ABSTRACT: Heterogeneous architecture is an underlining feature of 5G, however deployment and management of HetNets in 5G scenarios is yet to be explored. Given the need to satisfy overwhelming capacity demands in 5G, mm-wave spectrum (3-300 GHz) is expected to offer a very compelling long term solution by providing additional spectrum to 5G networks. Hence, the challenge is the integration of mm-wave in heterogeneous and dense networks as well as the backward compatibility and integration with legacy 4G/3G networks. Furthermore, Cloud radio access networks (C-RAN) contribution to 5G is considered as a cost effective and energy efficient solution for dense 5G deployment. From an energy point of view, cost and energy consumption are major considerations for 5G. C-RAN and energy efficiency techniques could help in performance improvements.

Although HetNets were introduced in 4G networks, their complexity has increased in 5G networks. In this paper, we will try to build a clear image of HetNets in 5G cellular networks. We consider different technologies with a special focus on mm-wave networks given its important role in 5G networks. We then address the available standards in HetNets that allow interworking and multihoming between different radio access technologies. Afterwards, we consider the virtualization of 5G HetNets and its benefits. Different resource allocation strategies in the literature are also presented for single-resource as well as for multi-resources. Finally, we give an overview of existing works addressing energy efficiency strategies in 5G networks.


1. INTRODUCTION

Fifth generation (5G) is not as previous generations, an evolution of the existing, but it is rather considered as a cellular network revolution that builds on the evolution of existing technologies. These technologies are complemented by new radio concepts that are designed to meet the new and challenging requirements of some use cases today’s radio access networks cannot support [2] [3].

This revolution is necessary to offer new services to 5G users with good quality of service (QoS). These services include:

- Good service even in very crowded places.
- Similar user experience for end-users on the move as for static users.
- The Internet of Things (IoT). Basically, anything that profits from being connected will be connected.
2. Heterogeneous networks/Multi-RAT

Today’s 3G and 4G networks are designed primarily with a focus on peak rate and spectral efficiency improvements. In the 5G era, we will see a shift towards network efficiency with 5G systems based on dense heterogeneous networks architectures. HetNets are among the most promising low-cost approaches to meet the industry’s capacity growth needs and deliver a uniform connectivity experience. A HetNet comprises a group of small cells that support aggressive spectrum spatial reuse coexisting within macro cells as shown in Fig. 1.2. However, HetNets will be architected to incorporate an increasingly diverse set of frequency bands within a range of network topologies, including macro cells in licensed bands (e.g., long term evolution network or LTE) and small cells in licensed or unlicensed bands (e.g., WiFi). New higher frequency spectrum (e.g., millimeter-wave or mm-wave) may also be deployed in small cells to enable ultra-high-data-rate services.
Architecture

HetNets are formed of macro cells and small cells. A macro cell is generally divided into several sectors in order to increase the spatial frequency reuse which increases the network capacity. Typically, a macro cell is implemented as a tri-sectorial base station (BS) with each sector of 120°. However, different definitions are considered for choosing the cell type, it can consider the radius of the cell, the number of connected users and the deployment options.

As their name indicates, small cells provide a smaller coverage area than a macro cell. As shown in Fig. 1.2, a macro cell overlaps several small cells. There are several types of small cells such as micro, pico, femto and relay cells, ordered in decreasing order of coverage and transmission power. These small cells can be managed by the same operator as a macro cell or by a different operator and require a lower installation cost. In addition, it is worth to note that small cells are mainly deployed in order to support the increasing rates of data services but can also support voice services.

WiFi small cells

Widely deployed WiFi systems are playing an increasingly important role in offloading data traffic from the heavily loaded cellular network, especially in indoor traffic hotspots and in poor cellular coverage areas. Very recently, the Federal Communications Commission (FCC) voted to make 100 MHz of spectrum in the 5 GHz band available for unlicensed WiFi use based on the IEEE 802.11ac standard [9], giving carriers and operators more opportunities to push data traffic to WiFi. WiFi access points have even been regarded as a distinct tier of small cells in heterogeneous cellular networks. Wireless local access networks (WLAN) technology evolution is mainly carried out within the WLAN IEEE 802.11 working group which released multiple set of standards for various operating frequencies and ranges specification.

LTE Small Cells

LTE small cell networks are highly dense networks constituting of home eNodeBs, indoor enterprise eNodeBs as well as outdoor deployed eNodeBs. Some of the major challenges of the LTE small cell networks are:

1) Maintaining the desired QoS with respect to downlink and uplink packet data transmission.
2) Efficient handover.

3) Interference co-ordination with neighbors. Especially in the uplink direction, i.e., from UE to small cell eNodeB, the task of delivering a wide variety of application layer packets is complicated due to limited transmission power of the UE, limited battery resources at UE and time-varying nature of wireless channels.

From an architectural point of view, two deployment scenarios were identified in [21], namely small cells co-existing with macro cells, known as Hot Spot, and small cells without macro cells known as Not spot. In such areas, only basic network coverage is needed, which can be adequately supported by lower cost small cells rather than more expensive resource from the macro site. Not-spot small cells are perfect for network coverage extension to reach the rural areas, both indoors and outdoors. The Not-spot scenario may potentially suffer however from high volume of handover signaling load, which may impact the users Quality-of-Experience (QoE).

**Figure 1.3: HetNet architecture with loose coupling (Source: [19]).**

**Figure 1.4: HetNet architecture with tight coupling (Source: [19]).**

**Mm-Wave Small Cells**
Capacity for wireless communication depends on spectral efficiency and bandwidth. It is also related to cell size. Cell sizes are becoming small and physical layer technology is already at the boundary of Shannon capacity [24]. It is the system bandwidth that remains unexplored. Presently, almost all wireless communications use spectrum in 300 MHz to 3 GHz band, often termed as “sweet spot” or “beachfront spectrum” [25]. In order to increase capacity, wireless communications cannot help facing the new challenges of high frequency bandwidth. The key essence of next generation 5G wire-less networks lies in exploring this unused, high frequency mm-wave band, ranging from 3 ∼ 300 GHz. Even a small fraction of available mm-wave spectrum can support hundreds of times of more data rate and capacity over the current cellular spectrum [26]. Thus, the availability of a big chunk of mm-wave spectrum is opening up a new horizon for spectrum constrained future wireless communications [26].

**Beamforming in mm-wave**

The main objective of adaptive beamforming is to shape the beam patterns (e.g., by beamsteering) so that the received signal-to-noise ratio (SNR) is maximized. Full control of beam pattern shaping requires changing both the amplitude and phase of transmitted signals. The need for low-cost and lowpower hardware, however, has pushed mm-wave towards a simpler analog architecture that contains only digitally controlled constant modulus phase shifters. Hybrid precoding proposed in [27] divides the required precoding processing between the analog and digital domains, and hence allows better control of the beam shape.

**Mm-wave Mobile Boradband Frame Structure**

As in 4G systems, mm-wave uses also OFDM and single-carrier waveform as multiplexing schemes. We show in Fig. 1.5 a mm-wave frame structure as described in [28]. The basic transmission time interval (TTI) is a solt of 62.5µs duration. Subframe, frame and superframe’s duration are chosen equal to those in LTE systems (1 ms, 10 ms and 40 ms, respectively) in order to facilitate the interworking between both technologies. The cyclic prefix (CP) is chosen to be 520 ns, which gives sufficient margin to accommodate the longest path, different deployment scenarios, and the potential increase of delay spread in the case of small antenna arrays or wider beams. The subcarrier spacing is chosen to be 480 kHz, small enough to stay within the coherent bandwidth of most multipath channels expected in mm-wave.

**Interworking between mm-wave and LTE**

A hybrid LTE/mm-wave system can improve coverage and ensure seamless user experience in mobile applications. In a hybrid LTE/mm-wave system, system information, control channel, and feedback are transmitted in the LTE system, making the entire millimeter-wave spectrum available for data communication. Compared with millimeter waves, the radio waves at < 3 GHz frequencies can better penetrate obstacles and are less sensitive to non-line-of-sight (NLOS) communication link or other impairments such as absorption by foliage, rain, and other particles in the air. Therefore, it is advantageous to transmit important control channels and signals via cellular radio frequencies, while utilizing the millimeter waves for high data rate communication.
Multihoming

HetNets were designed such that traffic can be offloaded between available access networks. However, concurrent multiple access to more than one network in wireless networks has recently been standardized in Release 12 under the name of “Dual Connectivity” [15]. In this section, we introduce the aspects and standards enabling multihoming’s concept implementation with focus on the dual connectivity standard. We also present a literature overview on interworking and network selection strategies in this context.

Multihoming aspects

Multihoming was first proposed as a redundancy solution for wired networks. Recently, the coexistence of different wireless access network technologies has renewed this concept and became an attractive topic for study during the past years. Wireless networks multihoming concept started with offloading [31, 32, 33], passing by load balancing [34], optimal distribution [35] [36] [37], as well as concurrent multiple access [38, 39, 40].

Load balancing concept was introduced in wired networks [31]. Such load balancing system must determine the available bandwidth through an access link, assign incoming and outgoing traffic, and detect access links failure. For this aim, a reliable routing protocol must be considered [32]. Similarly, load balancing management could be obtained in heterogeneous wireless networks by dynamically optimizing the packets’ split ratio between multiple access networks as shown in [34]. Such strategy might be based on the load information and channel quality information at each access network.

Multihoming Technology Enablers

Throughout the past years, 3GPP and IETF worked hard in order to standardize different HetNets interworking schemes. Their main interest was to standardize the users mobility between accesses, the transport layer support of multihoming, and frequency resource scheduling known as “Dual Connectivity” (DC).
Mobility in Heterogeneous Networks

Non-seamless offloading between LTE and WiFi is disturbing, especially for real-time applications that require the continuity of service (e.g., VoIP, Video Conference, HTTP page). It is highly desirable that mobile operators provide seamless service continuity between cellular and WiFi accesses with involving both user plane routing and control plane functions. This seamless continuity can be first supported by ensuring a service layer continuity even when the IP address has changed which is not supported in TCP/IP. In this section, we present some of the seamless continuity standardized technologies. Several mechanisms are proposed by 3GPP describing the offload management in 3GPP networks. I-WLAN is the first approach allowing local area network access to the 3GPP core.

Multihoming at Transport Layer

In addition to the mobility described above and maintaining the IP connection of a user when offloading, static multihoming of a user connected simultaneously to multiple access networks has multiple IP addresses. However, regular TCP can support only one flow which means only one IP address. For this reason, several transport protocols were proposed, we will present here an overview of multihoming-capable protocols.

Transport layer multihoming started with node multihoming which is an old concept defined as a device having more than one wired access interface. Two main standards were proposed: Stream Control Transport Protocol (SCTP) in 2000 [39, 40] and Multi-Path Transport Control Protocol (MPTCP) in 2010 [51]. SCTP uses only one path for transfer and keeps the other available paths for packet retransmission or for backup in case of handover or link failure. SCTP suffered however from the middleboxes blocking problem for SCTP packets.

Frequency Resources Aggregation

Since the operator’s first choice is to add more capacity on licensed spectrum, carrier aggregation (CA) technology [32] has been standardized in Long Term Evolution (LTE) Releases 10–12. CA was first proposed to aggregate multiple small band segments into maximum 100 MHz virtual bandwidth to achieve higher data rate in LTE small cells.

Frequency multi-connection is also being standardized by 3GPP. LTE dual connectivity is introduced in Release 12 [15] as a realization of different spectrum allocation between a macro cell and a small cell. Several work items in Release 13 differentiated between dual connectivity in LTE/LTE-A HetNets, the License Assisted Access (LAA), and in LTE/WLAN HetNets, the LTE/WLAN Aggregation (LWA).

Interworking Types

Several heterogeneous network types were considered in the literature. Heterogeneity in wired networks mainly consisted in accessing a server using more than one ISP, which means different routes. Generally, wired networks mutlihoming is considered as redundancy in case of failure. Few works tackled multihoming in such networks, we note [34] in which the authors conducted a study on multihoming
streaming in a residential context using a DSL and a cable connection. This study showed significant QoS improvement for connection splitting and migration in case of congestion.

Conversely, wireless networks interworking gained a huge reputation. Several HetNet models were proposed along with performance evaluation and interworking technologies standardization. Next, we present two main categories for wireless networks interworking:

(i) interworking between access networks with the same technology, mainly 3GPP, and

(ii) interworking between different wireless technologies with a focus on the interworking between 3GPP and WLAN networks.

Inter-3GPP interworking

Network densification using LTE small cells has been an important evolution direction in 3GPP, since LTE Release 10, to provide the necessary means to accommodate the anticipated huge traffic growth. Moreover, LTE small cells can be deployed both with macro coverage and standalone, indoor or outdoor, and can also be deployed sparsely or densely based on each case requirements. LTE interference coordination in such HetNets is widely studied and several radio coordination features are proposed. For example, we note downlink joint transmission, dynamic point blanking known as coordinated scheduling and enhanced inter-cell interference coordination (eICIC).

3. Heterogeneous Interworking

The ability to exploit different access network technologies while providing a seamless subscriber experience has a clear appeal for all service providers and network operators. This is why interworking between HetNets was adopted. Several combinations of access networks were studied including, but not limited to, UMTS/WiMAX [32], WiFi/UMTS [33], WiFi/WiMAX [44] WiFi/HSDPA [38], WiFi/LTE [32, 36], and recently in 2017 mmWave/LTE [37]. However, not too many studies considered simultaneous multihoming. In the following, we present an overview of research works concerning different cases of heterogeneous interworking.

Network Selection Decision

The network selection strategy in HetNets in the literature can be classified into three approaches: network centric, user centric, and hybrid decisions. We present here an overview for different research works in this domain and their contributions for network selection decision.

Network centric strategies generally propose a central scheduler managed by the operator. This central scheduler takes into consideration resource allocation between cell users. Several works addressed the interworking between HetNets using network centric scheduler, we note [36, 38, 40]. Alternatively, user centric strategies delegate the traffic splitting or offloading to the users. For example, the user equipment might decide based on the battery power level combined with the consumption on each access network with preferring to offload on WiFi networks in the battery saving mode [33, 37].

RAN Cloudification
Aiming to fill the blanks in the 5G’s complete image, we introduce in this section the virtual radio access network (V-RAN). The rationale behind VRANs starts with the emergence of cloud computing such as Amazon Web Services, Microsoft Azure and Google App Engine. In parallel, the rapid growth in mobile media applications and platforms was limited by energy and computational resources which imposed restrictions on the advancement of multimedia applications. That’s why cloud computing was proposed as a support for mobile platforms by leveraging the heavy-computational services by executing them on the cloud. The mobile cloud computing [38] was considered as the intersection between mobile computing and cloud computing. Cloud radio access networks (Cloud-RAN or C-RAN) architecture is considered as an innovation in HetNets. C-RAN allows scaling the mobile data network effectively under recent network challenges. C-RAN reduces both expenditures of mobile networks that are facing exponentially increasing data traffic demand [39] [40]. A logical evolution of C-RAN architecture is a V-RAN, a programmable architecture that is software definable and tuneable.

Macro cell

An LTE eNodeB is composed of one baseband unit (BBU) and up to three remote radio heads (RRHs) that can be connected. To connect the BBU and each RRH, an optical interface compliant with the common public radio interface (CPRI) specification, which is standard, is required (see Fig. 1.7). The BBU is responsible for digital baseband signal processing. IP packets received from the core network are modulated into digital baseband signals and transmitted to the RRH. The digital baseband signals received from the RRH are demodulated and IP packets are transmitted to the core network. As for RRH, an RRH transmits and receives wireless signals. An RRH converts the digital baseband signals from BBU that are subject to protocol-specific processing into radio frequency signals and power amplifies them to transmit them to UE. On the contrary, the RF signals received from UE are amplified and converted into digital baseband signals for transmission to the BBU.

C-RAN/V-RAN

In C-RAN, the RRHs are located at the cell site and the BBU is implemented separately and performs centralized signal processing for the RAN. The decentralized BBU enables agility, faster delivery, cost savings and improved coordination of radio capabilities across a set of RRHs. A number of BBUs can be aggregated to form a pool of baseband units (BBU pool).

![Figure 1.7: eNodeB hardware architecture (Source: [31]).](image-url)
In other words, V-RAN will open the door for many new applications in 5G. For example, it offers the possibility of using signal processing software dedicated to a special purpose based on the actual service. However, the realization of these benefits requires suitable strategies for an efficient usage of computing resources [25] [26], energy efficient resource allocation [27], sufficient fronthaul capacity [48] and effective BBU placement [29].

**Functional Splitting**

The C-RAN architecture can be divided into two types, based on the RRH and BBU functionalities: Full Centralization and Partial Centralization.

In full centralization, the functionalities of Layer 1, Layer 2, Layer 3 and signaling as well as operations and maintenance (O&M) are concentrated in the BBU, while RRH has only the radio functionalities as shown in Fig. 1.8. This provides optimum architecture for implementing network optimization techniques, however, it requires a large bandwidth and very low latency link to BBU hotel, to carry the baseband in-phase/quadrature (I/Q) signals.

Partial centralization’s baseband processing functions (Layer 1) are located in the RRH along with radio functions (see Fig. 1.8). This configuration greatly reduces front-haul bandwidth requirements as compared to full centralization. In return, bringing baseband processing in the RRH level makes the upgrade and multi-cell collaborative signal processing less convenient [40].

**Resource Allocation Strategies**

Resource allocation and scheduling is defined as the act of assigning resources to a set of tasks. A set of constraints must be met by any scheduler such as deadline and minimum resource allocation. The
decision and the scheduling problems address the feasibility of the scheduling. Resource scheduling started with the periodic scheduling in 1973 [22] by assigning zero or one resources at a time. Then another version allows sharing a resource or assigning more than one resource at a time. Among the proposed single-resource scheduling algorithms we note First In First Out (FIFO), Earliest Deadline algorithm (EDF) [22], Round Robin (RR), fair queuing (max-min fair scheduling), proportionally fair scheduling, and Scheduling optimization problems.

4. 5G and Energy Issues

The Information and Communications Technologies (ICT) account for a considerable portion of the total energy consumption. Statistics of 2017 tell that the annual average power consumption by ICT industries was over 200 GW, where telecommunication infrastructure and devices accounted for 25%. Moreover, it is expected that in 5G era, millions more base stations with higher functionality and billions more devices with ever higher data rates will be connected [31]. Therefore, dramatic improvements of Energy Efficiency (EE) are required to ensure sustainable energy consumption in ICT.

Various efforts are done to cut down the energy consumption of telecommunication networks. The Energy Aware Radio and Network Technologies (EARTH) project sponsored by EU, has built a framework to support the EE evaluation over the large scale and long term, which is named the EARTH Energy Efficiency Evaluation Framework (E3F). E3F offers the power consumption breakdown for eNodeB components of LTE wireless system. Meanwhile, a flexible power model is built to support the E3F evaluation, which considers differentiation of BSs types. Furthermore, each type of BS is divided into a group of hardware components. The power of each hardware component is affected by several scaling factors, including bandwidth, antenna, modulation, coding rate, and load as presented in [36].

- Energy consumption
- Energy consumption in cellular networks
- Energy consumption in cellular networks could be evaluated generally by considering the power consumed by all the components as well as the dynamic radio power used for transmission function of the load, or particularly by considering the power consumed by each allocated resources.

The consumed power at the base station follows the model provided by EARTH in generalized to all BS types, including macro, micro, pico and femto BSs. Different transceiver (TRX) parts power consumption is analyzed:

Antenna interface: The influence of the antenna type on the power efficiency is modeled by a certain amount of loss mainly at the feeder.

Power amplifier (PA): The power consumption in PA suffers from nonlinear effects which rises the poor power efficiency $\eta_{PA}$.

Radio Frequency RF: The RF power consumption depends of the required bandwidth, the allowable signal-to-noise-and-distortion ratio, and the resolution of the analog-to-digital conversion.

Baseband unit (BB): The BB unit power consumption includes the power consumed by functions such as filtering, modulation/demodulation, digital pre-distortion, signal detection, and channel coding/decoding.
Power supply and cooling: The power supply and active cooling consumption is presented as a loss that scales linearly with the power consumption of other components.

**Energy Consumption in WiFi**

The energy consumption in WiFi is less costly than cellular networks because of the reduced coverage and the lower number of users. The power consumption in this case depends of the AP’s two states: Idle or Dynamic. In a WiFi AP, the power consumption of PA, RF, BB, and power supply and cooling components are reduced or neglected.

**Energy Consumption in mm-wave**

In a mm-wave small cell, the power consumption includes the baseband functions, the RF chains and the phase shifters. The other power consuming part is the power amplifier (PA) which is the most power consuming part in a mm-wave access network. The power consumption in a mm-wave small cell depends on AP’s state: Idle or Dynamic.

**Energy Efficiency Maximization**

EE and sustainability of 5G networks have recently received significant attention from mobile operators, vendors and research projects.

A large amount of work has been reported on EE resource allocation in mobile networks. An energy efficient analysis was provided for LTE HetNets in using realistic power models defined in the EARTH project. Mainly, energy saving techniques such as sleep mode were proposed for idle femto cells. In the same way, authors in proposed small cells activation for the offloading from macro cells to small cells as a strategy to increase power savings.

As for HetNets with multihoming, authors in and developed an uplink and downlink energy efficient allocation model for bandwidth and power resources in a heterogeneous wireless network. In the downlink case, they adopted a win-win strategy that achieves cooperation between different operators. Similar works on network resource allocation with multihoming are presented in with power consumption minimization.

**Conclusion**

We presented in this paper a general overview of HetNets in 5G cellular networks. HetNets emerged as a promising low-cost approach for network densification. The interworking schemes range from load balancing, to offloading and multihoming; the latter being the focus of the present paper. We described multihoming aspects and technology enablers available in 3GPP releases and those proposed by IETF. These technologies mainly include mobility protocols, transport layer’s protocols, and dual connectivity mechanism in 5G. We reported on works on heterogeneous networks interworking, highlighting different network selection strategies.

We also described V-RAN’s architecture and defined BBU and RRH entities based on the different functional splitting types. We showed that BBU virtualization offers new efficiency and coverage enhancements by means of CoMP and eICIC. We reported on resource allocation works for both single type and multiple types of resources. We focused on proportional fairness and dominant resource fairness
strategies for single resource and multi-resource allocations, respectively. We finally presented energy consumption aspects in different wireless networks, described power consuming parts and reported different energy efficiency works in the literature, for HetNets, C-RAN and multihoming.

REFERENCES

12. Iso/iec/IEEE international standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks– specific requirements-part 11: Wireless lan medium access control (mac) and physi- cal layer (phy)


