, gNB functionality) can be split and configured for the implementation of the slices, which is not covered by the specifications. In this regard, the implementation of RAN slicing has been studied from multiple angles, ranging from the use of virtualization techniques and programmable platforms with slice-aware traffic differentiation and protection mechanisms (e.g. see [6–8]) to algorithms for dynamic radio resource allocation to slices (e.g. see [9–11]) and placement and configuration of the logical/physical resources for the deployment of end-to-end slices satisfying diverse service requirements (e.g. see [12]). Indeed, in our previous work [13], we analysed the RAN slicing problem in a multi-cell network in relation to how Radio Resource Management (RRM) functionalities can be used to properly share the radio resources among slices, while in [14], we defined a set of vendor-agnostic configuration descriptors that can be used to characterize the features, policies and resources to be put in place across the radio protocol layers of a NG-RAN node for the realization of concurrent RAN slices. Still in the context of the resource allocation problem, it is worth noting that several works have considered scenarios with different operators or tenants sharing the same wireless infrastructure (e.g. see [15, 16]). Moreover, the fulfilment of Service Level Agreements (SLA) via slice-aware Radio Resource Management has been also addressed [17] as well as the introduction of trading and pricing techniques in the slice resource allocation to cope with the business dimension [18].

However, less attention has received so far the implementation of solutions for the orchestration and management of network slicing in the NG-RAN, including automated and business agile provisioning of slices, which is still a challenging research area. In this respect,

a network slice lifecycle management solution for end-to-end automation across multiple resource domains was proposed in [19], including the RAN domain for completeness but not addressing it in details. More focused on the RAN domain, [20] introduced the notion of an on-demand capacity broker that allows a RAN provider to allocate a portion of network capacity for a particular time period and [21] provided some insight on the need to extend the current RAN management frameworks to support network slicing. Further progressing on this topic, in [22], we introduced a functional framework for the management of network slicing for a NG-RAN infrastructure, identifying the necessary information models and interfaces to support the dynamic provisioning of RAN slices. More recently, specifications for a new service-based overall management architecture for 5G systems and network slicing have been concluded by 3GPP as part of Release 15 specifications [23, 24], though the specifics of RAN slicing management have not been developed.

1.1 Contributions/methods

In this context, this paper extends our previous work in [22] by revisiting the proposed functional framework for slicing management in the NG-RAN so that the recently specified principles and features of the new service-based management architecture in 3GPP Release 15 specifications are incorporated. In addition, this paper proposes the specific management object classes and attributes that can be used for the provisioning of RAN slices and provides further insight on the applicability of the overall functional framework and information models in an illustrative NG-RAN deployment scenario. In particular, it is shown how the information models are used to represent the manageable aspects of a sliced NG-RAN operated by a neutral host provider and how the proposed functional framework operates through two examples: one illustrating the provisioning of a new RAN slice and another describing how the configuration of already activated RAN slices is modified in response to traffic demand variations.

The rest of the paper is organized as follows. Section 2 outlines the new service-based management architecture introduced by 3GPP. Section 3 describes the network slicing management architecture for the NG-RAN, including the elaboration on the information models. Section 4 provides a discussion on the complexity of the provisioning automation. Section 5 addresses the applicability examples, and, finally, Section 6 draws the concluding remarks.

2 3GPP work on the management of network slicing

3GPP Release 15 specifications come with a new management architecture for 5G systems based in the

service-oriented paradigm. Under such approach, the management of the 3GPP network is provided by *management services*. The management services can be produced by any entity, such as the network functions (NFs) of the 5G network, or the different network management functions that form part of the Operations support systems (OSS). The management services produced by an entity can be, for example, provisioning (i.e. configuration) management services, performance management services and/or fault supervision services. The management services can be consumed by another entity, which can in turn produce (expose) the service to other entities. The management services are built based on a combination of (1) Application Programming Interfaces (APIs) defining the management operations and/or notifications agnostic of managed entities; (2) information model definitions for representing the manageable characteristics of the managed entities (so-called Network Resource Models [NRMs]); and (3) other varied management information used to convey performance and fault management information of the managed entities (e.g. file formats for reporting performance data). For the implementation of the APIs, RESTful HTTP-based solutions are specified. Note that these management services are indeed a replacement of the so-called Integration Reference Point (IRP) information services (e.g. Itf-N interfaces [25]) used in prior 3GPP releases.

The specifications of this new 3GPP management architecture are organized around (1) the network concepts, use cases and requirements for management of 5G networks and network slicing (TS 28.530); the architectural framework and the functional and protocol specifications of generic management services (TS 28.533, TS 28.532); (3) the operations and procedures for network and network slicing provisioning management services (TS 28.531); and (4) the information model definitions for New Radio (NR) and NG-RAN, 5GC and network slice NRMs (TS 28.540, TS 28.541), performance management and assurance data (TS 28.550, TS 28.552, TS 28.554) and fault supervision data (TS 28.545).

A central aspect for the specification of the network slicing management system is the identification of the phases and high-level tasks involved in the management aspects of a Network Slice Instance (NSI), as illustrated in Fig. 1. A NSI represents the implementation perspective of a network slice. As such, a NSI is defined as the set of network function instances and the required resources (e.g. compute, storage and networking resources) that are deployed to serve the traffic associated with one or several S-NSSAIs. A NSI is composed of one or several Network Slice Subnet Instance(s) (NSSI(s)). For example, one NSI can be formed by a NSSI with the RAN functions and another NSSI with
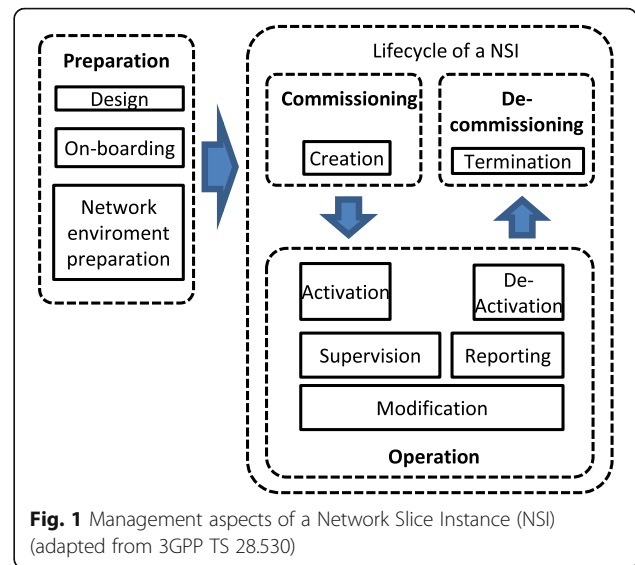

Fig. 1 Management aspects of a Network Slice Instance (NSI) (adapted from 3GPP TS 28.530)

the 5GC functions. As depicted in Fig. 1, the lifecycle management (LCM) of NSIs (applicable also to NSSIs) consists of four phases, with one or several tasks in each of them. The preparation phase includes all the tasks that are required to be done before a NSI can be created (e.g. network slice template design and on-boarding). The commissioning phase includes the creation task, in which the necessary resources are allocated and configured. The operation phase includes the activation, supervision, performance reporting (e.g. for KPI monitoring), modification and de-activation of the NSI. When no longer needed, a NSI may be decommissioned (i.e. termination task) and resources released or reconfigured accordingly. Within all these works, no specific aspects related to the management of slicing in the NG-RAN are covered with the exception of some attributes included in the NRM models that will be discussed in further detail in the next section.

3 Functional framework for management of network slicing in the NG-RAN

The operator of a NG-RAN infrastructure shall be able to flexibly deploy and operate a number of concurrent RAN slices that best serve its business needs. This includes the management capabilities for the creation, modification and termination of RAN slices in an automated and business agile manner, allowing the operator to speed up deployment time as well as flexibly adapt the allocated resources and configuration of the RAN slices to keep track of the scenario dynamics (e.g. dynamic scaling of the RAN slice capacity according to customer needs). Building upon the principles of the new 3GPP management architecture described in Section 2, a plausible functional framework for the realization of the network slicing management within the NG-RAN is depicted in Fig. 2.

Fig. 2 Architectural framework for the management of RAN slices
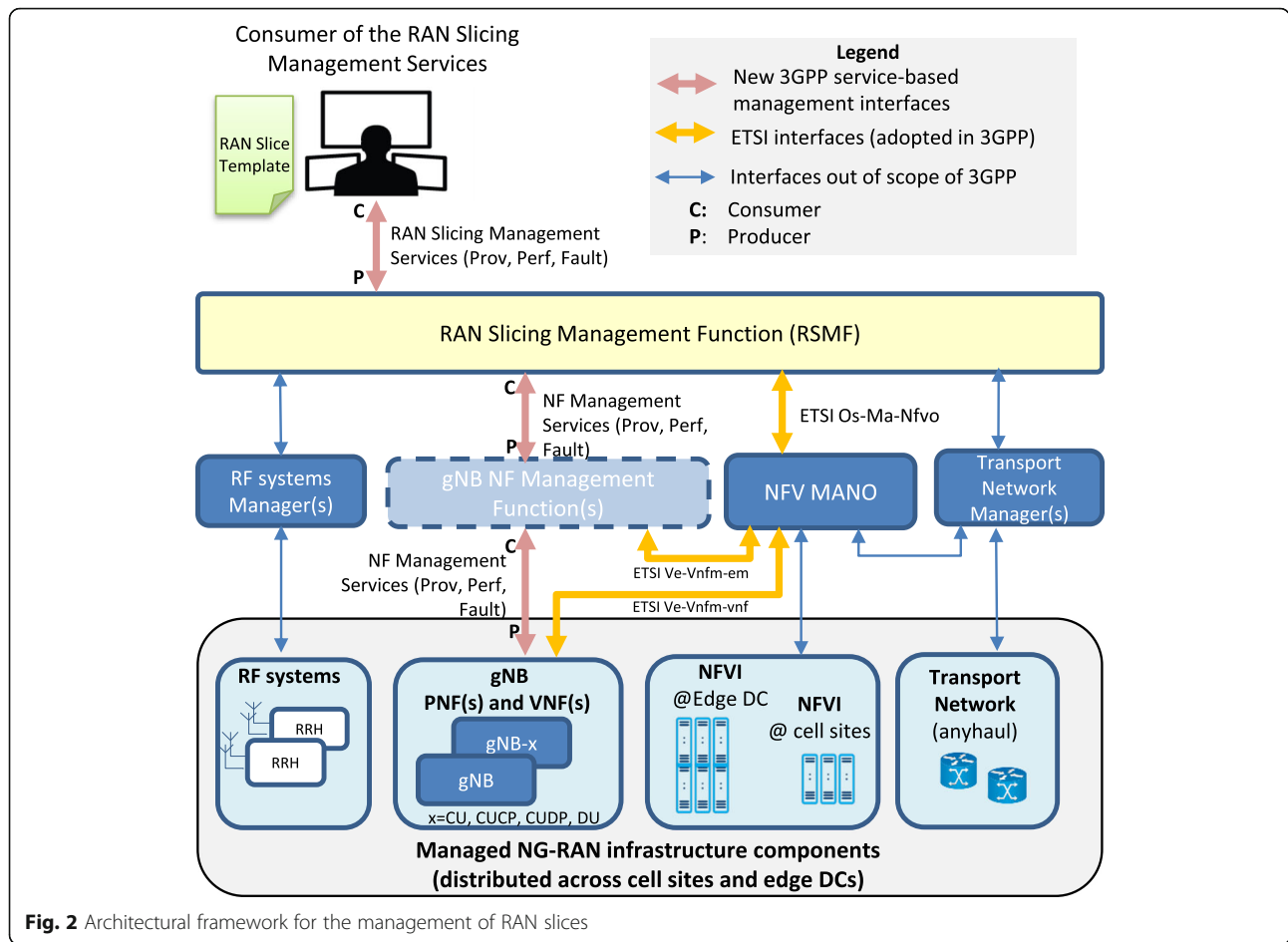
3.1 Managed NG-RAN infrastructure

The bottom part of Fig. 2 illustrates the main components of a NG-RAN infrastructure that will be affected by the management functions for RAN slicing. In a general setting, a NG-RAN infrastructure is to be formed by a collection of cell sites (i.e. NG-RAN Points of Presence [PoPs] where the antennas are located) and edge data centres (DCs), interconnected by means of a fibre and/or wireless-based transport network (TN). Focusing on the 5G NR interface, the NG-RAN functionality for 5G NR access can be implemented in the form of a single network function (NF), called gNB. A gNB hosts the full radio protocol stack functionality and supports 3GPP standardised interfaces for the interaction with the 5GC (i.e. NG interfaces) and with other gNBs (i.e. Xn interfaces). To introduce modularity and support different deployment options, 3GPP has also standardised interfaces (e.g. F1, E1) that are used to split the gNB into several NFs. In particular, F1 defines the split between a gNB central unit (gNB-CU) NF for upper protocol layer processing and a gNB distributed unit (gNB-DU) NF for lower protocol layer processing, while E1 is used to further split the gNB-CU into a gNB-CUCP for control

plane (CP) processing and gNB-CUDP for data plane (DP) processing. All these gNB-x NFs, with the full or part of the gNB functionality, can be implemented in dedicated hardware appliances (referred to as Physical Network Functions [PNF] in ETSI network functions virtualisation [NFV] terminology) or as virtualized network functions (VNFs) running on general-purpose hardware (e.g. NFV infrastructure [NFVI]), as could be the most likely case for the gNB-CU. The placement of the gNB and gNB-CUx NFs, virtualised or not, can be done at the cell sites or centralised in the edge DCs. In turn, gNB-DU functionality is most likely to be co-located with the RF systems (antennas, Remote Radio Heads [RRHs]) in cell sites. From a deployment perspective, the implementation of a RAN slice, denoted in the following as RAN Slice Instance (RSI), is likely to admit different possibilities as to how the NG-RAN components are deployed, connected and configured, including how radio spectrum is used by gNB to serve multiple RSIs (e.g. RSI-dedicated radio channels or channels shared by several RSIs) [13, 14]. In this regard, the 5G NR physical layer comes with new features such as the support of bandwidth parts (BWPs) that could facilitate

the implementation of RSIs with diverse physical layer requirements. The BWP feature in 5G NR allows for UEs to operate only in a portion of the channel, reducing UE processing complexity and power consumption [26]. Clearly, this could be exploited to implement slices for low-cost IoT devices. Moreover, each BWP can be associated with a specific numerology (i.e. subcarrier spacing and slot duration), which can be selected to meet certain slice-specific communication requirements, such as low latency [27].

3.2 Management functions and interfaces

Focusing now on the management components above the NG-RAN infrastructure illustrated in Fig. 2, the core functionality consists of a set of management functions, collectively referred to as *RAN Slicing Management Function (RSMF).* The RSMF is in charge of the LCM of the RSIs, including capabilities for the creation, modification and termination of RSIs. In this regard, the RSMF is expected to expose a set of management services for provisioning, performance monitoring and fault management of the RSIs. For example, the RSMF could offer services to request the creation of a RSI based on predefined templates. It could also offer services to create measurement jobs for collecting performance data of the operational RSIs. Accordingly, and in line with the terminology and service-based management concepts adopted by 3GPP, the RSMF plays the role of a producer of management services that can be accessed by one or multiple management service consumers. A consumer of the RAN slicing management services is likely to be the operator itself, that is, operator staff that access the RSMF through e.g. web-based graphical interfaces for RSI provisioning. Another consumer can be a management system in charge of end-to-end network slicing management, which in this case could consume the RSMF provisioning services via the 3GPP standardized RESTful HTTP-based APIs.

On the other hand, in order to interact with the underlying infrastructure components and carry out the LCM of the RSIs, the RSMF has to be able to consume the management services provided by the diverse management systems specific to each of the function types composing the NG-RAN infrastructure (i.e. gNB functions, NFVI, RF systems and TN nodes). In this respect, for the management of the gNB NFs, 3GPP Release 15 includes the NRM definitions that, apart from attributes used for configuring the operation of the NR cells (e.g. cell identifiers, channel frequencies), contain also attributes for the configuration of the network slicing behaviour of gNB NFs. In this way, the gNB NFs, irrespective of whether they are implemented as PNF or VNF, can expose management services for provisioning of network slicing by the RSMF based on the standardised gNB

NRM. More details on the gNB NRM definitions for the gNB NFs are covered in the next subsection.

As reflected in Fig. 2, the interaction between the gNB NFs and the RSMF might not necessarily rely on direct interfaces but it can be realised via intermediate gNB NF management functions (e.g. vendor-specific operation, administration and maintenance [OAM] tools). These intermediate management systems would actually be the ones exposing the management services using the standardized NRM definitions to the RSMF while proprietary interfaces and vendor-specific data models could be used for the direct interaction with the gNB NFs.

With regard to the management of the NFVI resource-level aspects of the gNB NFs delivered as VNFs, the RSMF could directly adopt the interfaces specified under the ETSI NFV management and orchestration (MANO) standards [28]. Indeed, the ETSI NFV MANO framework allows for the management of a distributed NFVI (e.g. lightweight NFVI at cell sites and more powerful NFVI in edge DCs), encompassing the LCM of individual VNFs as well as of groups of interconnected VNFs and PNFs (i.e. network services [NS]). In this respect, the RSMF layer would trigger the creation, modification or tear down of NSs and VNFs using ETSI standardised information models. In particular, ETSI NFV network service descriptors (NSDs) would be used as deployment templates to describe, among others, the components of a NS (e.g. VNFs, PNFs, virtual links [VLs]), their relationship (e.g. forwarding graphs) and diverse deployment attributes (e.g. redundancy, scaling configuration).

With regard to the management of the TN connectivity, the RSMF can benefit from the adoption of the protocols and information models being consolidated for connectivity management in the networking community, such as the IETF Network Configuration Protocol (NETCONF) and YANG data models specified for traffic-engineered networks [29]. Finally, but not less important, the RSMF could also have access to the management tools of the RF systems, which might expose management services through, e.g. web services APIs that can be integrated and consumed by the RSMF for RF system configuration (e.g. operating region and bands, transmit power allocation) [15].

While the interaction of the RSMF is given by the capabilities of its interfaces as discussed above, the internal design of the RSMF may admit a wide range of implementations with more or less complexity. For instance, the translation between the slice requirements included within the provisioning templates and the configuration parameters of the NG-RAN components may be implemented based simply on pre-provisioned/static mapping rules stored within the RSMF or rely on the introduction within the RSMF of SLA conformance monitoring and dynamic mapping functions. In another

example, the RSMF may include or not traffic forecasting functions in order to carry out admission control of new RSIs or RSI capacity modifications [30]. A discussion of the internal design of the RSMF falls beyond the scope of this paper. For the interested reader, a description of a specific RSMF implementation compliant with the framework presented here can be found in [33].

3.3 Information models

As previously introduced, the manageable characteristics and behaviour of the managed entities in 3GPP systems are represented by means of NRM definitions. NRMs are specified in Unified Modelling Language (UML) as a collection of information object classes (IOC) characterised by a set of attributes, together with different types of associations between the IOCs (e.g. inheritance, dependencies) and specific data types for describing the content of the attributes.

Based on the NRM definitions in 3GPP Release 15, Fig. 3 provides a simplified representation of the IOCs that could be leveraged to represent a RAN deployment with RSIs. In particular, this representation would allow (1) the provisioning of RSIs via the RAN management

services exposed by the RSMF (see Fig. 2) and (2) the configuration of the slicing features supported within the gNB functionality for the realization of the RSIs.

As depicted in Fig. 3, the root element of the whole construct is the *Subnetwork* IOC, which would represent the whole (o part of the) RAN domain. Within a subnetwork, the *NetworkSliceSubnet* IOC can be used to account for the properties of a RSI. These properties are defined in a data type known as *SliceProfile*. On the other hand, a subnetwork also includes the set of entities to be managed (e.g. *ManagedElement* IOC), each one providing specific functionality in the RAN (e.g. *gNBDU-Function* IOC, *gNBDUFunction* IOC). In particular, there is a pair of IOCs named as *NRCellCU* and *NRCellDU* IOC within a *gNBFunction* IOC, to which we refer simply as *NRCell* IOC in the following, that will be used to represent the characteristics of each of the cells handled by the gNB. Such a *NRCell* IOC, besides capturing the cell operational aspects such as the Absolute Radio Frequency Channel Number (ARFCN) and channel bandwidth, includes a number of attributes to manage the slicing features supported in the cell (e.g. S-NSSAI lists, RRM policies to split the cell resources between
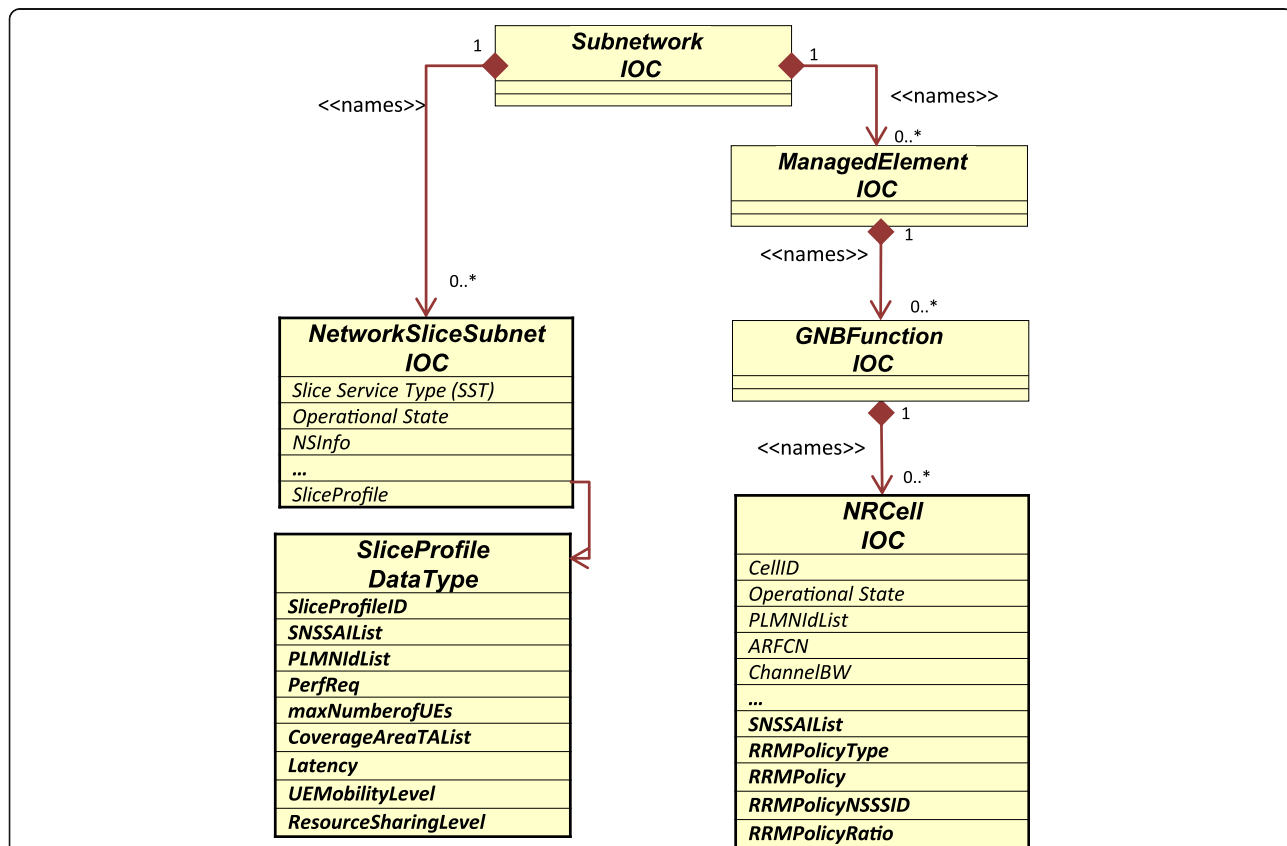
Fig. 3 NG-RAN Network Resource Model (NRM) for RAN slicing management. Arrows with an ending diamond and tagged with <<names>> represent associations between IOCs used to specify name containment composition (e.g. subnetwork#A->ManagedElement#B). Simple arrows indicate associations between attributes and their data types

the RSIs supported in the cell). A brief description of the most relevant attributes of the aforementioned IOCs for configuring the slicing capabilities is provided in Table 1.

From Table 1, it can be seen that the slicing modelling established by 3GPP for the *NetworkSliceSubnet* IOC with the *SliceProfile* attribute consists of a set of network identifiers associated with the slice (i.e. list of S-NSSAI(s) and PLMNId(s)) together with a set of high-level attributes that determine the expected slice behaviour and configuration. For example, the capacity of a slice can be limited by setting the '*maxNumberofUEs*' attribute and the coverage of the slice can be established by defining on which '*Tracking Areas*' the slice can be selected. With regard to the expected performance, the attribute 'PerfReq' contains the list of requirements, which indeed depends on the slice/service type (SST) that is encoded as part of the slice identifier S-NSSAI. Currently, 3GPP has standardised three SSTs: enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and Massive IoT (MIoT). As an example, for slices of type SST=URLLC, the 'perfReq' requirements list may include parameters such as the expected end-to-end latency, jitter, availability, reliability, data rate, payload size, traffic density, connection density and service area dimension.

On the other hand, with regard to the modelling of the slicing feature at NR cell level, the 3GPP model basically specifies the list of S-NSSAI(s) to be supported in a cell and the RRM policy to be enforced for the split of resources between multiples slices. In this regard, the attribute *RRMPolicyType* is used to select the RRM policy to be applied. As captured in Table 1, 3GPP has defined so far 3 different RRM policies: (1) *RRMPolicy* (*RRMPolicyType* = 0), *RRMPolicyRatio* (*RRMPolicyType* = 1) and *RRMPolicyRatio2* (*RRMPolicyType* = 2). The first one is just left as implementation dependent. The second one basically defines the percentage of Physical Resource Blocks (PRBs) to be used per slice in average over time. The third one, which is based on the second one but a bit more sophisticated, allows for the definition of maximum and minimum values for the percentage of PRBs to be used per slice as well as different types of quota to allocate resources: strict quotas, where resources are strictly usable only for the defined slices, and float quotas, where resources can be used by other slices when the defined slices do not need them. For the interested reader, further details can be found in [31].

Note, however, that the 3GPP model leaves a room for extensions in case that more fine-grained resource management becomes necessary. This could be accomplished for instance by using values higher than 2 for the *RRMPolicyType* parameter and defining new attributes to represent more complex RRM policies (e.g. policies intended to guide the operation of other RRM functions such as admission control and that can be specified

Table 1 Attributes for the management of the network slicing features

Attributes included in the SliceProfile DataType	
Attribute	Explanation
SliceProfileId	A unique identifier of the slice profile.
SNSSAIList	Set of supported S-NSSAI(s) in the NSSI. Each S-NSSAI is comprised of a SST (Slice/Service type) and an optional SD (Slice Differentiator) field. 3GPP has standardised three SST: enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and Massive IoT (MIoT).
PLMNIdList	Set of PLMN identifier(s) associated with the slice subnet.
PerfReq	It specifies the requirements to the NSSI in terms of the scenarios defined in 3GPP TS 22.261, such as experienced data rate, area traffic capacity (density) and information of UE density.
maxNumber ofUEs	It specifies the maximum number of UEs that may simultaneously access the slice subnet.
coverageArea TAList	List of tracking area(s) where the slice subnet can be selected.
Latency	It specifies the packet transmission latency (millisecond) through the RAN, CN and TN part of 5G network
UEMobility Level	It specifies the mobility level of UE accessing the slice subnet. Allowed values: stationary, nomadic, restricted mobility, fully mobility.
Resource SharingLevel	It specifies whether the resources to be allocated to the Network Slice Instance may be shared with another Network Slice Instance(s). Allowed values: shared, non-shared.

Attributes related to network slicing feature included in NRCell IOC	
Attribute	Explanation
SNSSAIList	Set of supported S-NSSAI(s) in the NR cell.
RRMPolicyType	Type of the RRM policy. Allowed values: 0 .. 65535. From these, three values are specified: The value 0 denotes the use of the RRMPolicy attribute. The value 1 denotes use of the RRMPolicyNSSIId and RRMPolicyRatio attributes. The value 2 denotes use of the RRMPolicyRatio2.
RRMPolicy	Represents RRM policy which includes guidance for split of radio resources between multiple slices the cell supports. The RRM policy is implementation dependent.
RRMPolicyNSSIId/ RRMPolicyRatio	Pair of attributes used to specify a percentage of Physical Resource Blocks (PRBs) to the corresponding slice, in average over time. The sum of the values shall be less or equal 100. Average time is implementation dependent.
RRMPolicyRatio2	Like the RRMPolicyRatio, this policy allows specifying the percentage of radio resources to be allocated to the corresponding SNSSAIList. However, this can be done with more details, including the selection of strict or flexible (i.e. float) sharing polices among slices as well as the definition of maximum, minimum and safety margin values for the percentages of radio resources (i.e. attributes referred to as quotaType, RRMPolicyMaxRatio, RRMPolicyMarginMaxRatio, RRMPolicyMinRatio and RRMPolicyMarginMinRatio).

separately for different QoS class identifiers [14]). A practical realization of an extended RRM policy is demonstrated in [33].

4 Complexity of RAN slicing provisioning automation

The feasibility to automate the provisioning of a RSI within the presented management framework depends on the elements and configurations within the NG-RAN infrastructure that are impacted. Hence, as depicted in Fig. 4, if the realization of a new RSI involves the installation of additional infrastructure or hardware modifications (e.g. adding new cell sites to extend coverage or increasing capacity or new RRHs to support new RF bands), several actions with human intervention have to be necessarily conducted before the RSI can be provisioned in an automated manner through the RSMF-exposed interfaces. Indeed, the problem of coping with infrastructure changes is not specific to network slicing but also applies to other network management (e.g. self-planning and network scaling). Otherwise, automation becomes feasible but with varying degree of complexity.

Automation is anticipated to be easily achievable with relatively low complexity if the provisioning of a RSI does not modify parameters that impact on the cell footprint, but it is mainly based on the configuration of parameters such as network and network slice identifiers within cells (e.g. S-NSSAI(s), PLMNId(s) that can be accessed in each cell) and the RRM policies used to split the capacity among the active RSIs. A more complex case in terms of automation comes when the provisioning process also embraces the reconfiguration of cell parameters that modify to some extent the capacity/coverage of the cells (e.g. transmission power, channel bandwidth, frequency band, mobility control configuration). In this case, automation feasibility mainly relies on the support of Self-Organizing Network (SON) functionalities to re-adjust the modified cell footprint for optimised performance (e.g. re-computation of neighbouring cells, correction of coverage holes and cell overshoot situations, interference mitigation, mobility optimisation and so on). Ultimately, this reconfiguration

of the cells might also involve the activation of new cells (e.g. new RF carriers), which could be served by the already operational gNBs (e.g. a gNB typically can handle multiple cells) or require the deployment of new VNF instances (e.g. instantiation of a new gNB VNF to handle the new cell).

5 Results and discussion: applicability examples

This section describes how the overall functional framework and the information models presented in Section 3 can be used for the lifecycle management of RAN slices. To that end, an illustrative NG-RAN infrastructure with a number of deployed RSIs is described first, showing how their manageable characteristics are captured into the previously introduced 3GPP NRM models and ETSI NFV NSDs. On this basis, two representative examples of RAN slicing lifecycle management procedures are illustrated, namely 'Activation of a new RSI' and 'Scaling up an operational RSI'. These examples have been chosen among the possible set of procedures (i.e., creation, activation, supervision, reporting, modification, de-activation and termination) in order to highlight the automation complexity associated with the procedure as discussed in Section 4. In this way, the first example shows the case that a new RSI is activated without modifying the cell footprint. This fits under the case of 'Configuration of slice identifiers and cell capacity guarantees/limits per RSI', showing the least automation complexity. In turn, the second example shows the case that the modification of a RSI impacts on the cell footprint to the point that a new cell has to be activated along with a new instance of a gNB VNF. Therefore, this case fits under the cases 'New cell activation' and 'Cell reconfiguration of transmission parameters' discussed in Section 4.

5.1 Illustrative configuration

Let us consider a NG-RAN infrastructure like the one illustrated in the bottom part of Fig. 5 with three PoPs (e.g. two cell sites and one edge DC) hosting the sort of infrastructure components described in Section 3.1 (i.e. RF systems, gNB VNFs/PNFs, NFVI and TN nodes).
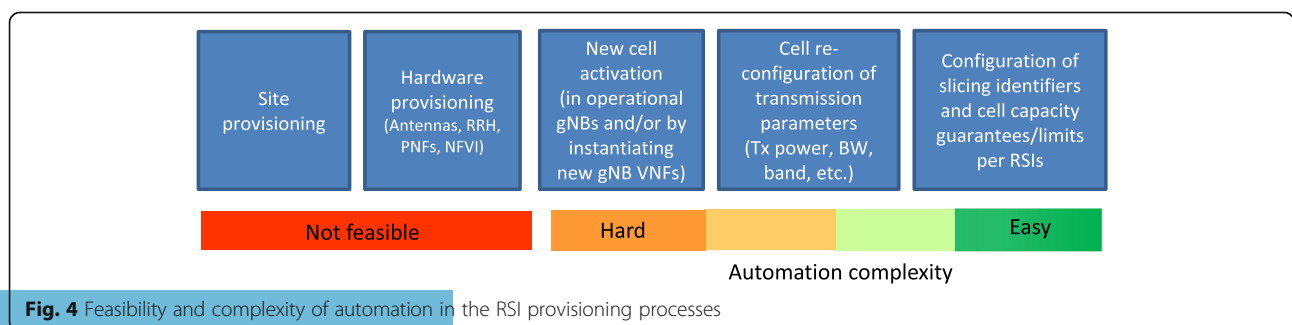

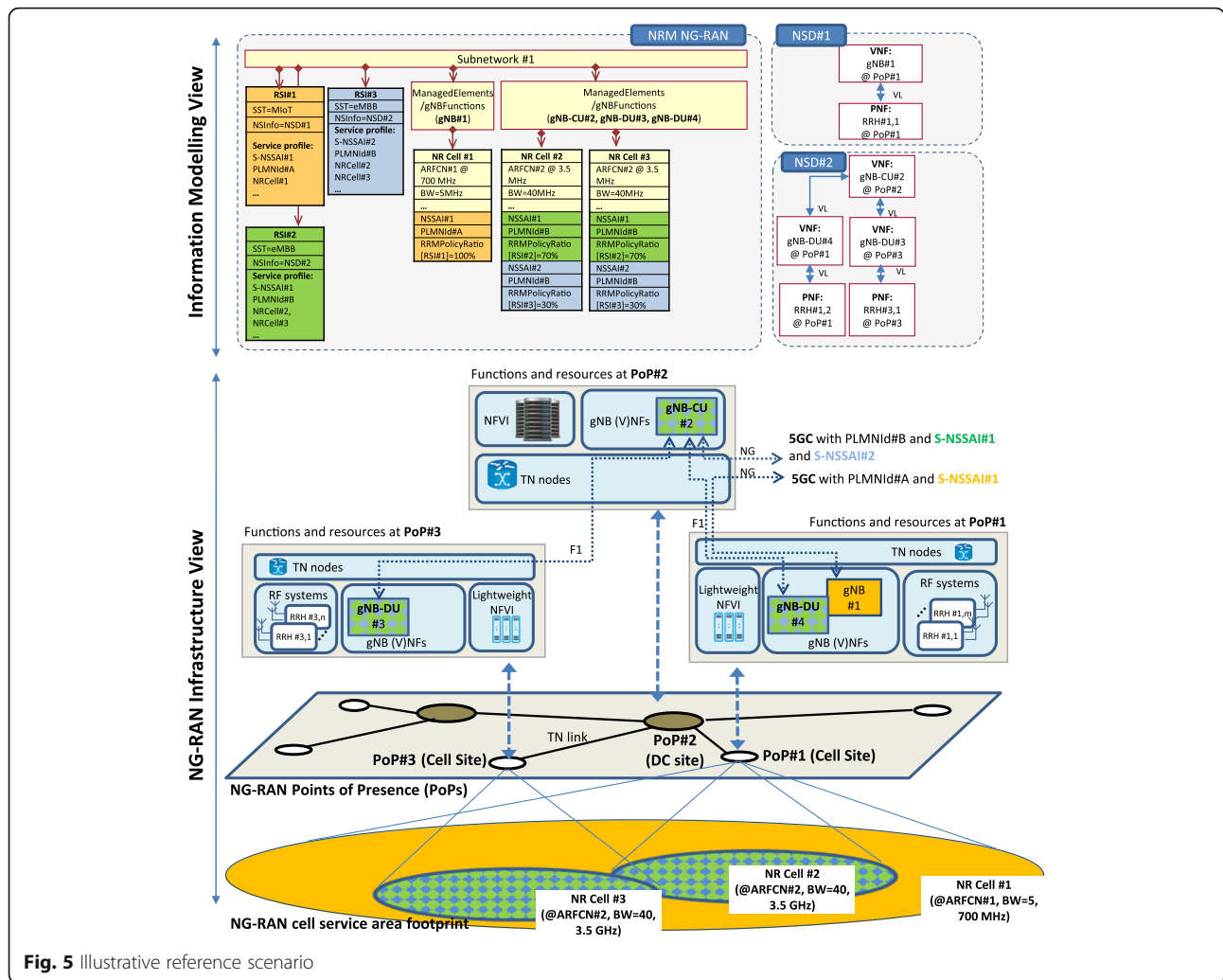**Fig. 4** Feasibility and complexity of automation in the RSI provisioning processes

Fig. 5 Illustrative reference scenario

The scenario in Fig. 5 may represent the deployment of a neutral host provider that offers access services to mobile network operators (MNOs) and other service providers (SP) in the form of RSIs. Each of these customer MNOs/SPs will operate its own 5GC, which would be connected to the neutral host provider infrastructure. Let us consider that 3 RSIs have been configured at a given stage. In particular, let us assume that RSI#1 is used by an IoT SP to offer MIoT services to utilities (e.g. utilities' metering application). Such MIoT services are accessed through S-NSSAI#1 in a private 5GC operated by the IoT SP and identified as PLMNId#A. On the other hand, let us consider that RSI#2 and RSI#3 are both configured to offer access to the public network of a MNO identified by PLMNId#B. In this case, it is considered that RSI#2 is used for eMBB services offered to the general public through S-NSSAI#1 in PLMNId#B and RSI#3 is used for mission-critical (MC) services restricted to public safety (PS) agencies through S-NSSAI#2 in PLMNId#B.

In terms of implementation, as highlighted in orange colour in the infrastructure view of Fig. 5, RSI#1 is implemented through a single macro cell (NR Cell#1) that operates in a radio channel (ARFCN#1) with 5 MHz bandwidth at 700 MHz band. This macro cell is handled by a virtualised gNB (gNB#1) deployed at the NFVI in PoP#1. Let us assume that such a configuration fulfils the requirements of the IoT SP, which counts with multiple IoT devices spread in the area covered by NRCell#1 that generate a very low aggregated throughput. On the other hand, the implementation of RSI#2 and RSI#3 is considered to be dictated by the need to serve much higher bandwidth applications and traffic demands, which results in this illustrative scenario in the deployment of two microcells (NR Cell #2 and #3) operated in a 40 MHz bandwidth radio channel (ARFCN#2) at the 3.5 GHz band. As illustrated in Fig. 5 now using green and blue colours for RSI#2 and RSI#3, respectively, it is considered that the two microcells are implemented with a centralised gNB-CU component in order to facilitate

coordination mechanisms among the two cells. In particular, the two microcells are handled by gNB-CU#2 VNF deployed at PoP#2, gNB-DU#3 VNF at PoP#3 and gNB-DU#4 VNF at PoP#1. Note that the two microcells and the associated gNB functionally are actually coloured in both green and blue to stress that, unlike RSI#1, all these resources are shared among RSI#2 and RSI#3.

From a management perspective, an illustration of the NRM and NSD models accounting for the configuration of the 3 RSIs is provided in the upper part of Fig. 5. Focusing first on the NS descriptors, NSD#1 accounts for the functions and the underlying NFVI resources used to implement RSI#1, as indicated by the attribute *NSInfo* in the RSI#1 IOC instance. NSD#1 includes in this case the gNB#1 VNF instantiated at PoP#1, a RRH (represented as a PNF in ETSI models) and the interconnection among the two (represented as a VL also as per ETSI models). Likewise, NSD#2 contains the functions and NFVI resources used to implement RSI#2 and RSI#3, which includes the gNB-CU#2 VNF instantiated at PoP#2, gNB-DU#3 VNF at PoP#3 and gNB-DU#4 VNF at PoP#1, together with the corresponding RRH PNFs and interconnection VLs. In this last case, as NSD#2 is shared among RSI#2 and RSI#3, the split of the radio capacity among the two slices is done through the configuration of the *RRMPolicyRatio* attribute of the IOC instances for NRCell#2 and NRCell#3 within the NG-RAN NRM model. As illustrated in Fig. 5, a capacity distribution of 70% and 30% has been configured for, respectively, the commercial eMBB (RSI#2) and the MC services (RSI#3). In contrast, note that the IOC instance of NRCell#1 in the NRM representation, which is used exclusively for RSI#1, is configured with a *RRMPolicyRatio = 100%*.

5.2 Activation of a new RSI
Let us assume that in a subsequent stage, another IoT SP with its own 5GC with identifier PLMNId#C wants to get some access capacity in the area covered by the neutral host provider infrastructure. Let us also consider that, after analysing the requirements at planning and business levels, the neutral host provider decides to implement a new RSI (RSI#4) over the capacity offered by the existing NRCell#1, since it is assessed during the planning process that this cell already satisfies the coverage requirements of the new IoT SP and shows very low utilisation levels by RSI#1, which ensures sufficient capacity is available for the new demand. At this point, the neutral host provider will use the RAN provisioning interface illustrated in Fig. 2 to create RSI#4, whose configuration is reflected in Fig. 6. For the sake of making easier the comparison between Figs. 5 and 6, all changes in Fig. 6 with respect to the initial configuration are highlighted in red colour. Focusing on the configuration of the new RSI#4, it can be seen that NRCell#1 IOC

instance is extended with new the identifiers S-NSSAI#1 and PLMNId#C and the *RRMPolicyRatio* attributes of both RSI#4 and RSI#1 are set to 50%. In this case, it is worth noting that no changes are required in NSD#1 to support the new RSI#4. From an automation perspective, this is an example of a process with low automation complexity since no modifications are required in the cell footprint.

5.3 Scaling up an operational RSI
Let us now consider that there is a need to increase the capacity provided by RSI#2 (commercial eMBB) in order to meet a temporary hotspot situation. Like in the previous example, the capacity increase requirement is first addressed at business and planning levels in order to determine the most convenient implementation. In this case, let us assume that the planning decision is to deploy a new cell (NR Cell#4) at PoP#3 on a new radio channel (ARFCN#3 at 3.7 GHz) with a bandwidth of 80 MHz. Accordingly, the provisioning process leads to the new configuration for RSI#2 illustrated in Fig. 6. As depicted in the figure, a new gNB-DU#5 VNF is instantiated at PoP#3 in order to handle the new NRCell#4. This change is also reflected in the modified NSD#2, which has been modified to include the new gNB-DU and its interconnections. With regard to the NRM attributes, note that NRCell#4 is configured only with the identifiers associated with RSI#2 (S-NSSAI#1 and PLMNId#B) and with *RRMPolicyRatio* set to 100%. In this case, SON functionalities shall be necessarily in place to automatically re-adjust the modified cell footprint and relationships between previous NRCell#2 and NRCell#3 and the new NRCell#4.

6 Concluding remarks
The NG-RAN is expected to be a flexible and versatile access platform where RAN slices can be dynamically created in response to changing business needs, thus radically transforming the legacy 3G/4G view of a static 'one-size-fits-all' access network. In order to achieve this goal, advanced management systems for automated RAN slicing provisioning are central components that need to be conceived, designed and implemented. Indeed, RAN slicing provisioning systems are anticipated to be important components within the future 5G network OSS, which are undergoing an important transformation towards service-based, model-driven architectures and technologies capable to rapidly automate new services and support complete lifecycle management.

In this respect, and consistently with the new service-based management architecture standardized in 3GPP Release 15 for the management of 5G systems and network slicing, this paper has described a plausible functional framework for the realization of network slicing management in the NG-RAN, discussing the functions,
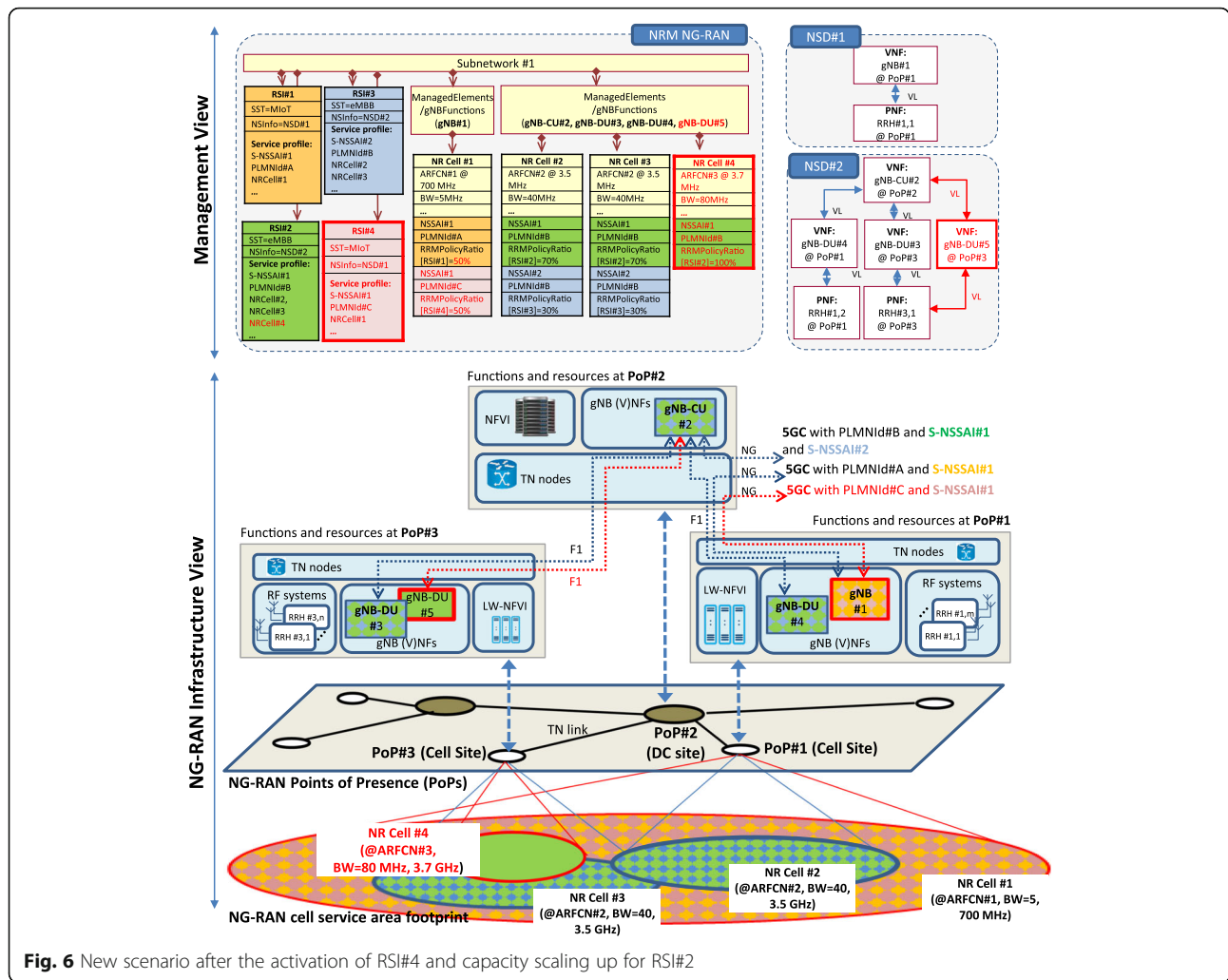
Fig. 6 New scenario after the activation of RSI#4 and capacity scaling up for RSI#2

interfaces and information models that would be necessary for the automation of the RAN provisioning process. Special attention has been paid to the use of the information models for representing the slicing features of a NG-RAN infrastructure from a management perspective.

From the analysis of the complexity of the RAN slicing provisioning process under the proposed framework, it has been anticipated that automation is achievable with relatively low complexity if the provisioning process relies centrally on the configuration of cell parameters such as the network slice identifiers and the RRM policy attributes used to dictate the capacity split of the cells per RSI. Otherwise, automation feasibility is highly dependent on the support of SON functionalities able to properly handle potential cell footprint modifications that may be triggered from the provisioning processes, including the automatic deployment of new cells.

To gain further insight into the applicability of the functional framework and information models as well as on the automation feasibility of the RAN provisioning process, the

representation from a management perspective of an illustrative NG-RAN infrastructure with a number of deployed RSIs has been analysed, together with two examples that show the creation and modification of RAN slices.

A proof-of-concept implementation of the proposed functional framework and information models is being developed based on open-source RAN distributions and the 5G-EmPOWER platform [32]. For the interested reader, a description of the implemented testbed can be found in [33]. In particular, the work in [33] presents the design of a software-defined RAN (SD-RAN) testbed with network slicing capabilities and showcases the operation of management services for the provisioning of RAN slices along with slice-aware RRM functions used to split the radio resources between RAN slices based on configurable RRM policy attributes. Future work is envisaged to implement RRM algorithms able to exploit the full potential of the RRM policy attributes and test their operation under more complex scenarios with multiple cells and a mix of applications demanding diverse QoS characteristics.

Abbreviations

3GPP: 3rd Generation Partnership Project; 5GC: 5G core; API: Application Programming Interface; ARFCN: Absolute Radio Frequency Channel Number; BW: Bandwidth; DC: Data centre; eMBB: Enhanced Mobile Broadband; ETSI: European Telecommunications Standards Institute; IOC: Information object class; IoT: Internet of Things; LCM: Lifecycle management; MANO: Management and Orchestration; MIoT: Massive IoT; MNO: Mobile network operator; NF: Network function; NFV: Network function virtualization; NFVI: NFV infrastructure; NG-RAN: Next-generation RAN; NRM: Network Resource Model; NS: Network service; NSD: ND descriptor; NSI: Network Slice Instance; NSSI: Network Slice Subnet Instance; PLMNId: Public Land Mobile Network Identifier; PNF: Physical NF; PoP: Point of presence; PRB: Physical Resource Block; PS: Public safety; RAN: Radio Access Network; RF: Radio frequency; RRH: Remote Radio Head; RRM: Radio Resource Management; RSI: RAN Slice Instance; RSMF: RAN Slicing Management Function; S-NSSAI: Single Network Slice Selection Assistance Information; SON: Self-organizing networks; SP: Service provider; TN: Transport network; UE: User equipment; URLLC: Ultra Reliable Low Latency Communications; VL: Virtual link; VNF: Virtualized NF

Authors' contributions

The views and ideas discussed in the paper are the result of joint work by all the authors. RF led the writing of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

References

1. NGMN Alliance, 5G White Paper, February 2015
2. 3GPP TR 22.864: Feasibility study on new services and markets technology enablers - network operation; stage 1 (release 15), September 2016
3. P. Rost et al., in *IEEE Communications Magazine, vol. 55, no. 5*. Network slicing to enable scalability and flexibility in 5G mobile networks (2017), pp. 72–79
4. 3GPP TS 23.501 V15.3.0, System architecture for the 5G system; stage 2 (release 15), September 2018
5. 3GPP TS 38.300 V15.3.0, NR; NR and NG-RAN overall description; stage 2 (release 15), September 2018
6. X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, S. Rangarajan, in *IEEE Communications Magazine, vol. 51, no. 7*. Radio access network virtualization for future mobile carrier networks (2013), pp. 27–35
7. C. Chang, N. Nikaein, in *IEEE Vehicular Technology Magazine, Vol. 13, No. 4*. Closing in on 5G control apps: enabling multiservice programmability in a disaggregated radio access network (2018), pp. 80–93
8. A. Ksentini, N. Nikaein, in *IEEE Communications Magazine, vol. 55, no. 6*. Toward enforcing network slicing on RAN: flexibility and resources abstraction (2017), pp. 102–108
9. E. Pateromichelakis et al., Service-tailored user-plane design framework and architecture considerations in 5G radio access networks. IEEE Access 5, 17089–17105 (2017)
10. M. Richart, J. Baliosian, J. Serrat, J.-L. Gorricho, Resource slicing in virtual wireless networks: a survey. IEEE. Trans. Netw. Service. Manag. **13**(3), 462–466 (2016)
11. P. Caballero, A. Banchs, G. de Veciana, X. Costa-Pérez, A. Azcorra, in *IEEE Transactions on Wireless Communications, Vol. 17, No. 10*. Network slicing for guaranteed rate services: admission control and resource allocation games (2018), pp. 6419–6432
12. W. Guan, X. Wen, L. Wang, Z. Lu, Y. Shen, A service-oriented deployment policy of end-to-end network slicing based on complex network theory. IEEE Access **6**, 19691–19701 (2018)
13. O. Sallent, J. Perez-Romero, R. Ferrús, R. Agusti, On radio access network slicing from a radio resource management perspective. IEEE Wirel. Commun., 166–174 (2017)
14. R. Ferrús, O. Sallent, J. Pérez-Romero, R. Agustí, On 5G radio access network slicing: radio interface protocol features and configuration. IEEE Communications Magazine **PP**(99), 2–10, Early access in IEEEXplore (2018)
15. Y.L. Lee, J. Loo, T.C. Chuah, L.-C. Wang, Dynamic network slicing for multitenant heterogeneous cloud radio access networks. IEEE Trans. Wirel. Commun. **17**(4), 2146–2161 (2018)
16. P. Caballero, A. Banchs, G. de Veciana, X. Costa-Pérez, Multi-tenant radio access network slicing: statistical multiplexing of spatial loads. IEEE/ACM Transactions on Networking **25**(5), 3044–3058 (2017)
17. B. Khodapanah, A. Awada, I. Viering, D. Oehmann, M. Simsek, G.P. Fettweis, in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring), Porto*. Fulfillment of service level agreements via slice-aware radio resource management in 5G networks (2018), pp. 1–6
18. O.U. Akguel, I. Malanchini, V. Suryaprakash, A. Capone, in *GLOBECOM 2017–2017 IEEE Global Communications Conference, Singapore*. Service-aware network slice trading in a shared multi-tenant infrastructure (2017), pp. 1–7
19. A. Devlic, A. Hamidian, D. Liang, M. Eriksson, A. Consoli, J. Lundstedt, in *2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris*. NESMO: network slicing management and orchestration framework (2017), pp. 1202–1208
20. K. Samdanis, X. Costa-Perez, V. Sciancalepore, in *IEEE Communications Magazine, vol. 54, no. 7*. From network sharing to multi-tenancy: the 5G network slice broker (July 2016), pp. 32–39
21. I. da Silva et al., in *2016 European Conference on Networks and Communications (EuCNC), Athens*. Impact of network slicing on 5G radio access networks (2016), pp. 153–157
22. R. Ferrús, O. Sallent, J. Pérez-Romero, R. Agustí, On the automation of RAN slicing provisioning and cell planning in NG-RAN. European Conference on Networks and Communications, Ljubljiana, June (2018)
23. 3GPP TS 28.530 v15.0.0 Management of network slicing in mobile networks; concepts, use cases and requirements (release 15), September 2018
24. 3GPP TS 28.533 v15.0.0, Management and orchestration; architecture framework (release 15), September 2018
25. 3GPP TS 32.102 v14.0.0: Telecommunication management; architecture, March 2017
26. J. Jeon, in *IEEE Communications Magazine, vol. 56, no. 3*. NR wide bandwidth operations (2018), pp. 42–46
27. Q. Li, G. Wu, A. Papathanassiou, U. Mukherjee, *Radio slicing* (IEEE Softwarization, 2017)
28. ETSI GS NFV-EVE 012 (V3.1.1): Network functions virtualisation (NFV) release 3; evolution and ecosystem; report on network slicing support with ETSI NFV architecture framework, 2017
29. X. Zhang et al., in *Draft-Zhang-Teas-Actn-Yang, Work in Progress*. Applicability of YANG models for abstraction and control of traffic engineered networks (2016)
30. L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, X. Costa-Perez, in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI*. OVNES: demonstrating 5G network slicing overbooking on real deployments (2018), pp. 1–2
31. 3GPP TS 28.541, 5G Network Resource Model (NRM); stage 2 and stage 3 (release 16), March 2019
32. 5G-Empower Platform. Website: https://5g-empower.io/. Last accessed in May 2019
33. K. Koutlia, R. Ferrús, E. Coronado, R. Riggio, F. Casadevall, A. Umbert, J. Pérez-Romero, "Design and Experimental Validation of a Software-Defined Radio Access Network Testbed with Slicing Support," Wireless Communications and Mobile Computing. **2019**(2361352), 17 (2019).

7 Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.