DOI: 10.3969/j.issn.1673-5188.2017.04.007
http://kns.cnki.net/kcms/detail/34.1294.tn.20171031.1653.002.html, published online October 31, 2017

Virtualization Technology in Cloud Computing Based Radio Access Networks: A Primer

ZHANG Xian and PENG Mugen
(Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract

Since virtualization technology enables the abstraction and sharing of resources in a flexible management way, the overall expenses of network deployment can be significantly reduced. Therefore, the technology has been widely applied in the core network. With the tremendous growth in mobile traffic and services, it is natural to extend virtualization technology to the cloud computing based radio access networks (CC-RANs) for achieving high spectral efficiency with low cost. In this paper, the virtualization technologies in CC-RANs are surveyed, including the system architecture, key enabling techniques, challenges, and open issues. The enabling key technologies for virtualization in CC-RANs mainly including virtual resource allocation, radio access network (RAN) slicing, mobility management, and social-awareness have been comprehensively surveyed to satisfy the isolation, customization and high-efficiency utilization of radio resources. The challenges and open issues mainly focus on virtualization levels for CC-RANs, signaling design for CC-RAN virtualization, performance analysis for CC-RAN virtualization, and network security for virtualized CC-RANs.

Keywords

network virtualization; CC-RAN; RAN slicing; fog computing

1 Introduction

With the explosive growth of wireless data traffic and differentiated performance requirements of mobile services and the Internet of Things (IoT), traditional networks are facing serious challenges in satisfying capacity demands, energy efficiency, and flexibility requirements [1]. Meanwhile, especially from 3G to 4G, many wireless networks have been deployed to satisfy data traffic demands and guarantee seamless coverage [2]. However, those heterogeneous wireless networks usually overlap the same areas, thus compromising the full use of infrastructure resources and radio resources [3]. In addition, based on the dedicated hardware and vertical network architecture, the wireless networks are designed for peak capacity demands and operated separately, with wireless resources and network resources being unable to be dynamically allocated. Those problems and challenges cause the capital expenses (CAPEX) and operation expenses (OPEX) not cost-efficient. From the perspective of network resources, 5G wireless communication networks are expected to intelligently integrate all kinds of resources from multiple resource owners to provide augmented and data-intensive services in a multi-vendor multi-proprietor scenario with maximizing resource utilization [4]. Traditional networks based on dedicated hardware and other resources with static or periodic allocation are unable to satisfy the requirements and facing great challenges. However, with virtualization, which can pool the resources as "virtual machines (VMs)" to realize dynamic sharing and reprogramming, the network functions and resources can be managed flexibly with higher utilization. Under different network environment, virtualization technology has different flexibility and complexity as well as challenges.

With the goal of creating a more flexible, cost-effective, and open wireless communication network, some ideas and concrete initiatives of the virtualization-driven 5G have been discussed on the 34th meeting of Wireless World Research Forum (WWRF34) [5]. Virtualization has been seen as one of the main evolution trends in the forthcoming 5G cellular networks [6]. Software defined networking (SDN), network function virtualization (NFV) and cloud computing are considered as the promising technologies to realize virtual networks, especially in core networks (CNs) and network controls. They are widely treated as the important virtualization technology in ICT area. Concretely, the fundamental principle in SDN is flexibly decoupling the control plane (i.e., configuration and management) and data plane (i.e., forwarding), therefore, SDN has been regarded as a crucial driver to virtualize wireless access and core networks [7]. NFV is mainly focused on functionality modular design and general-purpose platform replacing dedicated hardware. Through migrating network functions from the dedicated hardware to general-purpose networking and computing platforms, the cost of deploying and operating infrastructures can be reduced [8]. Through cloud computing, distributed computation infrastructures and services can also be quickly deployed, dynamically managed, and globally optimized [9].

With tremendous growth in wireless traffic and services, virtualization has been introduced into wireless networks, such as

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P1

mobile cellular network virtualization [10] and software-defined wireless mesh networks [11]. At the same time, as virtualization technology has been widely applied in CNs, in order to achieve its full potential and relief the pressure of RANs, it is necessary to extend virtualization technology to the radio access segment [8]. However, the inherent nature of wireless communications, such as broadcasting characteristic, stochastic fluctuation of wireless channel quality, severe interference and computational complexity, lets the traditional network architecture and resource allocation for high system capacity and data rates unable to satisfy diverse services with differentiated performance requirements. Fortunately, considering virtualization into RAN environments, resource virtualization with centralized and hierarchical control and management can provide more elaborate resource granularity and a greater degree of resource sharing. In addition, the signaling overhead and complexity can be decreased because of centralized processing gain and signaling-data split.

However, from the perspective of current related studies, network virtualization and wireless network virtualization are mainly focusing on mobile network virtualization based SDN and NFV. In [12], from the data plane and control plane perspective, a virtual server platform—NetVM was proposed for network functions, based on which the SDN NFV (SDNFV) for controlling the smarter data plane was introduced. The joint design of software-defined wireless networking (SDWN) and wireless network virtualization (WNV) was presented for addressing the crucial challenges in future networks and the NFV resource allocation (NFV-RA) problems were discussed in [13], while the network virtualization and resource description in SDWN were discussed in [14]. In addition, according to the requirements of 5G, a concrete approach for wireless network virtualization, including the framework model, control schemes were mentioned in [15]. The SDWN was discussed in [16], including use cases, generic architecture detailing in terms of modules, interfaces, and high-level signaling. In [17], given that SDNs mainly adopt forwarding rules as the basic control unit to manage network traffics, the rules increase dramatically after network virtualization. If the rule space is not large enough for processing new packets, network devices have to communicate to the SDN controller, which leads to latency for users and heavy load to controller. Therefore, the amount of social IoT groups vRANs under limited rule space and latency requirement has been optimized.

As to RAN virtualization, the current existing studies are mainly on software-defined and virtualized wireless access, concretely including the flow oriented perspective, protocol oriented perspective, and spectrum oriented perspective [18]. Furthermore, from the flow oriented perspective, the authors of [19] and [20] researched the network virtualization substrate (NVS) implemented in base stations, which mainly focused on the management, scheduling, and service differentiation of different data flows among different slices. In [21], multiple CN operators have accessed to a common RAN, which needs RAN virtualization to map virtual network elements onto wireless resources of the existing physical network. In [7], multiple isolated virtual networks are built on one or more physical network substrates and they are able to use customized network protocols, signal processing and network management functionalities to satisfy the intended services. Both [7] and [17] can be regarded as multiple wireless protocol instances that are isolated, customized and managed on the same wireless hardware from the protocol oriented perspective. As to the spectrum oriented perspective, the resources to be sliced are radio frequency (RF) bands and white spectrum [18] for extended sharing and high radio resource utilization, which can be treated as the spectrum virtualization.

To improve the spectral efficiency and energy efficiency, many kinds of cloud computing based RANs (CC-RANs) have been proposed, such as the cloud RAN (C-RAN), heterogeneous C-RAN (H-CRAN), and even fog RAN (F-RAN). Unfortunately, the virtualization technology in CC-RANs has not been widely and systematically researched, which can be a promising direction for future advanced RAN design. CC-RAN virtualization can fit well the type and amount, as well as the granularity of resources for the service performance requirements by dynamical and flexible abstraction and slicing.

The remainder of this paper is organized as follows. we provide a brief survey on the background of network virtualization in Section 2, including the related definitions, projects and business models, motivations, and contributions. In Section 3, we introduce the components of network virtualization with related realizing technologies. Three main CC-RAN architectures with virtualization are comprehensively analyzed in Section 4. Following, we discuss the key enabling technologies for CC-RAN virtualization to satisfy the requirements of isolation, customization and high-efficiency utilization of radio resources in Section 5. In Section 6, challenges and open issues has been discussed. In Section 7, we concluded the paper.

2 Background of Network Virtualization

In this section, we present a brief review of history and development of network virtualization. The related definitions are also given and compared, especially on CC-RAN virtualization. Then we investigate the main projects on network virtualization and realted business models, especially for the virtualization in CC-RANs. Finally, we give a brief summary of the existing related articles and summarize our contributions.

2.1 History Development of Network Virtualization

Virtualization has accelerated the development of information technology. In an information and communications technology (ICT) system, virtualization can realize the physical hardware, such as servers, network and storage devices in a software form, which simplifies IT management and enables ag-

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P2

ile IT services delivery with low costs. With a lot of benefits, virtualization has gradually become a popular concept in ICT area, such as virtual machines, virtual memory, virtual data centers and even virtual network functions. In wired networks, virtualization has been applied for decades. The authors of [22] described virtual local area networks (VLANs), virtual private networks (VPNs), active and programmable networks, and overlay networks. The common feature of them is that the wired network virtualization is limited to one or two layers, thus the benefits of virtualization have not been fully exploited. In order to take full advantages of virtualization, the networks need to be fully virtualized, and services be completely separated from their underlying infrastructure.

Since the introduction of SDN and NFV, the network virtualization has been widely researched. In [23], a network function virtualization proof of concept (PoC) was reported, which demonstrated that the dynamic SDN control coordinating with a cloud management approach could bring added value to telecommunication operators and service providers. The PoC also demonstrated how telecommunication services based on NFV can be made self-adaptive to the network conditional and user's changing requirements. Based on modular design, NFV combing with SDN control can satisfy well the differentiated scenarios and performance demands by CN slicing and further end-to-end network slicing. Accordingly, users can get better QoS because of customization. Network virtualization has been considered as one of the most significant technologies for the future Internet. With increasing mobile Internet and IoT wireless data traffic and services, wireless network virtualization has been widely discussed.

2.2 Definitions of Network Virtualization

In order to avoid confusion, in this part, the related definitions on network virtualization in the recent literature have been summarized, including virtualization, NFV, network virtualization, WNV, and RAN virtualization.

2.2.1 Virtualization

Virtualization is the process of creating a software-based (or virtual) representation of something rather than a physical one, and the heart of virtualization is the "VM", a tightly isolated software container with an operating system and application inside [24]. More importantly, a thin layer of software in platforms called hypervisor decouples VM from the host, and the hypervisor can dynamically allocates computing resources to each VM as needed. Virtualization can be applied to applications, servers, storage, and networks, which is the most effective way to reduce IT expenses while boosting efficiency and agility for all size businesses.

2.2.2 Network Function Virtualization

NFV was proposed by the European Telecommunications Standards Institute (ETSI) in October 2012, for minimizing or even eliminating the dependence on proprietary hardware [25]. Taking advantages of the evolution of IT virtualization, NFV is transferring network functions from dedicated hardware appliances to software-based applications running on commercial off-the-shelf (COTS) equipment [26], such as high-volume servers, switches, and storage. Through NFV, network functions can be instantiated in various locations such as data centers, network nodes, and even end-user equipment as the network requires. The NFV framework, state-of-the-art, and implementation as well as challenges for next generation mobile network can be found in [25].

2.2.3 Network Virtualization

Network virtualization can be regarded as a process of sharing the entire network system with an optimized approach [15]. In [27], network virtualization has been defined from the perspective of network resources. Network virtualization is any form of partitioning or combining a set of network resources, and abstracting it to users such that each user, through its set of the partitioned or combined resources, has a unique, separate view of the network. The resources can be fundamental (nodes, links) or derived (topologies), which can be virtualized recursively. The node and link virtualization involves resource partition, combination, and abstraction. NFV is a strong technology candidate toward network virtualization. Through NFV, network functionalities can be encapsulated into software packages that can be distributed through the network and performed in a homogeneous environment [28], which will simply the implementation of network virtualization.

2.2.4 Wireless Network Virtualization

WNV has a very broad scope ranging from infrastructure virtualization, spectrum sharing, to air interface virtualization [22], which can be easily understood as the network virtualization emphasizing on mobile networks, including mobile CNs and RANs. Wireless network virtualization as the technologies in which physical wireless network infrastructure resources and physical radio resources can be abstracted and sliced into virtual wireless network resources holding certain corresponding functionalities, and shared by multiple parties through isolating each other [10], [22], [29]. In [30], based on wireless network virtualization technology, several concurrent virtual networks could run on the shared wireless physical substrate and the physical nodes as well as physical links are virtualized into several virtual nodes and virtual links belonging to different virtual networks. Recently, the research of network virtualization mainly concentrates on WNV.

2.2.5 RAN Virtualization

To enable end-to-end network virtualization, both the wireless CNs and RANs have to be virtualized [19]. The network virtualization has been focused on the design of network substrate to support multiple virtual networks. However, with the relatively mature application of the virtualization technology such as NFV and SDN in mobile CNs, recent research attention has shifted to meeting the baseband-processing requirements of RANs on high-volume IT hardware as well as centralized resource virtualization and dynamical allocation to avoid over-provisioning [31].

Through radio spectrum virtualization, infrastructure sharing, virtualization of multiple radio access technologies (RATs), and virtualization of computing resources, virtualization can be applied in the context of RANs [31]. In [32], RAN virtualization is explained as leveraging SDN/NFV to virtualize a portion of the RAN functionalities onto standard IT or COTS hardware in a central location or in the cloud, which has the potential to offer advantages such as smaller footprint and energy consumption through dynamic load balancing and traffic steering. However, the radio resource virtualization and sharing were not taken into full consideration in [32].

The similar definitions of RAN virtualization, such as wireless access virtualization in [7], software-defined and virtualized wireless access in [18], RAN sharing in [19], and software-defined radio (SDR) and software-defined fronthaul (SDF) networks in [33] also have been discussed in the recent literature.

From the related definitions above, RAN virtualization is included in the wireless network virtualization, and wireless network virtualization is included in network virtualization. All three of them generally take virtualization technology, such as NFV, as the enablers. However, RAN virtualization has to deal with problems specific to the characteristics of wireless access links that usually have more dynamic sets of users, user mobility, and varying channel conditions, which makes it harder to virtualize the wireless resources across multiple entities [19]. Therefore, there exist several challenges in RAN virtualization:

- A virtualized RAN (or a RAN slice) should be able to use a permitted amount of resources by dynamical allocating, which easily varies with network load and is influenced by interference, user mobility and operator policies.
- It is harder to achieve the two conflicting goals of isolation and efficient resource utilization across slices, since the RAN has to consider the resource sharing for uplink traffic and the up-down direction of resource allocation [20].
- Wireless networks often incur considerable overheads due to signaling and retransmissions; however, the resources are rather limited.

In addition, unlike virtualization of CNs, where the virtualized functionalities are executed by VM under the control of a hypervisor on centralized general purpose processors (GPPs), RAN virtualization can only run before the RAN actually needs to interface with the physical network by transmitting an RF signal [32], which needs the system to respond in real time to the RF signal. In order to achieve the full potential of RAN virtualization, the definition of RAN virtualization needs to be extended wider than that of virtualization in the IT area, such as resource abstraction and sharing. In this paper, the virtualization technology in CC-RANs is comprehensively surveyed.

2.3 Projects and Business Models

Network virtualization has been one of the hottest topics in telecommunications. However, when it comes to RANs, it may seem technically complex, risky and hard to deploy. The authors of [22] have given a summary on network virtualization projects; however, those projects mainly focus on wireless network virtualization. In order to investigate the newest research and achievements, it's necessary to review the latest projects on RAN virtualization as well as business models.

2.3.1 Projects

Recently, the projects on virtualization of RANs have been launched, including the Flexible Architecture for Virtualizable Future Wireless Internet Access (FLAVIA) [19], [34], Small Cell Forum [35], SDN at the Edges [23], 5G Exchange (5GEx) project [36], and Small cEllS coordinAtion for Multi-tenancy and Edge services (SEAME) project [37], etc.

FLAVIA is a European Union 7th Framework Programme (FP7) project that fosters a paradigm shift towards the future wireless Internet: from pre-designed link services to programmable link processors. The significant concept of FLAVIA is to expose flexible programmable interfaces to enable service customization and performance optimization via software-based exploitation of low-level operations and control, transmission timing, frame customization and processing, spectrum and channel management, power control, etc.

SDN at the Edges is the framework of activity "SDN at the Edges" founded in 2015 by the EIT-Digital initiative under the "Future Networking Solutions" action line, which aims at investigating how to create the technical conditions for accelerating the practical development of SDN and NFV solutions.

Small Cell Forum provided a comprehensive analysis on small cell virtualization through its operator group in June 2014. The report [35] describes the findings of that activity, in particular, the benefits of centralization and virtualization of small cell RAN, including improved coordination, enhanced scalability, reduced cost, accelerated upgrade lifecycle, and flexibility. In addition, the different functional splits and related performance benefits and constraints are also discussed.

5GEx Project, through designing an agile exchange mechanism for contracting, invoking, and settling the wholesale consumption of resources and virtual network services, aims to enable cross-domain orchestration of services over multiple administrations or over one multi-domain administration and that the services can be provisioned in less than 90 minutes.

SEAME, by bringing virtualization, control and intelligence to the network edge, proposes the Cloud-Enabled Small Cell (CESC) concept, a new multi-operator enabled small cell that integrates a virtualized execution platform for deploying virtual network functions (NVFs), supporting powerful self-management and executing novel applications and services inside the access network infrastructure.

2.3.2 Business Models

With abundant services and applications unprecedentedly appearing, mobile network operators (MNOs) are under great stress because of being pipelined with little profits. Since MNOs are looking for new approaches to be cost-efficient and to relieve fast increasing financial investment burden, new business models based on an extension to existing standardized network sharing functionality have been provided in [19], including on-demand capacity, wholesale-only network, and over-the-top service providers.

Comparing with the business models developing for network sharing above, in the network virtualization environment, according to which party physical resources and virtual resources belong to, business models can be described as the roles (the business models themselves), the functions of the roles, and the relation between the roles [22], [38]. A two-level model and a three-level model as well as the functions of roles, including infrastructures provider (InP), MNO, mobile virtual network operators (MVNOs), and service providers (SPs) have been concretely explained in [22]. In [38], the business models of network slicing as a service (NSaaS) have been classified into three classes, including business to business (B2B), business to consumer (B2C), and business to business to consumer (B2B2C). In addition, according to different control levels on mobile network resources, MVNOs are further classified into three types, including resellers, services providers, and full MVNOs [10]. The detail functions of each roles and differences as well as typical representatives are discussed in [22] and [38], thus omitting them here for brevity.

Finally, the business models can also be summarized as "something" as a service (XaaS). Infrastructure as a service (IaaS) and network as a service (NaaS) are respectively provided by InPs and MVNOs. SPs can also provide software as a service (SaaS) and cloud as a service (CaaS). An example might be RAN-as-a-Service (RANaaS), where RAN is offered like a cloud-service [39].

2.4 Motivations and Contributions

Recently, RAN virtualization has been gradually concentrated on in literature. However, on the one hand, the related definitions may be in a muddle and the existing literature on network virtualization usually lacks consideration on concrete RAN architectures. RAN virtualization can be unified in different RATs, each of which is dedicated to different services and offering a different quality of service (QoS), since virtualization can simplify the management of different RATs [31]. On the other hand, the current network virtualization and wireless network virtualization are mainly emphasizing on mobile CN virtualization based on SDN and NFV, and jointly designing the generic network virtualization architecture.

To provide a better understanding of research opportunities and challenges on virtualization technology in CC-RANs, the survey gives the state-of-the-art CC-RAN virtualization and wants to attract more attention on it, unleashing the potential gains given by CC-RAN virtualization. The main contributions of the paper can be summarized as follows.

We systemically review the related definitions, and based on a brief investigation on the current network virtualization research, we give the components of network virtualization.

In order to take different advantages of CC-RANs and to satisfy the different performance requirements, we illustrate and classify the virtualization architecture into three trendy CC-RANs.

In order to satisfy the requirements of virtual networks, including isolation, customization and high radio resource utilization of CC-RAN virtualization, we review the key enabling technologies. Finally, we discuss the challenges and open issues as the broad perspectives in CC-RAN virtualization.

3 Components of Network Virtualization

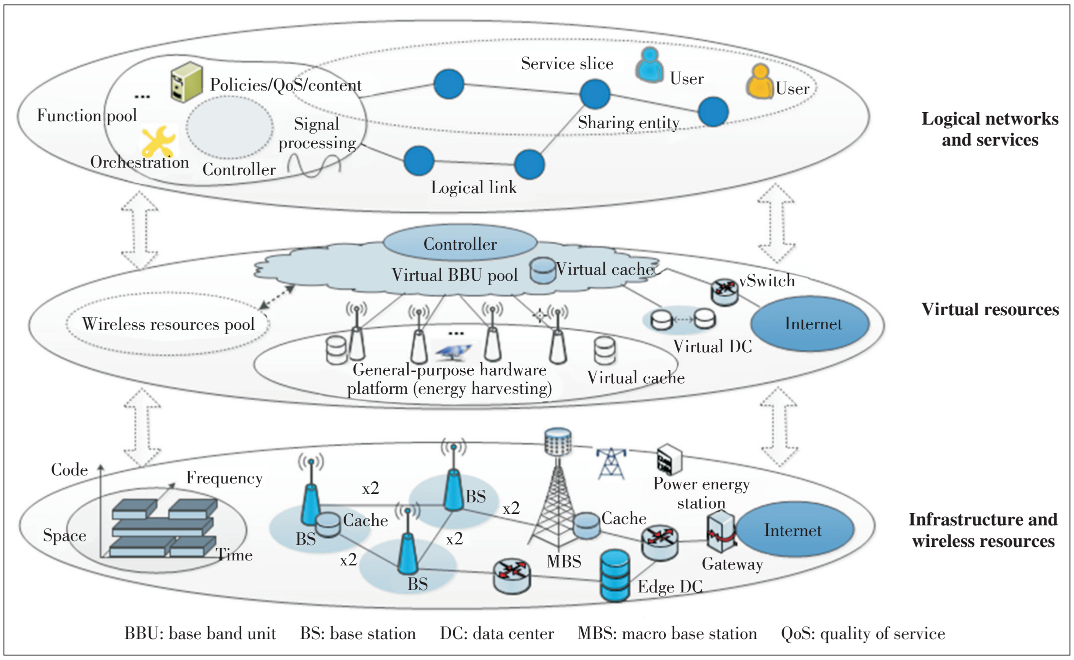
Network virtualization can be treated as the process of virtualizing a set of network resources. Since RAN virtualization is virtualization technology applied in RAN context, the components of network virtualization are same for RAN virtualization. From the perspective of resources, network virtualization can be treated as abstracting and dynamically allocating resources to efficiently share by multiple virtual networks via isolations. In **Fig. 1**, from the perspective of resource, the network virtualization architecture consists of three layers, including the infrastructure and wireless resources layer (L1), virtual resources layer (L2), and logical networks and services layer (L3). Through virtualization technology, the physical resources in L1 can be abstracted as virtual resources, which can be dynamically allocated to L3 and orchestrated for service slices.

3.1 Infrastructure Virtualization

The legacy network deployments are based on physical middle-boxes and vendor-locked equipment [23]. It may be impossible to achieve the scalability and flexibility in 5G network with higher resource utilization and satisfying various service demands. Meanwhile, through purchasing or developing software and running it on physical machines such as commodity services, all kinds of network functions can be realized. However, the gains of flexibility, dynamic resource scaling, and energy efficiency will be declined.

On the one hand, the physical network elements, such as antennas, BSs, processor hardware and routers, can be virtualized to support sharing of multiple operators by infrastructure virtualization [22]. Hardware and network sharing is beneficial for small cells in order to avoid massive over-provisioning [31], especially for the heterogeneous and dense deployments in 5G networks. In [40], a novel concept of Universal Intelligent Small Cell (UnISCell) has been proposed for enabling the

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P5



▲Figure 1. Network virtualization architecture from the resource perspective.

dense small cell via infrastructure re-engineering and sharing.

On the other hand, infrastructure virtualization also includes link virtualization as well as control and management. In [41], the processes of node virtualization and link virtualization, as well as management, have been concretely described. Specially, virtual links can be created by configuring Ethernet VLANs between the physical nodes hosting the virtual nodes. Actually, infrastructure virtualization may be directly observed in CC-RAN architecture in [42] and [43]; base band units (BBUs) are centralized and shared among different sites via virtualization, after which they are named a virtual BBU pool.

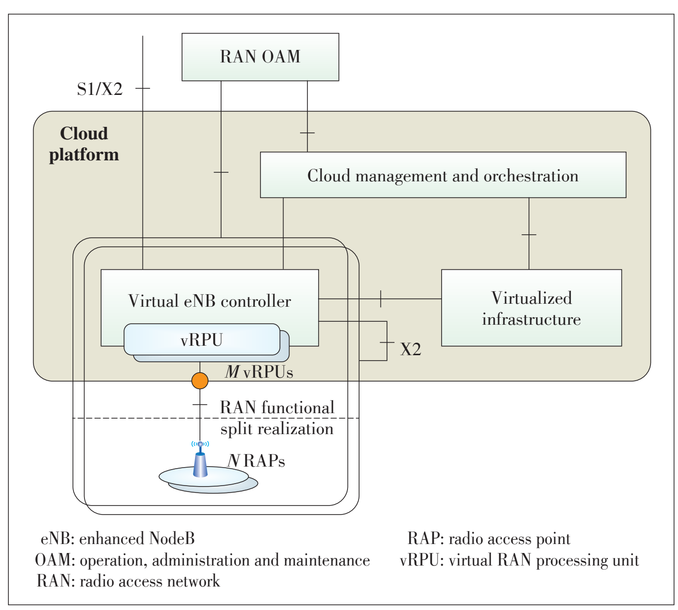
Since virtualization enables operation and management discrete, siloed infrastructure components toward a pooled infrastructure can be managed holistically. Adopting virtualized infrastructure can benefit a lot. For example, it can significantly reduce the complexity of deployment, and simplify operation and maintenance. Infrastructure virtualization can also create an elastic environment which helps business react more quickly to market demand changes and customization [24].

Fig. 2 gives an example for C-RAN architecture that applies the NFV architectural principle to an E-UTRAN based system. The enhanced NodeB (eNodeB) as a logical network entity is implemented both in the cloud platform and at physical radio access points (RAPs). The eNodeB consists of a number of virtual RAN processing units (vRPUs), responsible for executing the RAN protocol stack, and a virtual eNodeB controller that terminates 3GPP interfaces toward the core network and other eNodeBs. In the architecture, the virtualized eNodeB can be implemented as a virtualization network function (VNF), which is instantiated on a virtualized infrastructure.

Infrastructure virtualization can also achieve dynamical infrastructure resource allocation and traffic balancing. In this way, footprints and energy consumption can be lowered [44].

3.2 Radio Spectrum Virtualization

Radio spectrum resources are always scarce and significant for wireless communications. Generally, radio spectrum resources refer to the licensed spectrum or dedicated free spectrum [10]. With the emerging and developing of cognitive radio (CR) and edge computing technologies, the idle white spectrum can be used by others to alleviate the shortage, and relatively distributed computing for spectrum sensing has been applied for wireless networks. A novel wireless distributed com-



▲Figure 2. An example of infrastructure virtualization on C-RAN architecture [31].

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P6

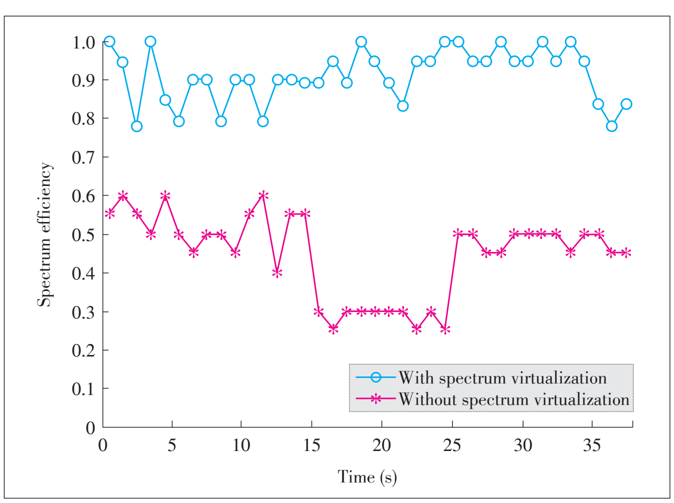
puting (WDC) scheme for cyclostationary feature detection spectrum sensing approach was proposed and investigated in [45]. Comparing with conventional Fast Fourier Transformation (FFT) time smoothing algorithms, the proposed scheme can reduce the computational complexity for each processing CR.

Spectrum virtualization considers the total available radio spectrum as a whole resource and virtualizes them as the abstracted access medium [22]. In [15], spectrum virtualization was viewed as an extension of dynamic access or sharing, which could achieve more spectrum sharing gain and improve wireless system capacity. A virtualization method and its procedures were also presented in [15], in which an advanced spectrum management (ASM) function obtains and updates related spectrum information, including available spectrum information from the cognitive plane and the latest spectrum usage information from eNodeBs, and the ASM function then allocates spectrum resource and coordinates among different eNodeBs. The evaluation results showed that after spectrum virtualization, the total utilization could increase by 30 percent (**Fig. 3**). In [18], a spectrum virtualization layer (SVL) under the physical layer was proposed to execute wireless spectrum virtualization. By spectrum reshaping techniques, the SVL could share the same RF front units on different ranges of the spectrum.

Spectrum virtualization can significantly promote full network virtualization, thus allowing extended radio spectrum shared by multiple operators [10] and higher spectrum efficiency (SE) and system capacity.

3.3 Cache/Storage Virtualization

Caching the multimedia contents at the network edge has been treated as one effective solution to dealing with the explosive growth of mobile multimedia traffic, improving the QoS of real-time data services, especially in strict-latency scenarios, and alleviating the heavy traffic burden on the backhaul and fronthaul. However, the effective in-network caching in traditional network architecture is not practical because of the dedicated hardware for signal processing in RANs and complex controlling and processing units in CNs, so the concept of "Cache-as-a-Service (CaaS)", a caching virtualization framework for cloud-based mobile networks was proposed and an illustration of the deployment of caching virtual machines was given as the Fig. 1 in [46]. CaaS is based on mobile cloud computing and a centralized caching controller is used to realize the caching VMs anywhere with properly allocated positions; by caching virtualization, the contents can be chunked, distributed, and stored based on its popularity, traffic diversity and the diversity of user requests.



▲**Figure 3. Evaluation results of the spectrum virtualization in [15].**

Meanwhile, the content-level slicing of virtual resources (**Fig. 4**) was proposed in [29], which can be considered as an extension of dynamic content access and sharing through time multiplexing, space multiplexing, etc. There are three physical contents (caches) and three services sharing the contents, and physical cache is sliced into several virtual contents used by a service without knowing the existing of other slices.

3.4 Energy Virtualization

Energy efficiency (EE) has become a key pillar in the design of communication networks [47]. EE requirement mainly lies in two aspects: the 1000x improvement comparing to 4G networks on the network side and on the device side, especially for machine-type devices requiring a very long battery life (e.g. more than 10 years) [48]. The concept of Energy-as-a-Service (EaaS) has been proposed for reducing energy costs for both users and network operators [49].

On the one hand, mobile users can offload energy-consuming applications to cloud servers for limitations of battery capacity and computing resources. The related research can be found in [50]. This is a tradeoff between energy consumption and traffic latency for mobile users. On the other hand, since electromagnetic wave is almost ubiquitous, simultaneous wireless information and power transfer (SWIPT) may be one promising way to achieve the "recycling" of transmit power when receiving data services; in order to minimize the energy cost of data transmission in the context of collaborative mobile cloud (CMC) with SWIPT, the resources allocation and user scheduling were studied in [51]. In [52], a framework named GreenDelivery was proposed, where energy harvesting (EH) based small cells (SCs) provide content delivery services with respect to the battery status and content popularity distribution. The case studies show that related power consumption is reduced. If the EH based SCs are deployed, traditional passive energy consumers may be shifted to active energy prosumers.

The energy resource virtualization mainly focuses on how to achieve energy efficiency, regardless of offloading or SWIPT, EH or any other advanced schemes, such as traffic-aware service provision and wireless multicasting.

3.5 Data Center virtualization

Data centers (DCs) are facilities that house computer sys-

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P7

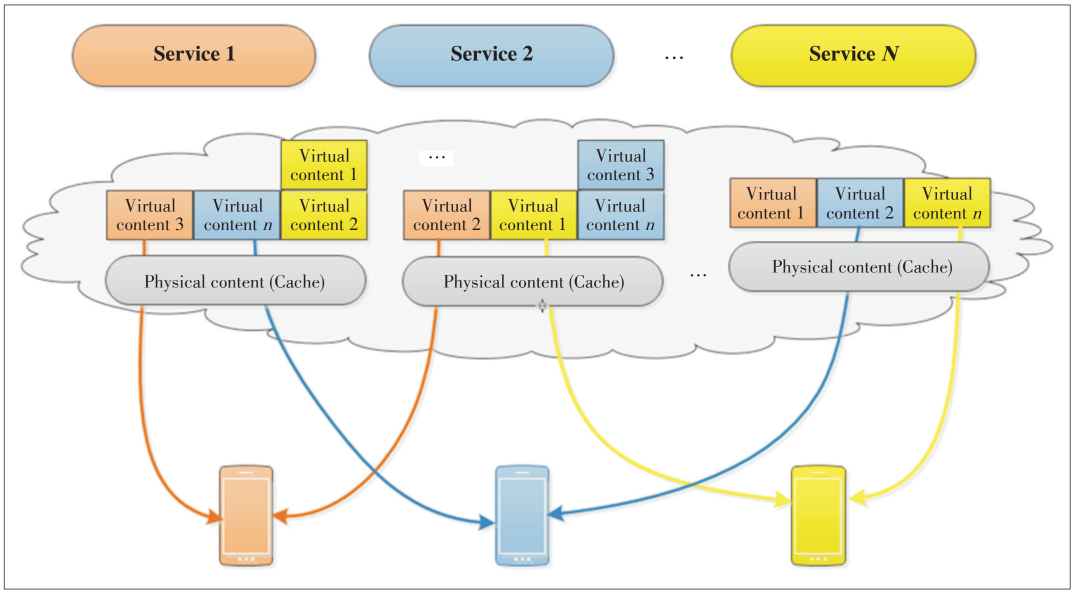


Figure 4.
An example of cache virtualization framework. The physical content (cache) is sliced into virtual contents that can be shared among different services dynamically and the slicing can be time multiplexing or space multiplexing [29].

tems and associated components such as networking and storage systems, and they are essential to satisfy ever-evolving computational demands around cloud computing, big data and IT infrastructure [53]. On the one hand, due to the delivery of explosively growing amount of data traffic over the wireless networks, considerable amount of wireless resources and costs will be occupied. Furthermore, it is expensive to build a lot of new DCs, so the best way is to improve usage of existing facilities with lower infrastructure overhead to achieve better resource management and cost efficiency [49]. On the other hand, data centers are facing some challenges such as the unguaranteed QoS, security risks, and management complexity and inflexibility, and the data center virtualization as an approach to address these challenges and virtualized data center embedding problem was studied in [54].

Data center virtualization is the process of designing, developing and deploying a data center through virtualization and cloud computing technologies, which encompasses a broad range of tools, technologies and processes that enable a data center to operate and provide services on top of virtualization layer/technology [55]. Using data center virtualization, an existing or a standard data center facility can provide/host multiple virtualized data centers on the same physical infrastructure, which can simultaneously be used by separate applications and/or organizations. This not only helps in optimal IT infrastructure/resource utilization, but also in reducing data center capital and operational costs. In addition, to reduce energy consumption, the approach of VM migration with server management and virtualization to increase the number of machines and switches that can be shutdown has been discussed [56].

4 Virtualization Architecture in CC-RANs

The RAN architecture has not experienced many changes since its birth in cellular system from the first generation (1G) to today's fourth generation (4G). Bases stations (BSs) are generally deployed and operated in a distributed way with the hardware and software tightly coupled [57]. Under the traditional RAN architecture, since control/user (C/U) planes are coupled, some BSs cannnot fall asleep or be switched off even though there is little traffic because of basic coverage, leaving some extra energy wasted. Therefore, capacity and coverage need to be separated via logically decoupling the data delivery and control signaling distribution in future R-RAN systems [58]. The network deployment and upgrade are also expensive and time consuming due to dedicated hardware and distributed deployment. In addition, facing the exponentially growing mobile Internet and IoT traffic as well as pursing 1000 times EE improvement, the RAN architecture and corresponding deployment are expected to achieve some breakthroughs.

On the one hand, cloud based network architecture takes the advantage of the centralized cloud principle to share storage and computational resources via virtualization [42], which can achieve powerful processing and dynamical resource allocation. It is worth to note that the C-RAN is a NFV instance on the RAN side to achieve soft RAN [59]. Actually, virtualization technology has been applied in CC-RANs. As mentioned in the Introduction section, the CC-RANs include the C-RAN, the H-CRAN, and even the F-RAN. All of them are integrated with virtualization technology but have some inherent bottlenecks or challenges.

On the other hand, the development of SDN is tightly connected to cloud computing, since cloud computing makes large-scale logical centralized control solutions feasible, including centralized data storing, processing, online accessing, etc. [60]. In addition, some researchers have introduced SDN into fronthaul networks named as SDF network in C-RANs [57], [61]. Therefore, the software-defined characteristic has a tight

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P8

relationship with the cloud computing, and we will give a brief review on virtualization architecture in SD-RANs, C-RANs, and F-RANs. We can discover that it is a trend that the virtualization architecture in F-RANs will be integrated with SDN, network virtualization and Cloud computing. Here, the cloud computing includes cloud computing and fog computing, since the fog computing is usually treated as an extension of classical cloud computing.

4.1 Virtualization Architecture in SD-RANs

SDN has brought a rethinking of routing and packet switching in the Internet [57]. Introducing the principles such as the C/U split, centralized control and software application programming interfaces (APIs) into RANs has been proposed in academia and industry. In [61], self-organizing networking (SON) solutions and the concept of SDF as the RAN optimization using above SDN principle were discussed. In [33], in order to enable multi-RATs interwork and heterogeneous deployment, one software-defined access (SDA) architecture was proposed and discussed. In [62], in order to achieve the conception of efficient spectrum management and sharing mechanisms, a software-defined 5G heterogeneous network (HetNet) system was introduced to support harmonized SDN-enabled approach.

In the SD-RANs, macro-cell BSs are directly connected to the controller, while the BSs of smaller cells connected with the macro-cells via a reliable backhaul link. In [63], SoftRAN was proposed as a software-defined RAN. By abstracting all base stations in a locally geographical area as a virtual big base station composed of a centralized controller and individual physical base stations just with minimal control logic, the SoftRAN was introduced as a software defined centralized control plane, which has refactored the control plane functionalities. The latency-sensitive decisions continue to be handled by the individual physical base station. In SoftRAN architecture, the global utility over a logical geographic area can be optimized and network management can be simplified by software programming.

In [58], based on the SD-RAN, a beyond cellular green generation (BCG2) architecture base on LTE system was proposed, which decouples the centralized control plane and data-forwarding, as well as decoupling coverage and capacity. The eNodeB functionalities are split between the signaling nodes and data BSs. If there is no demand, the data BSs are in asleep mode, while the signaling controller is always-on for coverage. Since the signaling node is designed based on general-purpose processors and can be sharing by operators via virtualization, the power consumption will not substantially increase. In [64], the SD-RAN controller for enterprise WLANs and the implementing platform were introduced, in which some of the main features of the SD-RAN controller were discussed, including slicing (a network slice is a virtual network), soft state, and modular architecture. **Fig. 5** shows the virtualization architecture in SD-RANs.

In Fig. 5a, the individual physical base stations include signaling nodes and data BSs. Signaling nodes are only responsible for control plane functionalities, while data BSs mainly focus on data transmissions and some delay-sensitive decisions with minimal control logic. However, through software-defined way, signal nodes and data BSs can be shifted to the others. Fig. 5b shows the centralized controller, referencing to [58], where the virtual big BS communicates with centralized controller via related APIs. It is worthy to note that the centralized controller has a global view of the SD-RAN. In the state-of-the-art proposals of SD-RANs, the centralized controller resides in either the core network or a centralized data center.

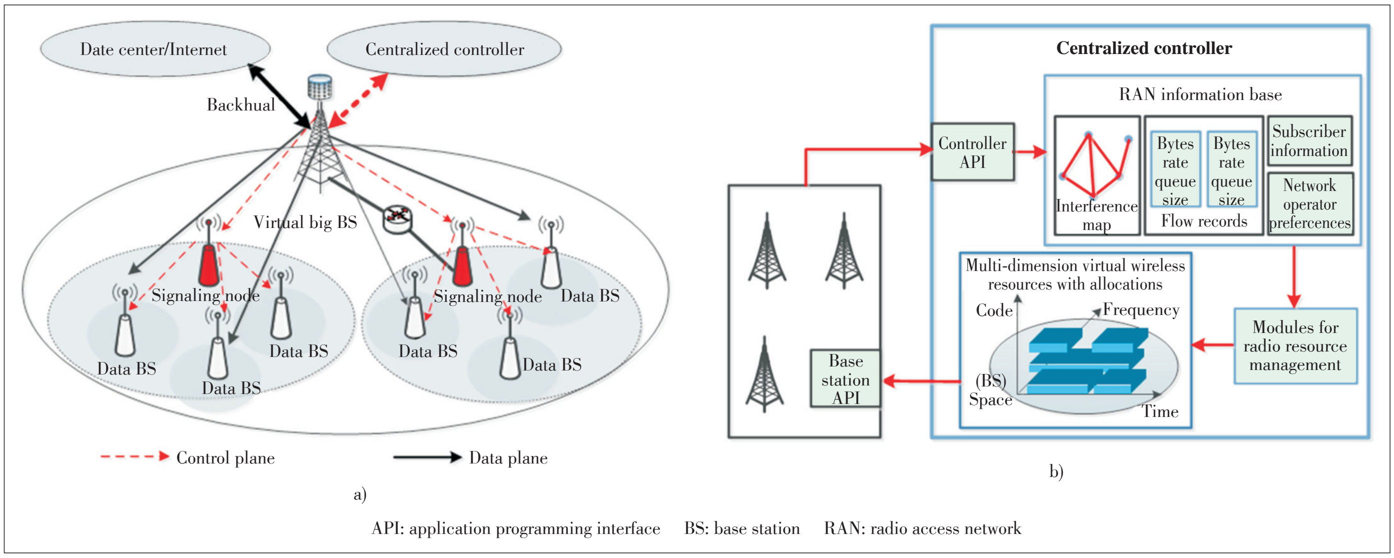
However, since the limited processing ability and flexibility of the centralized controller, the individual "virtual big BS" in the SD-RANs may fail to do well in large-scale cooperation processing (LSCP) and massive connection scenarios. It may also be caught in a dilemma in the latency-sensitive missions because of the simple control logic of data BSs and the delay between users and centralized controller. Through the signal nodes realizing control plane functionalities in a locally geographical area, the hierarchical design may perform well enough in the sparse deployment, but it is unable to effectively handle rapidly growing traffic and the dense base station deployment [63]. In addition, the virtualization gain of SD-RANs mainly comes from wireless resource virtualization and the centralized controller behind virtual big BS, leaving the CAPEX and OPEX saving not being considerable.

4.2 Virtualization Architecture in C-RANs

The SD-RANs are facing above challenges and urgent breakthroughs need to be made in RAN architecture for future network requirements. On the other hand, as a promising solution to reduce CAPEX and OPEX, the C-RAN architecture can provide high SE and EE [42]. The C-RAN architecture is a significant representative of cloud-based RANs, since underlay heterogeneous network (HetNet) with cloud computing (HetNet-CC) [65] and H-CRAN [66], [67] are based on cloud computing [68]. It offers novel schemes and architecture for centralized management and flexible network resource allocation.

The C-RAN architecture is the typical CC-RAN architecture with virtualization. Based on cloud computing, the traditional BSs are decoupled into two parts: the base band units (BBUs) migrated into a BBU pool and the distributed remote radio heads (RRHs) that are connected to the BBU pool via fronthaul links [69]. The authors of [42] have comprehensively surveyed the C-RAN architecture, including the general architecture, the systems proposed by industries and academia, and the architecture toward 5G network, so we omitted the C-RAN architecture here for brevity. Most functions of traditional BSs are virtualized as a centralized BBU pool [70], upon which multi-MNOs can share the physical infrastructures and VMNOs can share the virtual network resources according to the demands of SPs and users. Meanwhile, MNOs can quickly deploy RRHs

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P9

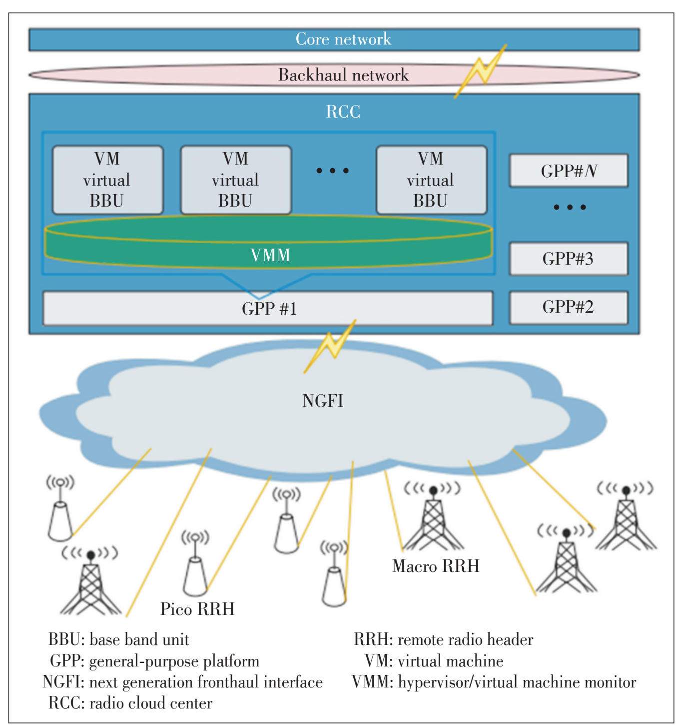


▲**Figure 5. Virtualization architecture of a) SD-RAN and b) the centralized controller [63].**

to expand and make upgrades to their cellular networks [42]. MNOs and VMNOs can also scale the computing resources of the cloud up and down depending on the live traffic demands, which can further improve the network resource utilization. However, a very important aspect to lower costs depends on how to realize the network functions of BBU pool in cloud [71]. Through virtualization technology, the BBU pool can be shared by different sites and among multiple MNOs. In [71], a BBU pool was built in the cloud based on SDN and network function virtualization concepts. In the pool, the hardware resources are virtualized and allocated to service manager by infrastructure manager on demand, which depends on the requirement of a BBU virtual instance. In [59], a C-RAN network architecture combing the ideas of SDN and NFV was introduced, and the idea that C-RAN is an NFV instance on the RAN side to realize soft RAN was illustrated. We treat the architecture as a typical virtualization architecture in C-RANs.

In **Fig. 6**, the C-RAN architecture consists of a radio cloud center (RCC), next-generation fronthaul interface (NGFI)-based fronthaul (FH) network, and RRHs. The RRC provides a cloud platform with all kinds of sharing resources and all the CC-RAN functions appear as software applications running in VMs under the management of one hypervisor/virtual machine monitor (VMM) [72]. The RCC may also provide isolated general-purpose platform (GPP) to network operators to develop new services or carry out an experiment without influencing existing operations. The baseband-related functions are processed by the BBU while the radio frequency related functions are processed by RRHs; therefore, the traditional common public radio interface (CPRI) is more and more unsuitable for future networks with networks evolving to 5G, because of its constant data rate, relatively fixed connection between the RRH and the BBU, and the sampling I/Q data rate dependent on antenna amount [59]. The next-generation FH interface has been designed. Rethinking the split BBU-RRH function, there will be several kinds of RRHs. One new RRH may contain RF functions and partial baseband processing functions. Especially, the authors have mentioned one RRH may be able to be automatically switched to another BBU pool for protecting communication reliability.

Some researchers have also introduced SDN into SDF network in C-RANs [57], [61], and the development of SDN is tightly connected to cloud computing, since cloud computing makes large-scale logical centralized control solutions feasible



▲**Figure 6. Virtualization architecture in C-RANs [59].**

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P10

[60]. Recent literature has started rethinking the C-RAN architecture with combining multiple virtualization technologies. Based on the idea of SDN, including SD-RAN [63], SDF [58], and the NFV on cloud infrastructure, the software-defined hyper-cellular architecture (SDHCA) has been proposed in [57] as its Fig. 1.

In the physical representation, the architecture can be divided into three subsystems. It is worthy to note that the RRHs have flexible functions. They can merely execute RF transmission/reception, or be dynamically configured to play the roles as control base stations (CBSs) and traffic base stations (TBSs) with some baseband processing functions. Especially, the deployment of RRHs can be compatible with the traditional network planning mechanisms, thus realizing the largely multiplexing of traditional BSs. On the other hand, the proposed SDHCA can provide one control coverage and multiple conceptual layers for different user traffic types as we can see in the logical presentation.

Concretely, through software-defined and function virtualization on the cloud infrastructure, the RRHs can be flexibly configured and the network functions can be split from hardware. The control and data planes of SDF network are also decoupled, which may be one direction to deal with the constraints of fronthaul capacity and coat-inefficiency [69]. By software applications running on VMs, the network functions, such as air interface control, service analysis, and fronthaul control, can be easily programmed and updated. In addition, the computing resource utilization may be higher since customization and the power consumption may be decreased in the sleeping model.

4.3 Virtualization Architecture in F-RANs

The traditional C-RAN architecture [42], [73], [74], with centralized processing architecture and fronthaul bandwidth constraints, may lead to time latency, unreliability, and jitter, especially when traffic load is heavy [69]. At the same time, since the emerging wave of the IoT and latency-sensitive applications, a new platform called fog computing has been proposed in [75] to satisfy such requirements as location awareness, mobility support, distributed deployment and low latency. Fog computing extends the cloud computing paradigm to the edge of networks, while fog and cloud use the same resources (networking, computing, and storage), and share many of the same mechanisms and attributes, such as virtualization and multi-tenancy [76]. Fog computing, cloudlet and mobile edge computing as three typical edge computing technologies were introduced and compared in [77], and the network slicing in edge computing was pointed out as a challenge, which may be realized via RAN slicing based on virtualization technology.

In [78], an integrated architecture for software-defined and virtualized RANs with fog computing was proposed to deal with IoT scenarios, since the tremendous number of communication devices connecting to wireless networks imposes huge challenges on network deployment, management, and data processing. The core idea of the architecture is that SDN decouples the control plane and data plane to provide network programmability, and virtualization realizes the sharing of network and radio resources among virtual slices, as well as the benefits of fog computing at the edge of networks. In addition, a virtual resource chain (or network-level virtualization) called Openpipe, and interface architecture was proposed. The Openpipe can be treated as an end-to-end network slice. Similarly, the SDN and network virtualization can be applied into F-RANs, which will bring great benefits into F-RANs.

The F-RAN has been presented in [79] as a promising paradigm to provide high SE and EE, while maintaining low latency for future wireless networks. Through local radio signal processing, cooperative resource management, and distributed storing capabilities in edge devices, F-RANs can decrease the heavy burden on fronthaul and avoid the large-scale radio signal processing in the centralized BBU pool, both of which can decline delay and improve resource utilization.

The related research on F-RANs has been widely done. In [79] and [80], the system architecture, key techniques such as transmission mode selection and interference suppression of F-RANs, and performance analysis based on cache, mode selection, radio resource allocation were studied for optimizing SE, EE and latency. In [81], combing the advantages of C-RAN and F-RAN, the F-CRAN was proposed as the harmonization architecture. In [82], an information-theoretic framework was introduced to capture the key tradeoff between delivery latency and the capacity of fronthaul and caching storage, with a conclusion that edge caching, fronthaul and wireless transmission should be jointly designed to leverage synergistic and complementary advantages of edge processing and virtualization.

However, as discussed in Section 1, F-RANs are also facing the challenges such as huge amount of power consumption and carbon dioxide CO2 gas [83], the complex management and mobility support requirements, and compatibility for multi-RATs due to its heterogeneity and distribution characteristics. Moreover, the cost efficiency, flexibility and high resource utilization are always the pursuing of network deployment. In addition, differentiated IoT application scenarios, such as connected vehicle, smart grid, wireless sensors and actuators networks, [75] need customized network performance, such as low latency and massive connectivity. The F-RAN virtualization may be a powerful way to deal with the above challenges.

Based on the fog computing architecture in [76], and the F-RAN architecture in [79] and [80], we design the virtualization architecture in F-RANs (**Fig. 7**). The architecture consists of the core cloud and edge cloud. In the core cloud, the centralized BBU pool is designed with virtualization technologies, such as NFV [25]. In the edge cloud, there are fog access points (F-APs), each of which integrates the front RF, the local distributed collaboration radio signal processing (CRSP), the cooperative radio resource management (CRRM) and caching capabilities, RRHs, UEs and other communication nodes in di-

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P11

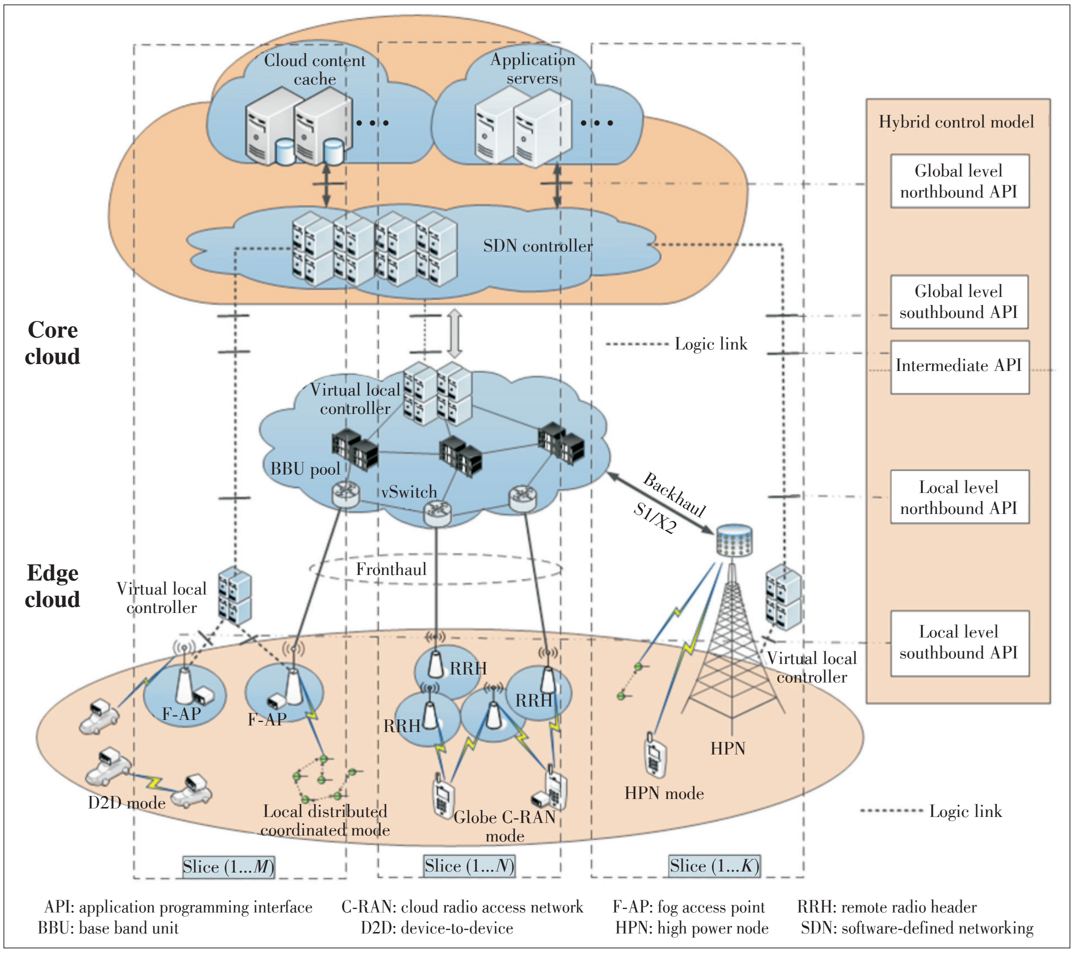


Figure 7. Virtualization architecture in F-RANs.

verse service scenarios. All these functions have their performance requirements, such as high rates for mobile Internet services, ultra-low delay and high reliability in vehicle networks, massive connections in intelligent metering and measurement with adaptive transmission mode selection. Based on virtualization technologies and network slicing, there are multiple slices coexisting on the same physical network.

In order to better deal with the distributed and hierarchical characteristics in F-RANs, the architecture in [84] adopted the hybrid control model of SDN controller as Kandoo. The model SDN controller (in the global level) defines the specific policies and sends them to local controllers (in the local level, which can be in a slice) through intermediate interface (intermediate API).

5 Key Enabling Technologies in Virtualized CC-RANs

The enabling technologies for wireless network virtualization have been presented according to different RATs in [22]. The isolation level, control method, and purpose were also mentioned to show that these technologies can be used as a taxonomy to classify key technologies. We have reviewed some related literature on virtualization in the context of RANs, such as [19]–[21] and [31].The requirements for RAN virtualization have been universally mentioned as customization, isolation, and high resource utilization. And since the RAN architecture with virtualization technology, such as sharing GPPs and other virtual resources, the C/U split and software-defined has broken the traditional networking and resource allocation ways, and new challenges and corresponding key enabling technologies are on the way to RAN virtualization.

5.1 Requirements of RAN Virtualization

In this section, the state-of-the-art key enabling technologies for virtualization in CC-RANs are presented as virtual resource allocation, RAN slicing, mobility management, and social-awareness, all of which are mainly encompassing on how to realize RAN virtualization properties and requirements, namely customization, isolation, and high-efficiency utilization of radio resources.

5.1.1 Customization

Customization mainly includes the resource allocation customization and the flexible architecture reconstruction to satisfy related network service demands. In [20], customization stresses that network NVS should provide simple and appropriate programming interfaces to enable customized solutions within slices (a slice consists of all the flows of an entity that requests virtualized resources), which is resource allocation customization. Similarly, in [19], through providing flexibility to different entities, these entities can program the vBS to optimize their service delivery. One resources negotiation for network virtualization (RENRV) algorithm has been proposed in [21], where resource customization is attained during the detection phase, and the requesting BS will be allocated resources according to its actual traffic load conditions. As to the architecture aspect, obviously, since the traditional "one-size-fits-all" network solution today is unable to satisfy future network service demands well, the Nokia analyzed dynamic end-to-end network slicing, which is introduced to establish multiple parallel network instances and realize customization in architecture reconstruction [85].

5.1.2 Isolation

Since multiple virtual network slices should coexist, isolation is the basic issue in virtualization. Isolation means that any changes in one virtual slice, such as the number of end users, mobility of end users and fluctuating of channel status should not cause any serious interference and negative influence on resource allocation for other slices [20]. In [86], isolation among slices was introduced as the essentiality about managing network and computing resources to realize a picture in which the performance of one slice is not affected by the operation of another slice. Moreover, the protection mechanisms were mainly concentrated on protecting common (control) channels or resources used for UEs accessing system.

However, each slice may be assigned with either shared or dedicated radio resources up to radio resource management (RRM) implementation and service level agreement (SLA) [87]. Therefore, different use cases and service requirements will need different isolation levels, which refers to the minimal resource granularity [22], in which the isolation ways, such as MAC layer, packet, flow, multilevel, and even traffic types and context, have been surveyed from sub-carriers, sub-channel, PRB, time-slot, space, upper layers and protocols. However, the ideas of isolation is mainly using orthogonal resources. In [21], the isolation is achieved during the transfer phase, through Resource nEgotiation for Network Virtualization (RENEV) algorithm ensuring a reserved portion of resources to each requesting BS, the isolation and resource demand can be satisfied. The two above methods may have low resource utilization with low complexity.

Actually, using completely orthogonal resources to realize isolation may be not practical, since the resources are limited, especially in RAN virtualization with scarce spectrum resources. Therefore, the interference or negative effects under one threshold may be able to treated as an isolation. From the discussion above, the isolation can be considered from the protocol design such as MAC scheduling protocols, and resource allocation.

5.1.3 Resource Utilization

Given the RAN virtualization is treated as the RAN resources mapping to multiple virtual RAN slices, resource utilization can be defined as the actually mapped resources of one slice divided by the total amount of available resources for all virtual slices [22], [88]. To evaluate the resource utilization of a virtual slice, it can be derived with the used virtualized infrastructure resources, bandwidth, power, time-slots, and processing capacity resources divided by the relatively available resources [22]. In [19] and [20], dynamically adjusting resources across virtual slices to maximize the resources occupied as much as possible was discussed. In [21], the RENEV guarantees the efficient physical radio resource utilization by reconfiguring the medium access of each pair of requesting-requested BSs with a rational signaling burden.

Finally, RAN virtualization should guarantee the efficient use of physical radio resources, computing resources, and other resources. The virtual resource allocation will play a significant role in higher resource utilization, which we will discussed in next subsection. At the same time, the virtualization technology needs to cope with the tradeoff between complexity and efficiency in dynamic resource allocation.

5.2 Virtual Resources Allocation in CC-RANs

The problem of embedding virtual networks in a substrate network is mainly the resource allocation in network virtualization [88]. Similarly, in CC-RANs, the virtualization can largely be treated as the allocation of virtualized infrastructure resources and wireless resources, thus the virtual resource allocation plays a vital role in the CC-RAN virtualization. The final goal of virtual resource allocation is to improve the resource utilization as much as possible with pursuing the higher system capacity and satisfying SLA requirements. However, the following has to be taken into consideration for different service scenarios: diverse service requirements, traffic fluctuation characteristics and the inherent characteristics of wireless environment such as broadcasting and stochastic fluctuation of channel quality, different resource granularity and resource priority, and the heterogeneity of IoT devices. For example, ultra-low latency services need more computing resources while high data rate services should be provided enough spectrum resources. Therefore, in virtualized RAN environment, among multiple virtual CC-RAN slices, the virtual resource allocation will be more challengeable. Fortunately, there have been some related works on virtual resource allocation.

Considering the fog computing at the edge of network and the increment of resource dimension (radio resources, edge caching and edge computing resources, and virtual network functionalities), the virtual resource allocation in virtualized F-RANs can be carefully designed from two significant aspects.

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P13

One is virtual resource allocation framework, and the other is joint resource allocation and resource granularity.

Virtual resource allocation has been widely discussed mainly based on a hierarchical architecture, which is suitable for virtualized F-RANs. since the diverse IoT devices and service scenarios as well as the distributed characters of F-RANs, the hierarchical architecture can be more efficient and flexible.

In [89], a hierarchical virtual resource allocation model was proposed to maximize the MVNO profit with the consideration of backhaul capacity of the infrastructure providers (InPs) and users' QoS requirements. In [90], a base station equipped with a large number of antennas serves users who belong to different virtual slices (a typical hierarchical scenario) was analyzed and joint power, sub-carrier and antenna allocation problems were studied to maximize the sum-utility while maintaining the minimum rate for each slice. In [91], a joint BS assignment, sub-carrier, and power allocation algorithm for dynamic resource allocation was proposed to maximize the network sum rates under the constraint of maintaining the minimum required rate of each slice (the users of different service providers) in multi-cell virtualized wireless networks.

Relatively, the models or methods to solve virtual resource allocation problems become more interesting. For example, the successive convex approximation (SCA) and complementary geometric programming(CGP) were adapted to develop an efficient two-step iteration approach to settle above virtual resource allocation problem in [91]. Since all involved parties, including InPs, MNOs, MVNOs, and SPs, want to maximize their own revenue, the game theoretic approach may be an effective tool [6], [92]. In [6], in order to address two-level hierarchical virtual resource allocation problem for achieving strict inter-slice isolation and the ability of intra-slice customization, a hierarchical combinatorial auction mechanism was mentioned. In [92], the dynamic interactions among the SPs and the network operators (NOs) were modeled as a stochastic game, where SPs are responsible for QoS management for their own users while NO manages the spectrum resources and regulates the game.

We can find the pervasive feature in virtual resource allocation in RANs is based on a hierarchical architecture with the introduction of MVNOs and slices for differentiated services and flexible management. The other outstanding characteristic is resource dimension increment in virtualized F-RANs. Because of the edge caching and edge computing ability of F-RANs as well as infrastructure and network function virtualization, traditional radio resource allocation may not be able to satisfy QoS requirements. The deployment problems of partial core network functionalities based on VMs and virtualized links and nodes mapping also need to be dealt with in virtual resource allocation.

In [13], virtual network embedding (VNE) and NFV resource allocation (NFV-RA) problems were discussed. VNE deals with the allocation of virtual resources in nodes and links mapping to substrate nodes and links, respectively. The VNE problem is known to be NP-hard [93] and most of the related work has been concentrated on designing heuristic algorithms and simplifying the network realization complexity. Meanwhile, VNE can be optimized by embedding cost, link bandwidth, QoS, and energy efficiency. On the other hand, NFV-RA has three different stages, including VNFs-chain composition (VNFs-CC), VNF forwarding graph embedding (VNF-FGE), and VNFs scheduling (VNFs-SCH). The VNFs-CC seeks to concatenate the VNFs efficiently for a network service in the most adequate way, and VNF-FGE seeks to find where the VNFs will be allocated in the NFVI in a suitable way, considering the requirements of individual requests and the overall requirements of all services. The VNFs-SCH seeks to determine when is better to execute each function into the NFVI to minimize the total execution time without degrading the service performance and respecting all the priorities and dependencies between the VNFs composing the service at the same time. In [8], the scheduling of wireless VNFs in the RAN was formulated as an integer programming problem. In [94], a management and orchestration (MANO) entity in virtualized infrastructure has been enabled. In addition to deriving the affinity scores for resources units (RUs) referencing to a specified RU for concrete VM instance, MANO can perform precise and efficient resource tailoring and dimensioning, which benefits the operation and management of the virtual network functions in virtualized infrastructure. Both [13] and [94] focus on virtualized infrastructure resource allocation that can be applied in that of F-RAN virtualization.

The joint resource allocation on virtualized CC-RANs also plays a vital role, since network performances such as service latency, energy efficiency and spectrum efficiency are influenced by multiple factors. When the resources are virtualized, the extended sharing and flexible allocation should be more elaborated. The joint optimization of cloud and edge processing, focusing on the strategy to populate the caches of enhanced RRHs (eRRHs) for F-RANs, can be found in [95].

Comparing with traditional static resource allocation, virtual resource allocation is expected to gain better performance, and the model and methods of resource allocation and optimization need carefully redesign with considering the resource granularity and dynamically sharing characteristic. In addition, virtual resource allocation should be further researched, especially for real-time and low-complexity demands since the virtual resource mapping processes and generic hardware with low cost may decrease specific performance.

5.3 RAN Slicing

The current relatively monolithic network architecture is not flexible and scalable enough to efficiently support future networks with diverse and sometimes extreme service requirements. In addition, the introduction of network services should be efficient and multiple use cases are expected to be operated

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P14

on the same operator network [96]. Network slicing has been regarded as one of the most significant technologies to deal with the above challenges.

In [38], the concept of network slicing as a service (NSaaS) was introduced to offer end-to-end cellular networks as a service. Three forms of networking slicing, including CN only, RAN only, and CN and RAN were also introduced. The RAN only slicing runs on wireless platforms, which consists of radio hardware and baseband resource pool. Both radio hardware and baseband resource pool can be realized on GPPs, where various parameters of air interfaces, such as symbol length, sub-carrier spacing, cycle prefix length, and the parameters of hybrids automatic repeat request (HARQ), as well as other parameters like cell selection, handover threshold, can be set up for each slice to realize logical network customization [38].

Recently, RAN slicing has been designed mainly from two aspects. One is CC-RAN resource management perspective, and the another is RAN architecture and protocol stack perspective. Through the architecture and protocol stack design, diverse logic RANs can be achieved based on the unified physical network according to SLA requirements. The SLAs can describe which QoS metrics are expected to be guaranteed, possibly denoting a minimum amount of spectrum.

From the RAN resource management perspective, the authors of [97] presented RAN slicing at four levels, including spectrum planning level, inter-cell interference coordination (ICIC) level, packet scheduling (PS), and admission control (AC), and compared them from the granularity of isolation (i.e. high level of radio-electrical isolation and low level of traffic isolation) and customization. It was concluded that the RAN slicing approaches providing high level isolation may lose the higher granularity and flexibility. In [98], the network slicing architecture featuring CC-RAN abstraction was proposed, which uses the FlexRAN concept [99] to enforce network slicing in the RAN and adapt the resource allocation policy according to the slice requirements. From the level of resource isolation, the dedicated resources and shared resources models were aslo mentioned, as well as a two-level MAC scheduler to abstract and share the physical resources among the slices.

From the protocol stack perspective, in [86], multiple RAN slices may multiplex into a common MAC, PHY and radio, while other slices may use dedicated spectrum or dedicated APs, and thus use a dedicated MAC/PHY that involves new protocol design and model selection. In [100], Chain Mobile and other companies have designed the RAN architecture with network slicing, which emphasizes that the RAN protocol stack can be tailored to meet diverse service requirements of different network slicing instances (e.g. enhanced mobile broadband (eMBB) slices, ultra-reliable and low latency communications (URLLC) slices, and massive machine-type communications (mMTC) slices, etc.).

In addition, due to the particularities of wireless environments, guaranteeing slice isolation is full of challenges [101]. As for AN slicing isolation, similar to the level of resource isolation mentioned above, there are mainly two ideas. One is that RAN slices occupy completely orthogonal resources at a resource granularity interval. The other is maximizing the network resource utilization with controlling the negative effect or radio interference under a certain threshold. The former can be easily deployed but with the low resource utilization, which can be used for low service traffic load and needs no extra overhead for interference management. The latter will be complex and needs other extra assisted technology such as intelligent spectrum sensing, interference cancellation, and other advanced virtualization or software migration technologies.

5.4 Mobility Management

Mobility management mainly includes location management and handoff management. The former enables content delivery and communications, while the latter mainly keeps service continuity when users move from one access point to another [102]. Generally, users move freely and independently, requiring ongoing communications and minimal disruptions. This may leads to stochastic and time-varied interference to others. In real network deployments, user mobility and interference in a multi-cell and multi-layer network need to be fully considered, which has a direct influence on the user experience and system capacity.

With virtualization applied in CC-RANs, virtual resource allocation, extended radio spectrum sharing, and hierarchical management architecture will leave more challenges on mobility management. For example, one user's location update may involve multiple different VMNOs or InPs, which makes the tracking location complex.

Related research has been done. In [103], based on wireless network virtualization technology, a handoff scheme was introduced to support handoff between three virtual networks, each of which has a network controller used to schedule users, in an integrated train-ground communication system.

In [104], a SDN-enabled authentication handover and protection mechanism was designed for dense and heterogenous small cell networks for 5G, which is mainly based on sharing of user-specific security context information among related APs. An authentication handover module was also implemented in the SDN controller to monitor and predict the location of users. The relevant cells prepare related resources in advance, and the user can get seamless handover authentication during mobility.

Since SDN has been introduced into virtualized F-RANs, the local controls are deployed in the edge cloud, which can support high level of SDN controller to make global handover decisions and can handle handovers in F-RAN slices. In [105], one scheme of mobility management for network slicing based on 5G system was proposed. **Fig. 8** shows the handover procedure, which is based on the SDN controllers in edge cloud as well as a hierarchical and cooperative control mechanism. The

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P15

mobility management can be employed in virtualized F-RANs.

5.5 Social-Awareness

Social-awareness is also known as social consciousness and originates from sociology, which is often used to describe the sociality and social behaviors of human beings [106]. Social-awareness has been introduced in communications, which is originated to enhance D2D communications for relieving energy consumption, data usage [107], cutting down transmission delay, etc. Recently, Social-aware energy harvesting D2D communications [108] and caching based socially-aware D2D communications [109] have appeared in literature.

Since rich diversity of mobile devices with powerful computation capacity, social-awareness has gradually been used for networking. By exploiting the users' social relationships and behaviors, the social properties of nodes can be analyzed and used. In [106], the socially aware networking (SAN) was surveyed. From the architecture of the SAN, we can see that the first two steps are to obtain social-awareness by sensing and analyzing data by using intelligent technologies, such as data mining and machine learning. Then, the SAN can deduce important social properties, including community, centrality, similarity, tie strength, and human mobility pattern. Based on these social properties, different networking protocols to satisfy the requirements of applications can be adaptively designed.

Social-awareness can be applied in F-RAN virtualization, especially in the cache/storage virtualization, energy virtualization, virtual resource allocation, and dynamically networking.

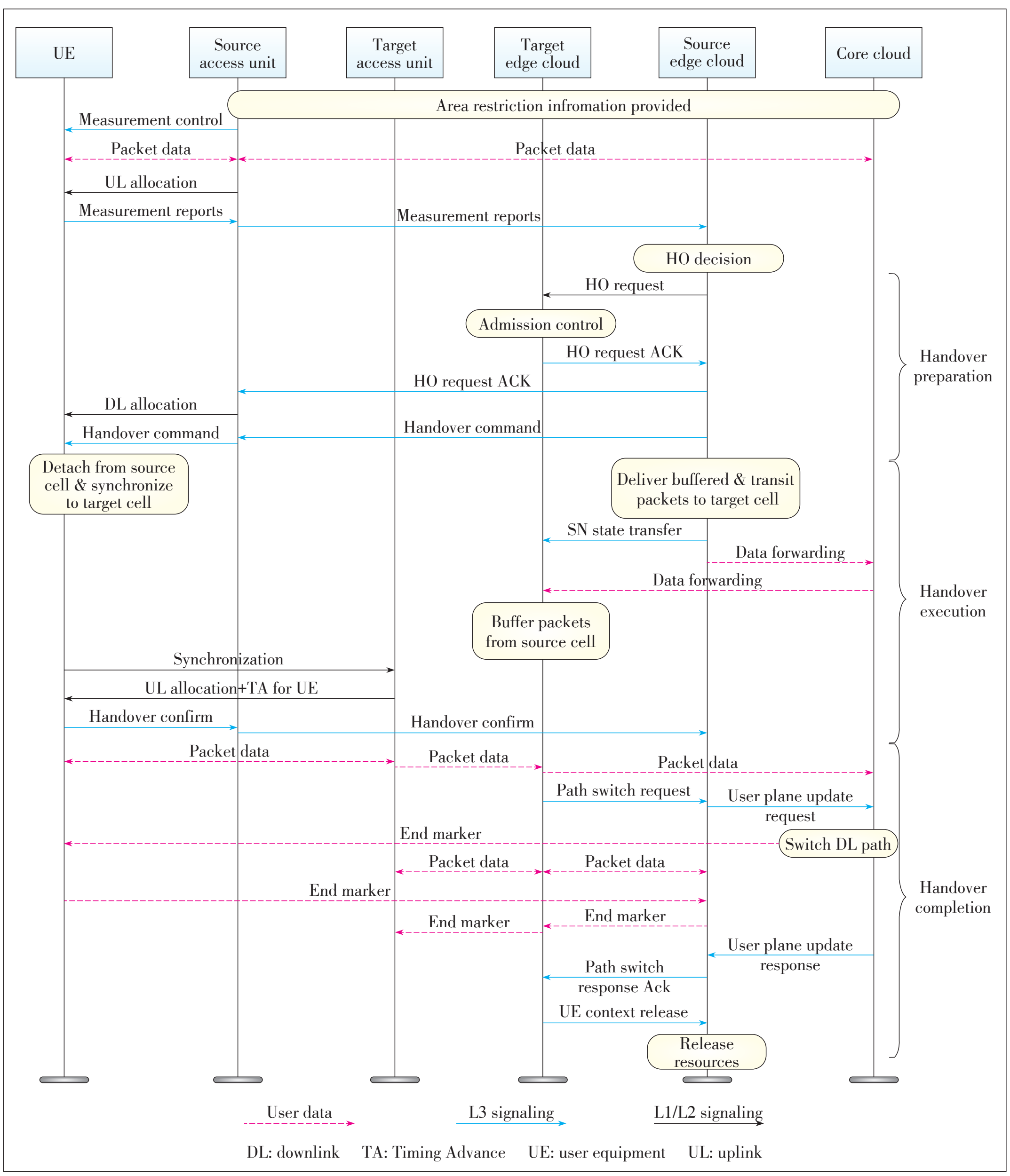


Figure 8. The handover process for virtualized F-RANs based on 5G network slicing system [105].

D:\EMAG\2017-08-58/VOL15\Review1.VFT——20PPS/P16

In a virtual environment, the entire system, including the operating system, the applications, and data, can be encapsulated as a set of VM files [24]. Based on the social-awareness, the probable locations of VMsand network resources can be predicted. With the deployment and migration in advance, the time of service request response can be declined and the resource utilization can be improved. In [110], the social-awareness was introduced into virtual network embedding (VNE) paradigm, which mainly contributes to virtual resources efficiently mapping onto physical ones. Since social-awareness provides useful information such as the importance of a node based on specific mapping algorithms, the physical resource allocation among the virtual node requests can be more efficient since the popular nodes will be initially selected.

6 Challenges and Open Issues

Virtualization technology has been applied in IT area for many years. With the explosively growing traffic of diverse services and the demand of quick network deployment with cost-efficiency, virtualization technology has been introduced into communication networks. However, there are still many challenges in the virtualization of CC-RANs before its wide and successful rollout.

6.1 Virtualization Levels for CC-RANs

Resource sharing has been fully discussed at the spectrum level, infrastructure level, and network level in H-CRANs in [28]. In [101], slicing could be done at different levels, from application and flow slicing to hardware and spectrum slicing. Different levels mean different benefits and challenges. Similarly, the virtualization levels represent different sharing and abstraction degrees, which influences the resource granularity and the system flexibility and complexity. In real deployments, it is necessary to decide up to which level virtualization should be applied to achieve an efficient virtualization solution. In CC-RAN virtualization environment, virtualization levels mean different protocol stack design. How to determine the virtualization levels and describe them are worth of further investigation.

6.2 Signaling Design for CC-RAN Virtualization

The process of virtualization maps multiple virtual networks to the same physical network with the logically centralized control function. The controller or hypervisor takes responsibility of the signaling generation and forwarding [72]. In addition, in each virtual network (or a slice), some more elaborate control signalings directly satisfy users' service requirements. Considering the differences among virtual networks, the differentiated signaling design or unified signaling set design should be analyzed. The unified design is easily generated, processed, and upgraded, while the differentiated design is more streamlined. The present common sense is that proper control signaling design needs considering delays and reliability and should be explored in a careful manner to enable the connectivity among different parties involved in CC-RAN virtualization [49].

6.3 Performance Analysis for CC-RAN Virtualization

With virtualization technology applied in wireless networks, the software-defined and functional modular design increases the difficulty in measuring the performances because of the flexibility and other constraints. As shown in [111], the initial virtualized network function (VNF)/VNF components (VNFCs) deployment strategy needs to satisfy the intra-functional constrains between multiple VNFCs as well as administrative constraints; the VNFC embodies a subset of network functions attributed to the respective VNF. The cost of the deployment is calculated in terms of the DC infrastructure resources, such as computation and networking. On the other hand, the radio resource abstraction and slicing need to be carefully consideration, especially for the virtualization in CC-RANs, which involves the radio resource sharing and the protocol design. What's more, virtualization for cloud computing brings in special challenges for cloud service performance [112], and the state of the art of cloud service performance evaluation has been presented from the system modeling perspective in [112]. The performance metrics will be more diversified, and the traditional system models and algorithms may be ineffective. Therefore, performance analysis urgently needs to be well dealt with not only for confirming the performance metrics, but also for the system models and algorithms according to different design and optimization problems.

6.4 Network Security for Virtualized CC-RANs

Due to the virtualization technology applying to CC-RANs, the extended radio spectrum sharing and infrastructure virtualization will let CC-RANs more flexible and open, which may be easily attacked. At the same time, a large number of intelligent IoT devices/nodes, even including malicious pseudo base-stations, may access networks with well self-adaption/context awareness capacities, which will bring lots of potential threats to network security. These threats may not only harm the service and information security of some users, but also restrict the entire network capacity. What´s more, the encryption often needs considerable computational resources and communication overheads [42]. In these contexts, traditional authentication, authorization, and even accounting strategies may need to be redesigned to satisfy diverse radio accesses and dynamic virtualized CC-RAN reconstruction.

7 Conclusions

This paper simply surveyed the state-of-the-art virtualization technology in CC-RANs, mainly concentrating on C-RAN, H-CRAN and F-RAN. The background of network virtualization, virtualization architecture in different CC-RANs, virtualization key enabling technologies, and open issues were presented. Since wireless network virtualization has attracted much attention, the benefits of virtualization, the diverse radio access demands and cost-efficient requirements render the introduction of virtualization technology into CC-RANs as an overwhelming trend. Meanwhile, the key enabling technologies for CC-RAN virtualization are summarized as virtual resource allocation, RAN slicing, mobility management, and social-awareness for satisfying the requirements of isolation, customization and high radio resource utilization. However, given the relative infancy of the CC-RAN virtualization, there are quite a number of outstanding problems that needs further investigation from the perspective of promising key technologies and advanced solutions. At last, considering the deployment of virtualized CC-RANs, we mainly discussed the virtualization levels, signaling design, performance analysis, and network security as the challenges and open issues. Besides, the advanced isolation technologies among multiple virtual CC-RANs networks, the network intelligence of CC-RAN virtualization, as well as the practical deployment schemes are also urgent to be discussed and further investigated.

References

[1] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Network*, vol. 29, no. 2, pp. 6–14, Mar. 2015. doi: 10.1109/MNET.2015.7064897.

[2] M. Peng and W. Wang, "Technologies and standards for TD-SCDMA evolutions to IMT-Advanced," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 50–58, Dec. 2009. doi: 10.1109/MCOM.2009.5350368.

[3] B. Cao, F. He, Y. Li, C. Wang, and W. Liang, "Software defined virtual wireless network: Framework and challenges," *IEEE Network*, vol. 29, no.4, pp.6–12, Jul.-Aug. 2015. doi: 10.1109/MNET.2015.7166185.

[4] H. Zhang, S. Vreic, G. Senarath, et al., "5G wireless network: MyNET and SONAC," *IEEE Network*, vol. 29, no. 4, pp. 14–23, Jul.-Aug. 2015. doi:10.1109/MNET.2015.7166186.

[5] S. Dixit, C. Politis, and A. T. Papathanassiou, "Network virtualization in the wireless world," *IEEE Vehicular Technology Magazine*, vol.10, no. 3, pp. 27–29, Sept. 2015. doi:10.1109/MVT.2015.2446452.

[6] K. Zhu and E. Hossain, "Virtualization of 5G cellular networks as a hierarchical combinatorial auction," *IEEE Transaction on Mobile Computing*, vol. 15, no. 10, pp. 2640–2654, Oct. 2016. doi:10.1109/TMC.2015.2506578.

[7] M. M. Rahman, C. Despins, and S. Affes, "Design optimization of wireless access virtualization based on cost and QoS trade-off utility maximization," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6146–6162, Sept. 2016. doi: 10.1109/TWC.2016.2580505.

[8] R. Riggio, A. Bradai, D. Harutyunyan, T. Rasheed, and T. Ahmed, "Scheduling wireless virtual networks functions," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 240–252, Jun. 2016. doi: 10.1109/TNSM.2016.2549563.

[9] Y. Li and M. Chen, "Software-defined network function virtualization: a survey," *IEEE Access*, vol. 3, pp. 2542–2553, Dec. 2015. doi: 10.1109/ACCESS.2015.2499271.

[10] C. Liang and F. R. Yu, "Wireless virtualization for next generation mobile cellular networks," *IEEE Wireless Communications*, vol. 22, no. 1, pp. 61–69, Feb. 2015. doi: 10.1109/MWC.2015.7054720.

[11] H. Huang, P. Li, S. Guo, and W. Zhuang "Software-defined wireless mesh networks: architecture and traffic orchestration," *IEEE network*, vol. 29, no. 4, pp. 24–30, Jul./Aug. 2015. doi: 10.1109/MNET.2015.7166187.

[12] T. Wood, K. K. Ramakrishnan, J. Hwang, G. Liu, and W. Zhang, "Toward a software-based network integrating software defined networking and network function virtualization," *IEEE Network*, vol. 29, no. 3, pp. 36–41, May-Jun. 2015. doi: 10.1109/MNET.2015.7113223.

[13] J. G. Herrera and J. F. Botero, "Resource allocation in NFV: a comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, Sept. 2016. doi: 10.1109/TNSM.2016.2598420.

[14] Q. Zhou, C. Wang, S. Mclaughlin, and X. Zhou, "Network virtualization and resource description in software-defined wireless networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 110–117, Nov. 2015. doi: 10.1109/MCOM.2015.7321979.

[15] Z. Feng, C. Qiu, Z. Feng, et al., "An effective approach to 5G: wireless network virtualization," *IEEE Communications Magazine*, vol.53, no.12, pp. 53–59, Dec. 2015. doi: 10.1109/MCOM.2015.7355585.

[16] C. J. Bernardos, A. D. L. Olive, P. Serrano, et al., "An architecture for software defined wireless networking," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 52–61, Jun. 2014. doi: 10.1109/MWC.2014.6845049.

[17] H. Li, M. Dong, and K. Ota, "Radio access network virtualization for the social internet of things," *IEEE Could Computing*, vol. 2, no. 6, pp. 42–50, Nov.-Dec. 2015. doi: 10.1109/MCC.2015.114.

[18] F. Granelli, A. A. Gebremariam, M. Usman, et al., "Software defined and virtualized wireless access in future wireless networks: scenarios and standards," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 26–34, Jun. 2015. doi: 10.1109/MCOM.2015.7120042.

[19] X. Costa-Prez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network for future mobile carrier networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 27–35, Jul. 2013. doi: 10.1109/MCOM.2013.6553675.

[20] R. Kokku, R. Mahindra, H. Zhang, and S. Ranfarajan, "NVS: a substrate for virtualizing wireless resources in cellular networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1333–1346, Oct. 2012. doi: 10.1109/TNET.2011.2179063.

[21] G. Tseliou, F. Adelantado, and C. Verikoukis, "Scalable RAN virtualization in multitenant LTE-A heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 6651–6664, Aug. 2016. doi:10.1109/TVT.2015.2475641.

[22] C. Liang and F. R. Yu, "Wireless network virtualization: a survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol.17, no. 1, pp. 358–380, first quarter 2015. doi: 10.1109/COMST.2014.2352118.

[23] F. Callegati, W. Cerroni, C. Contoli, et al., "SDN for dynamic NFV deployment," *IEEE Communications Magazine*, vol. 54, no. 10, pp. 89–95, Oct. 2016. doi: 10.1109/MCOM.2016.7588275.

[24] VMWARE. (2014, Jul.). Virtualization essentials [Online]. Available: https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/ebook/gated-vmw-ebook-virtualizationessentials.pdf

[25] H. Hawilo, A. shami, M. Mirahmadi, and R. Asal, "NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC)," *IEEE Network.*, vol. 28, no. 6, pp. 18–26, Nov.-Dec. 2014. doi: 10.1109/MNET.2014.6963800.

[26] ETSI. (2012, Oct.). Network function virtualization: an introduction, benefits, enablers, challenges, & call for action, *SDN and OpenFlow World Congress* [Online]: Available: https://portal.etsi.org/nfv/nfv-whitepaper. pdf

[27] A. Wang, M. Iyer, R. Dutta, G. N. Rouskas, and I. Baldine, "Network virtualization: technologies, perspectives, and frontiers," *Journal of Lightwave Technology*, vol. 31, no. 4, pp. 523–537, Feb. 2013. doi: 10.1109/JLT.2012.2213796.

[28] M. A. Marotta, N. Kaminski, I. Gomez-Miguelez, et al., "Resource sharing in heterogeneous cloud radio access networks," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 74–82, Jun. 2015. doi: 10.1109/MWC.2015.7143329.

[29] C. Liang, F. R. Yu, and X. Zhang, "Information-centric network function virtualization over 5G mobile wireless networks," *IEEE Network.*, vol. 29, no. 3, pp. 68–74, May-Jun. 2015. doi: 10.1109/MNET.2015.7113228.

[30] M. Yang, Y. Li, D. Jin, et al., "Software-defined and virtualized future and wireless networks: a survey," *Mobile Networking and Application*, vol. 20, no. 1, pp. 4–18, Feb. 2015. doi:10.1007/s11036-014-0533-8.

[31] P. Rost, I. Berberana, A. Maeder, et al., "Benefits and challenges of virtualization in 5G radio access networks," *IEEE Communications Magazine*, vol.53, no. 12, pp. 75–82, Dec. 2015. doi: 10.1109/MCOM.2015.7355588.

[32] N. Marshall and J. Hoffman. (2015, Jul.). NFV: radio virtualization in the RAN [Online]. Available: https://www.abiresearch.com

[33] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, "Software-defined networking in cellular radio access networks: potential and challenges," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 150–156, Jan. 2015. doi: 10.1109/MCOM.2015.7010528.

[34] EU FP7. (2017, Mar.). FLAVIA project [Online]. Available: http://www.ict-flavia.eu

[35] Small Cell Forum. (2015, Jun.). Virtualization for small cells: overview, Rep. version:106.07.01 [Online]. Available: http://scf.io/en/documents/106-Virtualization0-for-small-cells-Overview.php

[36] 5GPPP 5GEx project, "5GEx Multi-domain service creation-from 90 days to 90

minutes," A 5GEx White Paper V1, Mar. 2016.
[37] 5GPPP. (2017, Mar.). SEAME project [Online]. Available: https://5g-ppp.eu/sesame
[38] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146- 153, Jul. 2016. doi: 10.1109/MCOM.2016.7509393.
[39] D. Sabella, P. Rost, Y. Sheng et al., "RAN as a service: challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network," in *Proc. Future Network & Mobile Summit*, Lisboa, Portugal, 2013, pp. 1-8.
[40] M. Patwary, S. K. Sharma, S. Chatzinotas, et al., "Universal intelligent small cell (UnISCell) for next generation cellular networks," Digital Communications and Networks, vol. 2, no. 4, Nov. 2016, pp. 167- 174, doi: 10.1016/j.dcan.2016.09.003.
[41] R. P. Esteves, L. Z. Granville, and R. Boutaba, "On the management of virtual networks," *IEEE Communications Magazine.*, vol. 51, no. 7, pp. 80-88, Jul. 2014. doi: 10.1109/MCOM.2013.6553682.
[42] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: system architecture, key techniques, and open issues," *IEEE Communications Surveys &Tutorials*, vol. 18, no. 3, pp. 2282-2308, third quarter 2016. doi: 10.1109/COMST.2016.2548658.
[43] A. Checko, H. L. Christiansen, Y. Yan, et al., "Cloud RAN for mobile networks-a technology overview," *IEEE Communications Surveys &Tutorials*, vol. 17, no. 1, pp. 405-426, first quarter 2015. doi: 10.1109/COMST.2014.2355255.
[44] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 84-91, Apr. 2016. doi: 10.1109/MCOM.2016.7452271.
[45] M. I. M. Alfaqawi, J. Chebil, M. H. Habaebi, and D. Datla, "Wireless distributed computing for cyclostationary feature detection," *Digital Communications and Networks*, vol. 2, no. 1, pp. 47- 56, Feb. 2016. doi: 10.1016/j.dcan.2015.09.003.
[46] X. Li, X. Wang, C. Zhu, W. Cai, and V. C. M. Leung, "Caching-as-a-service: Virtual caching framework in the cloud-based mobile networks," in *IEEE INFOCOM 2015 Workshops on Computer Communications (INFOCOM WKSHPS)*, Hong Kong, China, pp. 372- 377, Aug. 2015. doi: 10.1109/INFCOMW.2015.7179413.
[47] S. Buzzi, C.-L. I, T. E. Klein, et al., "A survey of energy-efficient techniques for 5G networks and challenges Ahead," *IEEE Journal on Selected Areas in Communications.*, vol. 34, no. 4, pp. 697- 709, Apr. 2016. doi: 10.1109/JSAC.2016.2550338.
[48] ITU, "IMT vision—framework and overall objectives of the future development of IMT for 2020 and beyond," ITU-R M.2083-0, Sept. 2015.
[49] Z. Chang, Z. Zhou, S. Zhou, T. Ristaniemi, and Z. Niu. (2016, Apr.). Towards service-oriented 5G: virtualizing the networks for everything-as-a-service [Online]. Available: https://arxiv.org/pdf/1604.01739
[50] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, "Computing while computing: distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 45-55, Nov. 2014. doi: 10.1109/MSP.2014.2334709.
[51] M. Peng, Y. Yu, H. Xiang, and H. V. Poor, "Energy-efficient resource allocation optimization for multimedia heterogeneous cloud radio access networks", *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 879-892, May 2016.
[52] S. Zhou, J. Gong, Z. Zhou, W. Chen, and Z. Niu, "GreenDelivery: proactive content caching and push with energy-harvesting based small cells" *IEEE Communications Magazine*, vol. 53, no. 4, pp. 142- 149, Apr. 2015. doi: 10.1109/MCOM.2015.7081087.
[53] R. Khanna, H. Liu, and T. Rangarajan, "Wireless data center management: sensor network applications and challenges," *IEEE Microwave Magazine*, vol. 15, no. 7, pp. S45-S60, Nov./Dec. 2014. doi: 10.1109/MMM.2014.2356151.
[54] E. S. Correa, L. A. Fletscher, and J. F. Botero, "Virtual data center embedding: a survey," *IEEE Latin America Transactions*, vol. 13, no. 5, pp. 1661-1670, May 2015. doi: 10.1109/TLA.2015.7112029.
[55] Techopedia. (2017, Mar.). Data center virtualization [Online]. Available: https://www.techopedia.com/definition/29883/data-center-virtualization
[56] H. Shirayanagi, H. Yamada, and K. Kono, "Honeyguide: a VM migration-aware network topology for saving energy consumption in data center networks," in *IEEE Symposium Computers and Communications (ISCC)*, Cappadocia, Turkey, 2012. pp. 000460-000467. doi: 10.1109/ISCC.2012.6249339.
[57] S. Zhou, T. Zhao, Z. Niu, and S. Zhou, "Software-defined hyper-cellular architecture for green and elastic wireless access," *IEEE Communications Magazine*, vol. 54, no. 1, Jan. 2016. doi: 10.1109/MCOM.2016.7378420.
[58] Z. Zaidi, V. Friderikos, and M. A. Imran, "Future RAN architecture: SD-RAN through a general-purpose processing platform," *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 52- 60, Mar. 2015. doi: 10.1109/MVT.2014.2380632.
[59] C.-L. I, Y. Yuan, J. Huang, et al., "Rethink fronthaul for soft RAN," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 82-88, Sept. 2015. doi: 10.1109/MCOM.2015.7263350.
[60] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, "Software defined mobile networks: concept, survey, and research directions," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 126-133, Nov. 2015. doi: 10.1109/MCOM.2015.7321981.
[61] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, "Software-defined networking in cellular radio access networks: potential and challenges," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 150- 156, Jan. 2015. doi: 10.1109/MCOM.2015.7010528.
[62] A. M. Akhtar, X. wang, and L. Hanzo, "Synergistic spectrum sharing in 5G HetNets: a harmonized SDN-enabled approach," *IEEE Communications Magazine*, vol. 54, no. 1, pp. 40-47, Jan. 2016. doi: 10.1109/MCOM.2016.7378424.
[63] A. Gupta and R. K. Jha, "A survey of 5G network: architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206-1232, Jul. 2015. doi: 10.1109/ACCESS.2015.2461602.
[64] R. Riggio, M. K. Marina, J. Schulz-Zander, S. Kuklinski, and T. Rashed, "Programming abstractions for software-defined wireless networks," *IEEE Transactions on Network and Service Management*, vol. 12, no. 2, pp. 146-162, Jun. 2015. doi: 10.1109/TNSM.2015.2417772.
[65] M. Peng, C. Wang, J. Li, H. Xiang, and V. Lau, "Recent advances in underlay heterogeneous networks: interference control, resource allocation, and self-organization," *IEEE Communications on Surveys & Tutorials.*, vol. 17, no. 2, pp. 700-729, Second quarter 2015. doi: 10.1109/COMST.2015.2416772.
[66] H. Ibrahim, H. ElSawy, U. T. Nguyen, and M.-S. Alouini, "Mobility-aware modeling and analysis of dense cellular networks with C-Plane/U-Plane split architecture," *IEEE Transactions on Communications.*, vol. 64, no. 11, pp. 4879-4894, Nov. 2016. doi: 10.1109/TCOMM.2016.2609905.
[67] M. Peng, Y. Li, Z. Zhao, J. Jiang, J. Li, and C. Wang, "Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 126-135, Dec. 2014. doi: 10.1109/MWC.2014.7000980/ISSN: 1536-1284.
[68] M. Armbrust, A. Fox, R. Griffith, et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50-58, Apr. 2010. doi: 10.1145/1721654.1721672.
[69] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Communications.*, vol. 22, no. 2, pp. 152-160, Apr. 2015. doi: 10.1109/MWC.2015.7096298.
[70] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: virtualizing wireless access for dense heterogeneous systems," *Journal of Communications and Networks*, vol. 18, no. 2, pp. 135-149, Apr. 2016. doi: 10.1109/JCN.2016.000023.
[71] K. A. Meerja, A. Shami, and S. Refaey, "Hailing cloud empowered radio access networks," *IEEE Wireless Communications*, vol. 22, no. 1, pp. 122-129, Feb. 2015. doi: 10.1109/MWC.2015.7054727.
[72] A. Blenk, A. Basta, M. Reisslein, and W.Kellerer, "Survey on network virtualization hypervisors for software defined networking," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 655- 685, First quarter 2016. doi: 10.1109/COMST.2015.2489183.
[73] C.-L. I, J. Huang, R. Duan, et al., "Recent progress on C-RAN centralization and cloudification," *IEEE Access*, vol. 2, pp. 1030- 1039, Aug. 2014. doi: 10.1109/ACCESS.2014.2351411.
[74] C.-L. I, C. Rowell, S. Han, et al., "Toward green and soft: a 5G perspective," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 66-73, Feb. 2014. doi: 10.1109/MCOM.2014.6736745.
[75] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and it's role in the internet of things," in *Proc. First Edition of MCC Workshop on the Mobile Cloud Computing*, Helsinki, Finland, 2012, pp. 13-16.
[76] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: a platform for internet of things and analytics," in *Big Data and Internet of Things: A Roadmap for Smart Environments, N. Bessis and C. Dobre ed*. New York, USA: Springer, 2014, vol. 546, pp. 169-186. doi: 10.1007/978-3-319-05029-4_7.
[77] Y. Ai, M. Peng, and K. Zhang. (2017). Edge cloud computing technologies for internet of things: a primer, *Digital Communications and Networks* [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352864817301335
[78] K. Liang, L. Zhao, X. Chu, and H.-H. Chen, "An integrated architecture for software defined and virtualized radio access networks with fog computing,"

IEEE Network, vol. 31, no. 1, pp. 80- 87, Jan./Feb. 2017. doi: 10.1109/MNET.2017.1600027NM.
[79] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46–53, Jul.-Aug. 2016. doi: 10.1109/MNET.2016.7513863.
[80] M. Peng, and K. Zhang, "Recent advances in fog radio access networks: performance analysis and radio resource allocation," *IEEE Access*, vol. 4, pp. 5003–5009, Aug. 2016. doi: 10.1109/ACCESS.2016.2603996.
[81] S.-C. Hung, H. Hsu, S. -Y. Lien, and K. -C. Chen, "Architecture harmonization between cloud radio access networks and fog network," *IEEE Access*, vol. 3, pp. 3019–3034, Dec. 2015. doi: 10.1109/ACCESS.2015.2509638.
[82] R. Tandon and O. Simeone "Harnessing cloud and edge synergies: toward an information theory of fog radio access networks," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 44–50, Aug. 2016. doi: 10.1109/MCOM.2016.7537176.
[83] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the suitability of fog computing in the context of internet of things," *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1–14, Oct. 2015. doi:10.1109/TCC.2015.2485206.
[84] S. H. Yeganeh and Y. Ganjali, "Kandoo: a framework for efficient and scalable offloading of control applications," in *Proc. 1st Workshop on Hot Topics in Software Defined Networks*, New York, USA, 2012, pp. 19–24.
[85] Nokia, "Dynamic end-to-end network slicing for 5G," Whitepaper, Jun. 2016.
[86] 5GPPP, "Deliverable D2.2: draft overall 5G RAN design," Whitepaper, Jun. 2016.
[87] ZTE Corporation, "Some Clarification for NW Slicing," (Type: discussion), Reno, US, R3-162723, Nov. 2016.
[88] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, "Virtual network embedding: a survey," IEEE *Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1888- 1906, FourthQuarter 2013. doi: 10.1109/SURV.2013.013013.00155.
[89] T. LeAnh, N. H. Trane, D. T. Ngo, and C. S. Hong, "Resource allocation for virtualized wireless networks with backhaul constraints," *IEEE Communications Letters*, Oct. 2016, doi: 10.1109/LCOMM.2016.2617307.
[90] V. Jumba, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, "Resource provisioning in wireless virtualized networks via massive-MIMO," *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp.237–240, Jun. 2015. doi: 10.1109/LWC.2015.2402126.
[91] S. Parsaeefard, R. Dawadi, M. Derakhshani, and T. Le-Ngoc, "Joint user-association and resource-allocation in virtualized wireless networks," *IEEE Access*, Apr. 2016, doi: 10.1109/ACCESS.2016.2560218.
[92] F. Fu and U. C. Kozat, "Stochastic game for wireless network virtualization," *IEEE/ACM Transactions on Networking*, vol. 21, no. 1, pp. 84–97, Feb. 2013. doi: 10.1109/TNET.2012.2190419.
[93] E. Amaldi, S. Coniglio, A. M. C. A. Koster, and M. Tieves, "On the computational complexity of the virtual network embedding problem," *Electronic Notes in Discrete Mathematics*, vol. 52, pp. 213- 220, Jun. 2016. doi: 10.1016/j.endm.2016.03.028.
[94] F. Z. Yousaf and T. Taleb, "Fine-grained resource-aware virtual network function management for 5G carrier cloud," *IEEE Network*, vol. 30, no. 2, pp. 110–115, Mar.-Apr. 2016. doi: 10.1109/MNET.2016.7437032.
[95] S.-H. Park, O. Simeone, and S. Shamai(Shitz), "Joint optimization of cloud and edge processing for fog radio access networks" *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7621- 7632, Nov. 2016. doi: 10.1109/TWC.2016.2605104.
[96] NGMN Alliance, "Description of network slicing concept," Final Deliverable (approved), January 2016.
[97] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "On radio access network slicing from a radio resource management perspective," *IEEE Wireless Communications*, vol. PP, no. 99, pp. 2- 10, Apr. 2017. doi: 10.1109/MWC.2017.1600220WC.
[98] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on RAN: flexibility and resource abstraction," *IEEE Communications Magazine.*, vol. 55, no. 6, pp. 102–108, June 2017. doi: 10.1109/MCOM.2017.1601119.
[99] X. Foukas, N. Nikaein, and M. M. Kassem, "FlexRAN: a flexible and programmable platform for software-defined radio access networks," in *Proc. 12th ACM International Conference on Emerging Networking Experiments and Technologies*, Irvine, USA, Dec. 2016, pp. 427–441.
[100] China Mobile Communications Corporation, Huawei Technologies Co. Ltd., Deutsche Telekom AG, and Volkswagen, "5G service-guaranteed networking slicing white paper," White Paper V1.0, Feb. 2017.
[101] M. Richart, J. Baliosian, J. Serrat, and J. Gorricho, "Resource slicing in virtual wireless network: a survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 462- 476, Sept. 2016. doi: 10.1109/TNSM.2016.2597295.
[102] F. R. Yu, V. W. S. Wong, J.-H. Song, V. C. M. Leung, and H. C. B. Chan. (2011, Apr.). Next generation mobility management: an introduction, *Wireless Communications on Mobile Computing* [Online]. *11(4)*, pp. 446–458. Available: http://onlinelibrary.wiley.com/doi/10.1002/wcm.904/pdf
[103] L. Zhu, F. R. Yu, T. Tang, and B. Ning, "Handoff performance improvement in an integrated train-ground communication system based on wireless network virtualization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1165–1178, May 2017. doi: 10.1109/TITS.2016.2620171.
[104] X. Duan and X. Wang, "Authentication handover and privacy protection in 5G HetNets using software-defined networking," *IEEE Communications Magazine.*, vol. 53, no. 4, pp. 28- 35, Apr. 2015. doi: 10.1109/MCOM.2015.7081072.
[105] H. Zhang, N. Liu, X. Chu, et al. (2017, Apr.). Network slicing based 5G and future mobile networks mobility, resource management, and challenges [Online]. Available: https:// arxiv.org/abs/1704.07038v1
[106] F. Xia, L. Liu, J. Ma, and A. V. Vasilakos, "Socially aware networking: a survey," *IEEE System Journal*, vol. 9, no. 3, pp. 904- 921, Sept. 2015. doi: 10.1109/JSYST.2013.2281262.
[107] Y. Cao, C. Long, T. Jiang, and S. Mao, "Share communication and computation resources on mobile devices: a social awareness perspective," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 52- 59, Aug. 2016. doi: 10.1109/MWC.2016.7553026.
[108] L. Jiang, H. Tian, Z. Xing, et al., "Social-aware energy harvesting device-to-device communications in 5G networks," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 20–27, Aug. 2016. doi: 10.1109/MWC.2016.7553022.
[109] B. Bai, L. Wang, Z. Han, W. Chen, and T. Svensson, "Caching based socially-D2D communications in wireless content delivery networks: a hypergraph framework," *IEEE Wireless Communications.*, vol. 23, no.4, pp. 74–81, Aug. 2016. doi: 10.1109/MWC.2016.7553029.
[110] A. Leivadeas, C. Papagianni, and S. Papavassiliou, "Socio-aware virtual network embedding," *IEEE Network*, vol. 26, no. 5, pp. 35–43, Sept.-Oct. 2012. doi: 10.1109/MNET.2012.6308073.
[111] F. Z. Yousaf, P. Loureiro, F. Zdarsky, T. Taleb, and M. Liebsch, "Cost analysis of initial deployment strategies for virtualized mobile core network functions," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 60- 66, Dec. 2015. doi: 10.1109/MCOM.2015.7355586.
[112] Q. Duan, "Cloud service performance evaluation: status, challenges, and opportunities-a survey from the system modeling perspective," *Digital Communications and Networks*, vol. 3, no. 2, pp. 101- 111, May 2017. doi.org/10.1016/j.dcan.2016.12.002.

Manuscript received: 2017-07-14

Biographies

ZHANG Xian (zhangxian_email@163.com) received the B.E. degree in telecommunications engineering from Chongqing University of Posts and Telecommunications, China in 2016. He is currently pursuing Ph.D. with Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT), China. His research interests include resources allocation and optimization for fog computing based radio access networks (F-RANs) and software-defined radio access networks.

PENG Mugen (pmg@bupt.edu.cn) received the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), China, in 2005. Afterward, he joined BUPT, where he has been a full professor with the School of Information and Communication Engineering since 2012. His main research interests focus on cooperative communications, self-organization networking, heterogeneous networking, cloud communication, and the Internet of Things. He has authored/coauthored over 70 refereed IEEE journal papers and over 200 conference proceeding papers. He was a recipient of the 2014 IEEE ComSoc AP Outstanding Young Researcher Award, and the best paper award in JCN, IEEE WCNC 2015, WASA 2015, GameNets 2014, IEEE CIT 2014, ICCTA 2011, IC-BNMT 2010, and IET CCWMC 2009. He is on the Editorial/Associate Editorial Board of *IEEE Communications Magazine*, *IEEE Access*, *IEEE Internet of Things Journal*, *IET Communications*, and *China Communications*. He is the Fellow of IET.