Virtual Networking for Lowering Cost of Ownership

Table of Contents

F. Marzouk (✉)
University of Aveiro
Aveiro, Portugal
e-mail: fatma.marzouk@ua.pt

M. Lashgari, L. Wosinska, P. Monti
Department of Electrical Engineering, Chalmers University of Technology
Gothenburg, Sweden
e-mail: maryaml@chalmers.se

A. Radwan . J.P.Barraca J. Rodriguez
Instituto de Telecomunicações, Campus Universitário de Santiago,
3810-193 Aveiro, Portugal

Abbreviations

AP	access point
BBU	baseband unit
BPM	bin packing minimization
CCO	capacity and coverage optimization
CoMP	coordinated multipoint
CPRI	common public radio interface
CRAN	cloud radio access network
CSF	cost scaling factor
DAS	distributed antenna system
DC	data center
DeMUX	de-multiplexer
EE	energy efficiency
H-CRAN	hybrid-cloud radio access network
InP	infrastructure provider
JT-CoMP	joint transmission coordinated multi-point
MLB	mobility load balancing
MNO	mobile network operator
MRO	mobility robustness optimization
MUX	multiplexer
MVNO	mobile virtual networks operators
NFV	network function virtualization
NGFI	next generation fronthaul interface
ONF	open network foundation
OTO	one-to-one
OXC	optical cross connect
P	protected
PRB	physical resource block
PRS	preliminary resource sharing
RA	resource allocation
RAN	radio access network
RAU	radio aggregation unit
RCC	radio cloud center
RD	resource duplication
RGBM	RRH group based mapping
RIRS	reconfiguration and improved resource

RNC	radio network controller
RRH	remote radio head
RRU	remote radio unit
SD-RAN	software defined-ran
SD-UCP	software defined unified control plane
SDN	software-defined networking
SDR	software defined radio
SINR	signal to noise ratio
SO	self-organization
SON	self-organizing network
SoTA	state of the art
SP	service provider
SPP	set the partitioning problem
SPSCP	shared-path shared-compute planning
TP-Selection	transmission-point selection
UE	user equipment
UP	unprotected
UT	utility function
WDM	wavelength division multiplexing
WNV	wireless network virtualization

Keywords

C-RAN, Network Sharing, Resource Allocation, Cost Saving, BBU Minimization, Energy Efficiency, SDN, NFV, Optical Networking, Network Design, Network Resiliency, Cloud Computing, Latency, Availability, Backup Connectivity, Shared Protection.

Chapter 10

Virtual Networking for Lowering Cost of Ownership

Fatma Marzouk, Maryam Lashgari, João Paulo Barraca, Ayman Radwan, Lena Wosinska, Paolo Monti, and Jonathan Rodriguez[1]

5G and beyond mobile networks hold the promise of supporting a vast emergence of new services and increased traffic growth. This represents a challenge for mobile networks operators, which are faced with the pressure of providing a variety of these services according to the stringent requirements of future mobile network generations, while still being able to i) preserve service resilience ii) sustain profitability by reducing costs, and iii) ensuring minimal energy consumption in the infrastructure. Fortunately, emerging 5G and beyond networks are expected to adopt increasingly prominent technological drivers that can tackle the above challenges by pushing the planning and management operations logic to the limit.

10.1 Introduction

It is agreed that 5G and beyond mobile networks generations would increasingly include key paradigms such as virtualization and autonomous management, which promise more flexibility and reactivity for mobile network operators (MNOs). Along with this technical and architectural turning point, governing business models have also evolved to reinvent roles and relationships between players, while opening new opportunities for innovation.
These technological key paradigms have been either adopted or recognized to be relevant by the 5G 3GPP; however, approaches as to how to exploit their application for RAN technology in a bid to promote further efficiency for MNOs are still in their infancy.
To this extent, we devote this chapter to first highlight the benefits of virtualization and autonomous technologies, when applied to the RAN infrastructure. Particularly, we highlight how they have driven technical and business models for RAN sharing. In this context, we initially target meeting performance reliability in a cost-efficient way based on a proposed network planning strategy and subsequent performance evaluation. Then, we extend this study to investigate the trade-off between the savings introduced from the centralization of service processing and the additional cost due to the addition of backup paths/resources. Thereafter,

[1] F. Marzouk (✉)
University of Aveiro
Aveiro, Portugal
e-mail: fatma.marzouk@ua.pt

M. Lashgari, L. Wosinska, P. Monti
Department of Electrical Engineering, Chalmers University of Technology
Gothenburg, Sweden
e-mail: maryaml@chalmers.se

A. Radwan . J.P.Barraca J. Rodriguez
Instituto de Telecomunicações, Campus Universitário de Santiago,
3810-193 Aveiro, Portugal

we focus on energy efficiency RA (Resource Allocation) design. A literature review is conducted on existing approaches for energy efficient resource allocation, highlighting the existing gaps and open research challenges. Finally, we propose a design for hybrid Resource Allocation aiming at improving energy efficiency (EE) on the C-RAN, where we consider the optimal use of both computational and radio resources.

10.2 RAN Virtualization and Autonomous Management

The adoption of virtualization and autonomous management technologies represents one of the most promising approaches to handle the increased complexity in future RANs management operations.

These technologies are expected to coexist in synergy within the 5G landscape and beyond. In this section, we provide an overview of the main RAN virtualization and autonomous paradigms and highlight their impact on future mobile networking.

10.2.1 Enabling Paradigms

10.2.1.1 NFV

With virtualization techniques, networking functions are implemented in software that is able to run independently of the underlying hardware. As such, all devices can be virtualized by being abstracted to a virtualized functionality using Network function Virtualization, with the exception of the ones that handle the reception/transmission of wireless signal. The latter can be virtualized using Software Defined Networking concept. Although rising from the computing world, the concept of network function virtualization is today increasingly applied to wireless mobile networking, particularly towards radio access networks - baseband processing pooling is one form of enabling NFV, which is commonly referred to as Cloud-RAN or Virtual-RANs. Compared to legacy RANs, the baseband processing and scheduling tasks in C-RANs are migrated to the cloud. By generalising the application of NFV in future generation RANs, the capacity of every network function would be rendered available for expansion and reduction through an increase/decrease of virtual resources according to the network load conditions, which would enhance the elasticity of the network, improves resources utilization efficiency, and leads to CAPEX (Capital Expenditure) - OPEX (Operational Expenditure) savings. When virtualization techniques including NFV and SDR are applied beyond the RAN functionality coverage, to include either infrastructure, spectrum, and air interface virtualization, we refer to this as wireless network virtualization (WNV). WNV is the main enabler for future RAN sharing between multiple mobile virtual network operators (MVNOs), as it would allow them to share a common RAN on demand, potentially provided by a neutral operator.

10.2.1.2 SDN

To face the rapid evolution of networks, network configuration approaches had to evolve to include more innovation in the way complex networks are controlled. Since some years ago, research initiatives in that sense reached a theoretical concept answering this requirement. One fruitful outcome of these efforts is software-defined networking (SDN). The essence behind

the SDN logic is decoupling the control plane of networking devices from the forwarding plane. The Open Networking Foundation presents the main body responsible for the standardization of SDN. The proposed SDN architecture by ONF is composed of three planes: Data plane, Control Plane, and Application plane. In the data plane, the SDN enabled networking devices are abstracted to their simple data forwarding functions. The control plane contains the SDN controller that represents the control logic. The application plane hosts the various SDN-based business specific applications. Thanks to the abstraction of the underlying devices, new applications can easily be deployed at this layer through simple programming. The interfaces between the three different planes in SDN are open. The southbound is responsible for communications with the data plane, while the northbound is responsible for the communications with the application plane.

The application of SDN on the RAN presents a key enabler for addressing futures RAN complexity in terms of management and control operations. Indeed, SDN can complement C-RAN architectures by offering a software defined-RAN (SD-RAN) control plane that is programmable and configurable via an innovative and advanced SD-RAN application plane. The implemented application can include a panoply of evolved decision maker modules for the management and optimization of the complex RAN resources/operations. The output decisions at this layer will be ultimately translated via the SD-RAN control plane to a set of configurations to be adjusted by the data plane via the SDN southbound interface.

10.2.1.3 SON and Learning based Management

Self-Organization (SO) is a concept that was first developed in chemistry and physics and then applied to biology, physics, social sciences, computer science, and finally mobile networking systems [1]. With respect to the first field of application, the SO concept is defined by the emergence of pattern and order in a system by internal processes, rather than external constraints or forces [2]. Following the same basic principles, the application of SO to the field of mobile networking systems leads to Self-Organizing Networks (SONs), where traditional network management procedures are improved thanks to the automation capability, brought by the SO concepts and SON technology. SON driven automation is enabled by adding more intelligence to the network following the concept of self-configuration, self-optimization, and self-healing. For MNO, SON technology leads to: i) simplification of operational and management tasks in HetNets, ii) improvement of network performance, iii) reduction of time to market of new services, and iv) OPEX savings. Benefits include more than 50% reduction in dropped calls, OPEX savings of more than 30%, and an increase in service revenue by 5-10% [40].

Research in the area of SON, especially in the scope of HetNets, has attracted the attention of a large segment of the research community over the past decade. The contributions of the research efforts have spanned different challenges, including specifying SON challenges [3-5], proposing algorithms in support of a given self-function [6-9] or surveying the state of the art [10], [5], [11].

The first instance of SONs can be described as adaptive and autonomous systems, based on control loops and threshold comparison. In order to handle more complex scenarios, the current state of the art is investigating the application of advanced techniques, such as Machine Learning (ML), and data mining to SON [12-16]. ML algorithms can be categorized in multiple ways, with a stronger focus on supervised learning, unsupervised learning, and reinforcement learning. Several efforts of applying ML to SON are available in the literature to enhance

mobility robustness optimization (MRO), mobility load balancing (MLB), capacity and coverage optimization (CCO), self-healing, resource allocation, and energy saving.

SON in future networks will be considered as an integral part of the RAN, rather than a complementary part. Indeed, the evolved 5G and beyond Cloud RAN will require not only faster operation of SON, but also new, innovative and proactive operations. Hence, the coverage map of SON algorithms within the ongoing set of 5G and beyond standards will be extended to cater for new use cases. These include self-protection to cater for automated security [17-20], SON for mmWave [21-23], SON for MIMO to enable adaptability to the different propagation characteristics of the mmWave links [24-25], SON for NFV-based networking mainly to ensure optimal NFV placement and traffic steering problem [26-27], and SON for Multi-RAT optimization and spectrum sharing, towards overall improved networks performance [28]. Ultimately, SON for EE radio management is another use-case of SON application for 5G networks that is currently attracting lots of interest.

10.2.2 Impacts on Future Mobile Networking

In line with growing costs and declining revenues for mobile operators, network sharing is emerging as a disruptive mechanism that can recover significant OPEX/CAPEX costs, by creating new sources of revenues and new cost reduction solutions. Indeed, network sharing would allow an operator that does not have the infrastructure, nor spectrum resources, to dynamically share the physical networks operated with other mobile network operators and hence maximize resource utilization efficiency.

10.2.2.1 New Business Model

The concept of resource sharing has evolved over time and has recently experienced a major wave of revolution. Indeed, with the emergence of softwarisation and autonomic management technologies, resource sharing concept has transitioned from only hardware-based resource sharing to overall softwarized mobile network-based resource sharing. The first form of hardware-based sharing has appeared with 3GPP Rel.99, allowing operators to share non-active assets of the RAN such as site locations or physical supporting infrastructure of radio equipment. Later, with the advent of 3GPP Rel-6 (UMTS) [29], Rel-8 (LTE) [30], and Rel-10 (LTE-A) [31], active sharing appeared, where operators share BS elements like the RF chains, antenna, or even Radio Network Controllers (RNC). LTE brought also a growing interest among operators to additionally enable spectrum-based sharing to maximize spectrum efficiency. LTE spectrum sharing technologies consider three spectrum segments including the TV white space channels, the frequently unused service-dedicated 3.5GHz, and the 5GHz unlicensed band.

The aforementioned hardware and spectrum-based sharing schemes are based on fixed contractual agreements/sharing framework over long time periods (typically on a monthly/yearly basis) and entail sharing partial part of the Infrastructure provider/ MNO resources or the available unlicenced spectrum bands. This type of sharing complies with the rational of the traditional business model consisting of two entities: The infrastructure provider (InP), which has resources but no subscribers, and the MNO that in contrast has subscribers but has no infrastructure resources. The InP is the responsible entity of virtualizing resources to be used/shared by the different coexisting MNOs, with management operations of virtualized resources performed via interactions between both entities.

The expected shift to fully virtualized mobile networks with the ongoing and upcoming mobile network generations presents the key enabler for full sharing among coexisting mobile

network operators. This would entail an evolution of the governing business model to cater for new business opportunities for telecom/network operators, manufacturers, and solution providers as well as for a range of new stakeholders. Indeed, compared to the traditional two-level business model, the MNO in the evolved three-tier business model can be further separated into two different specialized categories: the service provider (SP) and the MVNO. The SP is the entity that has subscribers. The MVNO is the entity responsible for leasing the resource from one or multiple InP to satisfy the accumulated requests from each SP and hence the entity evolved in creating virtual resources based on these requests. The established sharing paradigm can be extended with other roles in future multi-tenants systems such as vertical segments/industries that lack network infrastructure, but opportunistically or periodically need to reach their customers or enable services orthogonal to the telecommunication industry. The multi-tenant resource allocation operation should ensure the SLA (Service Level Agreement) /QoS (Quality of Service) of the different slices and cater for the adaptive capacity allocation, to enable opportunistic sharing of the mobile network infrastructure between the different vertical segments and services providers on time scales shorter than the contract agreement.

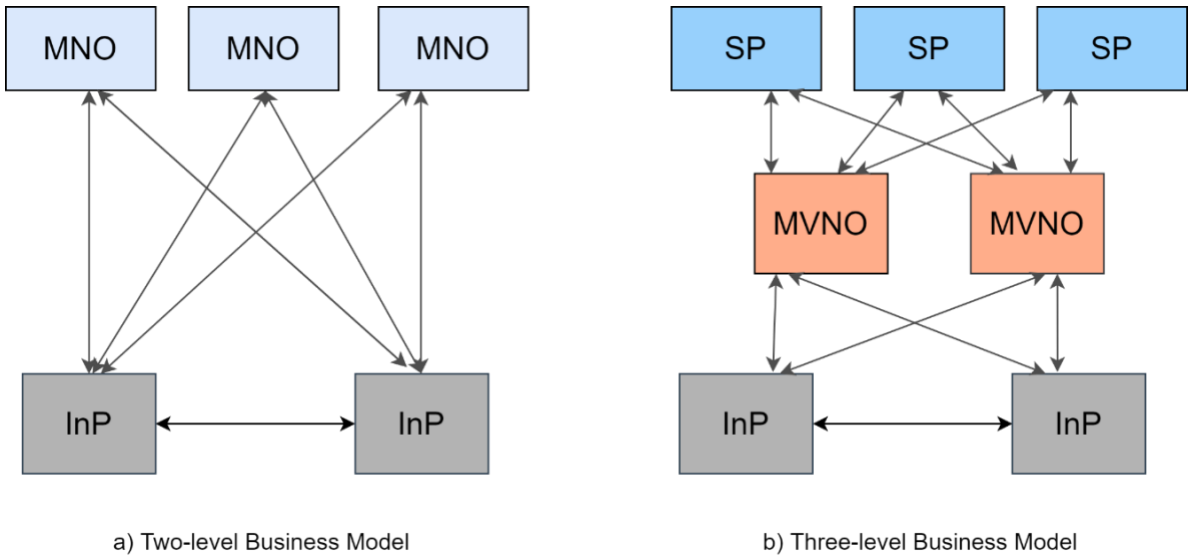

Fig. 10.1 Business model for virtualized mobile networks.

10.2.2.2 Enabling RAN Adaptive Sharing

The technical model for adaptive RAN sharing involves the main building blocks softwarization and autonomous management technologies, working in synergy for enabling adaptive RAN sharing/network slicing. Figure 10.2 depicts these building blocks. In this architecture, the RAN is fully virtualized thanks to the application of NFV for the virtualization of the RAN functions, SDR to slice the remote radio heads (RRHs), and SDN to manage the networking device. The RAN is controlled by a software defined unified control plane (SD-UCP). NFV/SDR enables RAN sharing by different tenants MVNOs, while the SD-UCP translates the decisions of the enhanced SO-VRM algorithms back to the radio physical nodes, to enable the dynamic allocation and the flexible management of resources according to SLA and load from each MVNO. The unified control plane would allow that all established slices c

share the available bandwidth on the different existing RATs depending on the MNO policy/MVNO SLA. Moreover, it allows greater efficiency, enabling rational use of resources. To ensure isolation of the traffic from the different MVNOs, dynamic resource allocation should include a minimum throughput for each isolated MVNO's application specific slice. Allocation of resources reflects the MVNO's policy and slice QoS, as well as the MNO's preference to optimize a certain metric, such as cost saving, energy efficiency, or a trade-off between both.

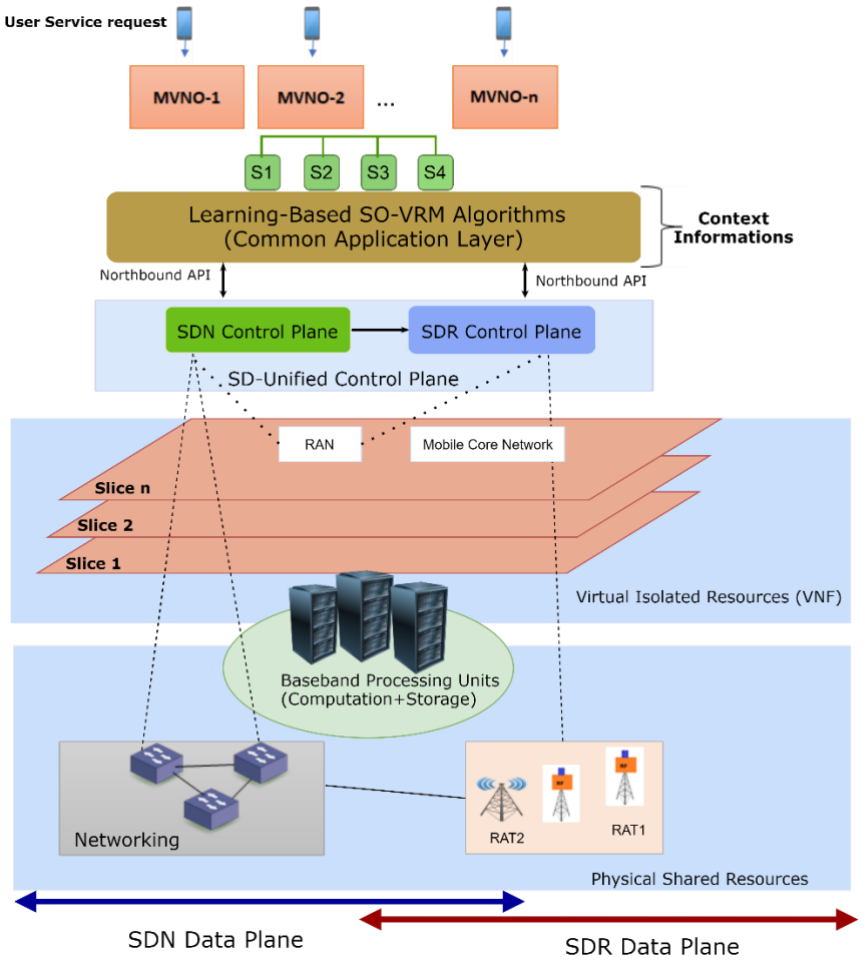

Fig. 10.2 Architecture for adaptive RAN sharing and application specific network slice

10.3 Cost Saving Design Strategies for 5G Network Infrastructures

Many 5G services have strict specifications in terms of latency and reliability performance [33]. Meeting these demanding requirements and designing a resilient network in a cost-efficient way are great challenges for the network operators. The hybrid-cloud radio access network (H-CRAN) architecture depicted in Fig. 10.3 is a promising option to meet 5G service constraints.

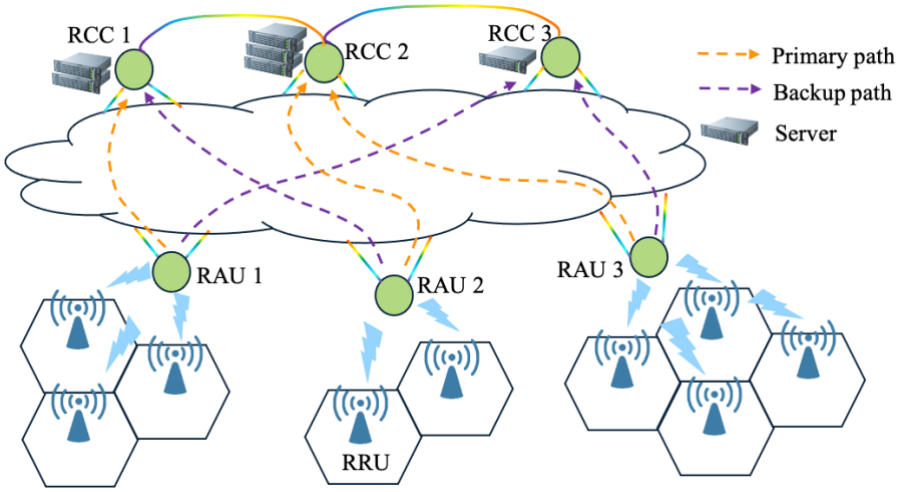

Fig. 10.3 An illustration of H-CRAN architecture with millimeter wave fronthaul and wavelength division multiplexing midhaul [34].

An H-CRAN architecture consists of three tiers: remote radio units (RRUs), radio aggregation units (RAUs), and radio cloud centers (RCCs). The RAU node is responsible for serving all the RRUs connected to it, and the RAU nodes are connected to the RCC. Part of the baseband processing is implemented at the RAU, and the rest of it is done at the RCC. The network segment connecting the RAU and RCC is called the *midhaul*, and the segment between the RRU and the RAU is referred to as *fronthaul*. The next generation fronthaul interface (NGFI) [35] and common public radio interface (CPRI/eCPRI) [36] are two options to transmit data over the fronthaul and midhaul segments.

There is a number of aspects that operators need to consider when designing their network infrastructures. Among them are resiliency and cost. More specifically, an H-CRAN architecture should support resiliency against the failure of the components or entire network nodes, while the cost of network deployment is minimized [34]. Additionally, from a cost perspective, it is beneficial for the operators to deploy services at centralized computing locations to be able to leverage the economy of scale of large data center sites [37]. However, meeting the service latency and availability requirements are challenges in a centralized deployment that are elaborated in the following.

In an H-CRAN architecture, a failure might happen at the RCC, RAU, RRU, or in any node/link in the fronthaul and midhaul segments. The number of users affected by a failure depends on the failure location. If the failure happens in the RCC, a large number of users will be affected, which makes significant impact on network reliability performance. Likewise, the failure in a midhaul node/link may cause service interruption for a subset of users associated with the RAU. Therefore, in order to improve the reliability performance, an H-CRAN

architecture should be designed to minimize disruption due to failure of a server in the RCC, the whole RCC (due to a catastrophic event), and any midhaul node/link.

A possible way to ensure the survivability of services is to provision backup resources for connections between the RAUs and RCC in a *dedicated* or *shared* fashion. Dedicated protection methods fall into two categories: a) 1+1, where backup resources are active and two live connections exist between the source and destination, and b) 1:1, in which the backup resources are not active until a failure occurs in the primary path [38]. On the other hand, meeting the reliability performance requirement of a service by duplicating network and compute resources can be very expensive. Therefore, when possible, it is preferable to use more cost-efficient solutions, i.e., shared protection where backup resources can be shared among multiple services.

The cost efficiency of network infrastructure can be further increased by centralizing service processing in a few large data centers (DCs). Some studies promote distributed RAN architectures, which can take a step towards satisfying the quality-of-service constraints (i.e., in terms of latency and reliability performance). However, this approach is losing benefits introduced by centralization, i.e., cost reduction and easy deployment of RAN features [39]. Indeed, processing services in large scale DCs will cost less than in the small and distributed DCs because of the economy of scale.

The main challenge to reach large and centralized DCs is guaranteeing the latency and reliability performance requirements which may be difficult due to the long distances to the large DCs. The only way to meet a latency constraint, using a specific technology (e.g., optical transport, millimeter wave), is to choose a DC close enough to the end user. However, the reliability performance can be improved by adding a redundant midhaul path. Unfortunately, the benefits of adding a protection path in terms of reliability performance will also introduce additional costs due to introducing backup resources that might adversely affect the overall cost savings of centralizing the service processing. For this reason, its impact needs to be analyzed carefully.

The research community has been looking into the aforementioned reliability and cost-efficiency challenges. A number of works in the literature studied resilient design methods and cost saving strategies [40-42]. The work in [40] considers centralized and distributed algorithms to place baseband unit (BBU) hotels in cloud radio access network (C-RAN) to ensure service continuity in case of single BBU hotel failure and compares their performance, scalability and adaptability to changes in the network topology. The work in [40] shows that a distributed approach helps to off-load the SDN orchestrator and is able to cope with the evolution of C-RAN topology whereas the changes in the original placement are limited. The authors in [41] bring up the benefits of "centralization" in terms of computational resources and power savings, as well as the importance of designing a survivable C-RAN network in case of failure. They proposed three approaches for survivable BBU hotel placement: (1) dedicated path protection, (2) dedicated BBU protection, and (3) dedicated BBU and path protection. Most of the interest in the literature is focused on designing a resilient C-RAN architecture. To provide survivability, the existing works either duplicate resources or consider sharing of the BBU ports in the BBU hotels. In order to improve cost savings, the authors in [42] assumed a given outage probability and used spatial traffic model and queuing theory to find the required number of transceivers in the considered scenario. Indeed, the transceivers in the fronthaul are used to work also at peak load, but the peak load conditions can be relatively rare. Therefore,

by accepting a reasonable outage probability, the work in [42] shows that the fronthaul can be designed with lower capacity and fewer transceivers, which leads to cost and energy savings.

In order to guarantee survivability of services in the event of a failure in the RCC or any node/link in the midhaul in a cost-efficient manner, this chapter introduces a strategy called shared-path shared-compute planning (SPSCP). The proposed strategy decides on the location of a primary and backup RCC for each RAU and respective midhaul paths while allowing the sharing of the backup connectivity and computing resources. The SPSCP strategy has lower cost than equivalent approaches that use dedicated computing and midhaul connectivity resources for the backup and shows a cost improvement compared to those approaches where sharing is not encouraged while deciding on the location of the primary RCC nodes and midhaul path [34].

In addition, to benefit from cost improvements of centralized network deployment, this chapter investigates the trade-off between the savings derived from centralizing service processing and the additional cost due to the protection path used to meet the availability requirement of a given service. The performance evaluation shows that by centralizing service processing with the help of a protection path to meet the service reliability requirement, savings in overall infrastructure cost can be achieved [37] while not violating any latency constraint.

In Section 10.3.1, the network planning strategy to meet reliability performance requirement is presented. The system architecture and use case are discussed in Sec. 10.3.1.1, while the proposed approach to design a resilient H-CRAN architecture and assessment of the results are presented in Sec. 10.3.1.2. Section 10.3.2 discusses the cost benefits of centralizing service processing. In particular, Sec. 10.3.2.1 presents the system architecture, latency and availability requirements, and cost model, and Sec. 10.3.2.2 discusses the economy of scale benefits and simulation results. Finally, conclusions are drawn in Section 10.3.3.

10.3.1 Shared-path shared-compute network planning strategy

In order to meet the quality-of-service requirements of 5G services, a resilient network infrastructure should be provided. One possible method is adding backup resources although it increases the network deployment cost. A possible solution to reduce this cost is maximizing sharing of the backup connectivity and computing resources. The intuition behind this section is to derive a cost-efficient resilient network design strategy by exploiting the potential of sharing backup resources.

10.3.1.1 System architecture and use case description

We consider an H-CRAN architecture with a mesh wavelength division multiplexing (WDM) network for the midhaul segment as shown in Fig. 10.3.

We assume a single failure scenario, i.e., at most one failure can happen in the network at a time. An RRU failure can be handled by handover to other RRUs and is defined by the operators' handover policy. Further, we assume that the RRUs are dual homed to two different RAUs. Accordingly, a failure of RRU or RAU will not affect the overall availability. We assume that a failure can happen in a server in the RCC, or the whole RCC can be down because of a catastrophic event. Also, a failure might happen in any node or link in the midhaul segment of the network. In order to satisfy the latency requirement, the number of hops between an

RAU and its RCC node cannot exceed a given value, denoted by h, which is dictated by the latency requirements.

The target is to design a resilient H-CRAN architecture. In order to achieve this goal, one primary and one backup RCC should be assigned to each RAU, and the primary and backup connectivity paths between the RAU and RCC should be found. The design of the resilient network and allocation of the resources should be done with the objective of minimizing the network deployment cost. The cost is the summation of the total deployment cost of RCC nodes, server units within the RCC, and connectivity units, which is defined as:

$$C = N_{RCC}.C_{RCC} + N_{Ser}.C_{Ser} + N_{Conn}.C_{Conn} \qquad (10.1)$$

where N_{RCC}, N_{Conn}, and N_{Ser} are the number of RCC nodes, connectivity units, and server units, respectively. C_{RCC}, C_{Conn}, and C_{Ser} are the cost of deploying one RCC node, one connectivity unit, and one server unit, respectively.

A key method to reduce the resilient network deployment cost is sharing the connectivity resources in the backup path and computing resources in the backup RCC. Two conditions should be met to enable the sharing of the computing resources, referred to as the *server sharing condition*. The RAUs can share a backup server in an RCC node if: 1) their primary servers are located in different RCC nodes, and 2) the paths to their primary RCC are node disjoint. Moreover, to share connectivity resources in the midhaul backup path two conditions should be satisfied, referred to as the *connectivity sharing condition*. The RAUs can share connectivity resources in the backup path if: 1) their primary servers are placed in different RCC nodes, and 2) the primary paths to their RCC nodes are node disjoint.
Therefore, the primary and backup RCC nodes and the midhaul paths should be found by considering the above sharing constraints to minimize the overall cost. The details of the proposed strategy are described in the next section.

10.3.1.2 General approach and performance evaluation

This section presents a strategy referred to as shared-path shared-compute planning (SPSCP) used to find the primary and backup RCC nodes of each RAU together with their connectivity paths.

Network planning strategy

The strategy is a heuristic algorithm that chooses the primary and backup RCC with the lowest cost for each RAU [34]. In this algorithm, first, all RAUs are sorted based on the increasing value of the nodal degree of the midhaul nodes where they are located. This set is called $\mathcal{A}_s$, which is used to choose the primary and backup RCC. The midhaul nodes without RAU, i.e., set denoted by $\mathcal{G}$, can be chosen to place the RCC, thus, we assign a tag to these nodes called combined degree. We define the combined degree of each midhaul node as the summation of the nodal degree of the node and the number of RAUs that are within h hops from that midhaul node. We choose the midhaul nodes that are within h hops from the RAU, i.e., set $\mathcal{P}$, and sort them based on the decreasing value of the combined degree and get set $\mathcal{P}'$. Then, we evaluate the total cost of the network for all different options of choosing primary and backup RCC and select the option with the lowest cost. The shortest path algorithm is used to find the primary

and backup connectivity paths. To calculate the cost, we use the cost function (10.1) described in Section 10.3.1.1 where we consider the possibility of sharing backup connectivity and computing resources. The cost of required resources for the backup is zero if they can be shared with backup resources of other RAUs. We repeat this procedure for all RAUs until we find primary and backup RCCs and their connectivity paths. The detailed steps of the algorithm are presented in Fig. 10.4 for a given value of the number of allowable hops (h).

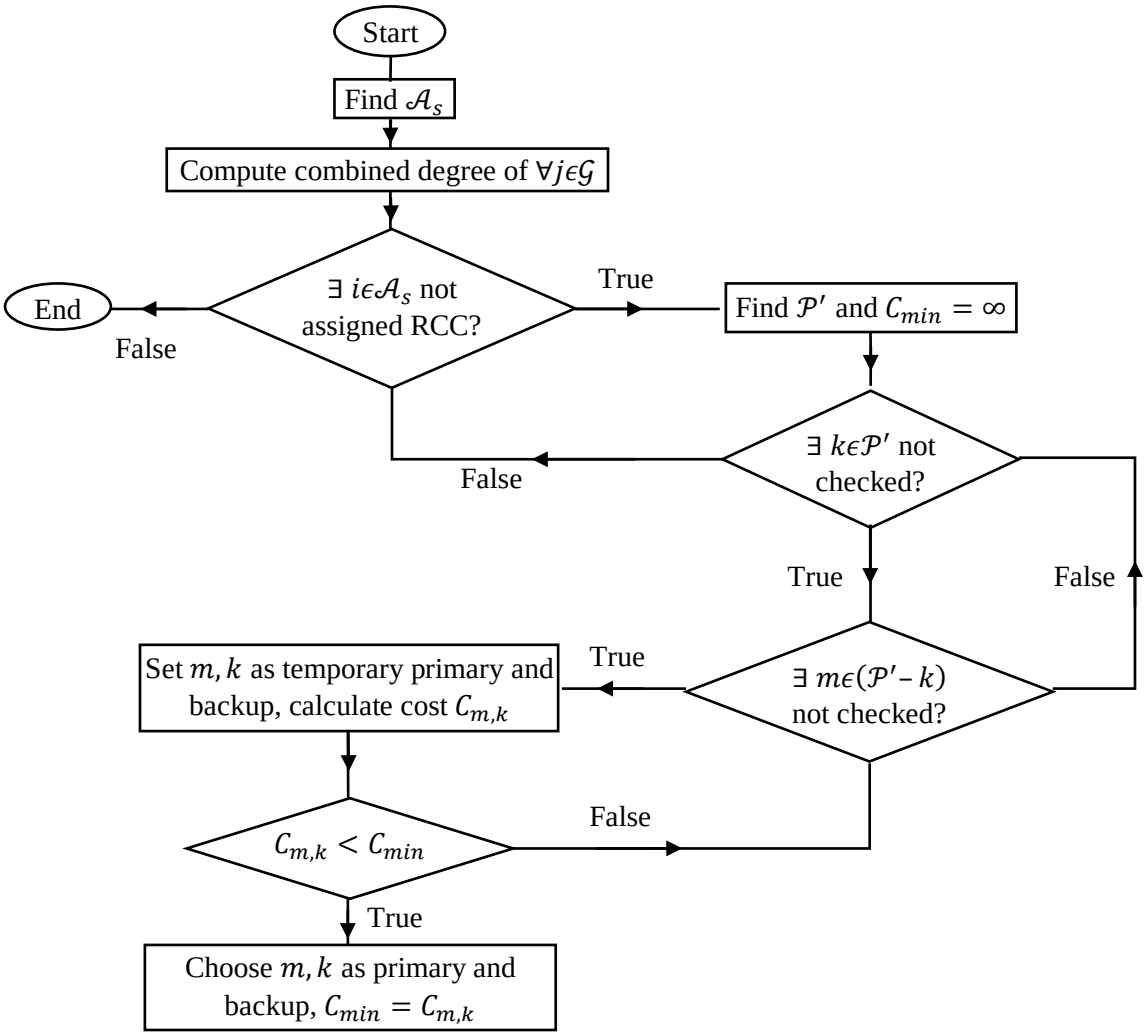

Fig. 10.4 The flowchart of SPSCP strategy.

Performance evaluation

In this section, the performance of the SPSCP strategy is evaluated via simulations. We assume the same network parameters as the ones described in [34]. To obtain the results we set $C_{RCC} = 120$ cost units [CU], and connectivity and computing cost are changed to show their impact on the total cost.

The performance is evaluated against three benchmark algorithms. The first one is referred to as resource duplication (RD) where connectivity and computing resources are duplicated for the backup. The second one is referred to as preliminary resource sharing (PRS). PRS works exactly as RD, but tries a-posteriori to share backup resources where possible, i.e.,

without changing the pairing between RAUs and RCCs. The third one is called reconfiguration and improved resource sharing (RIRS). RIRS aims at improving the cost performance of the network designed according to RD. RIRS revisits the pairing between RAUs and their backup RCC nodes and the connectivity paths in order to maximize sharing of the backup resources.

The total cost of the network as a function of the allowable number of hops h for $C_{Conn} = 1,\ C_{Ser} = 6$ [CU] is shown in Fig. 10.5 (a). By relaxing the hop count constraint, i.e., by increasing the number of allowable hops, the cost decreases. The breakdown of the total cost of SPSCP shows that the cost of RCC deployment and backup servers are decreasing with the increasing number of hops between an RAU and its RCC node. This is the direct result of concentrating RCC nodes on a few midhaul nodes, although SPSCP considers the potential of sharing backup resources when it chooses the place of deploying the primary and backup RCC nodes. However, RD, PRS, and RIRS try to deploy RCC on fewer midhaul nodes without considering the shareability potential, which results in a lower cost reduction than SPSCP for higher h. Therefore, SPSCP shows better cost savings, which increases with increasing values of h.

The total cost of the network as a function of the number of allowable hops for $C_{Conn} = 6$, $C_{Ser} = 1$ [CU] is shown in Fig. 10.5 (b). The results show a similar trend as in Fig. 10.5 (a) when the value of h increases. However, the cost savings of SPSCP with respect to RD, PRS, and RIRS are smaller than in the case shown in Fig. 10.5 (a). This is because in the scenario under investigation, the connectivity resources are the bottleneck, and the opportunity of sharing connectivity resources is limited. Therefore, increasing the connectivity unit cost results in lower cost savings. For the sake of comparison, the cost savings of SPSCP with respect to three other approaches for low and high values of h are shown in Tab. 10.1. It is evident that the cost savings of SPSCP with respect to all strategies when $C_{Conn} = 6\ ,\ C_{Ser} = 1$ [CU] are lower compared to the case when $C_{Conn} = 1,\ C_{Ser} = 6$ [CU].

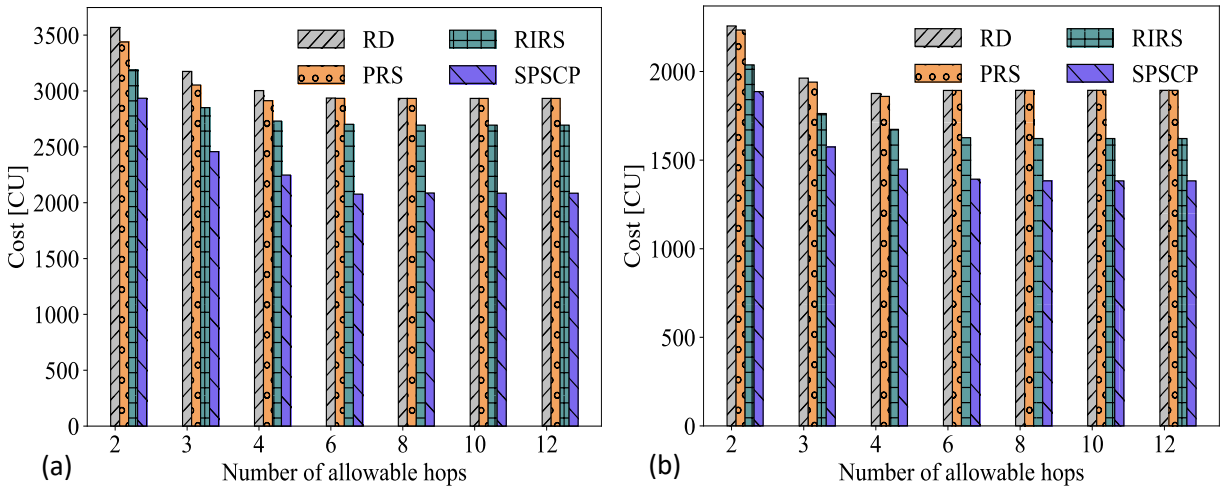

Fig. 10.5 Overall cost of network deployment as a function of number of allowable hops between RAU and RCC: **a** $C_{Conn} = 1,\ C_{Ser} = 6$ [CU], **b** $C_{Conn} = 6,\ C_{Ser} = 1$[CU].

Table 10.1. Cost savings of SPSCP with respect to RD, PRS, and RIRS for different connectivity and computing unit cost.

Allowable hop count	Method	SPSCP cost saving, $C_{Conn} = 1,\ C_{Ser} = 6$	SPSCP cost saving, $C_{Conn} = 6,\ C_{Ser} = 1$
$h = 3$	RD	22.63%	19.79%
	PRS	19.55%	18.86%
	RIRS	13.81%	10.71%
$h = 12$	RD	28.91%	26.95%
	PRS	28.91%	26.95%
	RIRS	22.61%	14.71%

10.3.2 Cost benefits of centralizing service processing

Processing services in large-scale data centers is more cost efficient than using distributed small computing nodes. On the other hand, large-scale data centers may be placed far from the users in centralized locations. Therefore, the potentially long propagation delay should be considered in the provisioning phase in order to make sure that the latency requirements are met. In addition, the availability requirements of services should be met, regardless of which computing nodes are used for service processing. In this section, we leverage upon the cost-effectiveness of the centralized deployment and propose a strategy to maximize the involvement of large-scale data centers, which also guarantees the quality of service in terms of latency and availability requirements.

10.3.2.1 Infrastructure model, service requirements, and cost

In this subsection, the network architecture, latency, availability, and cost models are presented.

Network architecture

The considered network architecture is presented in Fig. 10.6. The RRU and RAU functionalities are placed in the same network element, referred to as access point (AP), while the RCC functionalities are deployed in the DC. The user equipment (UE) and AP are connected through wireless links. It is assumed that the AP is connected to a number of servers residing in the DC through an optical transport network. The transport network is composed of three segments with different transmission capacities. In this architecture, we have two points for aggregating and grooming traffic. The first aggregation point is on the boundary of the local and province segments. The second aggregation point is on the boundary of the province and regional segments.

We assume to have four types of DCs located in various segments of the network and with different characteristics in terms of size (amount of computing resources and the number of users that can be served) and cost-efficiency, (the overall DC cost vs. the total number of users that can be supported). The DC types are the following: local, province, regional, and national. Local DCs are small, while the national DCs have the largest scale and are the most cost-efficient among other types.

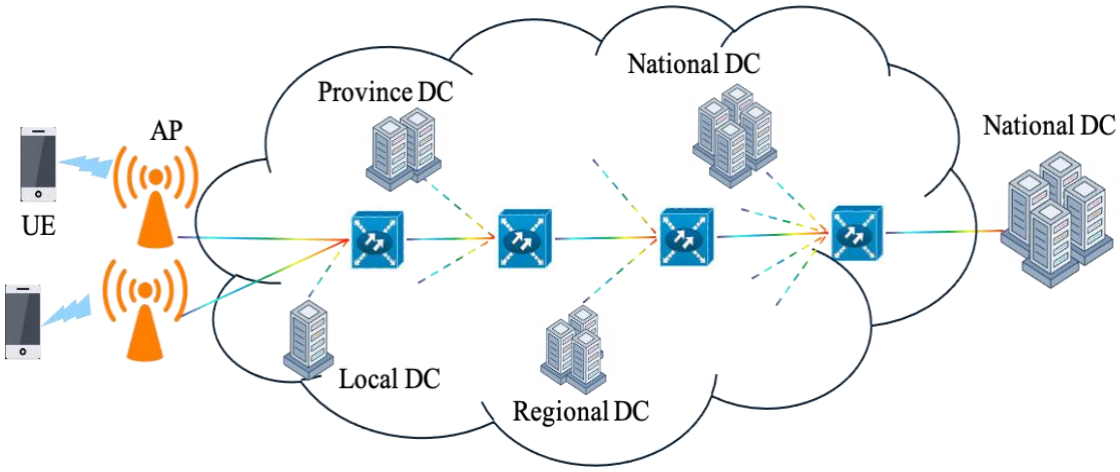

Fig. 10.6 The network architecture with three segments in the transport network [37].

Latency and availability requirements

The number of APs connected to small DCs is lower than the number of APs served by the large DCs. Small DCs are often close to the end user, and a lower number of nodes and links should be traversed to reach them. This option is offering lower latency and higher availability compared to processing services in large DCs. On the other hand, by centralizing service processing and using large scale DCs, a higher number of APs can be served at one location, which has cost savings because of the opportunity to leverage on a better economy of scale. However, in this case, meeting the latency and availability requirements of the services can be challenging because large DCs are normally deployed far from the users, and services must traverse more components to reach those DCs. Therefore, the latency and availability constraints should be considered when designing the network.

The latency is the summation of latency of the UE, RAN, server, propagation, and switching latency of links and devices in the transport network. The availability is modelled as the product of availability of UE, RAN, all the nodes and links along the path from the AP to the DC, and the server. The latency value can be decreased only by choosing a server in a DC close to the UE. However, the availability can be improved by adding a protection path, referred to as protected (P) scenario, while in the unprotected (UP) scenario only one path between AP and DC can be used. More details on the latency and availability computation are provided in [37].

Cost computation

Our cost model includes the cost of computing, i.e., server in the DC, and connectivity resources in the transport network. We do not model the cost of the radio access network.

The total computing cost is a function of the required number of DCs, the number of servers in each DC, and the cost scaling factor (CSF) of a DC. CSF is used to calculate how many cost units need to be spent for the DC infrastructure (i.e., cooling, power, and networking equipment) out of each cost unit spent on servers. Clearly, the CSF of the national DC is the lowest among all types of DCs because of their economy of scale. Two different categories of switching nodes exist in the transport network. The cost of the nodes which are not performing traffic aggregation is modeled as the cost of their optical cross connects (OXCs), multiplexers (MUXs), and de-multiplexers (DeMUXs). For the switching nodes performing traffic aggregation and grooming, the cost of packet switches and transceivers are added to the OXCs, MUXs, and DeMUXs costs. Furthermore, the cost difference for transceivers with different transmission rates is reflected in the cost model. More details on the cost modeling assumptions can be found in [37].

10.3.2.2 Trade-off evaluation

In this section, the trade-off between cost savings of centralizing service processing and the extra cost of providing a protection path to meet availability requirements is investigated.

Characteristics of the network and devices

We consider five use cases corresponding to five types of services along with their different latency and availability requirements as included in Tab. 10.2. For each use case, the maximum distance between the AP and DC should be calculated, according to the latency and availability requirements.

Table 10.2. Considered use cases and their requirements

Use Case	Description	Latency	Availability	Reference
1	Augmented Reality, collaborative gaming	12 ms	99.9%	[43]
2	Remote control for smart manufacturing	5.5 ms	99.99%	[43]
3	Discrete automation	20 ms	99.99%	[44]
4	Process automation / Monitoring	20 ms	99.9%	[44]
5	V2X for short term environment modelling	10 ms	99.99%	[45]

Table 10.3 provides the value of the key parameters for each type of DC considered in the study, i.e., CFS, DC distance from the AP, number of deployed servers in the DC, and the number of APs that can be connected to a DC (referred to as service density). The value of the availability of the RAN is assumed to be 99.999%. For the information on the value of the simulation parameters and the cost of the devices in the network infrastructure, we refer to [37].

Table 10.3. DC characteristic. The CSF of the regional, province and local DCs are a function of η, i.e., national DC cost scaling factor. d_{CN} is the distance between DC and AP.

DC type	d_{CN} range [km]	Service density	Num. server	Cost scaling factor
National	$d_{CN} > 100$	1000	250	$N_{EF} = \eta$
Regional	$100 \geq d_{CN} > 10$	100	25	$R_{EF} = 3 \times N_{EF}$
Province	$10 \geq d_{CN} > 1$	10	3	$P_{EF} = 2 \times R_{EF}$
Local	$1 \geq d_{CN}$	2	1	$L_{EF} = 2 \times P_{EF}$

Cost assessment and results

Figure 10.7 shows the maximum distance between AP and DC as a function of number of transport network (TN) nodes for the different use cases listed in Tab. 10.2.
By adding one protection path, the maximum distance between the AP and DC for use cases 1 and 4 can be increased by up to 300 km while still meeting their latency requirements. For use cases 2, 3, and 5 in the protected scenario, a service can be deployed in DCs even further from the AP, i.e., in the 1000 km range.

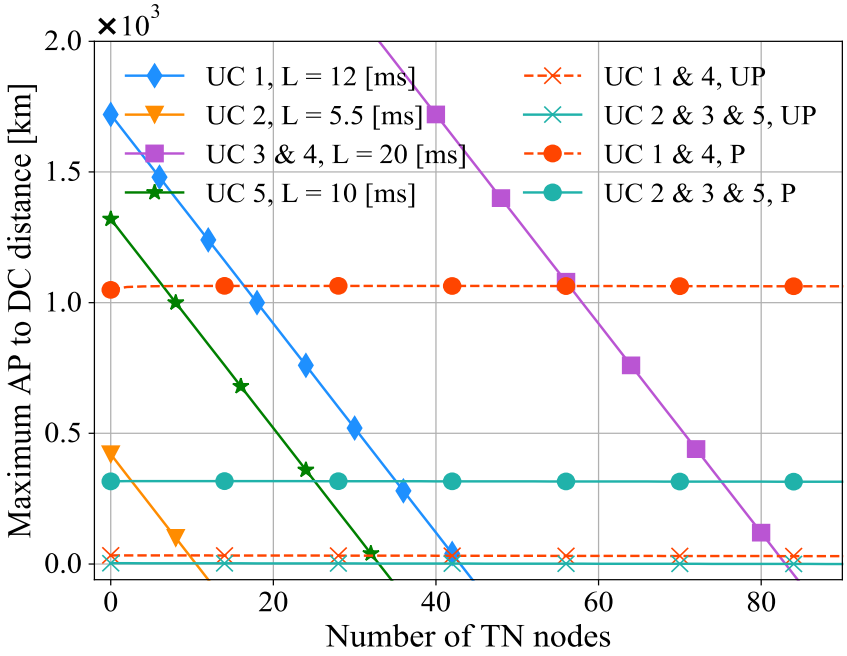

Fig. 10.7 Maximum distance between AP and DC for protected and unprotected cases.

Figure 10.8 depicts cost saving offered by a more centralized service processing. Use case 3 presents the highest cost saving (up to 63%) because it has a relaxed latency requirement which allows centralization, whereas the strict availability requirement can be met by adding a protection path. In addition, use cases 3 and 4 have the same latency requirements, but the cost saving of use case 3 is higher which shows availability requirement is the determining factor for the achieved gain. This result is also evident by comparing cost savings of use cases 1 and 4 since they have the same availability requirements.

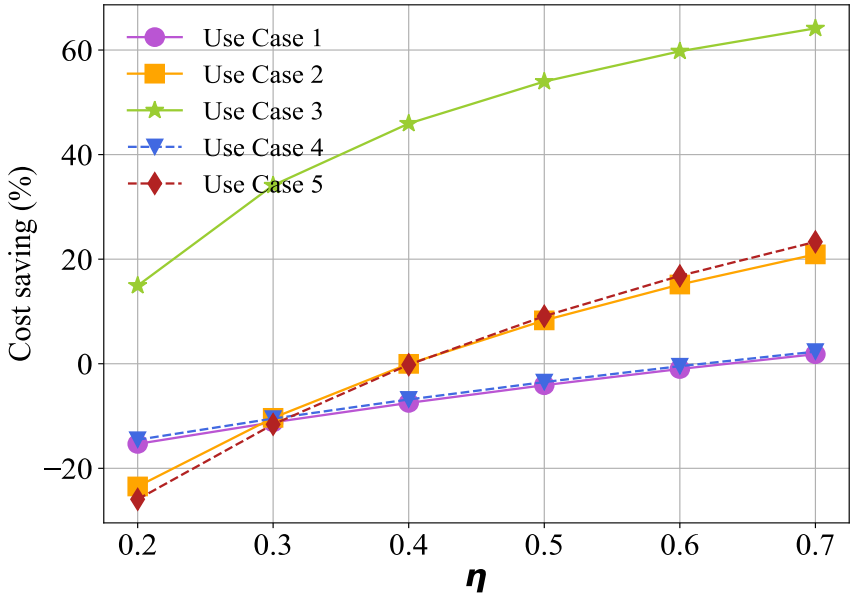

Fig. 10.8. Cost savings by adding a protection path in TN as a function of national DC CSF.

10.3.3 Conclusion

In this chapter, cost saving strategies considering the latency and reliability performance requirements in 5G network were proposed.

In the first scenario, the goal was to design a resilient H-CRAN architecture in which the failure can happen in any server of the RCC nodes, the whole RCC, or any link or node in the midhaul segment. The proposed cost-efficient strategy, referred to as shared-path shared-compute planning (SPSCP), assigns a primary and a backup RCC node to each RAU. To decrease the overall cost of the network, the SPSCP strategy tries to maximize sharing of the backup connectivity and computing resources. Adopting SPSCP strategy results in 26.9% cost savings compared to the approach that uses dedicated resources for the backup, and 14.7% cost savings compared to the method that does not consider shareability potential of the backup resources when assigning the primary RCC.

In the second scenario, in order to increase cost-efficiency of the 5G network infrastructure, we investigated the benefits of processing services in large data centers in contrast to the small, distributed edge computing nodes. By considering services with different constraints, we guaranteed that all reliability and latency requirements of the deployed services are met. One protection path was added in the transport network to meet the availability requirements and processing services in centralized and large DCs. In spite of the extra cost of the redundant path, it still leads up to 63% cost savings because of the economy of scale offered by centralized service processing.

10.4 Energy Efficient Virtual Resource Management for 5G and Beyond

We devote this section to energy efficient virtual resources management approaches for 5G RANs and beyond. After a thorough analysis of existing research works, we shed the light on some perspectives and some opportunities arising in that sense. Interestingly, we present a performance evaluation of an EE efficient virtual resource management that tackles some of the identified challenges.

10.4.1 Review on Energy Efficient Resource Allocation

Energy efficiency has always been in the spotlight of the cellular networking, especially with the shift of the design and planning logic from traditional rigid planning considering the peak traffic demand to more adaptive schemes leveraging from the flexibility of the virtualization or self-organization to bring more energy efficiency to the RAN. Most of the research contributions in that sense focused on the optimization of the RRH resource elements. Other works, targeted the optimization of EE through the design of computational resource allocation schemes to map load from RRHs to an optimal number of baseband Units in the cloud. The joint radio and computational resource allocation has recently been addressed by several research works, aiming to cater for the hybrid nature of resources in Cloud based RAN environments.

10.4.1.1 EE Radio Resources Allocation

Radio resource management research works include dynamic radio resource allocation on physical resource blocks (PRBs) and/or a power allocation to users, depending on their channel state information and the required data rate, with the aim of optimizing EE. Other radio related tasks include user pairing to an optimal RRH or to a set of RRH antenna resources, in a coordinated multipoint (CoMP) and/or beamforming design. In particular, recent research works on CoMP [46-50] provide evidence on the traditional benefit of interference mitigation where interfering signals from neighboring RRHs are used constructively to provide diversity gain enhancing the reliability of the received signal. Moreover, applying self-organization to the RRHs by selectively and automatically powering on/off according to the load variation has also been extensively addressed. In particular, the allocation tasks have been formulated into mixed combinatorial or non-convex optimization problems, and solved after decomposition to elementary steps. For instance, RRH selection and RRH on/off problems have been solved via heuristic algorithms, such as greedy activation [51-53]. Power/bandwidth allocation has been solved by relaxation and decomposition techniques [54] or game theory [55]. Table 10.4 provides a summary of these EE radio resource allocation works.

Table 10.4 EE Radio Resource Allocations Works

Ref	[46], 2014	[47],2014	[56], 2014	[57],2015	[58],2015	[59], 2016	[51],2016	[60], 2016	[61], 2016	[52], 2016	[53], 2016	[62], 2016	[50], 2016	[55], 2016	[63], 2017	[64],2017	[65], 2017	[66] , 2017	[67], 2017	[68], 2018	[48], 2018	[54],2018
Subchannel/ PRB allocation			✓			✓	✓		✓	✓					✓	✓						✓
Power allocation		✓	✓						✓					✓	✓					✓		✓
CoMP	✓	✓															✓				✓	
Beamforming				✓				✓			✓	✓	✓				✓		✓		✓	
User-RRH pairing/			✓	✓	✓		✓	✓		✓	✓			✓	✓	✓		✓		✓		✓
RRH ON/OFF	✓				✓		✓			✓	✓		✓				✓	✓	✓	✓		
InP/MVNO																			✓			
Simulation Based Evaluation	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓
Evaluation Based Implementation																						

10.4.1.2 EE Computational Resources Allocation

A second existing approach for energy efficiency improvement is to minimise the power consumed at the BBU pool by minimizing the number of BBUs at the cloud, considering RRH that can be grouped into a cluster and mapped to one BBU. In the literature, the problem has been mostly formulated as bin packing minimization (BPM) problem and solved via [69-70], or meta-heuristic [71]. Although these efforts proved to be efficient for minimizing the active number of BBUs, they do not account for the user QoS requirement and the level of interference in the network, when forming RRHs clusters [72-74]. Different from the constraints-oblivious behaviour of these approaches, recent works included the QoS constraint by formulating the problem to a modified BPM [75] or to set the partitioning problem (SPP) [76]. Other research efforts considered service time constraint along with the power minimization problem through the design of a BBU workload-scheduling scheme [77], whilst [78] presents another research strand toward more realistic BBU resource allocation by reshaping the problem into a virtual BBU minimization problem. Table 10.5 provides a summary of these BBU resource allocation works.

Table 10.5 EE BBU Resource Allocations Works

Ref	[79], 2012	[72],2013	[71], 2015	[80], 2015	[69], 2015	[70], 2016	[81],2017	[73], 2017	[77],2017	[74],2017	[82],2018	[83], 2018	[78], 2018
BBU Scheduling		✓							✓				
RRH-BBU Mapping	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	
Virtualized BBUs		✓	✓					✓		✓			
Load dependent BPS Power Consumption	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
BBU ON/OFF				✓	✓	✓		✓	✓	✓	✓		✓
InP/MVNO													
Simulation Based Evaluation	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		
Implementation Based Evaluation		✓											

10.4.1.3 EE Hybrid Resources Allocation

In contrast to the aforementioned approaches that considered computational or radio optimization aspects independently, recent approaches have targeted the design of EE multi-resource allocation taking into account both resources, as summarized in Tab. 10.6. We refer to these collectively as hybrid resource allocation; for e.g. [84] aims to minimize the overall system power consumption for C-RAN by optimizing the number of virtualized BBU, set of selected RRHs, and the beamforming vector at active RRHs. It is worthy to note, that none of the hybrid research works considered the distributed antenna system behaviour of a set of RRH mapped to the same BBU, and consequently the associated relationship between the radio and computational resource allocation. QoS constraints consideration has also been overlooked. Furthermore, none of these schemes considered leveraging on the baseband servers' virtualization gain, and including this in the cross-layer optimization problem by enabling the dimensioning of the optimal number of virtual BBUs as a trade-off between QoS and reduced energy consumption.

Table 10.6 EE Hybrid Resource Allocations Works

Ref	[85], 2014	[84], 2015	[86], 2016	[87], 2016	[88],2016	[89], 2017	[108], 2017	[91], 2017	[92], 2018	[93], 2018	[94],2019
Subchannel/PRP allocation	✓		✓							✓	✓
Power allocation		✓	✓								✓
CoMP							✓	✓			
Beamforming		✓			✓		✓			✓	
User -RRH pairing		✓	✓		✓		✓	✓	✓	✓	✓
BBU Scheduling		✓			✓						
RRH-BBU Mapping			✓				✓		✓		✓
Virtualized BBUs	✓	✓	✓	✓	✓		✓	✓		✓	✓
Load dependent BPS Power Consumption			✓	✓	✓	✓	✓		✓	✓	✓
RRH ON/OFF		✓	✓	✓				✓	✓	✓	
BBU ON/OFF	✓			✓			✓	✓	✓		✓
InP/MVNO											
Simulation Based Evaluation	✓	✓	✓	✓	✓		✓	✓		✓	✓
Implementation Evaluation Based									✓		

10.4.2 Challenges towards Virtual Resource Management for C-RANs

Despite the valuable contributions that the research community brought in terms of hybrid resource allocations, there are still some open challenges that need to be addressed. One primary challenge is related the design of EE Multi-Operator (EE-MO) hybrid RA. It is worthy to note, that [67] presents the only work about multi-operator resource allocation. The scheme consider cooperating MVNOs with inter-band non-contiguous carrier aggregation, by segmenting the licensed spectrum of each into private and shared bands, in which UE access is mutually exclusive. Results of the proposal evaluation proved an improvement in terms of energy efficiency and spectrum efficiency. The work however, caters for RRH antenna resource sharing among operators solely, and not for the BBU sharing. In this context, an interesting challenge here lies in the design of energy aware multi-operators RAN resource management operations that relies on powerful SO algorithms and the fully virtualized RAN infrastructure to allow multiple operators to coexist, and adaptively share the RAN resources while reducing the energy consumption on the RAN. Precisely, the unified software defined control plane would leverage big data and ML algorithms and use as input context information collected from all network resources while considering each of the MVNOs' SLA, instantaneous load, and minimum required QoS. Using ML capability, the collected data will be analysed, and operation specific optimizations models developed. Ultimately, these optimization models will be applied, and the output is a set of optimal parameters to be adapted by the different radio and virtual RAN resources.

Moreover, the aforementioned approaches should cater for the hybrid nature (radio and computational) of the resource and include new optimization schemes that cater for not only intra-operators'-scale, but also inter-operators'-scale. Regarding the multi-operator radio resource operations, a promising idea to improve energy efficiency and coverage is the use of UEs belonging to a given MVNO to act as a small cell upon coverage whole detection [95]. As for the multi-operator computational resource operations, it would ensure VNFs placement at the various physical network locations (for NFV enabled resources), and their efficient migration from under-utilized and high energy cost to low energy physical locations, targeting the minimization of the energy consumption. This requires proactive ML frameworks that can predict the future traffic among MVNO's users, and consequently pre-organize optimal NFV.

10.4.3 EE Hybrid Resource Allocation for C-RAN

Given the aforementioned state-of-the-art, we identified several open research challenges, among others, the need to explore more EE hybrid resource allocations that cater for both types of resources and maintain QoS aware behaviour when performing RRH to computational BBU mapping. To this extent, we propose an EE resource allocation scheme that aims to minimize the power consumption at the BBU pool when mapping RRH group to BBUs, while considering users' QoS and BBU capacity constraints. The objective of our RRH group based mapping (RGBM) scheme is twofold. First, it aims to improve the throughput of users experiencing bad radio conditions, while meeting a minimal required throughput for all users in the network. Second, it targets the minimization of the number of BBU units aiming to reduce power consumption and increase the spectral efficiency, while maintaining the minimal throughput required for users. To achieve this, the proposed scheme uses two key steps: i) the formation of cooperative RRH groups aimed at improving the QoS of weak users, and ii) the formation of RRH cluster to be mapped to a minimal number of BBUs without violating the minimum user QoS demands.

The considered system is a C-RAN composed of a set of N distributed RRHs, R where, $R = \{r_1, r_2, .. r_N\}$ connected to a BBU pool through high performance links, for centralized baseband processing and resource block scheduling tasks.

The cooperative RRH grouping performed in the first step of our scheme targets the improvement of weak users' conditions. Formed groups are denoted as the set $G = \{g_1, g_2, .. g_N\}$. The joint transmission coordinated multi-point (JT-CoMP) and transmission-point selection (TP-Selection) is considered among a group of cooperative RRHs. The group formation procedure starts by identifying weak users, i.e., users experiencing a signal to noise ratio (SINR) below a fixed threshold. Then, it considers for every weak user a list of the most interfering RRHs for group formation. An RRH candidate on that list would potentially join the group if the grouping conditions with respect to the minimal required throughput are met for other users.

Once the groups are formed, the second step aims to form clusters of RRH groups. The set of formed clusters is denoted $B = \{b_1, b, .. b_N\}$. These clusters are mapped according to a one-by-one association relation to a set of baseband units at the pool level. Within one cluster, members are RRHs groups including singleton RRH groups.

We consider also that all RRHs start operating with a frequency reuse of one. Then, the scheme relies on attributing the distributed antenna system (DAS) behaviour to formed groups and then

formed clusters. That is, attempts for bandwidth sharing are first considered among small scale groups, which consist of cooperative RRHs following the RRH group formations steps, then for larger scale groups which are composed of RRHs cluster (BBU units) during the clustering formation procedure.

10.4.3.1 Defined as a Linear Programming Problem

In this section, we provide the mathematical modelling of the RGBM scheme.
We start by deriving the throughput, after and before the group formation procedure for weak and normal users.
Initially, the SINR of a weak user u_e connected to an RRH r on resource block rb is:

$$\Gamma_{e,r}^{rb} = \frac{Pt_r \, l_{e,r} \, h_{e,r,rb}}{\sum_{r' \in R, r \neq r'} Pt_{r'} \, l_{e,r'} \, h_{e,r,rb} + N_0} \qquad (10.2)$$

Where Pt_r , $l_{e,r}$ $h_{e,r,rb}$ denotes the transmission power of RRH r, the path loss and the small-scale fading between user u_e and RRH r, respectively.
The SINR and throughput of a weak user being served by a formed group of cooperative RRHs (g), would be improved according to (10.3-4), respectively.

$$\Gamma_{e,r}^{rb} = \frac{\sum_{r \in g} Pt_r \, l_{e,r} \, h_{e,r,rb}}{\sum_{r' \in g', g \neq g'} Pt_{r'} \, l_{e,r'} \, h_{e,r',rb} + N_0} \qquad (10.3)$$

$$Th_{e,r} = B_0 log_2(1 + \Gamma_{n,g,r}^{rb}) \qquad (10.4)$$

Where B_0denotes the sub-channel bandwidth.
The SINR and throughput of a normal user part of a formed group of cooperative RRHs, g, are provided by Eq. 10.5-6, respectively.

$$\Gamma_{n,r}^{rb} = \frac{Pt_r \, l_{e,r} \, h_{e,r,rb}}{\sum_{r' \in g', g \neq g'} Pt_{r'} \, l_{e,r'} \, h_{e,r',rb} + N_0} \qquad (10.5)$$

$$Th_{n,r} = B_0 log_2(1 + \Gamma_{n,g,r}^{rb}) \qquad (10.6)$$

We assume that the initial bandwidth allocated to each RRH is B_g and that upon the group formation procedure, cooperative RRHs part of the group share this bandwidth. We assume that there are a number RB_g of resource blocks per TTI to be assigned to all users of group, U_g given the bandwidth B_g.
The cumulative throughput experienced by a group g is hence:

$$Th(g) = \sum\nolimits_{u \in U_g} \sum_{r \in RB_g} Th_{u,r} \qquad (10.7)$$

$$\text{Where } Th_{u,r} = \begin{cases} Th_{e,r} & \text{if u is weak user} \\ Th_{n,r} & \text{if u is normal user} \end{cases} \qquad (10.8)$$

We use a binary variable denoted $x_{b,g}$ which determines the association of a group g to a BBU b such as:

$$x_{b,g} = \begin{cases} 1 & \text{if group } g \text{ is associated to BBU } b \\ 0 & otherwise \end{cases} \qquad (10.9)$$

That said, the throughput experienced by cluster b is defined by (10.10).

$$Th(b) = \sum\nolimits_{g \in G} x_{b,g} \, Th(g) \qquad (10.10)$$

Regarding the power model at the BBU pool, we consider a power consumption at each BBU that varies linearly as a function of the load processed in terms of offered throughput. The power consumed by a BBU b is, hence, provided by Eq. 10.11.

$$PC(b) = \tau + \mu\, Th(b) \quad (10.11)$$

Where τ and μ reflect the power consumption by active BBU b with no traffic and the coefficient varying with the traffic, respectively.
We formulate in (10.12), the optimization utility function (UT) for cluster formation reflecting the targets of the RGBM scheme in terms of power minimization and meeting the minimum required throughput for all users.

$$Minimize(UT) = y_b \sum_{b \in B} PC(b) \quad (10.12)$$

Where

$$y_b = \begin{cases} 1 \text{ if cluser } b \text{ is chosen} \\ 0 \ otherwise \end{cases} \quad (10.13)$$

Subject to:

$$\text{C1:} \quad \sum_{b \in B} y_b x_{b,g} = 1 \quad (10.14)$$

$$\text{C2:} \quad \frac{Th(b)}{N_b} \geq Th_{min} \quad (10.15)$$

Constraint C1 ensures that a group g can be mapped to only one cluster b. Constraint C2 reflects that the number of users N_b processed by the same cluster b must satisfy a minimum required throughput.
This problem formulation belongs to the integer linear programming class and is NP- Hard.

10.4.3.2 Optimization Vs Heuristics

In this section, we provide a low complexity solution to the problem defined in the previous section. The solution relies on the use of an efficient greedy approach for the RRH group based mapping [96]. Algorithm 1 depicts the proposed solution. The inputs for the algorithms are the set of groups G, the number of users at each group U_g, the number of resource blocks used by each group and the number of resource blocks available at each cluster, R_c.

At each iteration, the RGBM algorithm selects an RRH group that minimizes the utility function (UT). The group in only mapped to the currently filled cluster b, if it meets C1 and C2.

Algorithm 1 RRHs groups-BBU Mapping

Inputs: RRH groups $\mathcal{G} = \{g_1, g_2, .., g_i.., g_n\}, U_g, RB_g, RB_c$
Outputs: Set of optimal clusters $\mathcal{B}$ **AND** BBU_r
Initialization: $G_{unmapped} = \mathcal{G}$ and $\mathcal{B}$ = $\varnothing$;
while $G_{unmapped} \neq \varnothing$ **do**
  Map first group g_1, $(\mathcal{B} \bigcup g_1)$ **AND** set **b** $= \{g_1\}$;
  Update: $Queue = \mathcal{G} - g_1$ **AND** $G_{unmapped} - g_1$;
  while $Queue \neq \varnothing$ **do**
    Choose g_i, such that $UT(\mathcal{B} \bigcup g_i)$ is minimal;
    if *C2 and C3 are satisfied* **then**
      b = **b** $\bigcup g_i$ **AND** $Queue$=$\mathcal{G}$ - g_i;
      Update: $Sum(RB_g)$ **AND** $Sum(U_g)$;
    else
      $Queue$=$\mathcal{G}$ - g_i;
    end
  end
end

10.4.3.3 Performance Evaluation

We evaluated the performance of the RGBM approach with Matlab based simulations, comparing it to two state-of-the-art schemes for mapping, those being: the bin packing minimization (BPM) based mapping and the conventional One-To-One (OTO) mapping. The BPM represents the classical resolution method of the state-of-the-art which accounts only for a fixed per BBU capacity as a constraint and overlooks the users' QoS requirement. On the other hand, the OTO represents the conventional scheme used in distributed RAN where each RRH is mapped to one BBU. The considered C-RAN architecture comprises 19 RRHs along with a uniform distribution of users [96]. The performance evaluation accounts for different performance metrics, being: the number of required BBUs, total power consumption, average users' throughput, and energy efficiency.

Figure 10.9 depicts the comparison in terms of required BBUs as a function of the number of UEs per RRH. As can be shown, the OTO scheme exhibits the highest number of BBUs. This number reflects the total number of RRHs in the network. The use of the RGBM (proposed approach) leads to the lowest number of BBUs, considering all users densities, whereas the BPM based mapping shows intermediate usage that increasingly worsens as the user density per cell increases. Particularly for user densities equal to or higher than 14 UEs/ cell, the BPM reaches its limit by activating one BBU for each RRH similar to the OTO scheme. The outperformance of the RGBM scheme in maintaining the lowest BBUs can be justified by its interference and QoS aware nature, that tends to maximize the number of RRHs groups belonging to the same BBU, as long as the minimum required throughput of processed users is satisfied. Hence, our scheme succeeds to maximize the spectral efficiency by maximizing the share of the BBU, which leads to the use of the lowest number of BBUs.

Figure 10.10 illustrates the comparison between the three schemes in terms of induced BBU pool power consumption as a function of the number of UEs per RRHs. As shown by the figure, the results obtained with RGBM and BPM are well in accordance with the linear power consumption model used and explained in section 10.4.3.1. Indeed, for both schemes, the total power consumption increases as the number of UEs increases. Nevertheless, the power

consumption is fixed for the non-adaptive scheme OTO, where each RRH is assigned to a BBU operating at its maximal power. The results illustrated in this figure, prove that our proposed scheme succeeds to maintain the lowest power consumption for all user densities, followed by the BPM and the OTO. The better power saving capability of our proposed scheme directly emanates from the ability to effectively reduce the number of instantiated BBUs. In particular, for user densities higher than 12 UE/cell, the adaptive BPM scheme reaches the limit in terms of power savings compared to the conventional OTO. This is due to the activation of all BBU as reported in Fig 10.8. In contrast, our proposed mapping scheme succeeds to bring significant power savings, even for this high-density scenario, in contrast to the two other schemes.

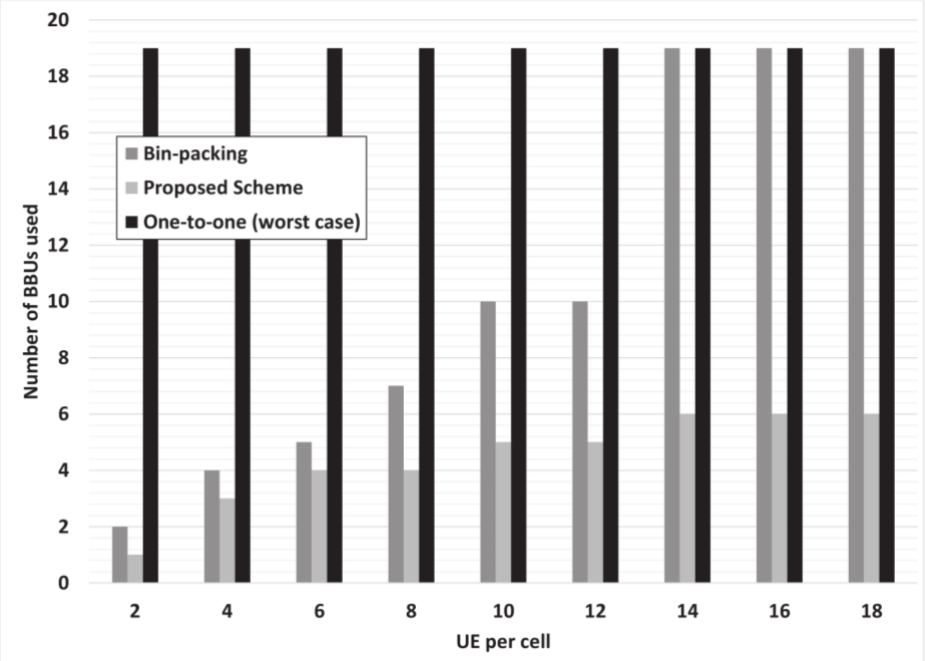

Fig. 10.9 Comparison of number of required BBUs

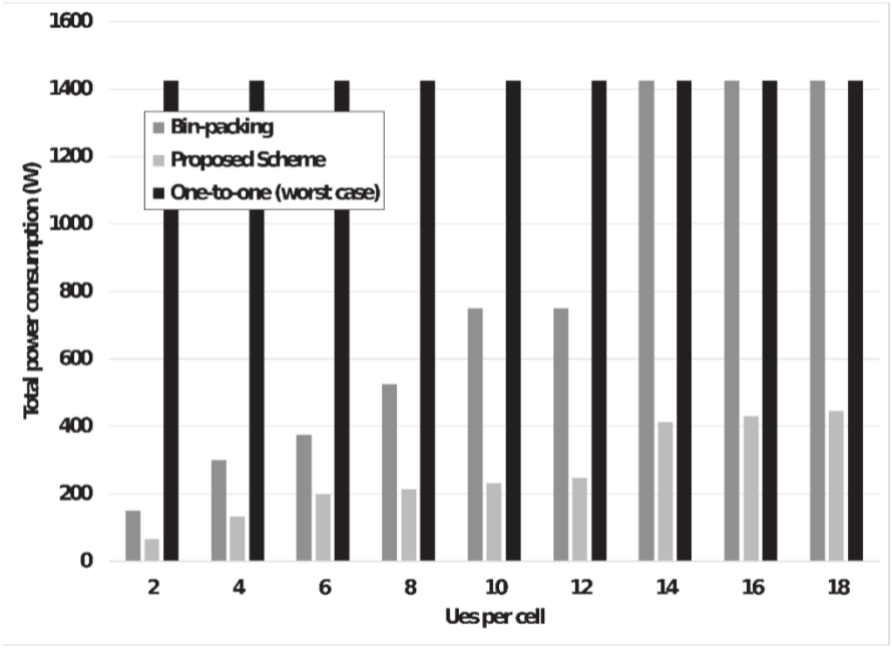

Fig. 10.10 Comparison of total power consumption

Figure 10.11 shows the comparison results with respect to the average user throughput. As one can deduct, all schemes succeed to maintain an average throughput equal to or higher than the required minimal throughput, with OTO showing the highest throughput followed by BPM, and lastly the RGBM. Indeed, the increase in the spectral reuse and energy saving achieved by our scheme comes at a cost, in terms of a slight reduction in the average user throughput while still satisfying users' target QoS.

Figure 10.12 comes to quantify the trade-off between acceptable throughput addressing the user QoS requirements, and achieving low power consumption. An ideal metric for capturing this trade-off is energy efficiency. In this context, this metric (bits per joule) reveals that the OTO is the least energy efficient due to its extreme power consumption. Both optimized mapping schemes (RGBM and OTO) show competitive results for low user densities, with our scheme achieving a distinctive trade-off for medium and high user densities.

It is worth noting that the minimization of the number of required BBUs does not only impact the overall power consumption and energy efficiency on the C-RAN, but also creates less overhead in the front-haul of the network. Indeed, according to the RGBM approach, groups are formed in a distributed way and the information required for cluster formation would only be sent per group, instead of per RRH.

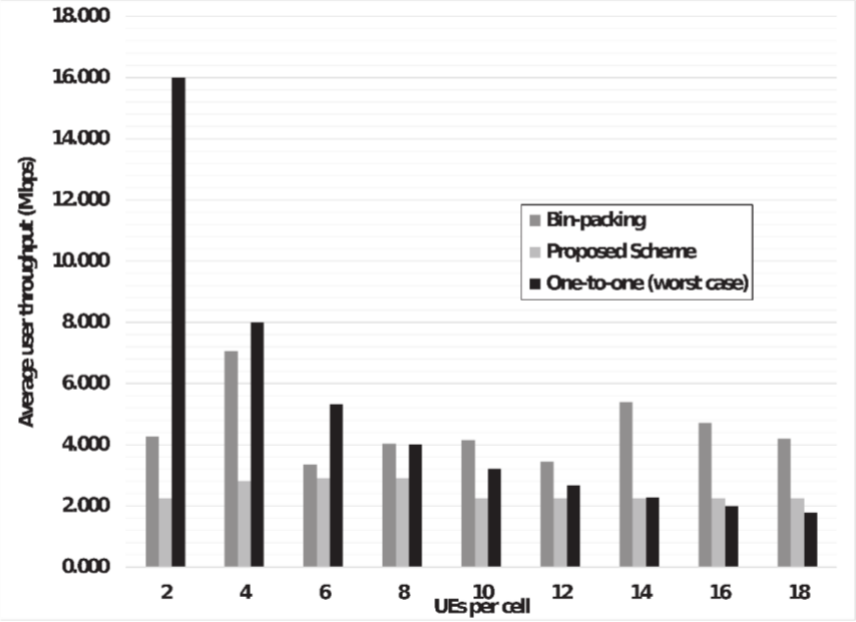

Fig. 10.11 Comparison of average user throughput

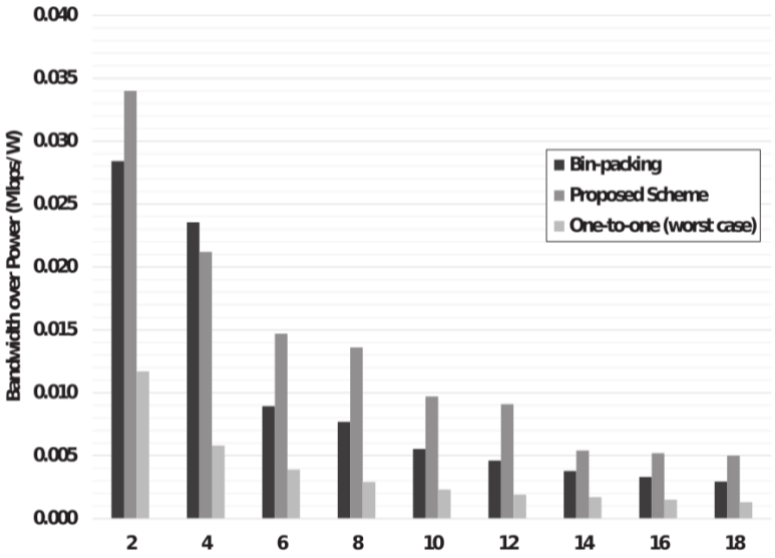

Fig. 10.12 Comparison of bandwidth over power

10.4.3.4 Conclusions

In this section, we presented a review on efficient resource allocations works with respect to energy efficiency. Based on the presented review, we identified some of the open research challenges and future research directions. Then, we proposed a group based RRH to BBU mapping (RGBM) for minimizing the number of BBUs instantiated in the C-RAN. Different from state-of-the-art approaches, the BBU minimization heuristic is further optimized, thanks

to the first stage of cooperative RRH groups formation based on QoS requirement of weak users. By considering the DAS behaviour on formed BBU (clusters), we can establish that our proposed solution leads to the formations of more optimal clusters, that is, a better adjustment of the level of interference on the network, when compared to other SoTA schemes. This leads to a better improvement of the initially detected weak users' radio and hence to the capability of consolidating more RRHs to a BBU while still satisfying the individual users QoS, when compared to the interference and QoS oblivious approaches for mappings. Simulations results have demonstrated that the aforementioned features endows our proposed solution with a higher gain in terms of power saving capability and energy efficiency, when compared to two SoTA schemes. The presented results prove that catering for the users' radio quality conditions as well as their QoS requirement in the RRH to BBU mapping can bring considerable power savings and energy efficiency to the C-RAN.

10.5 Conclusions

Softwarization and autonomous management technologies are expected to play major roles in future emerging mobile networks (5G and beyond). In this chapter, we have provided an overview of these technologies and elaborated on how they can be harnessed to provide in a broad sense greater revenue for mobile stakeholders. Moreover, the chapter elaborated design targets for the virtual infrastructure, which include greater resiliency, cost saving, and energy efficiency.

Toward designing a resilient and cost-efficient H-CRAN architecture, we presented two novel network-planning strategies. In the first instance, we proposed the SPSCP strategy, as a resilient design strategy that assigns a primary and a backup RCC node to each RAU. Besides providing resiliency, the strategy is also capable of achieving cost efficiency thanks to the maximization of the sharing in the backup connectivity and in the computing resources. Simulation results have shown that our proposed strategy demonstrates 26.9% and 14.7% of cost savings compared to the dedicated resources and the non-shareable potential of the backup resources, respectively, when assigning the primary RCC. Secondly, we investigated the benefits of processing services in large-scale data centers in contrast to the small scale, and proved that the addition of a protection path in the transport network, besides providing reliability, had the potential to provide 63% cost savings thanks to the economy of scales obtained from centralized service processing.

Toward pushing further energy saving gains in the network, we investigated EE-MO-RA considering both computational and radio resources. In this context, we proposed an RRH group based mapping (RGBM) scheme that aims to first improve weak users' radio conditions through the formation of cooperative RRH groups, followed by the subsequent minimization of the C-RAN power consumption through an efficient greedy heuristic for mapping. The simulation results presented that factoring in the level of interference in the network and the user QoS, leads to a considerable gain in terms of power saving and energy efficiency when compared to the baseline schemes (BPM based mapping and the conventional OTO).

The set of presented solutions throughout this chapter provides a foundation toward more efficient mobile network planning and resource allocation solutions for B5G systems in terms of resiliency, cost, power consumption, and energy efficiency.

Acknowledgments

This work was funded by the MSCA-ITN project 5G STEP FWD with funding from the European Union's Horizon 2020 research under grant agreement number 722429.

The research leading to these results has also received funding from the European Union's Horizon 2020 research and innovation program under grant agreement H2020-MSCA-ITN-2016 SECRET-722424.

References

[1] E. Karsenti, "Self-organization in cell biology: a brief history," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 3, pp. 255–262, 2008.

[2] F. E. Yates, *What Is Self-Organization?* Princeton, USA: Princeton University Press, 1983.

[3] M. M. S. Marwangi *et al.*, "Challenges and practical implementation of self-organizing networks in LTE/LTE-Advanced systems," in *Proc.Int. Conf. Inf. Technol. Multimedia*, 2011, pp. 94–100.

[4] H. Gacanin and A. Ligata, "Wi-Fi Self-Organizing Networks: Challenges and Use Cases," *IEEE Commun. Mag.*, vol. 55, pp. 158–164, Jul. 2017.

[5] N. Marchetti, N. R. Prasad, J. Johansson, and T. Cai, "Self-Organizing Networks: State-of-the-art, challenges and perspectives," in *Proc. Int.Conf. Commun.*, Jun. 2010, pp. 503–508

[6] T. Alsedairy, Y. Qi, A. Imran, M. A. Imran, and B. Evans, "Self organising cloud cells: a resource efficient network densification strategy: T. Alsedairy et al.," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 8, pp. 1096–1107, Aug. 2015.

[7] J. Park and Y. Lim, "Adaptive Access Class Barring Method for Machine Generated Communications," *Mobile Inf. Syst.*, vol. 2016, p. 6, Aug. 2016.

[8] J. Ramiro and K. Hamied, Eds., *Self-Organizing Networks: Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE*. Chichester, UK: John Wiley & Sons, Ltd, 2011.

[9] J. A. Fernández-Segovia, S. Luna-Ramírez, M. Toril, and C. Úbeda, "A Fast Self-Planning Approach for Fractional Uplink Power Control Parameters in LTE Networks," *Mobile Information Systems*, 2016, p. 11, Oct. 2016.

[10] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A Survey of Self Organisation in Future Cellular Networks," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 1, pp. 336–361, 2013.

[11] H.-H. Choi, S.-C. Kwon, Y. Ko, and J.-R. Lee, "Self-Organization in Mobile Networking Systems." 2016.

[12] J. Moysen and L. Giupponi, "From 4G to 5G: Self-organized network management meets machine learning," *Comput. Commun.*, vol. 129, pp. 248–268, 2018.

[13] Z. M. Fadlullah *et al.*, "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems," *IEEE Commun. Surv. Tutor.*, vol. 19, no. 4, pp. 2432–2455, Fourth quarter 2017.

[14] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks," *IEEE Access*, vol. 6, pp. 32328–32338, 2018.

[15] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, Nov. 2014.

[16] K. Sultan and H. Ali, "Where big data meets 5G?", in Proc. Int. Conf. Internet Things Data Cloud Comput., 2017, pp. 1–4.

[17] J. Chen, X. Cheng, R. Du, L. Hu, and C. Wang, "BotGuard: lightweight real-time botnet detection in software defined networks," *Wuhan Univ. J. Nat. Sci.*, vol. 22, no. 2, pp. 103–113, 2017.
[18] A. H. Celdrán, M. G. Pérez, F. J. G. Clemente, and G. M. Pérez, "Towards the autonomous provision of self-protection capabilities in 5G networks," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–14, 2018.
[19] C. C. Machado, L. Z. Granville, and A. Schaeffer-Filho, "ANSwer: Combining NFV and SDN features for network resilience strategies," in *Proc. IEEE Symp. Comput. Commun.*, 2016, pp. 391–396.
[20] M. G. Pérez *et al.*, "Dynamic Reconfiguration in 5G Mobile Networks to Proactively Detect and Mitigate Botnets," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 28–36, 2017.
[21] F. Ahmed, J. Deng, and O. Tirkkonen, "Self-organizing networks for 5G: Directional cell search in mmW networks," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–5.
[22] J. M. Arana, J. P. Han, and Y. S. Cho, "Random-Access Technique for Self-Organization of 5G Millimeter-Wave Cellular Communications," *Mobile Inf. Syst.*, vol. 2016, p. 11, Sep. 2016.
[23] R. Amiri and H. Mehrpouyan, "Self-organizing mm wave networks: a power allocation scheme based on machine learning," 2018, pp. 1–4.
[24] S. Adeshina Busari, K. Mohammed Saidul Huq, G. Felfel, and J. Rodriguez, "Adaptive Resource Allocation for Energy-Efficient Millimeter-Wave Massive MIMO Networks," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 1–6.
[25] N. Chen, B. Rong, X. Zhang, and M. Kadoch, "Scalable and Flexible Massive MIMO Precoding for 5G H-CRAN," *IEEE Wirel. Commun.*, vol. 24, no. 1, pp. 46–52, Feb. 2017.
[26] A. Mohammadkhan, S. Ghapani, G. Liu, W. Zhang, K. K. Ramakrishnan, and T. Wood, "Virtual function placement and traffic steering in flexible and dynamic software defined networks,"
in *Proc. IEEE Int. Workshop Local Metropolitan AreaNetw.*, 2015, pp. 1–6.
[27] A. G. Tasiopoulos *et al.*, "DRENCH: A semi-distributed resource management framework for NFV based service function chaining," in *Proc. IFIP Netw. Conf. (IFIP Netw.) Workshops*, Jun. 2017, pp. 1–9.
[28] M. Nekovee, Y. Qi, and Y. Wang, "Self-organized beam scheduling as an enabler for coexistence in 5G unlicensed bands," *IEICE Trans. Commun.*, vol. 100, no. 8, pp. 1181–1189, 2017.
[29] *Network Sharing; Architecture and Functional Description (Release 6)*, 3GPP Standard TS TS 36.104, 2009
[30] *Network Sharing; Architecture and Functional Description (Release 8)*, 3GPP Standard TS 23.251, 2010.
[31] 3GPP TS 23.251, "Network sharing; Architecture and functional description (Release 11)", 2011.
[32] F. Marzouk, J. P. Barraca, and A. Radwan, "On Energy Efficient Resource Allocation in Shared RANs: Survey and Qualitative Analysis," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1515-1538, thirdquarter 2020.
[33] S. E. Elayoubi, M. Fallgren, P. Spapis, G. Zimmermann, D. Martín-Sacristán, C. Yang, S. Jeux, P. Agyapong, L. Campoy, Y. Qi, and S. Singh, "5G service requirements and operational use cases: Analysis and METIS II vision," in European Conference on Networks and Communications (EuCNC),2016, pp. 158–162.
[34] M. Lashgari, L. Wosinska, and P. Monti, "A Shared-Path Shared-Compute Planning Strategy for a Resilient Hybrid C-RAN," in *Proc. International Conference on Transparent Optical Networks (ICTON)*, Jul. 2019, pp. 1–6.

[35] Ericsson AB, Huawei Technologies Co. Ltd, NEC Corporation and Nokia, “eCPRI Specification”, May 2019, Version 2.0.
[36] C. I, Y. Yuan, J. Huang, S. Ma, C. Cui, and R. Duan, “Rethink fronthaul for soft RAN,” *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 82–88, Sep. 2015.
[37] M. Lashgari, C. Natalino, L. M. Contreras, L. Wosinska, and P. Monti, “Cost Benefits of Centralizing Service Processing in 5G Network Infrastructures,” in *Proc. Asia Communications and Photonics Conference (ACP)*, Nov. 2019, pp. 1–3.
[38] Jane M. Simmons (2014), “Optical Network Design and Planning”, 2nd ed. Springer, Chapter 7.
[39] M. Jaber, D. Owens, M. A. Imran, R. Tafazolli, and A. Tukmanov, “A joint backhaul and RAN perspective on the benefits of centralised RAN functions,” in *Proc. IEEE International Conference on Communications Workshops (ICC)*, Kuala Lumpur, 2016, pp. 226-231.
[40] B. M. Khorsandi, F. Tonini, and C. Raffaelli, “Centralized vs. distributed algorithms for resilient 5G access networks,” *Photonic Netw. Commun.*, vol. 37, no. 3, pp. 376–387, Jun. 2019.
[41] M. Shehata, F. Musumeci, and M. Tornatore, “Resilient BBU placement in 5G C-RAN over optical aggregation networks,” *Photonic Netw. Commun.*, vol. 37, no. 3, pp. 388–398, Jun. 2019.
[42] J. K. Chaudhary, J. Zou, and G. Fettweis, “Cost Saving Analysis in Capacity-Constrained C-RAN Fronthaul,” i*n Proc. 2018 IEEE Globecom Workshops (GC Wkshps)*, 8, pp. 1–7, Dec. 201
[43] NGMN Alliance: 5G extreme requirements: End-to-end considerations. *White Pap*. Version 2.5, (2019).
[44] 3GPP: Service requirements for the 5G system; Stage 1. TS22.261, Version 16.8.0, (2019)
[45] N. Alliance, “Perspectives on vertical industries and implications for 5G,” *White Pap. Jun*, 2016.
[46] Q. Zhang, C. Yang, H. Haas, and J. S. Thompson, “Energy Efficient Downlink Cooperative Transmission With BS and Antenna Switching off,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 9, pp. 5183–5195, Sep. 2014.
[47] J. Li, M. Peng, A. Cheng, Y. Yu, and C. Wang, “Resource Allocation Optimization for Delay-Sensitive Traffic in Fronthaul Constrained Cloud Radio Access Networks,” *IEEE Syst. J.*, vol. 11, no. 4, pp. 2267–2278, Dec. 2017.
[48] K.-G. Nguyen, O. Tervo, Q.-D. Vu, L.-N. Tran, and M. Juntti, “Energy-efficient transmission strategies for CoMP downlink—overview, extension, and numerical comparison,” *Eurasip J. Wirel. Commun. Netw.*, vol. 2018, no. 1, 2018.
[49] P. Chand, R. Mahapatra, and R. Prakash, “Energy Efficient Coordinated Multipoint Transmission and Reception Techniques-A Survey,” *Int. J. Comput. Netw. Wireless Commun.*, vol. 3, no. 4, pp. 370–379, 2013.
[50] J. Li, J. Wu, M. Peng, and P. Zhang, “Queue-Aware Energy-Efficient Joint Remote Radio Head Activation and Beamforming in Cloud Radio Access Networks,” *IEEE Trans. Wirel. Commun.*, vol. 15, no. 6, pp. 3880–3894, Jun. 2016.
[51] Y. L. Lee, L.-C. Wang, T. C. Chuah, and J. Loo, “Joint Resource Allocation and User Association for Heterogeneous Cloud Radio Access Networks,” in *Proc. Int. Teletraffic Congr. (ITC)*, Sep. 2016, pp. 87–93.
[52] H. M. Soliman and A. Leon-Garcia, “QoS-aware Joint RRH activation and clustering in cloud-RANs,” in *2016 IEEE Wireless Communications and Networking Conference*, Apr. 2016, pp. 1–6.

[53] P. Luong, L. Tran, C. Despins, and F. Gagnon, “Joint Beamforming and Remote Radio Head Selection in Limited Fronthaul C-RAN,” in *Proc. IEEE VTC-Fall*, 2016, pp. 1–6.
[54] T. Kim and J. M. Chang, “QoS-Aware Energy-Efficient Association and Resource Scheduling for HetNets,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 650–664, Jan. 2018.
[55] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, “Energy-Efficient Resource Allocation for D2D Communications Underlaying Cloud-RAN-Based LTE-A Networks,” *IEEE Internet Things J.*, vol. 3, no. 3, pp. 428–438, Jun. 2016.
[56] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, “Energy-Efficient Resource Assignment and Power Allocation in Heterogeneous Cloud Radio Access Networks,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5275–5287, Nov. 2015.
[57] S. Luo, R. Zhang, and T. J. Lim, “Downlink and Uplink Energy Minimization Through User Association and Beamforming in C-RAN,” *IEEE Trans. Wirel. Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.
[58] G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki, “Dynamic cell activation and user association for green 5G heterogeneous cellular networks,” *in Proc. IEEE PIMRC*, 2015, pp. 2364–2368.
[59] K. Wang, K. Yang, and C. S. Magurawalage, “Joint Energy Minimization and Resource Allocation in C-RAN with Mobile Cloud,” *IEEE Trans. Cloud Comput.*, vol. 6, no. 3, pp. 760–770, Jul. 2018.
[60] Z. Yu, K. Wang, H. Ji, X. Li, and H. Zhang, “Joint User Association and Downlink Beamforming for Green Cloud-RANs with Limited Fronthaul,” in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.
[61] B. Fan, H. Tian, and X. Yan, “Resource allocation in a generalized LTE air interface virtualization framework exploiting user behavior,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2016, no. 1, p. 107, Dec. 2016.
[62] M. Peng, Y. Yu, H. Xiang, and H. V. Poor, “Energy-Efficient Resource Allocation Optimization for Multimedia Heterogeneous Cloud Radio Access Networks,” *IEEE Trans. Multimed.*, vol. 18, no. 5, pp. 879–892, May 2016.
[63] S. Wang and Y. Sun, “Enhancing performance of heterogeneous cloud radio access networks with efficient user association,” in *Proc. IEEE ICC*, May 2017, pp. 1–6.
[64] B. Wang, Q. Yang, L. T. Yang, and C. Zhu, “On minimizing energy consumption cost in green heterogeneous wireless networks,” *Comput. Netw.*, vol. 129, pp. 522–535, Dec. 2017.
[65] D. Zeng, J. Zhang, S. Guo, L. Gu, and K. Wang, “Take Renewable Energy into CRAN toward Green Wireless Access Networks,” *IEEE Netw.*, vol. 31, no. 4, pp. 62–68, Jul. 2017.
[66] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, “Joint Precoding and RRH Selection for User-Centric Green MIMO C-RAN,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 5, pp. 2891–2906, May 2017.
[67] J. Opadere, Q. Liu, and T. Han, “Energy-Efficient RRH Sleep Mode for Virtual Radio Access Networks,” in *Proc. IEEE Globecom*, Dec. 2017, pp. 1–6.
[68] N. A. Chughtai, M. Ali, S. Qaisar, M. Imran, M. Naeem, and F. Qamar, “Energy Efficient Resource Allocation for Energy Harvesting Aided H-CRAN,” *IEEE Access*, vol. 6, pp. 43990–44001, 2018.
[69] T. Sigwele, A. S. Alam, P. Pillai, and Y. F. Hu, “Evaluating Energy-Efficient Cloud Radio Access Networks for 5G,” in *Proc. IEEE Conf. Data Sci. Data Intensive Syst.*, Dec. 2015, pp. 362–367.

[70] H. Guo, K. Wang, H. Ji, and V. C. M. Leung, "Energy saving in C-RAN based on BBU switching scheme," in *Proc. IEEE Int. Conf. Netw. Infrastruct. Digit. Content (IC-NIDC)*, Sep. 2016, pp. 44–49.
[71] M. Qian, W. Hardjawana, J. Shi, and B. Vucetic, "Baseband Processing Units Virtualization for Cloud Radio Access Networks," *IEEE Wirel. Commun. Lett.*, vol. 4, no. 2, pp. 189–192, Apr. 2015.
[72] Z. Kong, J. Gong, C. Xu, K. Wang, and J. Rao, "eBase: A baseband unit cluster testbed to improve energy-efficiency for cloud radio access network," in *Proc. IEEE ICC*, Jun. 2013, pp. 4222–4227.
[73] S. R. Aldaeabool and M. F. Abbod, "Reducing power consumption by dynamic BBUs-RRHs allocation in C-RAN," in *Proc. Telecommun.Forum (TELFOR)*, 2017, pp. 1–4.
[74] W. Al-Zubaedi and H. S. Al-Raweshidy, "A parameterized and optimized BBU pool virtualization power model for C-RAN architecture," *in Proc. IEEE EUROCON*, Jul. 2017, pp. 38–43.
[75] K. Boulos, M. El Helou, and S. Lahoud, "RRH clustering in cloud radio access networks," *in Proc. Int. Conf. Appl. Res. Comput. Sci.Eng. (ICAR)*, Oct. 2015, pp. 1–5.
[76] K. Boulos, M. E. Helou, M. Ibrahim, K. Khawam, H. Sawaya, and S. Martin, "Interference-aware clustering in cloud radio access networks," in *Proc. IEEE Int. Conf. Cloud Netw. (CloudNet)*, Sep. 2017, pp. 83–88
[77] L. Ferdouse, W. Ejaz, A. Anpalagan, and A. M. Khattak, "Joint Workload Scheduling and BBU Allocation in cloud-RAN for 5G Networks," in *Proc. Symp. Appl. Comput.*, 2017, pp. 621–627.
[78] R. S. Alhumaima, R. K. Ahmed, and H. S. Al-Raweshidy, "Maximizing the Energy Efficiency of Virtualized C-RAN via Optimizing the Number of Virtual Machines," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 992–1001, Dec. 2018.
[79] S. Namba, T. Warabino, and S. Kaneko, "BBU-RRH switching schemes for centralized RAN," in *Proc. Int. Conf. Commun. Netw. China*, Aug. 2012, pp. 762–766.
[80] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in C-RAN," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, 2015, pp. 169–174.
[81] M. Y. Lyazidi, L. Giupponi, J. Mangues-Bafalluy, N. Aitsaadi, and R. Langar, "A Novel Optimization Framework for C-RAN BBU Selection based on Resiliency and Price," in *Proc. VTC-Fall*, Sep. 2017, pp. 1–6.
[82] J. Liu and O. E. Falowo, "Traffic-Aware Heuristic BBU-RRH Switching Scheme to Enhance QoS and Reduce Complexity," in *Proc. IEEE PIMRC*, Sep. 2018, pp. 1–7.
[83] S. Guo, D. Zeng, L. Gu, and J. Luo, "When Green Energy Meets Cloud Radio Access Network: Joint Optimization Towards Brown Energy Minimization," *Mob. Netw. Appl.*, pp. 1–9, Feb. 2018.
[84] J. Tang, W. P. Tay, and T. Q. S. Quek, "Cross-Layer Resource Allocation With Elastic Service Scaling in Cloud Radio Access Network," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 9, pp. 5068–5081, Sep. 2015.
[85] K. Wang, M. Zhao, and W. Zhou, "Traffic-aware graph-based dynamic frequency reuse for heterogeneous Cloud-RAN," in *Proc. IEEE Globecom*, Dec. 2014, pp. 2308–2313.
[86] M. Y. Lyazidi, N. Aitsaadi, and R. Langar, "Dynamic resource allocation for Cloud-RAN in LTE with real-time BBU/RRH assignment," in *Proc. IEEE ICC*, May 2016, pp. 1–6.
[87] A. Al-Dulaimi, S. Al-Rubaye, and Q. Ni, "Energy Efficiency using Cloud Management of LTE Networks Employing Fronthaul and Virtualized Baseband Processing Pool," *IEEE Trans. Cloud Comput.*, pp. 1–1, 2018.
[88] K. Wang, W. Zhou, and S. Mao, "Energy Efficient Joint Resource Scheduling for Delay-Aware Traffic in Cloud-RAN," in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.

[89] M. Vincenzi, A. Antonopoulos, E. Kartsakli, J. Vardakas, L. Alonso, and C. Verikoukis, "Cooperation incentives for multi-operator C-RAN energy efficient sharing," in *Proc. IEEE ICC*, May 2017, pp. 1–6.
[90] K. Wang, W. Zhou, and S. Mao, "On Joint BBU/RRH Resource Allocation in Heterogeneous Cloud-RANs," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 749–759, Jun. 2017.
[91] N. Iardella *et al.*, "Flexible dynamic coordinated scheduling in virtual-RAN deployments,"in *Proc. IEEE ICC Workshops*, May 2017, pp. 126–131.
[92] J. Yao and N. Ansari, "QoS-Aware Joint BBU-RRH Mapping and User Association in Cloud-RANs," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 4, pp. 881–889, Dec. 2018.
[93] Q. Liu, T. Han, and N. Ansari, "Energy-Efficient On-Demand Cloud Radio Access Networks Virtualization," in *Proc. IEEE Globecom*, Dec. 2018, pp. 1–6.
[94] N. Amani, H. Pedram, H. Taheri, and S. Parsaeefard, "Energy-Efficient Resource Allocation in Heterogeneous Cloud Radio Access Networks via BBU Offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1365–1377, Feb. 2019.
[95] J. Rodriguez *et al.*, "SECRET — Secure network coding for reduced energy next generation mobile small cells: A European Training Network in wireless communications and networking for 5G," A European Training Network in wireless communications and networking for 5G," in *Proc.Internet Technol. Appl. (ITA)*, 2017, pp. 329–333.
[96] F. Marzouk, T. Akhtar, I. Politis, J. P. Barraca, and A. Radwan, "Power Minimizing BBU-RRH Group Based Mapping in C-RAN with Constrained Devices," *in Proc. IEEE ICC,* Dublin, Ireland, 2020, pp. 1-7

Indexing