Wireless Backhaul Architectures for 5G Networks

Jialu Lun

PhD

University of York

Electronic Engineering

October 2017

Abstract

This thesis investigates innovative wireless backhaul deployment strategies for dense small cells. In particular, the work focuses on improving the resource utilisation, reliability and energy efficiency of future wireless backhaul networks by increasing and exploiting the flexibility of the network. The wireless backhaul configurations and topology management schemes proposed in this thesis consider a dense urban area scenario with static users as well as an ultra-dense outdoor small cell scenario with vehicular traffic (pedestrians, bus users and car users). Moreover, a diverse range of traffic types such as file transfer, ultra-high definition (UHD) on-demand and real-time video streaming are used.

In the first part of this thesis, novel dynamic two-tier Software Defined Networking (SDN) architecture is employed in backhaul network to facilitate complex network management tasks including multi-tenancy resource sharing and energy-aware topology management. The results show the proposed architecture can deliver efficient resource utilisation, and QoS guarantee.

The second part of the thesis presents wireless backhaul architectures that serve ultra-dense outdoor small cells installed on street-level fixtures. The characteristics of vehicular communications including diverse mobility patterns and unevenly distributed traffic are investigated. The system-level performance of two key technologies for 5G backhaul are compared: massive MIMO backhaul using sub-6GHz band and millimetre (mm)-wave backhaul in the 71 – 76 GHz band.

Finally, innovative wireless backhaul architectures delivered from street fibre cabinets for ultra-dense outdoor small cells with vehicular traffic is proposed, which can effectively minimise the need for additional sites, power and fibre infrastructure. Multi-hop backhaul configurations are presented in order to bring in an extra level of flexibility, and thus, improve the coverage of a street cabinet mm-wave backhaul network as well as distribute traffic loads.

Contents

List of Figures

List of Tables

Acknowledgement

First, I would like to express my sincere gratitude to my supervisor Prof. David Grace, without whom my work would not have been possible. He has always been very supportive ever since I began working on wireless communications as an MSc student. He has helped me not only by providing lots of insightful ideas and valuable feedback, but also encouraging and supporting me when I had difficult times.

I would also like to acknowledge my thesis advisor, Mr. Tim Clarke for giving many useful suggestions and assistance for my work.

During my studies in York, I have been fortunate to work with various members of the Communications group, with special mention to Dr. Paul Mitchell, Prof. Alister Burr and Dr. Yuriy Zakharov, for providing me with useful comments; Zhehan Li, Shipra Kapoor and Sunghyun Park, for creating a friendly and cooperative atmosphere at work.

I am grateful for the project funding sources that allowed me to pursue my studies: GREENBACK Project from Huawei Technologies Sweden and GCRF Nigeria-HAP Project from EPSRC.

Finally, my deepest thanks go to my family and friends. Thanks to my parents for their unconditional love and support all these years, and close friend Xiao Zhang, for the countless visits to York. Additionally, Shuo Sun, Yahui Wu, Cheng Chen, Hyekyung Seo, and Melissa Young have made my time here a lot more fun.

Declaration

This thesis is a presentation of original work to the best knowledge of the author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Some of the work presented in this thesis has been published in, or planned to submit to conferences, journals and project deliverables, which are listed as follows:

Conference Papers:

J. Lun, D. Grace, A. Burr, Y. Han, K. Leppanen, and T. Cai, "Millimetre wave backhaul/fronthaul deployments for ultra-dense outdoor small cells," in *2016 IEEE Wireless Communications and Networking Conference*, Doha, Apr. 2016, pp. 1-6.

J. Lun and D. Grace, "Software defined network for multi-tenancy resource sharing in backhaul networks," in *2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, New Orleans, LA, June 2015, pp. 1-5.

J. Lun and D. Grace, "Cognitive green backhaul deployments for future 5G networks," in *IEEE International Workshop on Cognitive Cellular System*, Germany, 2014.

Journal Articles:

J. Lun, D. Grace, K. Leppanen, and T. Cai, "Providing millimetre wave backhaul diversity utilising street fibre cabinets," Submitted to *IEEE Transactions on Networking.*

J. Lun, D. Grace, K. Leppanen, and T. Cai, "Design aspects of millimetre wave multi-hop backhaul using street fibre cabinets," Submitted to *IEEE Transactions on Network and Service Management.*

Project Deliverables:

J. Lun, Y. Han, A. Burr, and D. Grace, GREENBACK D1: Use-case Definition and Traffic Assumptions, 2015

J. Lun, R. Lei, A. Burr, and D. Grace, GREENBACK D2: Proposed System Architecture and Technical Enablers for Optimal KPIs, 2016

J. Lun, R. Lei, A. Burr, and D. Grace, GREENBACK D3: Protocols, Algorithms and System Parameters and Their Evaluation with Simulation, 2016

Invention Disclosures Submitted to Huawei:

J. Lun and D. Grace, "Handover decision based on angle for reducing handover frequency in small cells," 2015.

J. Lun and D. Grace, "Providing backhaul/fronthaul path diversity utilising street fibre cabinets," 2016.

D. Grace and J. Lun, "Wireless Backhaul/Fronthaul aggregation node," 2016.

Chapter 1. Introduction

Contents

1.1 Motivation

Future networks are expected to be richer and more complex than those of today, supporting an increasingly diverse range of new applications and services such as device-to-device (D2D) communications, Ultra High Definition (UHD) video streaming, and cloud remote gaming. It is foreseen that 5G networks will provide high data rates and area capacity, reduced end-to-end latency, seamless mobility, and lower energy consumption compared to current 4G LTE networks [1]. The ever increasing capacity demand together with the introduction of new emerging applications is creating a significant burden on the existing networks.

Denser deployment of infrastructure nodes will be a key enabler to boost the capacity of wireless systems [2]. A series of techniques such as coordinated multipoint (CoMP), enhanced inter-cell interference coordination (eICIC) and massive multiple-input multiple-output (MIMO) have been proposed to aid the small cell networks in achieving high capacity and spectrum efficiency [3]. These technologies present new challenges for backhaul networks as additional traffic and coordination information is needed among the base stations (BSs). Therefore, innovative backhaul architectures as well as networking mechanisms are needed to support the future flexible and cost effective networks.

Although fibre can provide a high-speed and high reliability backhaul to support the small cells, installing adequate fibre in every small cell for future networks may not be feasible due to the associated high deployment costs as well as operational challenges. Instead, wireless backhaul technologies have become a major focus of attention due to their high deployment flexibility and cost efficiency. The choice of wireless backhaul solutions for small cells depends on a number of factors, such as the required network capacity, access node density, data rate requirements, as well as energy consumption. Hence, 5G backhaul architecture should embrace a mix of technologies that could be deployed and scaled on demand to accommodate various use cases.

Some of the revolutionary technologies can be applied to wireless backhaul networks, e.g. the introduction of Software Defined Networking (SDN) to manage the network resources, and employing Network Function Virtualisation (NFV) technologies to dynamically isolate and configure resources; the adoption of massive MIMO techniques to improve the spectral efficiency; as well as the use of millimetre wave (mm-wave) band to dramatically increase the bandwidth and therefore the capacity of the backhaul network. In addition, the system level evaluation of the impact of realistic vehicular traffic distribution on wireless backhaul performance, and QoS aware topology management in a highly dynamic backhaul topology need to be investigated.

1.2 Hypothesis

The hypothesis guiding the research presented in this thesis is as follows:

"Introducing and exploiting an appropriate level of flexibility in wireless backhaul networks can improve network utilisation and user experience, enabling improvements in capacity, reliability and latency."

Flexibility has become a fundamental requirement for wireless backhaul network and will have a significant impact on the design of new backhaul network architectures.

The impact of increasing the flexibility in the backhaul network is assessed by evaluating the QoS performance of the proposed backhaul architectures in different large-scale small cell deployment scenarios.

1.3 Thesis Outline

The rest of the thesis is organised as follows:

Chapter 2 provides a literature review on the established work related to this thesis. The use cases and key requirements for 5G are discussed. Small cell networks and the associated challenges are then introduced as a key enabler to 5G. It then reviews the wireless backhaul technologies including sub 6-GHz, mm-wave and free space optics (FSO). Finally, the cloud technologies including C-RAN, NFV and SDN are discussed.

In Chapter 3, a SDN architecture for a 5G dense multi-infrastructure provider deployed wireless backhaul network is introduced. The architecture contains a two-tier controller in a hierarchical setup aiming to offload some of the control functionalities from the central controller to a collection of logically distributed, dynamically configured, local controllers in order to balance the trade-off between scalability and system performance. A multi-tenancy dynamic resource sharing algorithm, together with a backhaul link scheduling strategy based on the proposed SDN architecture are introduced as case studies.

Chapter 4 firstly introduces an ultra-dense outdoor small cell scenario with vehicular traffic (pedestrians, bus users and car users). It then proposes a mm-wave backhaul architecture, and compares it with a massive MIMO backhaul approach whereby the upper bound capacity and access network load balancing are considered.

Chapter 5 presents a novel backhaul architecture where Edge Nodes (ENds) are co-located with street fibre cabinets to provide a cost-effective and robust mm-wave backhaul network for outdoor dense small cells. An analytical framework for backhaul link outage and diversity analysis is then introduced. Next, the backhaul

path management procedure both in a distributed configuration and in a centralised configuration is explained. Finally, a case study is discussed to illustrate the feasibility of deploying existing street cabinets to deliver backhaul.

Chapter 6 investigates various aspects of deploying multi-hop mm-wave backhaul network for outdoor small cells including the choice of relay node and the impact of full duplex capability. Next, several multi-hop topology management schemes for a street cabinet ENd multi-hop network are introduced. The schemes take into account traffic load at the access network, handover frequency, network resource utilisation conditions and QoS control.

Chapter 7 presents the conclusions, summarises the orginal contributions and gives a number of recommendations for future work.

Chapter 2. Literature Review

Contents

2.1 Introduction

The responsibility of a backhaul network is to connect access networks to their core networks via a wired (e.g. fibre and copper) or wireless medium (e.g. microwave and mm-wave). Therefore, the design and optimisation of wireless backhaul networks

play a crucial role in future 5G networks. The purpose of this chapter is to provide the background knowledge related to this thesis.

The rest of this chapter is organised as follows: the vision of 5G, its associated features and small cell networks are discussed first in Section 2.2. Next, the wireless backhaul design considerations are summarised in Section 2.3. Some of the key technologies for backhaul network design including C-RAN, SDN, and NFV are introduced in Section 2.4

2.2 Small Cell Deployment for 5G Network

This section first presents the emerging use cases in 5G era and a set of key performance indicators (KPIs) identified by the research initiatives. It then introduces one of the key technologies to boost the network capacity – small cells, and discusses their associated deployment challenges.

2.2.1 5G Use Cases and Key Performance Indicators

The emerging wide range of new applications is a major driving force which motivates moving towards 5G systems. UHD video streaming, video conference, augmented reality, virtual gaming, intelligent farming, and connected vehicles are examples of the services that the 5G system should support [4]. Among these, UHD video streaming on mobile devices experiences the most rapid increase [4]. Although this use case generally does not have stringent requirement for latency, bandwidth and end-to-end user perceived throughput requirements are relatively high. For example, 4K and 8K on-demand video streaming can require data rates up to 20 and 100 Mbps respectively, and they are mainly characterised by a unidirectional continuous stream [5]. Real-time services, such as video conference and online gaming have also started to gain in popularity. In addition to high data rates, they have generic requirements for low latencies due to the conversational nature. Machine to Machine communication (M2M) such as smart cities, environmental

monitoring and autonomous vehicle control is another trend with billions of devices connected anywhere at any time [6]. While these applications have a broad range of requirements in terms of capacity, latency and information loss, the common challenge is to support a large amount of mobile devices/sensors over a wide area without affecting the performance of other services.

Based on the requirements of the new applications and the ever growing number of mobile devices, industrial and research initiatives have identified a set of KPIs [7] [8] [1] [9]. Figure 2.1 shows a notable example on the expected enhancement of KPIs suggested for IMT-2020 [1].

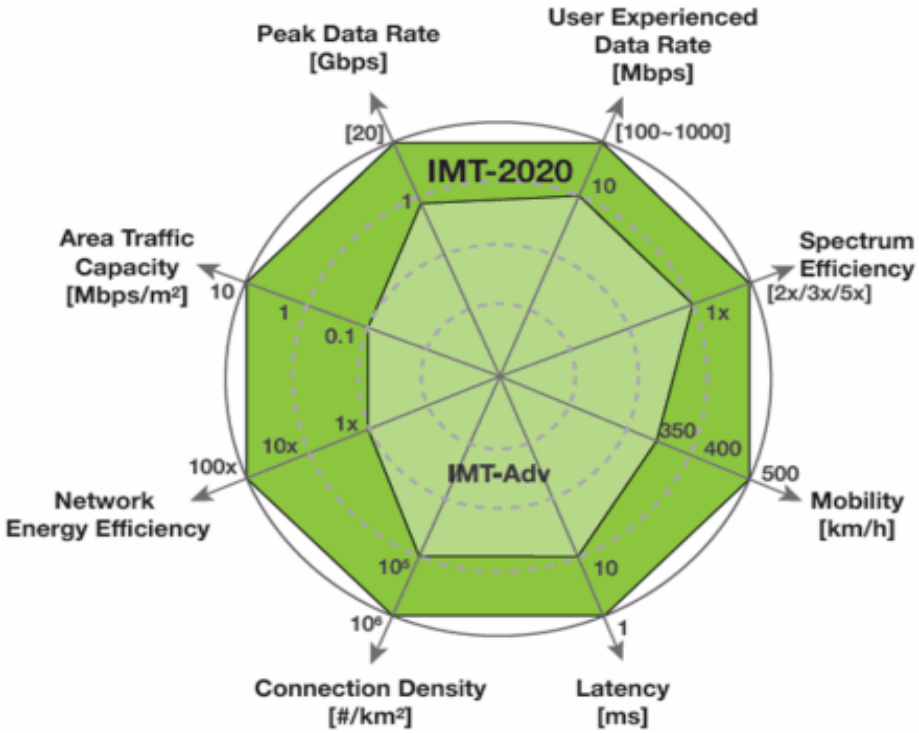

Figure 2.1: Enhancement of key parameters in 5G (directly reproduced from [1])

Some of the KPIs most relevant to this thesis are described in the following:

- *User perceived throughput* measured in Mbps is the experienced data rate can be guaranteed for a given use case of a user.

- *Area traffic capacity* is the total amount of traffic capacity of a network in a given area and is measured in Mbps/km^2 (or Gbps/km^2).

- *Latency* is the time it takes for a small data packet to be transferred over the network from initial generation to final usable reception in ms.

- *Mobility* in km/h is the maximum speed between a vehicle and a radio node, at which the network is able to support the required QoS.

2.2.2 Dense Small Cell Networks

Dense small cell deployments which utilise low transmit power base stations (BSs) to provide localised coverage and capacity will be a key enabler to boost the capacity and coverage of future networks especially in city centres and other high traffic areas [2] [10]. Small cells usually have a coverage range of tens of metres to several hundred metres, and are referred to as femto-cells, pico-cells and micro-cells depending on the cell sizes and transmission power. They can be deployed both indoors and outdoors. A mix of different cell sizes and radio technologies results in a heterogeneous network (HetNet). Despite the benefits brought by small cells, there are many challenges need to be resolved.

Interference management

Interference management is a challenging issue due to the large-scale deployment of small cells. One of the key interference management technologies firstly defined in 3GPP release 10 is known as enhanced Inter-cell Interference Coordination (eICIC) [11] [12]. It introduced the concept of Almost Blank Subframes (ABSs) and Cell Range Expansion (CRE) to reduce the inter-cell interference. ABSs sent by a macro eNodeB only contains reference signals but no data traffic, so that a pico/femto cell can send its data traffic to the UEs typically in the CRE region at a much higher data rate during the ABS periods.

Another emerging state-of-art technique for interference management is Coordinated Multipoint (CoMP) transmission and reception which can also improve the spectral efficiency and cell edge throughput [13] [14]. Unlike the traditional settings where a user equipment (UE) is connected to a single cell at a given time, a cooperating set consisting of geographically separated nodes can be used in CoMP to transmit data to a UE simultaneously. This technique give rise to the new concept of virtual cells, meaning that the traditional cell boundary is eliminated, and the UEs can be served by one or several access nodes to achieve an always cell-centre experience. A preferred set of ANds for a UE would form a virtual cell to track the UE and provide a similar experience throughout the entire network. In the later chapters of this thesis, this advanced CoMP technique is assumed in the access network to allow for a continuous user experience envisioned in future networks.

3GPP has categorized CoMP schemes into the following mainly two categories depending on the amount of shared user data and the level of scheduling [15].

- Joint Processing (JP): The data packets need to be available at all the coordinated points. The most promising implementation of JP is Joint Transmission (JT). In JT, multiple points can jointly transmit or receive the same data to/from UEs simultaneously. The signals are combined at the receiver in order to improve SINR gain. This technique requires the exchanging of channel state information (CSI) and user data among coordinated points, which leads to increased traffic in backhaul networks. In addition, latency and synchronization requirements are very tight. For example, for multi-user MIMO with a JT scheme, the UE perceived throughput gain with a backhaul delay of 15 ms is 26.6% less than the one with zero backhaul delay. Another form of JP is Dynamic Point Selection (DPS)/Muting. In this scenario, only one transmission point in the cooperating set can transmit data to a UE on a given time-frequency resource. The system can choose the most suitable transmission point and resources to convey the data.

- Coordinated Scheduling/Beamforming (CS/CB): The user data is available only at one coordinated point. Only CSI and scheduling information may be exchanged among coordinated points. CS/CB is an effective approach to minimise interference. In CS, transmission points and resources for the users in each time slot are carefully scheduled to reduce the unnecessary interference. In CB, transmit power and beamforming weights for each UE set are calculated to maximise the SINR and reduce inter-cell interference. In the later chapters, perfect coordinated beamforming is assumed in the access network, i.e., a data traffic flow is available and will be served from only one of the access nodes from the coordinated points.

The backhaul capacity required for the networks with CoMP-JP increases linearly with the cooperating cluster size. When high data rate backhaul (fibre) is available, and the processing is centralised to a single entity, which is essentially a C-RAN architecture, CoMP is able to achieve a tight coordination. On the other hand, if the capacity of backhaul is limited, the high signalling overhead in wireless network may result in outdated CSI which can significantly reduce the efficiency of CoMP. In [16], possible solutions to alleviating backhaul bottleneck problem was discussed. Mode switching between CoMP-JP and CoMP-CB/CS based on backhaul constraints can improve overall throughput. This is because in high traffic and limited backhaul scenario, the performance of CoMP-CB/CS is better that that of CoMP-JP and vice versa. In [17], the impact of different backhaul topologies and clustering levels on the feasibility of CoMP was analysed. The results show that a more densely connected backhaul network can improve the cluster feasibility, at the expense of higher costs. Sometimes, even small wireless clusters are not feasible due to the constraints of backhaul network. Hence, they further proposed a backhaul network pre-clustering scheme which targets to exclude the BSs that is unable to participate in the data transmission based on the current conditions of backhaul. The signalling overhead can be significantly reduced, and the savings are between 55% and 90%.

User association

User association is another issue which has drawn much attention in the wireless communication research community. The existing user association schemes use a mix of criteria such as link quality (Reference Signal Received Quality), available bandwidth and traffic type for improved network performance and low service interruption probability [18] [19] [20]. Two realistic factors have often been neglected: *backhaul bottleneck* and *user mobility*.

Most of the researchers have assumed a reliable backhaul with infinite capacity and negligible latency. However, the emerging wireless backhaul solutions for small cells are far from ideal due to the limited bandwidth and reliability issues. Therefore, in addition to the traffic load and interference environment in the access network, new user association criteria are required taking into account backhaul capabilities in terms of capacity, latency. Examples addressing the backhaul issues can be found in [21] [22]. The authors of [21] presented a distributed load balancing user association scheme for fibre-wireless HetNets taking into account backhaul reliability and delay. In [22], a joint scheduling problem for user rate maximisation was investigated for a two-tier HetNet with backhaul bandwidth constraint.

Mobility management is essential in order to provide a seamless uniform service when users move in or out of small cells. Most of the state-of-art algorithms for handover decision is received signal strength based. They compare the received signal strength of the serving and target cells directly or with a hysteresis margin to minimise the handover probability and ping-pong effect. In LTE, the UE measures the signal strength as an RSRP (Reference Signal Received Power) value which is the average power of all resource elements [23]. Another parameter can be calculated from this value is called Reference Signal Received Quality (RSRQ). It is the ratio between the RSRP and the Received Signal Strength Indicator (RSSI). RSSI is the total received wideband power including interference and thermal noise. These schemes perform well in conventional macro cell networks since the coverage of a macro cell is normally up to several kilometres. However, when a mobile user moves between small cells, handovers will be much more frequent even for low speed users

because of the much smaller coverage of each cell. The frequent handovers can cause significant signalling load to the core network entities. Hence, location and speed based handover decision algorithms have become increasingly popular in the small cell research communities. For example in [24], the impact of user mobility in multi-tier heterogeneous network was investigated. Then a speed dependent bias factor was proposed in the tier association scheme to improve the coverage probability and the overall network performance. In [25], a reactive handover decision policy, based on the prediction of user movement and target handover cell, is proposed in order to eliminate frequent and unnecessary handovers.

2.3 Wireless Backhaul Design Considerations

In order to support the discussed new features in 5G, and roll-out a large number of small cells, more cost-effective and easy to install backhaul solutions are needed. In this section, wireless backhaul technologies and their deployment options are presented.

2.3.1 Wireless Backhaul Technologies

There are comprehensive studies that describe the evolution of backhaul, including wired and wireless backhaul technologies such as [26] [27]. With regard to the medium of backhauling, while fibre has been considered as an ideal backhaul solution that is able to provide almost unlimited capacity and high reliability, installing adequate fibre in every small cell for future networks is not feasible due to the associated cost issues as well as operational challenges. Instead, wireless backhaul technologies have become a major focus of attention due to their high deployment flexibility and cost efficiency. There are multiple candidate frequency bands for wireless backhaul, including sub-1 GHz band, 1-6 GHz band, and above 6 GHz band especially mm-wave band. Besides, Free Space Optics (FSO) has also been considered as a promising technology for wireless backhaul [28].

2.3.1.1 Sub-1 GHz band

Frequency bands below 1 GHz have significant potential in providing broadband services to underserved rural communities. Due to the favourable propagation characteristics, they can cover a wider area compared to those at higher frequencies. In particular, Television White Space (TVWS) technology is a promising low cost solution for data provision, which utilises the unused TV broadcasting spectrum on a secondary basis without causing harmful interference to the primary TV receivers [29]. Interference management is an important aspect in developing a TVWS based network. A classic approach to obtain the channel usage information is to use a geo-location data base [30]. Unlike conventional cellular bands, TVWS bands have strict sharing criteria (e.g. transmit power) with the primary users (TV system). This is a limiting factor for small cell backhaul deployment.

2.3.1.2 1-6 GHz band

The 1-6 GHz band has been widely deployed in cellular mobile systems because of its propagation characteristics that can be used in non-line-of-sight (NLOS) point-to-point and point-to-multipoint scenarios. The largest contiguous spectrum allocated for IMT is 200 MHz in the 3.5 GHz band [31]. This frequency band is chosen in the sub-6 GHz wireless backhaul deployment scenario in this thesis. As the rest of the frequency bands of this range are generally in high demand, it is difficult to provide the capacity and QoS required by 5G. This is why extensive research has been focusing on increasing the spatial spectral efficiency of the traditional frequency bands.

In the FP7 BuNGee project, a dual-hop architecture was proposed in order to provide a capacity density of 1 Gbps/km^2. A series of emerging technologies have been applied to the backhaul network to improve the spectral efficiency and resource utilisation, including the use of advanced antenna array and in-band backhaul, whereby the access network spectrum can also be used by backhaul. Some of the work in this thesis is based on the BUNGee backhaul architecture which will be discussed in detail in Chapter 3.

One of the key technologies to drastically increase the spectral efficiency is massive MIMO. It proposes to use a much larger number of antennas at the base stations than its serving devices and aggressive spatial multiplexing techniques to achieve high capacity [32]. Although originally proposed for access networks, the authors of [33] applied massive MIMO technology to in-band wireless backhaul for small cells. The results demonstrated that higher throughput can be expected with the massive MIMO technique.

2.3.1.3 Millimetre-wave Band

Millimetre-wave (mm-wave) in bands 60 GHz and 70 - 80 GHz are very promising Line-of-Sight (LOS) wireless backhaul solutions for future 5G networks, as they can offer abundant spectrum, therefore multi-gigabit data rates. In addition, “license exempt” or “light licensing” is possible in these bands, which can ensure efficient frequency reuse, allowing a significantly lower spectrum cost. For example, apart from 3.5 GHz band, BuNGee architecture also promoted to adopt mm-wave band for backhaul to achieve higher capacity [34].

High attenuation caused by oxygen absorption in 60 GHz and rainfall in 70 -80 GHz limits the propagation range. However, the total attenuation by atmospheric gases at sea level between 71 – 86 GHz is between 0.37 – 0.5 dB/km, which is negligible considering the dense deployment of small cells [35]. Rain attenuation is the dominant concern for mm-wave propagation. The rainfall rate in the UK that is exceeded for 0.01% of the average year as specified by ITU-R P.837 is 35 mm/hr, and this is categorised as very heavy rain [36]. The rain attenuation in this rainfall rate calculated according to ITU-R P.838 is between 14 – 15 dB/km [37]. For a 100 m backhaul link, the attenuation due to rain is approximately 1.5dB. This is not an insurmountable challenge compared with the path loss over distance. Intensive work on the propagation modelling of mm-wave, through field measurements has been presented in [38]. Measurements showed that mm-wave in LOS environments has almost identical path loss as free space.

Two types of mm-wave link failure are usually considered because of its propagation characteristics: weather based outages and beam misalignment. The current attempts to increase the reliability of mm-wave links include using a microwave link as a backup link, deploying multiple nodes to provide link diversity, and multi-hop relay backhauling [39] [40] [41]. Using hybrid links which combine microwave and millimetre wave transceivers can improve reliability of the network when the millimetre wave links suffer from outages caused by rain or obstacles. However, switching to microwave wave when a millimetre wave link fails may not be able to support the capacity and QoS required in dense urban areas. These existing methods also assume several candidate links/paths are already available and focus on the link/path selection algorithms. The feasibility of deploying the backhaul aggregation nodes is often neglected. Due to the narrow antenna beamwidth, millimetre wave links are prone to misalignment outages caused by wind. The impact of wind sway on mm-wave beamforming misalignment has been investigated in [42]. It is shown that beam tracking on the order of milliseconds is required to overcome outage. Beam steering allows beam alignment during the initial stage and is resilient to the misalignments due to wind.

2.3.1.4 Free Space Optics

FSO uses laser light to transmit data through free space propagation [28]. FSO systems are able to provide point-to-point high bandwidth links without licensing requirements, and the installation and maintenance costs are relatively low. The transmission windows centred on the wavelengths of 850 nm and 1550 nm are suitable for FSO transmission because of their low attenuation, as well as the inexpensive transmitter and detector components. However, systems operating at wavelength around the 1550 nm band have lower risk in terms of the eye safety which allows approximately 55 times more transmit power than those at 850 nm [43]. In addition, wavelength division multiplexing (WDM) FSO is also feasible for 1550 nm.

The major disadvantages of the FSO are the vulnerability to certain weather conditions such as fog and snow. Although 99.999% availability is generally

achievable for FSO links ranges less than 140 m as suggested in [44], it requires a considerable link margin. For the dense fog situation for example, at least a 31.5 dB link margin needs to be included in the link budget. This may not be a cost-effective solution as a higher transmit power is needed. Besides, it is difficult to determine the availability of the FSO system based on the worst possible weather conditions which only apply to certain areas. Hybrid architectures of FSO and mm-wave have been proposed to increase availability because of their complementary propagation characteristics, but such benefits are only realised over long length links where rain and fog attenuation is significant. In a dense deployed small cell architecture, the rain attenuation at mm-wave is not a significant issue. Hence introducing FSO to the system may not able to bring benefits in terms of availability and energy efficiency. However, in the case where a future mm-wave access network is deployed, where increased bandwidths and data rates are being proposed, a FSO backhaul would be one way of providing the high link capacities needed to support this increased capacity.

2.3.2 Topology

The next generation wireless backhaul networks also need a rethink of the topologies, and the impact of the available options needs to be determined. In [45] and [46], tree topology which is a combination of star and chain topologies, and ring topologies for wireless backhaul were compared in terms of capacity, resiliency, latency, and costs. The mentioned topology options are shown in Figure 2.2. A mesh topology whereby each node in the network is interconnected with one another was also analysed. It can offer better robustness to traffic fluctuation and availability compared to the other two due to the path redundancy. However, these benefits come with associated costs, complexity of topology management, and scalability issues. Hence, there is no clear cut of which topology is superior as the deployments of the wireless backhaul are usually based on a combination of factors, such as geographical and business requirements.

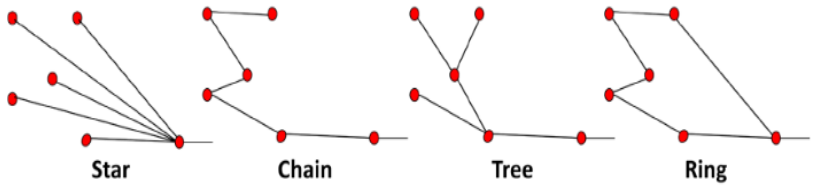

Figure 2.2: Backhaul topology options (directly reproduced from [46])

2.3.3 Green Wireless Networks

The densification of the network driven by increasing traffic volumes can lead to increased CAPEX and OPEX. Therefore, energy efficiency has become an important research topic. New designs of 5G need to reduce the energy consumption by considering all layers, from physical components up to network level.

For hardware components, novel active antennas for MIMO can efficiently improve the data rates or reduce the transmit power [47]. Energy efficient power amplifiers and baseband processing are needed, since they contribute to the highest energy consumption of the BS. Innovative radio resource management (RRM) solutions and MAC protocols that can increase the utilisation of resources are also required. In [48] for example, variable power/bandwidth modulation schemes based on different spectrum occupancy probabilities are used in order to improve bandwidth efficiency and therefore reduce the energy consumption. From a network level perspective, since the traffic of mobile networks is usually non-uniform in both time and spatial domain, it is worth introducing novel network management paradigms. BS cooperation techniques such as CoMP we mentioned in Section 2.3.1 can be used to exploit the transmission diversity. Self-organizing capability is crucial for maximizing the resource utilisation in future complex network architectures. Energy efficient topology management strategies have been investigated on a large scale in wireless access networks. In [49], BSs can be activated or deactivated based on local traffic conditions, so that the energy consumption can be reduced.

In [50], the overall network power consumption under different backhaul architectures has been assessed. The results showed that the power consumption of

the backhaul network can reach up to 50% of that of the wireless access network, and deployments of macro base stations (BS) with femto BSs being more cost-effective than a macro-cell architecture only in very high traffic scenarios. Since backhaul networks play a key role in terms of the overall energy consumption, joint energy efficiency optimisation of access network and backhaul network need to be addressed.

2.4 Cloud Technologies for 5G Backhaul

The emerging diverse traffic patterns in small cells, both spatial and temporal, together with the increasing need in processing power, make the cloud technologies very attractive for facilitating complex network management tasks [51]. Backhaul networks need to adopt new flexible management mechanisms as the rest of the system in order to improve the flexibility and resource utilisation of the network. In this section, revolutionary technologies including centralised radio access network (C-RAN), NFV and SDN are introduced.

2.4.1 Cloud-RAN

By centralising baseband processing functions, C-RAN allows a full joint processing version of Coordinated Multipoint (CoMP) transmission to be readily implemented, which can significantly improve system performance and resource utilisation. The network between the remote radio heads (RRHs) and the baseband unit (BBU) carries digitised signals: for this reason, it is referred to as fronthaul rather than backhaul as in the conventional Distributed–RAN (D-RAN) architecture (as shown in Figure 2.3). However, high data rates and extremely low latency are required for fronthaul. In contrast to distributed-RAN (D-RAN) architectures, where BBUs are located close to antennas, C-RAN is able to reduce CAPEX and OPEX because of the simplified hardware and management control. It enables joint processing and scheduling, increased BBU utilisation, and reduced power consumption due to the enormous centralised functionalities and flexibility introduced to the networks.

Besides, advanced techniques, such as CoMP, eICIC, virtualisation, and multi-RAT, can be supported by C-RAN.

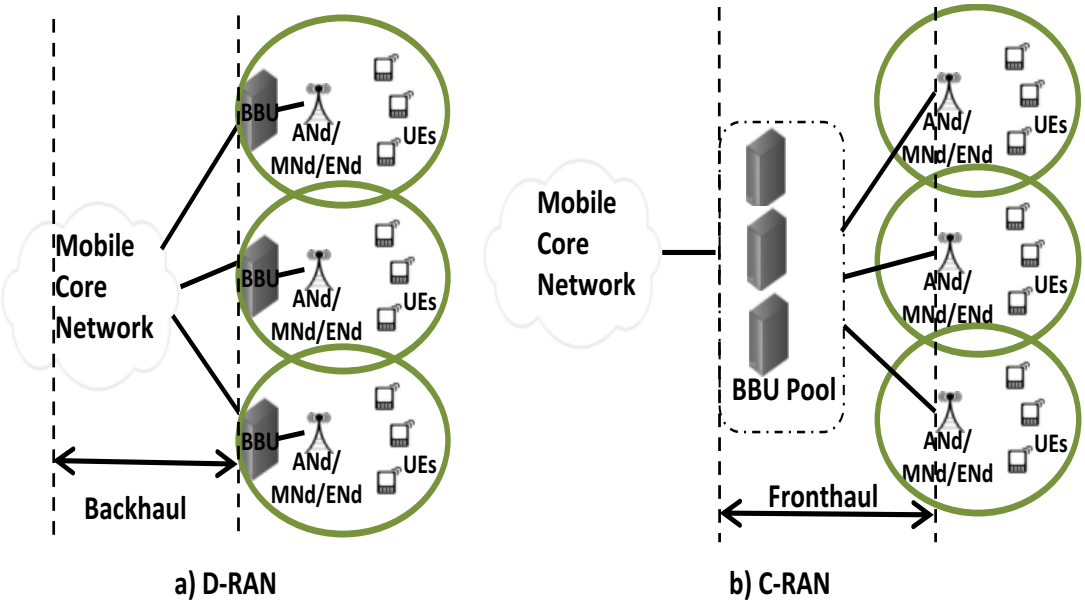

Figure 2.3: Example D-RAN and C-RAN Architectures [51]

In [52] and [53], it has been predicted that the maximum statistical multiplexing gain achieved in C-RAN by dynamically allocating BBU resources across the network compared to a D-RAN configuration is 4 based on their mobile traffic forecast. Although the result was obtained under the assumptions of specific proportion of different types of BSs, traffic distribution and modelling, and scheduling algorithms, it demonstrated the potential of C-RAN as a promising solution to CAPEX and OPEX reduction. After adding TCP/IP and Ethernet protocol processing, the multiplexing gain can reach up to 1.6. The authors also pointed out that C-RAN is more advantageous in a densely deployed urban area than in a sparsely populated one.

In [54], the Next Generation Mobile Networks (NGMN) alliance described possible deployment scenarios for C-RAN, followed by main functionalities and

requirements. They then presented some reference C-RAN solutions to facilitate different network conditions. For example, it may not be economical to deploy a full mesh C-RAN architecture as a network grows, since the multiplexing gain achieved from resource sharing becomes marginal. Instead, a hierarchical BBU architecture, where the BBUs were split into several BBU sub-clusters, was encouraged due to scalability and flexibility reasons. They also illustrated that the connection between fronthaul networks could be in a centralised or distributed fashion. For the fronthaul, which is the link between RRHs and BBU pool, topology suggestions depending on different levels of resource availability were discussed. Last but not least, they suggested how traditional RAN could gradually evolve to C-RAN, which include three main phases: BBU centralisation, BBU pooling and Virtual RAN.

High data rates are required for fronthaul, which is why the existing C-RAN architectures all rely on fibre. Approximately 10 Gb/s for TD-LTE with 20 MHz bandwidth is needed [55]. In addition, the interconnection between BBUs requires low latency and high reliability. As the size of C-RAN becomes bigger, more data needs to be backhauled, and more processing needs to be performed within the cloud. To strike a balance between centralised processing in C-RAN and decentralised processing using traditional backhaul, the iJOIN project has introduced the concept of RAN-as-a-Service (RANaaS), which enables flexible RAN functionality split between radio access point (RAP) and central entity based on the actual demand and conditions of the network [55]. This is particularly beneficial for backhaul networks as it is difficult to find a backhaul technology that is suitable for all the deployment scenarios. Heterogeneous backhaul solutions are needed. With a RANaaS platform, the system can adjust the functionality split according to the availability of backhaul resources as well as data rate and latency requirements. The challenges are to develop flexible Medium Access Control (MAC) protocols and radio resource management algorithms that can adapt to fluctuating backhaul requirements and access load conditions.

2.4.2 Network Virtualisation

In [56], a unifying definition of network virtualisation is given as:

Network virtualisation is any form of partitioning or combining a set of network resources, and presenting (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. Resources can be fundamental (nodes, links) or derived (topologies), and can be virtualized recursively.

With network virtualisation, the networks can achieve increased resource utilisation, reduced deployment and operational investment costs, and more flexibility in network management. While the concepts of network virtualisation were originated from wired networks, such as overlay networks, Virtual Private Networks (VPNs), and Virtual Sharing Networks (VSNs), there are good reasons to consider wireless network virtualisation [56] [57]. The overall costs can be reduced by sharing the network. According to [58], for urban dense area, sharing the wireless network can reduce 25% - 48% of CAPEX, and 16% - 18% of OPEX depending on different sharing configurations. The sharing elements could be antennas, BSs, Radio Network Controllers (RNCs), backhaul, or spectrum. The authors also stress that the spectrum sharing contributes the most to the difference of OPEX saving. Besides, network virtualisation opens the door for small companies to lease virtual networks with flexible and customized services without huge investments. Furthermore, it is more convenient to evaluate new technologies in the system due to the isolation among the virtual networks.

In terms of the business roles of wireless network virtualisation, there are usually two types of entities in the network: infrastructure providers and service providers. An example of this model is shown in Figure 2.4, where the infrastructure providers own the physical network resources and provide an interface that can create virtual networks. On the other hand, service providers lease resources from one or more infrastructure providers, and offer end-to-end services to the end users, and do not necessarily have the knowledge of the underlying physical architecture. The model

can be further split up into more fine-grained models [57-60]. In 4WARD project for example, three roles including Infrastructure Provider, Virtual Network Provider (VNP), and Virtual Network Operator (VNO) are used [60]. With the more specific roles, new use scenarios can be supported, and the providers may have more flexibility in choosing different business strategies.

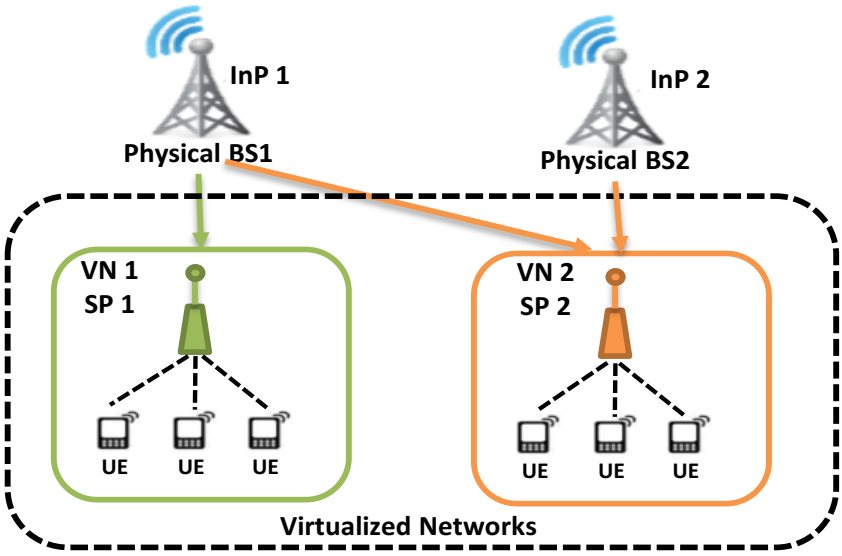

Figure 2.4: An Example of Wireless Network Virtualisation Model

Based on different types of resources that can be virtualised and the levels of virtualisation, three main perspectives of wireless network virtualisation are presented in [61]: flow-based virtualisation, protocol-based virtualisation, and spectrum-based virtualisation. In flow-based virtualisation, the virtualized resources could be radio resource blocks or traffic/ packet flows between uplink and downlink, hence more flexible and efficient flow provisioning can be achieved. However, it requires all the virtual slices share the same wireless protocol stack. Protocol-based wireless virtualisation on the other hand, focuses mainly on MAC/PHY layer, where multiple wireless protocols, such as MAC decision and configuration logic, PHY processing blocks can be virtualized. In this paradigm, different protocol stacks are allowed to run on the same radio hardware. In spectrum-based wireless virtualisation,

dynamic spectrum allocation and RF frontend virtualisation are addressed. It involves cognitive radio, dynamic spectrum access techniques, as well as RF frontend circuit enhancements. In the case study presented in Section 3.4 of this thesis, both spectrum and backhaul infrastructures are virtualised to cater for different service providers.

Several wireless network virtualisation proposals have been introduced in the last few years. A LTE wireless virtualisation framework was proposed in [62], which shows better performance and resource utilisation can be achieved by adding a "Hypervisor" on top of the physical resources allowing dynamic spectrum sharing among virtual service providers. The Hypervisor is responsible for virtualizing the LTE base stations, as well as scheduling the air interface resources among different virtual operators. Despite the benefits this framework can bring, there are still some issues: the coordination among service providers, the scalability of the framework, and the impact of extra signalling overhead.

In [63] and [64], a network virtualisation substrate (NVS) based on a WiMAX test-bed was presented. It is essentially a hierarchical scheduler which decouples the flow scheduling from the slice scheduler in order to facilitate different resource provisioning (resource-based and bandwidth-based) requirements simultaneously. It provides a higher level of flexibility and customization in slices. Here, a "slice" means a virtual network which consists of a collection of flows. The authors also analysed the trade-offs between the level isolation and virtualisation. Higher level virtualisation can achieve increased resource utilisation, but reduced isolation and flexibility, whereas virtualisation at a lower level leads to better isolation at the cost of resource utilisation efficiency.

Although many advantages of wireless network virtualisation are identified, there are still some challenges to be addressed for 5G [56] [57] [59] [61]. The first one is the control framework. It should have a well-defined common interface with high flexibility and programmability that is able to coordinate among multiple service providers over different wireless technologies. High level control of wireless resources can offer efficient design, but also increased system overhead and latency.

Hence, it is important to balance the trade-off of flexibility and performance, all taking the QoS requirements from service providers into account. The second one is resource allocation fairness. Due to the variability of air interfaces, topologies, user mobility, power control, and interference of different service providers, it is more complicated to ensure the fairness of resource allocation in wireless network virtualisation compared to the wired ones. Robust resource scheduling and allocation mechanisms are needed to accommodate different provisioning requirements. Furthermore, due to the complex interference environment of wireless network, the isolation among different service providers is not straight-forward. The solution to wireless network virtualisation should be resilient to the fluctuation of channel conditions and traffic.

2.4.3 Software Defined Networking

With the conventional proprietary and distributed integrated infrastructure, launching new services and managing the networks are becoming exceedingly challenging for network operators, since they require significant investment in network deployments and fast reactions to network-wide dynamic changes with low-level vendor-specific configuration [65]. The characteristics of SDN, such as a global network-wide view, unified network management and programmability, make it a suitable candidate for these new network management tasks. One of the main benefits of SDN is enhanced capability of deploying different devices and services. With the decoupled unified control plane, SDN is able to dynamically reconfigure the network from a centralised entity [66]. SDN can also improve the resource utilisation as well as network performance due to the ability of acquiring global network status instantaneously. Furthermore, the joint intra-tenant and inter-tenant control of multi-tenancy is possible with the help of SDN. Recently, researchers have applied SDN concepts to various scenarios, such as network virtualisation, and quality of experience (QoE) guarantee [67] [68]. The most common scenario of SDN is network virtualisation. OpenFlow for example, can be used to build virtual

networks by dynamically configuring the forwarding rules in physical and virtual switches.

In [68], the authors envisioned that through the use of software defined wireless networking (SDWN) technology, an API may be able to dynamically change the traffic paths and priorities based on the network conditions and the agreements between infrastructure providers and service providers. It can also be aware of the user preference, as well as the service experienced by the user, and then make personalised adjustments according to the requirements. They then proposed a high-level SDN architecture for wireless networks. Key interfaces for control plane functions regarding resource sharing of backhaul/access networks, third party applications, and programmability on the mobile node were defined. In [69], a SDN architecture for co-primary spectrum sharing was presented, which abstracts the control functionalities of all the base stations into one centralised entity. A SDN-controlled topology-reconfigurable optical mobile fronthaul (MFH) architecture was proposed in [70], which can be used for bidirectional CoMP communication in 5G. With the centralised SDN-controller and tuneable switching, only one BBU is needed to send/receive the same message to/from RRHs for CoMP, whereas multiple BBUs are needed in the X2 interface with legacy distributed control and fixed topology. Computational complexity as well as latency can be reduced dramatically with the presented architecture when CoMP with massive MIMO is enabled in the system.

Most SDN architectures we mentioned above utilise centralised SDN controller, and make decisions reactively which raises scalability concerns, moreover, some functionalities may be less time-critical, and just need a global view of the network. A multi-controller setup can be used [71] [72]. Kandoo use a two-level controller architecture consisting of a root controller dealing with the requests that need network-wide view, and a collection of local controllers processing time-critical requests [73]. Another concern of the all-centralised architecture is the resiliency to single point failures. A fast recovery mechanism is a must for the control applications.

2.5 Conclusion

This chapter has provided the background information related to this thesis. The use cases and key requirements for 5G have been discussed. Small cell networks were then introduced as a key enabler to 5G. The challenges in deploying a small cell network including interference management and user association have also been discussed.

Wireless backhaul technologies using different frequency bands and topologies have been summarised. Furthermore, the energy efficiency of wireless networks has been discussed. Enabling technologies to reduce the energy consumption have been reviewed.

Innovative networking technologies including C-RAN, network virtualisation, and SDN have been extensively reviewed. To enable those techniques, high data rate, low latency, and scalable wireless backhaul solutions are needed.

Chapter 3. Software Defined Networking for Wireless Backhaul Networks

Contents

3.1 Introduction

As discussed in Section 2.4, new flexible management mechanisms need to be adopted in backhaul as the rest of the network in order to improve the flexibility and resource utilisation of the overall system. To this end, SDN has emerged as a promising paradigm to facilitate new network management tasks. Most of the SDN architectures for wireless networks as discussed in Section 2.4.3 utilise an all-centralised architecture which raises scalability concerns. This is particularly challenging if such networks are deployed across different infrastructure providers, as there will be additional coordination messages and signalling overhead.

In this chapter, a two-tier dynamic controller SDN architecture for wireless backhaul networks is applied to the backhaul network in order to strike a balance between scalability and system performance. Two use cases are presented using the SDN concept to address two increasingly popular research area in wireless backhaul network: multi-tenancy dynamic resource sharing with QoS guarantee and energy-aware topology management for backhaul networks. These applications of the SDN architecture are not meant to be exhaustive and can be extended to other network management tasks discussed in the later chapters.

The rest of the chapter is organised as follows: in Section 3.2, the proposed two-tier controller SDN architecture is introduced. The detailed backhaul network model is discussed in Section 3.3. In Section 3.4, two use cases leveraging the concept of SDN are presented. Finally, conclusions are given in Section 3.5.

3.2 Software Defined Networking Architecture with Two-tier Dynamic Controllers

SDN proposes to decouple the control plane from the data plane enabling a centralised control over the entire network. Figure 3.1 shows a high-level view of a

widely accepted SDN reference model [66]. It has three layers: an infrastructure layer, a control layer, and an application layer, described as follows:

- *Infrastructure Layer*. This layer consists of SDN enabled switches and network devices. Each switch processes the incoming packets based on a flow table defined by the controllers. It also collects the network usage statistics such as traffic load and topology information, and then reports them back to the controllers via a south-bound interface which is responsible for the communication between the controllers and the switches.

- *Control Layer*. A centralised controller is responsible for controlling the network behaviour via a standardized protocol such as OpenFlow, which is a widely adopted mechanism for SDN technology, whereby the switches can be programmed by open application program interfaces (APIs) [66] [71]. The central controller is located between the infrastructure layer and application layer. The responsibilities of this layer are reporting network status and deciding forwarding rules.

- *Application Layer*. This layer allows third parties to program control algorithms using the APIs provided by the controller, and then push the policies to the controller via a northbound interface.

In order to alleviate the scalability problem at the central controller, especially for a densely deployed network, we further split the control layer of the reference model into two hierarchical tiers: a logically centralised controller and a collection of reconfigurable local controllers as shown in Figure 3.2. The reconfigurable local controllers act as supplements to the systems that are required to scale down the control burden from the central controller if necessary, and can be disabled depending on the system requirements. This enables offloading the control functionalities, which require fast reactions (low latency) but not necessarily a global view of the network, to the local controllers. They reside near to the base stations separately or logically partitioned by the central controller based on the scalability requirements, and are periodically reconfigured to control and record the network

behaviours in the specific assigned areas. The central controller deals with the requests that are less time-critical yet need network wide knowledge. It provides the necessary network status to, and obtains the network policies from, the application layer using a northbound interface.

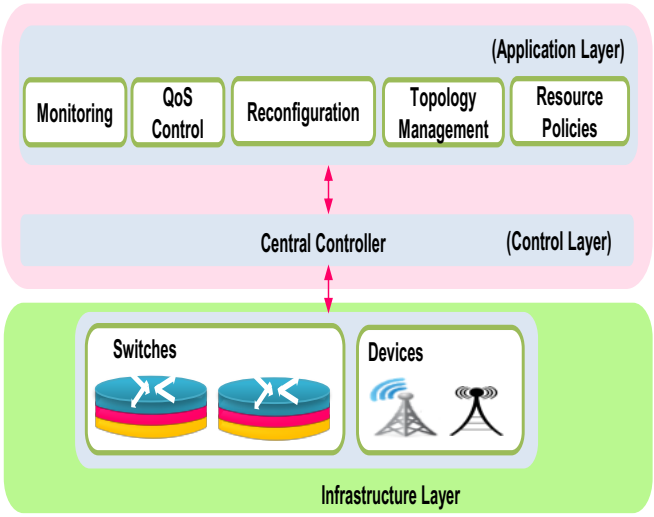

Figure 3.1: SDN reference model

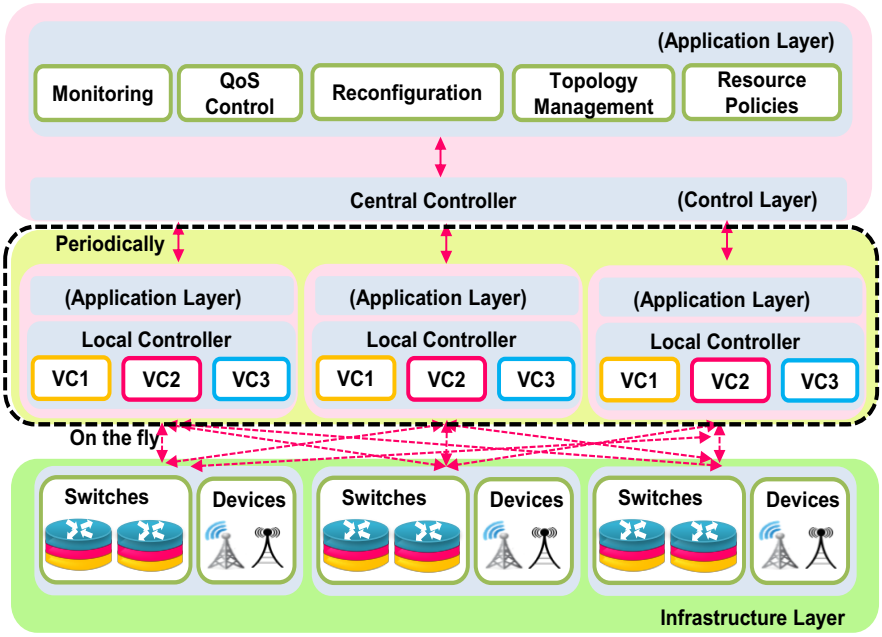

Figure 3.2: SDN architecture with two-tier controllers

3.3 Network Modelling

The proposed SDN architecture is applied to the wireless backhaul network derived from the BuNGee architecture proposed in the FP7 Beyond Next Generation Mobile Broadband project, which has been considered as a future 5G architecture providing a capacity density of 1 Gbps/km^2 [34]. The details of the network modelling are described in the rest of this section.

3.3.1 Network Architecture

In order to achieve the high capacity density promised in 5G, a dual-hop system as illustrated in Figure 3.3 is used as the network architecture in this chapter and throughout the thesis. It has been shown that this architecture can significantly improve the coverage and capacity of a wireless network [34]. The Edge Nodes (ENds) providing the coverage and connection to the core network for the system are equipped with antenna arrays that can generate highly directional beams. Access Nodes (ANds), acting as relays from ENds to User Equipment (UE), are used to deliver high capacity density. They are low-cost, easy to install devices, and can be deployed below rooftops. Each ANd is co-located with a Mesh Node (MNd), which is used to provide connections between ENds and ANds. They can also be used to form a wireless mesh network which will be discussed in the later chapters.

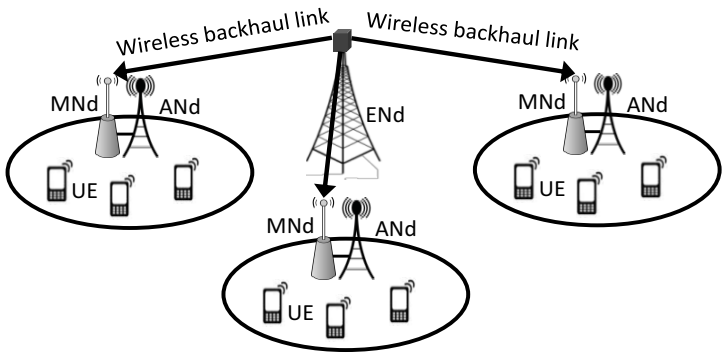

Figure 3.3: Dual-hop network architecture

3.3.2 Spectrum and Antenna Model

In this chapter, each infrastructure provider in the system is allocated a 10 MHz bandwidth in the 3.5 GHz frequency band, which is divided into 30 channels. All simulations reported only focus on the downlink transmissions from ENds to MNds. The uplink transmission is not considered since a separate frequency band can be used, and it does not affect the key results investigated.

The ENds are equipped with highly directional multi-beam antennas which can generate up to 24 beams pointing to their surrounding MNds. Each beam has a 15° azimuth beamwidth. 20 beams are used in this architecture [74]. At the MNd receiving end, an electronically steerable antenna array is deployed. A maximum of two beams can be generated by adjusting the antenna weight vectors on multiple antenna ports. The antenna model used in this chapter is derived from a practical product, which is designed specifically for the BuNGee architecture [34]. The beam pattern of a MNd antenna is shown in Figure 3.4.

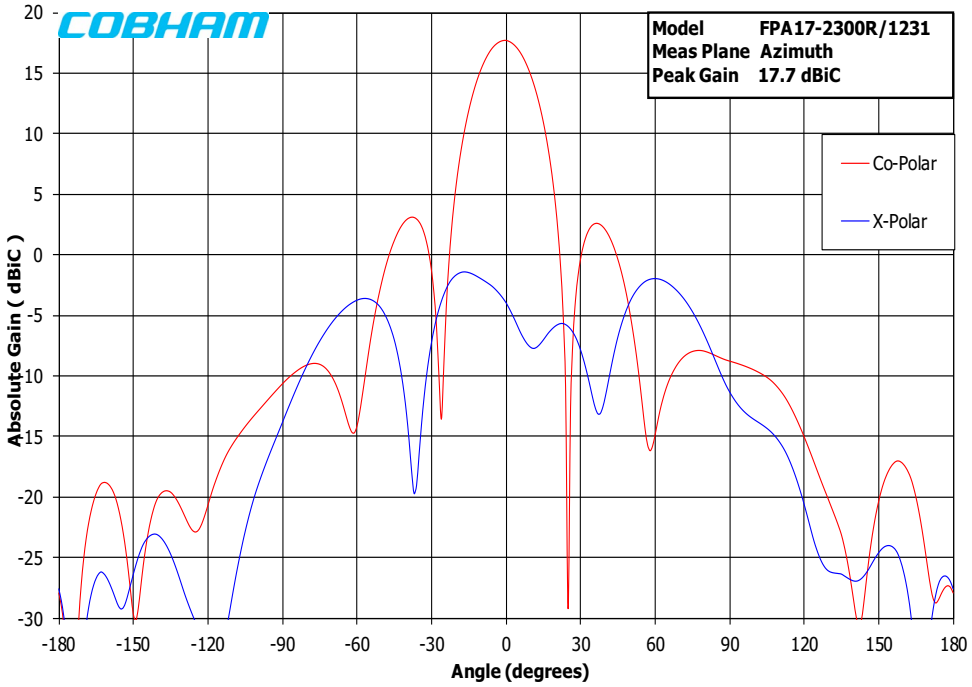

Figure 3.4: Antenna beam pattern of MNd antenna (directly reproduced from [74])

3.3.3 Radio Propagation Model

The propagation model used to calculate the path loss between the ENd and MNd in this chapter is WINNER II B5f, which describes the scenario of stationary BSs (rooftop-to-below/above rooftop) [75]. Specifically, the path loss can be calculated by:

$$PL = 23.5log_{10}(d) + 57.5 + 23log_{10}\left(\frac{f_c}{5.0}\right) + \gamma \quad (3.1)$$

where PL is the path loss in dB, d is the distance in metres between an ENd and a MNd, $f_c = 3.5$ is the carrier frequency in GHz, and γ is the shadow fading loss which is log-normal distributed with a 0 dB mean and a standard deviation of 8 dB.

3.3.4 Link Model

Channel capacity is the upper bound of the data rate that can be transmitted over a communication channel. In reality, the data rates are always constrained by physical layer. A Truncated Shannon Bound model proposed in [76] and is used in our system to determine the data rate of links. The throughput in bps/Hz of a link can be calculated by:

$$C = \begin{cases} 0, & SINR < SINR_{MIN} \\ \alpha B log_2(1 + SINR), & SINR_{MIN} \leq SINR \leq SINR_{MAX} \\ \alpha B log_2(1 + SINR_{MAX}), & SINR > SINR_{MAX} \end{cases} \quad (3.2)$$

where α=0.65 is the attenuation factor. $SINR$ is the Signal-to-Interference plus Noise Ratio at the receiver. $SINR_{MIN}$=1.8 dB is the minimum SINR threshold to

maintain transmission, and $SINR_{MAX}$=21 dB is the SINR when the maximum capacity of a link can be reached. Here, the SINR in linear form at a given receiver i on a channel can be obtained by:

$$SINR_i = \left(\frac{P_{Ti}G_{Ti}G_{Ri}PL_i^{-1}}{\sum_{j=1}^{N} P_{Tj}G_{Tj}G_{Rj}PL_j^{-1}+\sigma^2}\right) \quad (3.3)$$

where P_{Tx} is the signal received from the transmitter x ($x = i$: signal of interest; or $x = j, (j = 1, \dots, N)$: all the other interfering signals using the same channel), G_{Tx} and G_{Rx} are the transmitter and receiver gain respectively, PL_x is the path loss between a transmitter and receiver in linear form. Assuming a noise temperature of 290 K and a noise figure of 8 dB, the noise floor σ^2 is -111 dBmW.

3.3.5 Traffic Model

The 3GPP File Transfer Traffic (FTP) model 1 with a fixed file size of 2 Mbytes is used in this chapter to generate the traffic for the downlink [77]. It can capture the random bursty behaviour of the network traffic. The file arrival rate is modelled as a Poisson process with a mean arrival rate of λ, which can be adjusted to obtain different offered traffic densities.

3.3.6 Resource Allocation Assumptions

The network requires fast reaction upon changes in the current interference environment, hence the channel state information is collected and sent to the local controllers by SDN enabled switches. The local controller is assumed to assign a channel with the highest SINR from an ENd to its serving MNd which requires a new traffic flow. A traffic flow is assumed to occupy its assigned channel until the file transmission is completed, even when the data rate drops to zero temporarily

during the transmission. Hence, file dropping and retransmission are not considered. A new flow is blocked if no channel has a SINR above the minimum threshold of 1.8 dB. In addition, the handovers of current transferring files are not considered, i.e., if a backhaul link is used for a MNd, it will remain connected until the current transferring files have finished transmitting. For the purposes of evaluating the performance of backhaul network, a perfect access network which simply delivers the data from the serving ANd to the UEs (or *vice versa* for the uplink) is assumed. In this circumstance, the throughput and capacity are constrained by the backhaul network alone.

3.4 Use cases

In this section, two use cases are presented using the SDN concept to address two increasingly popular research area in wireless backhaul networks in order to improve the flexibility and resource utilisation: multi-tenancy dynamic resource sharing with QoS guarantee and energy-aware topology management for backhaul networks. These applications of the SDN architecture are not meant to be exhaustive and can be extended to other network management tasks discussed in the later chapters.

3.4.1 Multi-tenancy Resource Sharing in Backhaul Networks

For a wireless network, the resources to be virtualised contain infrastructure as well as spectrum. As introduced in Section 2.4, increased resource utilisation can be achieved with inter-infrastructure virtualisation strategies as they can introduce new possibilities to optimise the resource allocation across the infrastructure providers in the same coverage area [56]. In this case study, we consider a scenario where the backhaul resources of two infrastructure providers are shared among 3 service providers in the same coverage area.

3.4.1.1 Deployment Scenario

Figure 3.5 shows the network architecture, where 9 ENds and 60 ANds/MNds are distributed across a service area of 1.35×1.35 km^2. Here, 5 of the ENds belong to InP1, and the other 4 belong to InP2. This type of independently deployed architecture by different providers having overlapped service coverage is very common in the existing wireless networks. Each infrastructure provider owns 10 MHz of spectrum, which is divided into 30 logical channels. The service providers dynamically request resources; hence the network needs to be virtualised into maximum of three virtual networks. As 5G networks will introduce a wide range of new services, installing independent relay nodes for all the service providers is certainly not a cost-effective and space-saving solution for multi-hop systems. The trend towards 5G will be to increase the integration of multiple radio access technologies (RATs) into devices. Hence, we assume the MNds are multi-RAT enabled provided by a relay node infrastructure provider, and can be shared among different service providers. The antenna beams on the MNds point to the directions of the nearest two ENds from different infrastructure providers.

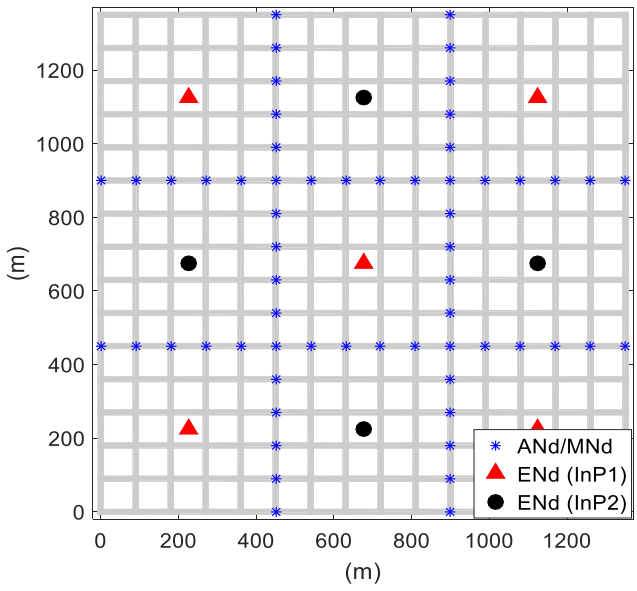

Figure 3.5: Network architecture with 2 infrastructure providers providing the backhaul

3.4.1.2 Different Control Level Comparison

In the proposed architecture, the local controllers are reconfigured by the central controller periodically, and are responsible for requests which require immediate responses but not necessarily a global view of the entire network. During each time resolution, they take their local backhaul link and channel state information as input, and assign the resources allocated by the central controller to different traffic flows.

The central controller is in charge of the issues which require coordination across infrastructure providers and service providers, such as mapping and allocating physical resources including infrastructure and radio resources to service providers, and configuring and managing the local controllers. It allocates a trunk of resources which may belong to either or both of the infrastructure providers to each service provider periodically or upon Packet-In messages indicating a new request cannot be handled by the current forwarding rules, for instance, a new service provider requiring access. In order to differentiate different service providers, a service provider identifier (e.g., PLMN-ID) can be used for data forwarding as well as measurement report. Since the pricing and business model of the involving partners in a virtualised network is out of the scope of this thesis, we assume the service providers are assigned half of the total allocated spectrum from each infrastructure provider.

Assuming all the service providers (SPs) are served on best-effort basis, a proportional resource sharing scheme is used in the proposed SDN architecture. At each time interval T_s, the estimated demands of service providers $D_{SP}(j), (j = 1,2,3)$ which are the average number of channels they have requested during the last T_s s are evaluated. Since the number of allocated channels is always an integer, the channel provision $A_{SP}(i), (i = 1,2)$ for SP1 and SP2 at a MNd is rounded to the nearest integer and the remaining channels are allocated to SP3. $A_{SP}(i)$ can be calculated by:

$$A_{SP}(i) = \begin{cases} \left\lfloor \frac{D_{SP}(i)}{\sum_{j=1}^{3} D_{SP}(j)} R_T \right\rceil , \ i = 1,2 \\ R_T - A_{SP}(1) - A_{SP}(2), \ i = 3 \end{cases} \quad (3.4)$$

where R_T is the total amount of channels on both of the backhaul links of a MNd.

The multi-tenant resource sharing scheme enabled by the proposed two-tier controller SDN architecture is compared with two other control approaches: *no control* and *all-centralised control*. In the no control setup, all the service providers obtain equal shares of the spectrum from both of the infrastructure providers regardless of the demand, which is similar to the situation where the operators owning exclusive bandwidths share the same sites. The other extreme benchmark is that the central controller manages all the control functionalities and handles the requests on the fly.

Results

The described SDN framework and schemes were simulated in the wireless backhaul network introduced in Section 3.3 and 3.4.1.1. The simulation lasts for 10,000s, and the measurements are taken after 100 s when the system is relatively stable. Each data points are obtained by averaging 50 different simulations with randomly generated user location and traffic. Further parameters used in the simulation are listed in Table 3.1.

Table 3.1: Simulation Parameters

Parameter	***Value***
MNd max antenna gain	17 dBi
Carrier frequency	3.5 GHz
ENd max antenna gain	20 dBi
SINR threshold	1.8 dB
Transmit power of ENd	5 W
Number of users	400 users for each SP
T_s for QoS guarantee scenario	1 s
Thr_{thres}	1.2 Mbit/s/CH
TL_{thres}	15 channels
R_T	60 channels
A_{pre}	5 channels

Figure 3.6 shows the performance of the 3 multi-tenant resource sharing schemes enabled by different controllers. In order to evaluate the performance of the described control approaches in the scenario where the service providers have different capacity requirements, a relatively high user arrival rate $\lambda = 0.2$ is used for SP1, and a relatively low value $\lambda = 0.05$ is chosen for SP2 and SP3. For the no control fixed allocation approach, SP1 experiences extremely congested traffic and therefore high blocking probability, despite the fact that there are available resources within the local area. The two-tier controller scheme performs significantly better than the no control approach in terms of the blocking probability (approximately 20-21% lower) due to the efficient utilisation of the spectrum. Compared to the all-centralised scheme, the proposed scheme only experiences marginal performance deterioration, and yet delivers a more scalable control plane solution. With a shorter time resolution $T_s = 1\ s$, the performance of the proposed scheme approximates the all-centralised one better. However, there is no notable performance gain introduced by this shorter time resolution compared with that of $T_s = 20\ s$ due to the relative static offered traffic load in the system. In a more dynamic scenario where the traffic load fluctuates rapidly over time and space, a short time resolution can be beneficial to

improving the resource utilisation. The time resolution also needs to be carefully controlled based on the scalability and performance requirements of the system.

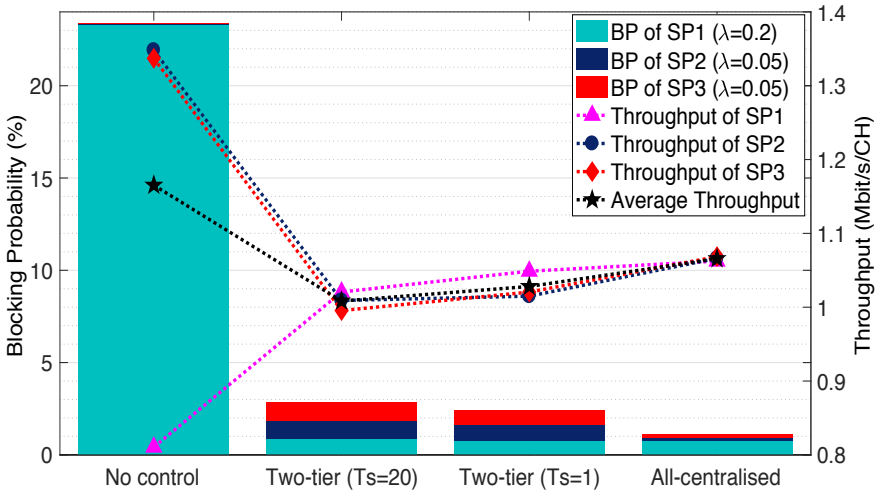

Figure 3.6: Blocking probability (%, left axis) and perceived throughput (Mbit/s/channel, right axis) against schemes

3.4.1.3 Multi-tenant Resource Sharing – QoS Guarantee

Since 5G needs to accommodate use cases with different performance requirements, in this subsection, the proposed two-tier SDN approach is applied to a backhaul network with service providers having different QoS requirements. As an early work of this thesis, we assume SP1 requires an average guaranteed user perceived throughput $\mathrm{Thr}_{\mathrm{thres}}$, whereas SP2 and SP3 agree on best-effort services. The central controller can obtain the network usage statistics including user perceived throughput and traffic load, and the locations of the ENds and MNds are fixed, therefore the required user perceived throughput of SP1 can be mapped to a maximum traffic load threshold TL_{thres}, above which the required user perceived throughput of SP1 cannot be guaranteed. The resource sharing algorithm which performs at each time interval T_s or upon a Packet-In message contains four entries

including different traffic load levels as well as resource request from a new service provider. They are explained in the following:

1) *Low traffic demand, i.e.* $\sum_{j=1}^{3} D_{SP}(j) \leq TL_{thres}$. The resource provision $A_{SP}(i), (i = 1,2,3)$ for a SP at one MNd is proportional to its estimated demand $D_{SP}(i)$, and can be calculated using equation (3.4).

2) *Medium traffic demand, i.e.* $\sum_{j=1}^{3} D_{SP}(j) > TL_{thres}$ && $D_{SP}(1) \leq TL_{thres}$. In this situation, the overall resource demand is higher than the maximum traffic the system can handle to maintain the acceptable QoS level for SP1. Meanwhile, there are still available resources that can be used by SP2 and SP3 after the requirement of SP1 is fulfilled. Hence, $A_{SP}(i)$ can be calculated by:

$$A_{SP}(i) = \begin{cases} \lfloor \frac{D_{SP}(i)}{TL_{thres}} R_T \rfloor, & i = 1 \\ \lfloor \left(R_T - A_{SP}(1)\right) \frac{D_{SP}(i)}{\sum_{j=2}^{3} D_{SP}(j)} \rfloor, & i \neq 1 \end{cases} \tag{3.5}$$

In other words, the central controller allocates the remaining channels which will not fail the QoS of SP1 to the other two SPs proportional to their demands.

3) *High traffic demand, i.e.* $D_{SP}(1) \geq TL_{thres}$. The requests from SP2 and SP3 will be blocked, since the system is hardly able to provide the agreed user perceived throughput for the public safety provider SP1.

$$A_{SP}(i) = \begin{cases} R_T, & i = 1 \\ 0, & i \neq 1 \end{cases} \tag{3.6}$$

4) *A new service provider requires access.* The switch will immediately forward a Packet-In message to the central controller, and the central controller decides

whether the request can be accepted. The resource allocation of the new SP is determined by:

$$A_{SP}(new) = \min\left(R_T - A_{SP}(1) - \sum D_{SP\neq 1}, A_{pre}\right) \quad (3.7)$$

where A_{pre} is a pre-defined value used to control the amount of resources that can be assigned to the new SP.

Results

The described scheme was simulated in the wireless backhaul network introduced in Section 3.3 and 3.4.1.1, with the parameters listed in Table 3.1. Figure 3.7 shows the number of demanded and actual allocated channels of the three service providers. Figure 3.8 and Figure 3.9 show the performance comparisons of the three operators in different time periods. During 100 – 700 s, only SP1 and SP2 are in the system. The number of allocated channels for each service providers are proportional to their demand due to the low level of traffic load. The performance of the two service providers is approximately the same in terms of blocking probability and perceived throughput. A new service provider SP3 joins the network from 700 s. Although the estimated demand of SP1 is lower than that of SP2 and SP3 during 700-1300 s, it receives more resource provision than the other two service providers because of the QoS protection. As a result, the performance of SP2 and SP3 deteriorate rapidly while the throughput of SP1 remains above 1.2 Mbit/s/channel. As the estimated demand of SP1 further increases, the system will not allocate any resources to the other two service providers to maintain the QoS for SP1.

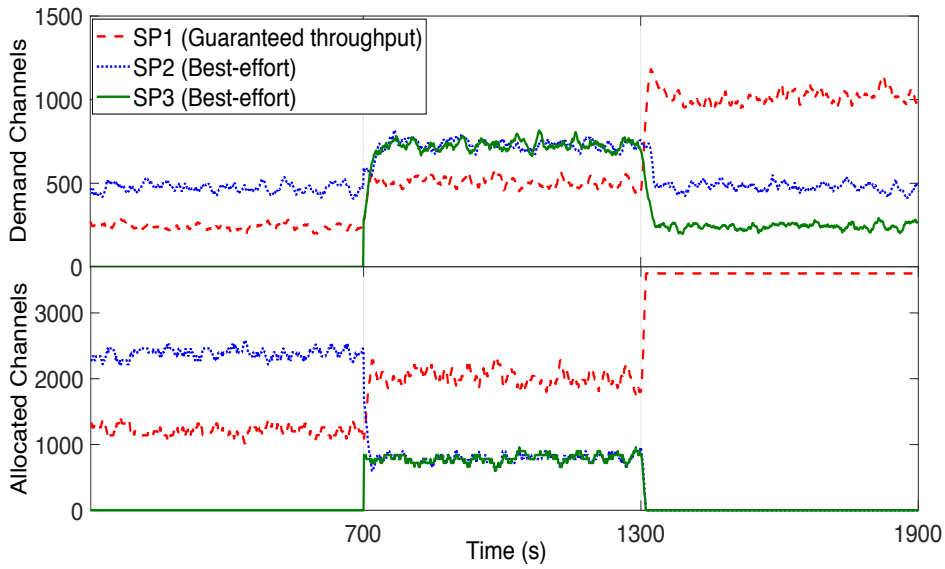

Figure 3.7: Number of demanded/allocated channels against time (s)

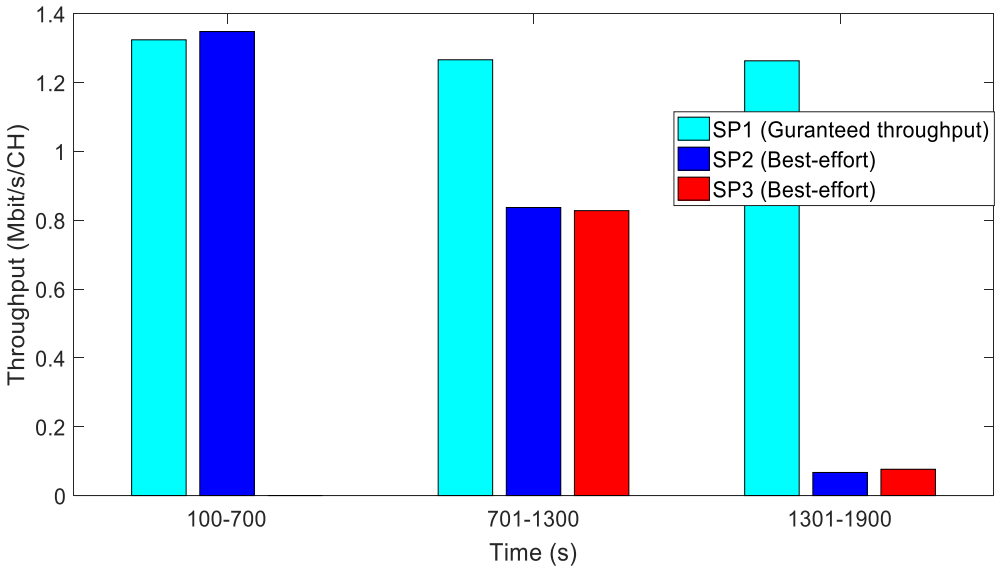

Figure 3.8: Average perceived throughput (Mbit/s/channel) against time (s)

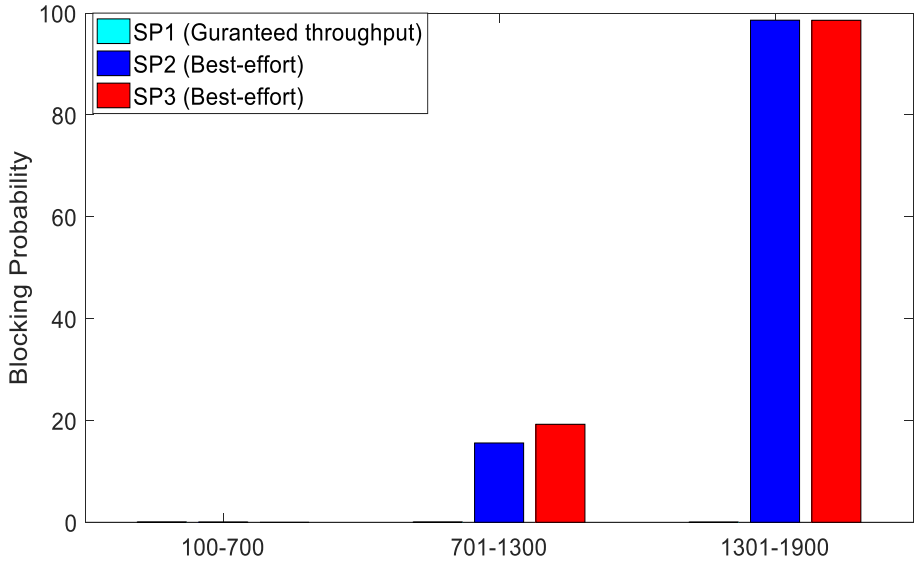

Figure 3.9: Blocking probability (%) against time (s)

3.4.2 SDN Opportunities for Green Wireless Backhaul Network Topology Management

As discussed in Section 2.3.3, dense network deployment is characterised by high ratio of peak to average mobile data usage and it is more cost-effective than a macro-cell architecture only in high traffic scenarios. Despite the extensive efforts that have been put into energy-efficient topology management strategies in wireless access networks such as [47] [78], relatively little attention has been paid to backhaul energy consumption management aspects. In this subsection, the SDN framework is applied to the backhaul network based on BuNGee architecture to facilitate a topology management strategy which can exploit the backhaul diversity in order to reduce the energy consumption of the backhaul network in low to medium traffic scenarios. We consider a network that is equipped with backhaul link diversity when the ANds in the system are able to connect to different available backhaul links.

3.4.2.1 Deployment Scenario

In order to allow the use of a distant ENd and exploit the backhaul diversity, the network model used is similar to an extended version of Manhattan grid topology proposed in the BuNGee Project as shown in Figure 3.10 [34]. There are 13 ENds, 244 ANds and their associated MNds serving a 2.25×2.25 km^2 area. With the steerable antenna introduced in Section 3.3.2, the beam on each MNd antenna can be reconfigured as one to two beams. One of the beams is directed to the ENds in their own cell (home ENd) and the other one can be oriented to the direction of the beam belonging to another ENd on demand. The wireless connection between a home ENd to its associated MNd is called a main link and the connection between an available ENd from other cell to the MNd concerned is called an alternative link. In the conventional BuNGee architecture on the other hand, the MNds are only directed towards their home ENds.

Assuming the same geographical layout as Figure 3.10, Figure 3.11 shows an example of a MNd using an alternative link. Here, the MNd is served by an alternative link and the main link is deactivated due to a low traffic load.

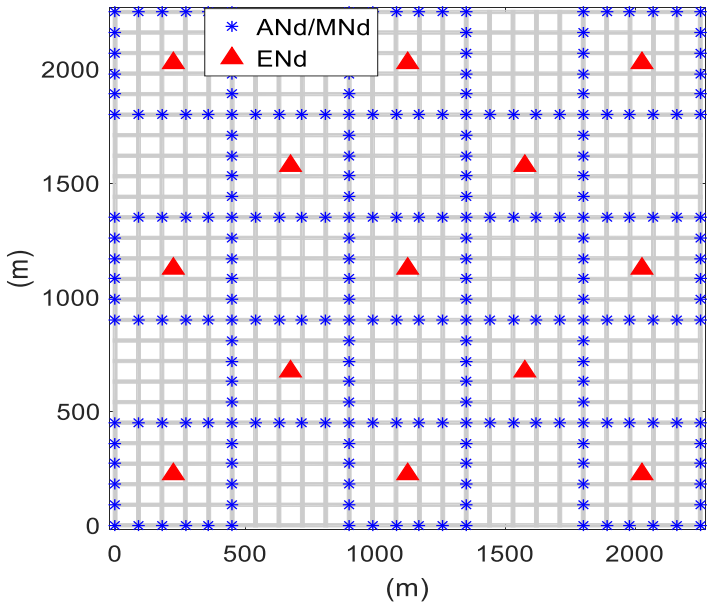

Figure 3.10: Network architecture

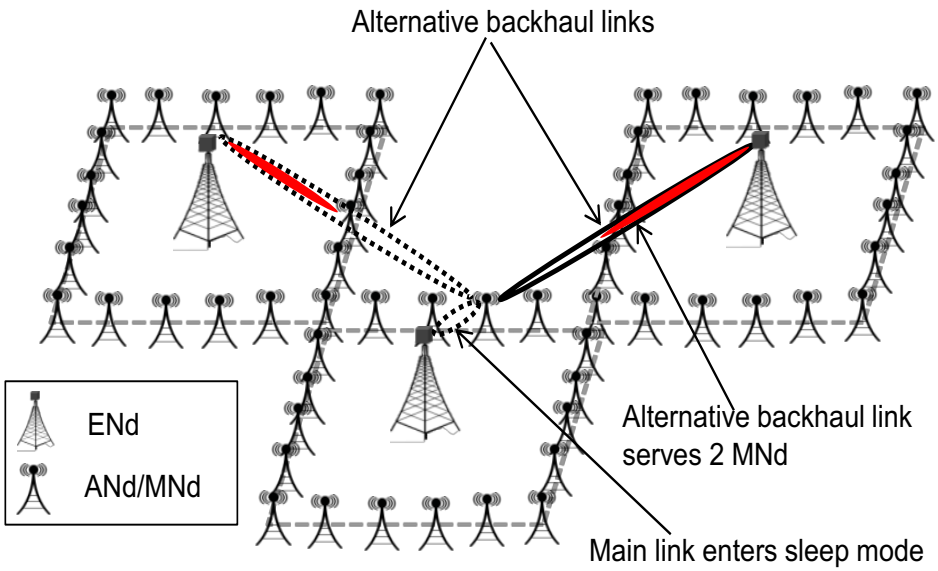

Figure 3.11: An example of a MNd using an alternative backhaul link

3.4.2.2 Backhaul Link Scheduling

In order to minimise the backhaul network energy consumption, a backhaul link selection algorithm, which concentrates the distributed traffic on fewer beams within the network and puts the idle beams into sleep mode is developed. Using the SDN framework can provide an effective offloading evaluation process as well as reducing the need of handover.

Since the position of the nodes are practically always known, the central controller is preconfigured with a list of candidate backhaul links for each MNd. Periodically, the network status information including active backhaul links, traffic load, link quality and resource utilisation is forwarded to the central controller. The central controller then runs the topology management application and carries out an evaluation of all the beams of the ENds to decide whether a certain MNd attached to an ENd should be offloaded to another backhaul link. We denote T_s=*10s* as the minimum time a MNd has to stay on a backhaul link. This parameter can be adjusted in order to alleviate the problem of frequently switching backhaul links from one to

another. An ENd may offload a MNd to another backhaul link in the following two situations:

- Offload a MNd from its main link to an alternative link in order to deactivate a beam when the average number of active users on the main link over T_s is below the low traffic threshold and there are no other MNds from other cells attached to the concerned beam.

- Offload a MNd from its alternative link to another alternative link or its main link (if no alternative link is available) when the average number of the active users on the concerned beam and either of its two adjacent beams is above the threshold of $(1 - capacity\ margin)$ which leaves a safety margin for the future file transferring request from the local MNds. Note that the total average number of active users of the concerned beam and either of its two adjacent beams are evaluated respectively since an approximate beam frequency reuse of two can be achieved using the multi-beam antennas in the BuNGee project [79].

After evaluation, the central controller checks the traffic load on the candidate backhaul links for the MNd which require offloading, and chooses the alternative link which has the maximum signal strength, and the active users on the link and either of its two adjacent beams are below the threshold of $(1 - low\ traffic\ threshold - capacity\ margin)$. If no such link exists, the MNd will be offloaded to its main link. The central controller also runs the reconfiguration application to reconfigure the beams on the MNds pointing to their serving ENds. During each time interval, the local controllers coordinate and assign the available resources to individual flows in the network on event basis.

3.4.2.3 Backhaul Energy Consumption Model

Since the increase of energy consumption of the access network caused by the antenna configuration change of MNds is negligible, only the energy consumption of ENds is evaluated, as the energy consumption of the access networks remain

approximately constant irrespective of the schemes used. To quantify the energy savings of backhaul networks, it is necessary to use a suitable energy model. The EARTH project has provided a power consumption model for various BS types [47]. For the backhaul architecture we use, where highly directional array antennas are equipped on the ENds, it is similar to the scenario where multiple Remote Radio Heads (RRHs) are connected to a common central entity we introduced in Section 2.3.3. Therefore, the base station energy consumption models described in EARTH project are used, and the parameters of RRH are chosen to calculate the energy consumption. In this case, the feeder and active cooling loss can be neglected [47]. The energy consumed by the ENds at the maximum load can be calculated by:

$$P_{in} = N_{TRX} \cdot \frac{\frac{P_{out}}{\eta_{PA}(1-\sigma_{feed})} + P_{RF} + P_{BB}}{(1-\sigma_{DC})(1-\sigma_{MS})(1-\sigma_{cool})} \tag{3.8}$$

where N_{TRX} is the number of TRX chains (transmit/receive antennas per site), P_{out} is the power measured at the input of the antenna element in Watts, P_{RF} is the power consumed in the small-signal RF TRX module, P_{BB} is the power consumed in the BBU, η_{PA} is the efficiency of the power amplifier (PA). σ_{feed},σ_{DC}, σ_{MS} and σ_{cool} are the loss factors of antenna feeder, DC-DC power supply, main supply (MS) and active cooling respectively. In our model, each beam of the ENd antenna can be considered as a RRH TRX chain.

A linear approximation of the BS power model has also been formulated in [47] to calculate the energy consumption at variable traffic loads:

$$P_{in} = \begin{cases} N_{TRX} \cdot (P_0 + \Delta_P P_{out}), & 0 < P_{out} \leq P_{max} \\ N_{TRX} \cdot P_{sleep}, & P_{out} = 0 \end{cases} \tag{3.9}$$

where P_{max} is the maximum RF output power at maximum load, P_0 is the linear model parameter to represent power consumption at the zero RF output power (estimated) and Δ_P is the slope of the load dependent power consumption. The sleep mode energy consumption P_{sleep} is introduced for future BSs with fast deactivation capability.

Here, the input power for one beam of an antenna on an ENd is a fixed 15 W which makes the power consumption of a TRX chain at variable traffic loads and maximum load, approximately the same level as in the active mode [79]. We choose to use (3.8) to calculate the energy consumption in the active mode and (3.9) to calculate the energy consumption in the sleep mode.

3.4.2.4 Results

The developed topology management was simulated in the network we described in Section 3.3 and 3.4.2.1. In this section, we compare the performance of the developed topology management scheme with the existing BuNGee architecture without the backhaul link diversity capability to illustrate the potential energy savings from backhaul diversity in low traffic scenarios. There are 2000 users randomly distributed across the service area. Parameters used in the simulation are listed in Table 3.1 and Table 3.2. Fast deactivation of ENd's TRX chains is assumed in all the schemes, i.e. the beams which have no data to transmit will enter sleep mode.

Table 3.2 Simulation Parameters

Parameter	***Value***
P_{RF}	6.8 W
P_{BB}	29.6 W
η_{PA}	31.1%
σ_{DC}	7.5%
σ_{MS}	9.0%
P_{sleep}	56 W
P_{max}	20 W
Low traffic threshold	9 (30% of the capacity)
Capacity margin	3 (10% of the capacity)

Figure 3.12 shows the energy savings of the backhaul link diversity scheme with different numbers of maximum allowable alternative link requests and the existing BuNGee architecture only with sleep mode enabled scheme. Firstly, both of the schemes can achieve relatively high energy savings from the sleep mode at low traffic loads, and the energy reduction reaches zero at high traffic loads since all the links need to be active to deal with the traffic demands. Secondly, the backhaul link diversity scheme can achieve much higher energy savings (35% maximum) by allowing more links to enter sleep mode at low to medium traffic loads. Moreover, further exploitation of the backhaul link diversity by increasing the number of maximum allowable alternative links can only provide marginal benefits on energy saving (the scheme with 10 alternative links is only 3% higher than that with 3 alternative links), but it will lead to increasing the size of the flow table for each MNd.

Figure 3.13 and Figure 3.14 show the blocking probability and perceived throughput per channel of the schemes respectively. The performance of the schemes with backhaul link diversity is slightly worse than the one without backhaul link diversity, and the ones with a higher number of higher number of alternative links have worse performance. This is most obvious at medium traffic loads as the interference is also centralised to fewer beams in backhaul link diversity schemes. Such a decrease in performance is marginal compared with the level of energy savings obtained. The blocking probability of the scheme with 10 allowed alternative links increases almost linearly with the traffic load up to medium level. As the traffic further increases, the performance gets closer to the other schemes due to the less use of alternative links.

Performance of the system has also been tested with different low traffic thresholds (20%-40% of the total capacity), capacity margins (10%-20% of the total capacity) and T_s(0.01 s, 10 s, 100 s). Generally, the higher the low traffic thresholds and the lower the capacity margins the more energy reduction gain and worse performance achieved. Note that when T_s is very small, the backhaul link switching frequency of the MNds is much higher than that with a relatively larger value without providing much energy saving gain, especially when the traffic is at medium level.

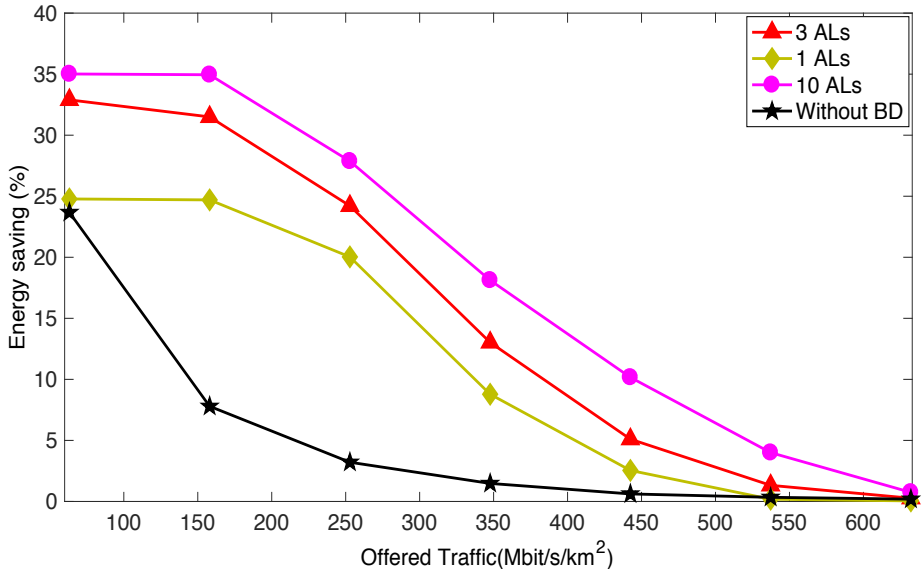

Figure 3.12: Energy Saving (%) against Offered Traffic (Mbit/s/km^2)

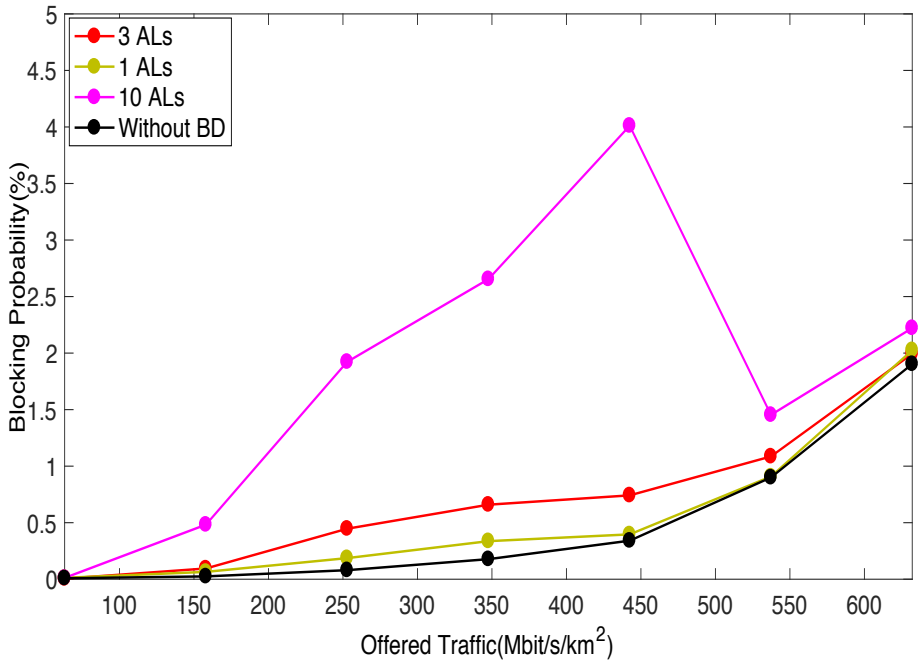

Figure 3.13: Blocking Probability (%) against Offered Traffic (Mbit/s/km^2)

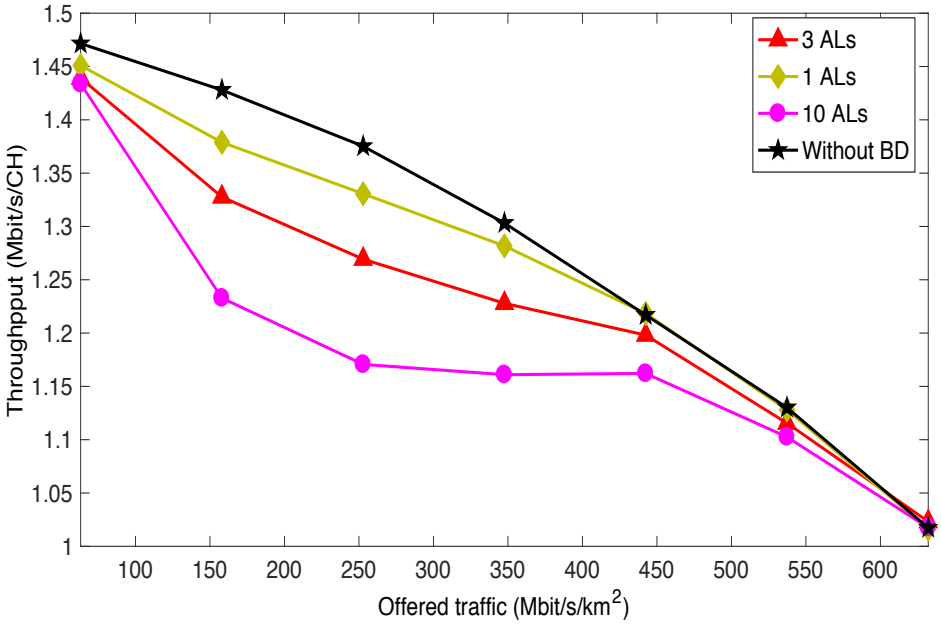

Figure 3.14: Perceived Throughput (Mbit/s/CH) against Offered Traffic (Mbit/s/km^2)

3.5 Conclusion

In this chapter, a novel SDN architecture has been proposed to facilitate the flexible deployment of a 5G dense deployed wireless backhaul network. The architecture contains a two-tier controller in a hierarchical setup aiming to offload some of the control functionalities from the central controller to a collection of logically distributed dynamically configured local controllers in order to balance the trade-off between scalability and system performance.

A multi-tenancy dynamic resource sharing algorithm based on the proposed SDN architecture were introduced as a case study. The results demonstrated the dynamic resource sharing scheme enabled by the proposed SDN architecture for multi-tenancy cross-infrastructure scenarios can achieve good performance while maintaining a scalable control plane compared to the all-centralised architecture. The trade-off between scalability and performance of the system can be controlled by

adjusting the time resolution. The architecture is capable of providing fairness in terms of resource allocation and supporting QoS guarantee functionality.

As cost and energy efficiency will also be important factors in 5G backhaul networks, the proposed SDN framework has been used in the context of backhaul link diversity. Periodically, the central controller re-schedules the backhaul links for the access networks allowing the unused components in the idle links to be switched off. It significantly reduces the energy consumption by up to 35% at low to medium traffic loads by concentrating distributed traffic on fewer backhaul links. It has been found that the energy savings increase with the number of allowed alternative links up to a maximum of 3, after which it tends to saturate. In addition, the simulation results revealed the trade-off between energy saving and system performance including blocking probability and throughput.

As discussed in the case studies, SDN paradigm provides a global view of the entire network and can facilitate network management tasks without a device-centric configuration. Such a decoupling of control and data plane results in improved network performance in terms of network management and resource utilisation. SDN is one of the key ingredients of future backhaul architectures to elastically support network functional demands.

Chapter 4. Wireless Backhaul Deployments for Ultra-Dense Outdoor Small Cells

Contents

4.1 Introduction

The wireless backhaul deployment strategy and network management technique such as SDN investigated in Chapter 3 have been shown to be a promising approach to increasing resource utilisation and energy efficiency. In addition to providing cost-efficient broadband service everywhere, the backhaul architectures in 5G should also be able to support a diverse range of new use cases, one of them being the emerging deployment of ultra-dense small cells. Although ultra-dense small cells are mainly deployed for indoor environments such as offices and shopping malls, operators have also started to investigate outdoor small cell deployments to accommodate the increase in outdoor traffic [80]. The reason is that people tend to access internet more

and more on the go: whether it is streaming HD videos or playing online games on a bus or using mobile banking on a street. It is becoming a requirement that a similar data experience should be achieved for both indoor and outdoor users. However, currently the satisfaction with the outdoor connectivity experience is significantly lower than that of indoor. In order to achieve high flexibility and capacity in future outdoor ultra-dense small cells while accommodating various use cases and deployment constraints, a number of novel backhaul architecture options are proposed in the following chapters.

In this chapter, the performance of two of the key technologies for 5G backhaul are compared: massive MIMO with conventional frequencies and mm-wave backhaul at the 71 – 76 GHz band. The purpose is to show the upper bound capacity that can be achieved by applying massive MIMO technology and access network load balancing techniques to a capacity constraint network band, as well as to present the performance improvement by using mm-wave backhaul.

4.2 Network Modelling

In order to demonstrate the viability of the backhaul architectures, the simulation model chosen needs to be appropriately dimensioned capturing the relevant details of a realistic outdoor small cell scenario. However, since it is impossible to cover every possible situation, a dense urban street canyon scenario suggested by Huawei GREENBACK project is used. It reflects a typical densely populated urban area, where thousands of people per km^2 are either walking or in moving vehicles using internet for data-hungry applications envisioned for 5G. The mobility model used enables to highlight the key challenges of designing a communication network for a dynamic vehicular environment including moving hot spot in buses and potential congested data traffic at the intersections. The goal of the project is to develop a high capacity and low latency wireless backhaul network solution that serves a dense 5G radio access network installed on street fixtures (e.g. street lights, building walls, and billboards).

Since 5G is expected to provide up to 1000 times higher area capacity than the current systems, the proposed backhaul architectures are tested against a wide range of offered area capacity (100 Gbps/km^2 – 1000 Gbps/km^2) in this thesis, whereas for the BuNGee architecture discussed in Chapter 3, a capacity density of 1 Gbps/km^2 was targeted. It is believed that the chosen test cases and parameters described in this section can be used to establish backhaul design requirements as well as compare the capabilities of different backhaul architectures. The detailed environment scenario and network modelling are described in the rest of this section.

4.2.1 Environment Scenario and Network Architecture

Figure 4.1 illustrates the environment scenario. Here, the small cell ANds are deployed on street light fixtures with a height of 8 m and inter-site distance of approximately 30 m to provide ultra-high capacity to the mobile users. A Manhattan grid topology with streets lined with high buildings is used in as depicted in Figure 4.2, where 459 users (density: 5000/km^2) are located across the outdoor area of three building blocks (block size: 150 m×150 m). The area is served by 72 ANds/MNds (density: 800/km^2) uniformly distributed along both sides of the streets aiming at providing a capacity density of up to 1000 Gbps/km^2.

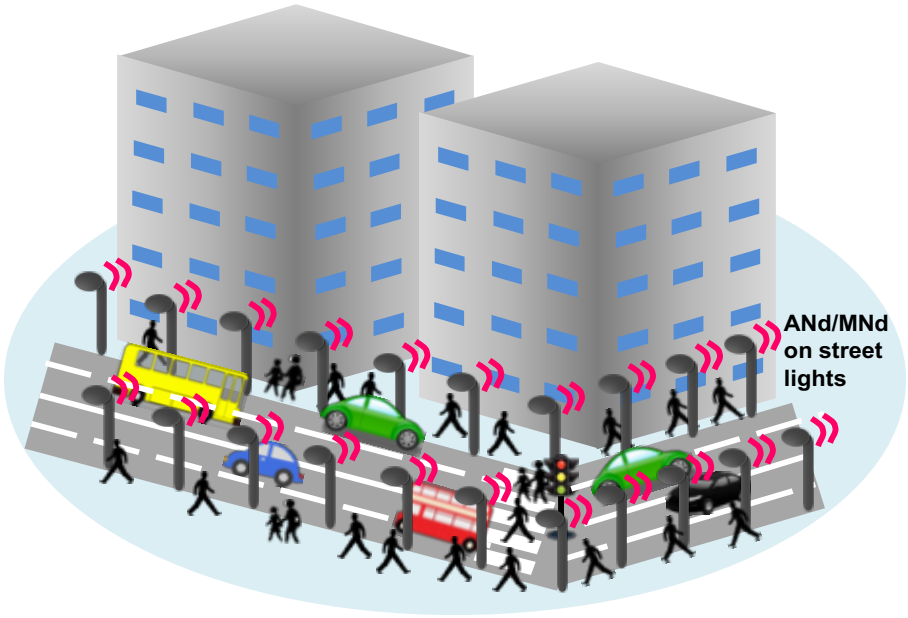

Figure 4.1: Urban street canyon with ultra-small cells scenario

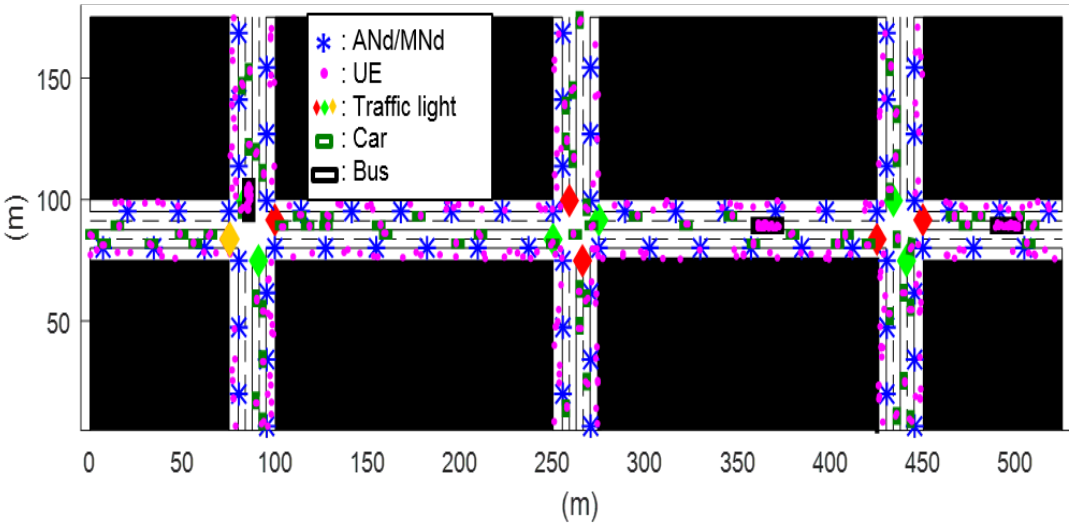

Figure 4.2: Urban ultra-small cell network architecture with traffic lights at the intersections

Vehicular communication in urban outdoor areas is characterized by diverse mobility patterns and unevenly distributed traffic. Hence three user groups are considered here: pedestrian users (50%), car users (30%, 2 users per car), and bus users (20%, 25 users per bus). Each street (width: 15 m) has four lanes (two in each direction) with cars

and buses driving with speed of 0-50 km/h. On the pavements (width: 5 m), pedestrians walk with speed of 0-6 km/h. Developing a microscopic mobility model is a complex research field in its own right, which involves defining car following model, lane changing model, and movement restrictions etc. [81] [82]. Hence a mobility simulation tool developed by Huawei, with which the mobility trace of each user can be extracted, is used in this thesis, and the traces are then fed into the network simulator. The mobility model parameters are listed in Table 4.1.

Initially, the status and the remaining time of traffic lights, the speed of pedestrians and vehicles, and their turning directions in the next crossing are randomly initialised. The locations of the users are updated every 0.1 s. Periodically, the tool checks whether the vehicles need a lane changing, an update of acceleration/deceleration, or stopping. A safety distance of 1 m is assumed, below which the following vehicle would stop moving. If a vehicle is in the wrong lane of turning (driving on the right), a lane changing function is performed, allowing the vehicle to change to the next lane.

The mobility traces including the geographic position of each user is recorded every 1 second. This allows for the estimation of UEs' candidate serving ANds which can be based on signal quality measurement (RSRP/RSRQ) reported. In order to avoid border effect and obtain statistically acceptable results from a limited simulated area, the UEs are assumed to be vertically and horizontally wrapped around at the edges of the service area, meaning that if a UE leaves the service area from one edge, it will re-enter the service area from the opposite edge. This makes the mobility pattern in the edge cells the same condition as that in the middle cells.

Table 4.1: Mobility parameters

Types	Description	
Car	Length	5 m
	Width	2 m
	Maximum turning speed	15 km/h
	Maximum acceleration	10 km/h/s
	Turning probability(left/right)	25%
Bus	Length	13.5 m
	Width	2.5 m
	Maximum acceleration	5 km/h/s
	Turning probability(left/right)	0
Pedestrian	Maximum acceleration	1.2 km/h/s
	Turning probability(left/right)	25%
Traffic light	Red/green light duration	100 s
	Yellow light duration	5 s

4.2.2 Communications Traffic Models

Measurements and analysis have shown that the traffic inter-arrival times and file sizes exhibit a heavy tail behaviour in a network which has unexpected and dynamic traffic demand (e.g. multimedia traffic, moving hot spots), providing bursts of traffic arrivals, long inter-arrival times and large files with finite probability [83]. This load statistics is typically represented by the Pareto distribution. In practical applications

however, generally there is an upper bound of the inter-arrival times and file sizes. A truncated Pareto distribution with an upper limit cut-off value has been considered to be a better fit to the internet traffic [84]. Hence, it is used in this thesis to model the outdoor traffic. The cumulative distribution function (CDF) is given by [85]:

$$F(x) = \frac{1-x_{min}^{\alpha}x^{-\alpha}}{1-(\frac{x_{min}}{x_{max}})^{\alpha}}, \qquad x_{min} \leq x \leq x_{max} \tag{4.1}$$

where α is the shape parameter, x_{min} and x_{max} are the minimum and maximum possible value of x respectively. The mean value of the truncated Pareto distribution is:

$$E(X) = \frac{x_{min}^{\alpha}}{1-\left(\frac{x_{min}}{x_{max}}\right)^{\alpha}}\left(\frac{\alpha}{\alpha-1}\right)\left(\frac{1}{x_{min}^{\alpha-1}} - \frac{1}{x_{max}^{\alpha-1}}\right), \ \alpha \neq 1 \tag{4.2}$$

Given a random variable U which is uniformly distributed on $[0,1]$, then the random samples of the inter-arrival times and file sizes which have the CDF of (4.1) can be generated by applying inverse transform sampling [86]:

$$x = \frac{x_{min}}{[1-U+U(\frac{x_{min}}{x_{max}})^{\alpha}]^{\frac{1}{\alpha}}} \tag{4.3}$$

It is expected that mobile videos will account for over 78% of the total mobile traffic by 2021 [4]. In order to evaluate the capability of backhaul deployments of supporting these resource hungry applications for outdoor users, two different traffic types are modelled, including on-demand video streaming and real-time services.

The detailed traffic model parameters are listed in Table 4.2, which are based on the streaming requirements of YouTube and Netflix [87] [88], and the recommendations suggested in [89].

Table 4.2: Traffic Model Parameters [87] [88] [89] [5]

Traffic Type	Description	
On-demand video	Encoding rate	4K: 20 Mbps; 8K: 80 Mbps
	Mean video length	300 s (max 1000 s)
	Buffer size	40 s of a video
	Proportion of different video qualities	4K: 25%; 8K: 25%
	Speed factor in steady state	1.25
Real-time service	Encoding rate/ playback rate	4K: 20 Mbps; 8K: 80 Mbps
	Mean video length	300 s (max 1000 s)
	Proportion of different video qualities	4K: 25%; 8K: 25%

4.2.3 Resource Allocation Schemes

For each backhaul link, the mixed traffic flows arrive at random times and share the whole bandwidth. It is critical to meet the minimal QoS requirements of different traffic classes. The resource allocation scheme described in this section is based on LTE-Advanced Pro which aims at pushing LTE capabilities closer towards 5G. In 3GPP TS 23.203, 15 QoS Class Identifiers (QCIs) and Allocation and Retention Priority (ARPs) are standardised which can also be pre-configured by the operators [90]. The QCI values are used to determine the packet forwarding treatment

including required resource type (GBR or non-GBR), packet delay budget (ranging from 50 ms to 300 ms), packet error loss rate (ranging from 10^{-6} to 10^{-2}) and priority level. For example, real-time services such as video conferencing and mobile cloud gaming have very stringent requirements for data rate and latency, hence GBR bearers should be assigned to them. On the other hand, on-demand video streaming has a relatively high tolerance for delay, and can be carried by non-GBR bearers. ARP which defines the priority level, the pre-emption capability and the pre-emption vulnerability can be used in admission and congestion control [90]. In the situation when there is a lack of resource in the network, a service of a high priority can be configured to pre-empt the resource allocated to the service of a low priority whose ARP information element (IE) 'pre-emption vulnerability' is 'pre-emptable'. In this thesis, the real-time services are considered as high priority. The low priority is given to the on-demand video. Based on the above information and the traffic characteristics, the resource allocation and QoS adaption for each of the traffic types are described as follows:

On-demand videos

An on-demand video is served with dynamic data rate based on its service type and the network condition. An on-demand video transmission begins with a buffering phase, where the video is downloaded with the highest data rate possible. After the buffer is full (40 seconds of the video), the transmission enters a steady phase with ON-OFF cycles in order to limit the data rate. A constant average data rate $r_m^{O\_ste} = f_m r_m^{O\_req}$ $(m = 1, \dots, M)$ can be assumed as shown in Figure 4.3 [5]. Here, f_m is called speed factor which is a factor of the video encoding rate $r_m^{O\_req}$. $f_m \geq 1$ is required in order to guarantee a smooth video playback.

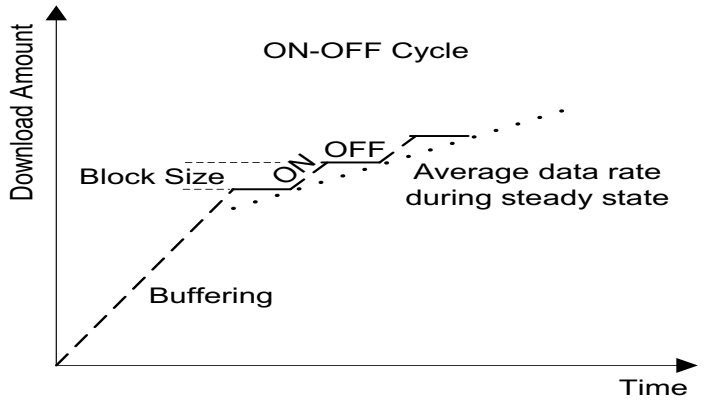

Figure 4.3: Phases of on-demand video streaming

If there are sufficient resources on the considered backhaul link, we assign the resources $r_m^{O\_allocate}$ to the m_{th} on-demand video request according to the following:

$$r_m^{O\_allocate} = \begin{cases} r_m^{O\_ste} + r_m^{O\_ex}, & m \in \boldsymbol{F} \\ r_m^{O\_ste}, & m \in \boldsymbol{B} \end{cases} \tag{4.4}$$

where $\boldsymbol{F}$ and $\boldsymbol{B}$ denote the index sets of video streaming flows with their buffers full and buffers not full respectively. $r_m^{O\_ex}$ represents the amount of remaining resources on the link which can be used by the m_{th} on-demand video in its buffering phase, and can be expressed as:

$$r_m^{O\_ex} = \frac{w_m}{\sum_{j \in \boldsymbol{F}} w_j} \left(R^{ava\_R} - \sum_{g \in \boldsymbol{B}} r_g^{O\_ste}\right) \tag{4.5}$$

where w_m denotes the weight of flow m which is equal to $r_m^{O\_req}$ to ensure that the excess resources are allocated proportional to the requested demand. R^{ava_R} is the

amount of remaining resources after allocating resources to the real-time services, which is explained in the real-time service section.

In case of congestion, a service can pre-empt resources from other services with the same or lower priorities. Hence here, no blocking or dropping is assumed for the on-demand videos, and a new on-demand video can pre-empt resources from the ongoing on-demand traffic on the link. The resources are re-scheduled according to the weights of each flow, and the allocated resources r_m^O can be obtained by:

$$r_m^{O\_allocate} = \frac{w_m}{\sum_{j=1}^{M} w_j} R^{ava\_R} \tag{4.6}$$

Real-time services

Real-time services are served with their requested resources $r_k^R (k = 1, \dots, K)$. The ENd checks the resource usage of the ongoing K GBR services on the considered backhaul link to determine whether it is possible to admit the $(K+1)_{th}$ real-time service connection request. For a backhaul link with a total amount of resource of R^{tot}, the amount of resources R^{ava_R} available or can be pre-empted from the on-demand videos can be calculated by:

$$R^{ava\_R} = R^{tot} - \sum_{k=1}^{K} r_k^R \tag{4.7}$$

The request is blocked if $R^{ava\_R} < r_{k+1}^R$, meaning that there is insufficient resource that can be pre-empted from the services with lower priorities to support the connection request. However in this way, it is possible that the real-time services occupy significantly more if not all the resources on a congested link than that of the on-demand videos. In order to ensure a fair resource allocation between the two

traffic types, a factor representing the deterioration level of the allocated resources compared to the requested resources of the on-demand videos is incorporated into the decision process to block or drop some of the real-time services, and can be expressed as follows:

$$\alpha = \frac{\sum_{j=1}^{M} r_j^{O\_req} - \sum_{m=1}^{M} r_m^{O\_allocate}}{\sum_{j=1}^{M} r_j^{O\_req}} \tag{4.8}$$

A real-time service request is accepted if $R^{ava\_R} \geq r_{k+1}^{R}$, and there are sufficient resources on the considered link that guarantee smooth playback for the on-demand videos, i.e. $\alpha < 0$. Otherwise if $\alpha \geq 0$, new/handover real-time connection requests are randomly blocked/dropped with the probability of:

$$\gamma = \alpha\beta \tag{4.9}$$

where β $(0 \leq \beta \leq 1)$ is an adjustable factor used to control the priority level of real-time services in terms of blocking and dropping probability, and a smaller value indicates that a real-time request is less likely to be blocked/dropped and thus increased priority. Generally, dropping an ongoing service during a handover process is far more frustrating than blocking a new service, hence a relatively smaller value ($\beta = 0.2$) and a larger value ($\beta = 0.8$) are used for dropping and blocking decision making respectively. The values chosen only affect the relative difficulty in terms of real-time service admission, and they can be changed without the loss of generality.

The resource allocation scheme for mixed traffic is summarised as follows:

Algorithm: Resource allocation and QoS adaption for mixed traffic types

Calculate R^{ava_R} using Equation (4.7)
if new flow is real-time service **then**
  if $R^{ava\_R} < r_{K+1}^{R}$ **then**
    Block/drop the new flow
    return
  else
    Calculate α
    if $\alpha \geq 0$ **then**
      Calculate γ using Equation (4.9)
      Generate a random number x from a uniform distribution over [0,1]
      if $x \leq \gamma$ **then**
        Block/drop the new flow
        return
      end if
      Accept the new flow; assign the requested resource r_{K+1}^{R}
      Update $R^{ava\_R} \leftarrow R^{ava\_R} - r_{K+1}^{R}$
    end if
  end if
end if

Calculate aggregated resource requested for video streaming in the steady state: $R^{O\_ste} \leftarrow \left(\sum_{m=1}^{M} r_m^{O\_ste}\right)$
Streaming video flow weights: $w_m \leftarrow r_m^{O\_req}$
Find the index sets of flows with buffers full $\boldsymbol{F}$, and buffers not full $\boldsymbol{B}$
if $R^{ava\_R} > R^{O\_ste}$
  Assign resources using Equation (4.4) and (4.5)
else
  Assign resources using Equation (4.6)
end if

4.3 Choice of Technologies

As discussed in Section 2.3.1, improving the spectrum efficiency and resource utilisation are the main approaches to provide high capacity in sub-6 GHz bands since the available spectrum is scarce. Massive MIMO proposes to use much a larger number of antennas at the base station than its serving devices in order to achieve high spectrum efficiency. It provides high capacity to the underlying tier of small cells. Another promising solution is to use millimetre wave, which is intended to significantly improve system performance and user experience. In this section, the

performance of the two backhaul options are compared, and thus, the appropriate backhaul technology for the ultra-dense outdoor small cell network introduced in Section 4.2 is proposed and used in the rest of this thesis.

4.3.1 Massive MIMO Backhaul

Currently, a relative small bandwidth such as 10 MHz used in Chapter 3 is often allocated to operators. However, with some of the dense urban use cases in 5G requiring an area capacity of over 500 Gbps/km^2, it is envisioned that a wider bandwidth and more aggressive approaches improving spectral efficiency are needed. As introduced in Section 2.3.1, a maximum of 200 MHz bandwidth in the 3.5GHz band is available, and it is assumed in this chapter to provide higher data rates [31].

By utilising a large number of antennas at the base station and aggressive spatial multiplexing, a massive MIMO system is able to push the system closer to a noise limited environment, and drastically increase the spectral efficiency [32]. This is achieved by appropriately shaping the signals emitted from different antennas so that they add up constructively at the desired terminals and destructively at other locations. In [91], a 128 element massive MIMO system operating at 3.5GHz is shown to support a simultaneous transmission of 22 user streams in a single 20 MHz channel and provide a spectral efficiency of 79.4 bits/s/Hz. With the rapid ongoing development of massive MIMO technology in physical layer, it is highly relevant to examine the potential capacity capability that a massive MIMO system can achieve on a system level. Here, the upper bound performance of a massive MIMO backhaul system is investigated to show whether it is possible to use massive MIMO technology for an ultra-dense small cell network.

Assuming the same geographical layout as Figure 4.2, Figure 4.4 shows a possible backhaul network architecture with a centralised massive MIMO site on a rooftop. However, distributed antenna arrays across the geographical area can also be deployed. We assume that the intra cell interference can be mitigated, and the optimal max-min power control suggested in [32] which enables an equal SNR of 21 dB on

each of the backhaul link is applied to the system. The truncated Shannon Bound throughput estimation described in Chapter 3 is used here which upper bounds a fixed capacity of 900 Mbps for each massive MIMO site-MNd link. The aim here is to examine the capability of massive MIMO in serving an ultra-dense small cell network. Therefore, the physical layer aspects, for example the imperfections of the beamforming schemes and channel estimation errors are out of the scope of this thesis. However, it is expected that the performance of the massive MIMO backhaul approach presented later in this chapter may degrade when taking the physical layer implementation into consideration.

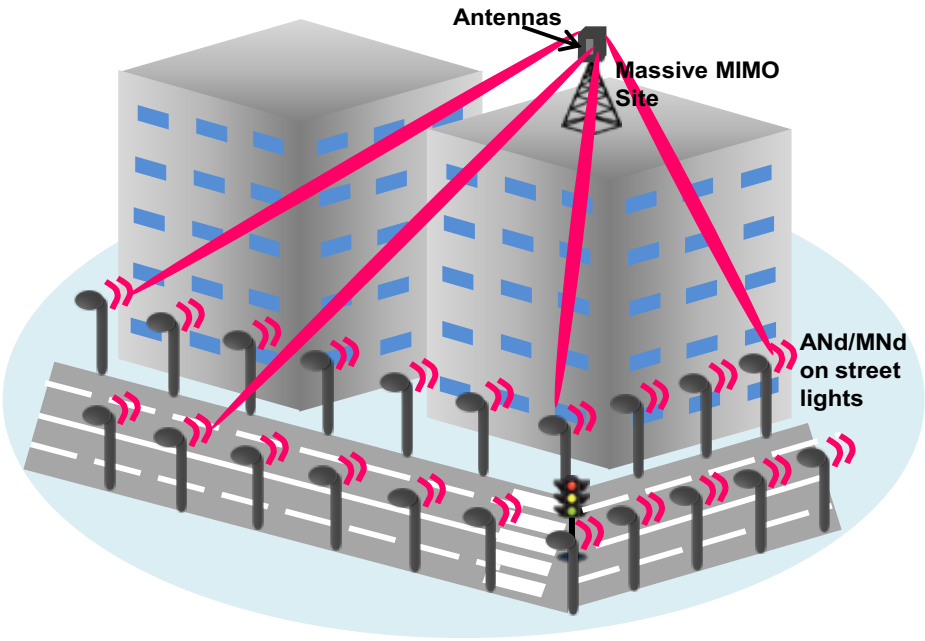

Figure 4.4: Massive MIMO network architecture

Bottleneck Issue and Access Network Load Balancing

Moving crowd users, especially bus users will generate temporary hot spots in the network. This will bring a significant burden to the local access and backhaul network with constraint capacity if the crowd users only connect to their nearest ANd which can potentially offer the best link quality. For example in Figure 4.5 (a), the

users on the bus can only connect to the two green ANds depending on their locations. Since each ANd has a dedicated backhaul link, the traffic loads on these two backhaul links will become extremely congested, whereas some of the nearby ANds and their corresponding backhaul links may still have available resources.

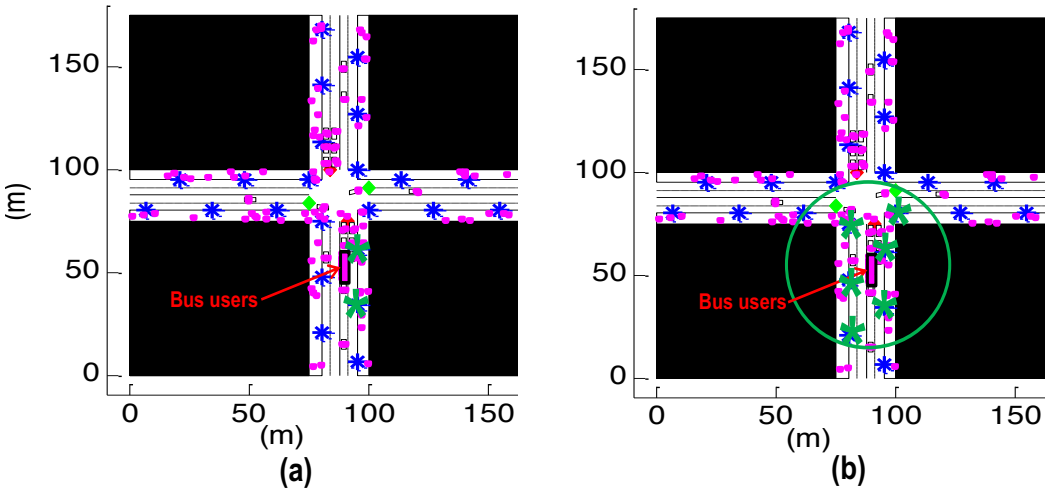

Figure 4.5: Example of bottleneck issue caused by temporary hot spot

In order to improve the utilisation of the available network resources, we apply a load balancing scheme which can dynamically allocate ANd to users depending on their locations, and current traffic loads on their associated backhaul links. As illustrated in Figure 4.5 (b), an ANd service range is defined to reflect the acceptable signal quality threshold for a UE. Due to the dense deployment, a UE may be within the serving range of several ANds and backhaul link options to the core network. Since the access network is not the main focus of this thesis, we assume advanced interference management techniques such as eICIC mentioned in Section 2.2.2 can be used, and a UE is able to connect to any ANd within their service area. The access network simply delivers the data from the serving ANd to the UEs. In this circumstance, the throughput and capacity are constrained by the backhaul network alone. When a user moves outside of its serving ANd's service area, a handover procedure will be triggered and it will connect to another ANd which has the lightest load within the connecting area. In Figure 4.5 (b), the users on the bus can be served

by up to 6 ANds with the load balancing scheme which can effectively alleviate the bottleneck problems brought about by the moving hot spots.

4.3.2 Millimetre Wave Backhaul

Another emerging state-of-art technology for alleviating the backhaul bottleneck problem is the use of mm-wave band. Mm-wave in bands 71 - 76 GHz and 81 - 86 GHz are very promising line-of-sight (LOS) wireless backhaul solutions for future 5G networks, as they can offer abundant spectrum, and therefore multi-gigabit data rates [92]. As discussed in Section 2.3, distance-dependent path loss, high attenuation by atmospheric gases and rain are not insurmountable challenges since the distance between the nodes considered here is typically a few tens of metres. The street canyons with high buildings in the network model also add benefits, such as interference reduction and higher frequency reuse. A possible mm-wave backhaul deployment scenario is depicted in Figure 4.6. The ENds are mounted on tall structures such as building walls which ensure high LOS probability of the backhaul links.

In accordance with the channelization regulation from ECC, it is assumed that each backhaul link uses a LOS 1 GHz bandwidth with a carrier frequency of 73.5 GHz [93]. In theory, it is also possible to use NLOS links by employing reflections from building walls and other objects. However, it is highly sensitive to node placement, antenna pattern and the local environment especially with the narrow beams. Hence, only LOS links are considered to deliver the backhaul connectivity. Electronically steerable antenna arrays are deployed at both the transmitting and receiving ends of a backhaul link with a 2 degree beamwidth [94] [95]. Beam steering allows beam alignment during the initial stage and is resilient to the misalignments due to wind. It will also allow the antenna beams to point at other nodes dynamically, without angular constraints. It is assumed that a maximum of 8 beams can be generated from the antenna module at the ENd end which is achievable based on the results demonstrated in MiWaveS project [96]. Although the maximum output power of E-band specified by ETSI is 30 dBmW, most of the commercial E-band backhaul

output power ranges from 10 – 25 dBmW [97]. Here, the ENds use a fixed transmit power of 16 dBmW.

The use of highly directional beams implies that both interference and multipath should be negligible, and the link is largely unaffected by its environment. Hence a simplified antenna model is used where only the main lobe with a fixed gain of 38 dB (assuming the antenna efficiency is 70%) is considered. Moreover, since the transmitters and receivers are fixed and in known locations, the path loss can be treated as deterministic rather than stochastic. Measurements showed that mm-wave in LOS environments has almost identical path loss as free space [38]. The exception is the effect of rain and fog: however these effects are small and can readily be allowed for by means of a fade margin. Therefore, the path loss can be calculated by:

$$PL = 20log_{10}(d) + 20log_{10}(f_c) + 32.45 + \gamma d \tag{4.10}$$

Where PL is the LOS path loss in dB, d is the distance in metres between an ENd and a MNd, $f_c = 73.5$ is the carrier frequency in GHz, and $\gamma = 0.015$ is the fade margin in dB/m which is consistent with the rain attenuation for heavy rainfall rate defined in ITU-R.P.838 [37]. The LOS propagation paths are also assumed to be vertically and horizontally wrapped around the edges of the service area in order to reduce border effects in the simulation and collect statistically acceptable results from all the cells. Moreover, for the purpose of avoiding beam overlapping, a shorter distance between a MNd and an ENd within 100 m is assumed. Assuming an 8 dB noise figure at the MNds and temperature of 290 K, then the noise floor is -166 dBm/Hz, which gives a SNR of 56.2 dB for a 100 m link with a bandwidth of 1 GHz.

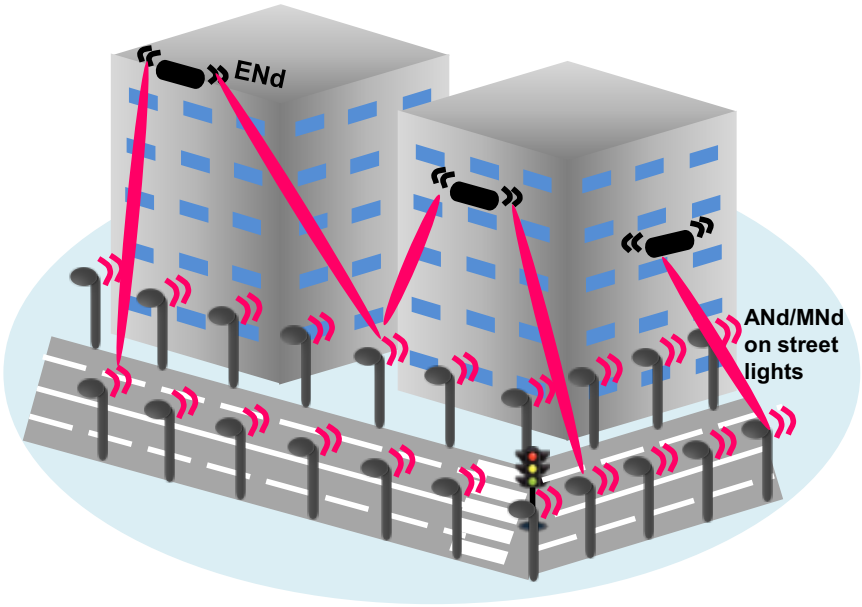

Figure 4.6: Backhaul deployment scenario: ENds installed on building walls

4.3.3 Results

The urban street canyon scenario with mixed traffic described in Section 4.2 is used to test the performance of the discussed backhaul deployment options for ultra-dense outdoor small cells. The simulation lasts 10,000 s for each traffic load, and the measurements are taken after 100 s when the system is relatively stable. The results are obtained by averaging 50 different simulations with randomly generated mobility traces and user traffic. To demonstrate the performance improvement of using mm-wave backhaul compared to the sub-6 GHz band approach, it is assumed that each MNd has one backhaul link connection to an ENd. More realistic ENd deployment options including the location, density and their topology management strategies will be presented in Chapter 5 and Chapter 6. For the massive MIMO scenario, the scheme without load balancing, where a user always connects to its closest ANd, is compared with the load balancing scheme over different ANd service ranges ranging from 25 m to 100 m which covers a low coverage range and a relative high coverage range of small cells. For the mm-wave backhaul scenario, in order to evaluate the benefits purely brought by using mm-wave, no access network load balancing is

assumed in this chapter. The access network aspects in terms of load balancing and handover control will be discussed in Chapter 6.

Average Performance

Figure 4.7 and Figure 4.8 show the system throughput and perceived throughput (8K on-demand videos shown, 4K on-demand videos follow the same trend) for massive MIMO backhaul and mm-wave backhaul. Both plots demonstrate significant performance improvement by using mm-wave backhaul due to the wide bandwidth it can offer: a 21-53% increase in perceived throughput is achieved compared to the massive MIMO approach with an ANd coverage of 100m at medium to high traffic loads (above 500 Gbps/km^2). Even the upper bound performance of the massive MIMO backhaul approach is not able to provide sufficient capacity due to the lack of resources. The upper bound perceived throughput achieved by using massive MIMO without load balancing decreases rapidly as the offered traffic increases. The deterioration actually starts at relatively low traffic load due to the uneven traffic loads. The average performance is improved with the lightest load balancing scheme. By increasing the coverage range of ANds, the congested traffic loads can be offloaded to more ANds and their associated backhaul links, therefore much higher data rates can be achieved. However, the improvement brought by further increasing the coverage range becomes marginal as the traffic loads on different backhaul links becomes more and more balanced. Moreover, the more coordinated ANds for a user, the tighter interference coordination and more control signalling exchange among the ANds are required, which may be challenging for a distributed control plane.

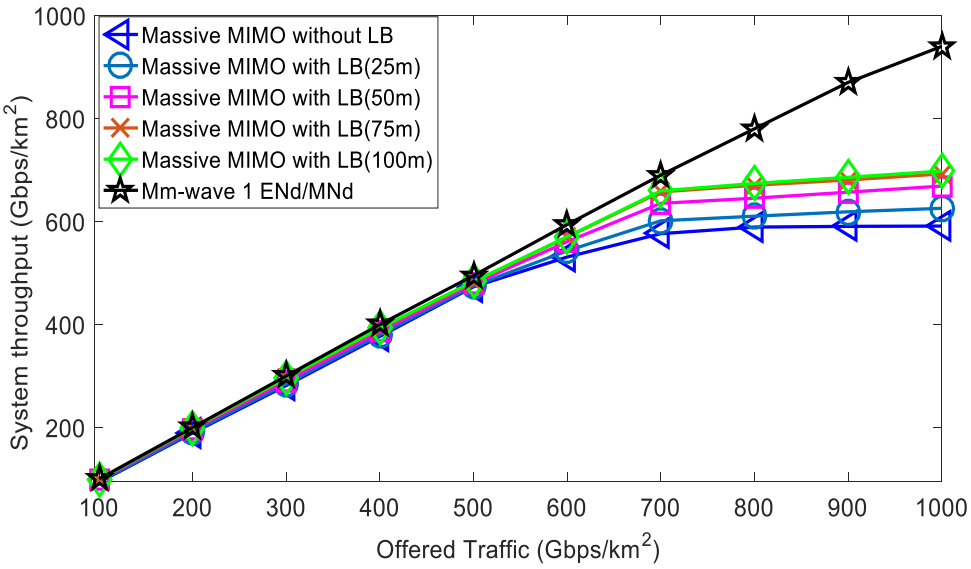

Figure 4.7: System throughput – mm-wave and massive MIMO with access network load balancing (upper bound)

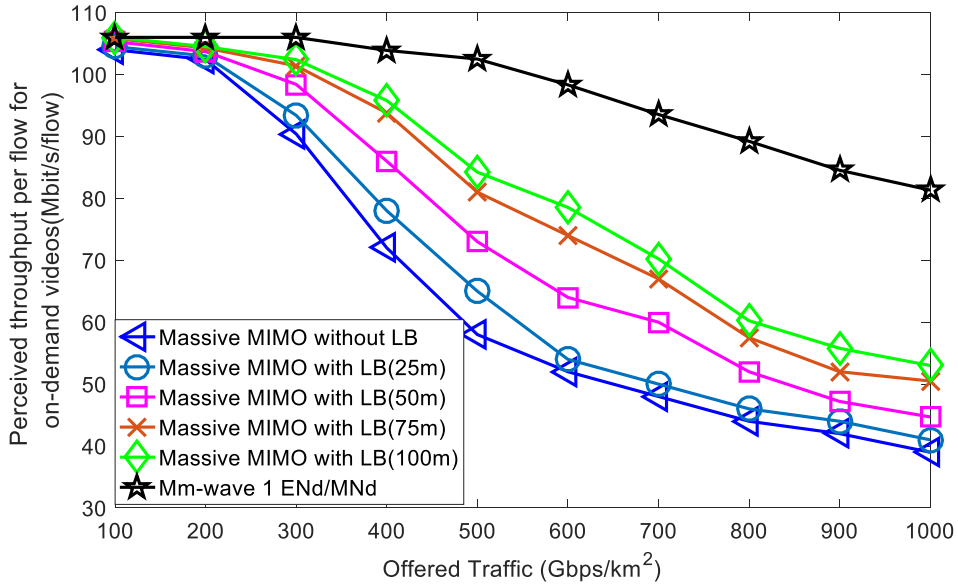

Figure 4.8: Average flow perceived throughput (8K on-demand videos) - mm-wave and massive MIMO with access network load balancing (upper bound)

Quality of Service for different user groups

In order to evaluate the QoS for on-demand video in terms of the experienced interruptions throughout media playback, two states are defined: smooth playback state and interrupted state, and the number of interruptions represents the number of transitions from smooth playback state to the interrupted state. Figure 4.9 and Figure 4.10 show the box plots of perceived throughput and number of interruptions of bus users versus car and pedestrian users at both low traffic load (200 Gbps/km^2) and high traffic load (1000 Gbps/km^2). The box plots are used to demonstrate the distribution information of the data points in terms of their range, median, and quartiles from all the 50 different simulations. The edges of the box are the 25th ($Q1$) and 75th ($Q3$) percentiles of the distribution, the line in the middle of each box shows the median value. The maximum whisker lengths equal $1.5\times(Q3-Q1)$. Any results smaller than $Q1-1.5\times(Q3-Q1)$ or larger than $Q3+1.5\times(Q3-Q1)$ are considered as outliers, and are plotted as individual points. Figure 4.13 and Figure 4.14 show the blocking and dropping probability comparison of real-time videos.

Without the load balancing scheme, there is a significant performance difference in terms of perceived throughput and blocking probability between the bus and the other two types of users. This is due to the fact that the bus users generate temporary hot spots and bring significant burdens to the local backhaul links, whereas the performance of car and pedestrian users depends on their locations. The performance improves with the increase of ANd coverage in the load balancing schemes, and the performance difference between bus and the other two types of users becomes smaller. This can also be observed from Figure 4.12, which presents snapshots of the contour plots showing the spatial distribution of the user perceived throughput across the area (the corresponding user distribution of the snapshots is shown in Figure 4.11). This illustrates the importance of generating flexibility within the communication system to help average out spatial traffic fluctuations.

Although the number of interruptions of the massive MIMO scenario is far less than that of mm-wave, the majority of the videos are interrupted only once, and these videos remain in the interrupted state due to congestion, never returning to the

smooth playback state during the lifetime of the simulation. This can also be observed from the perceived throughput box plot, where the perceived throughput of the massive MIMO scheme are much lower than the required playback rate (20 Mbps). In reality, the users may choose to pause or leave the video in this circumstance, however for simplicity, it is assumed that the users stay and play the videos as soon as there is sufficient content for playback.

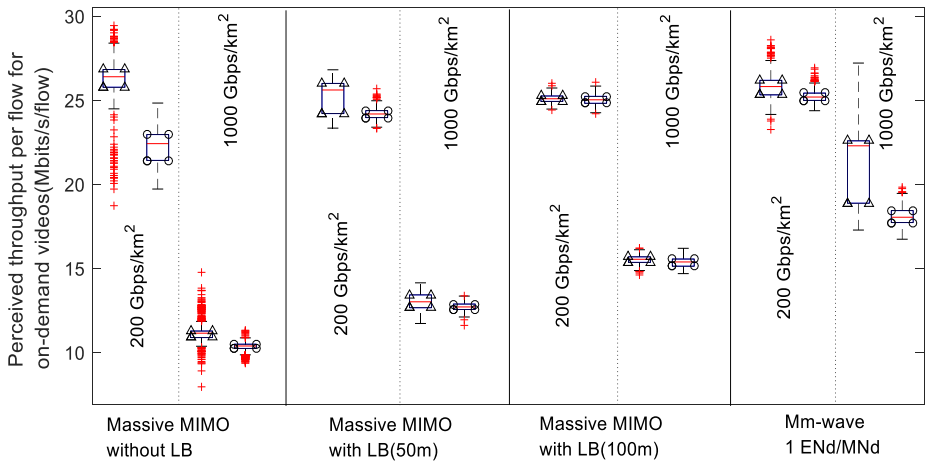

Figure 4.9: Perceived throughput (on-demand 4K videos) box plot - mm-wave and massive MIMO with access network load balancing (upper bound)
(For each boxplot pair: left side: perceived throughput for car and pedestrian users; right side: perceived throughput for bus users)

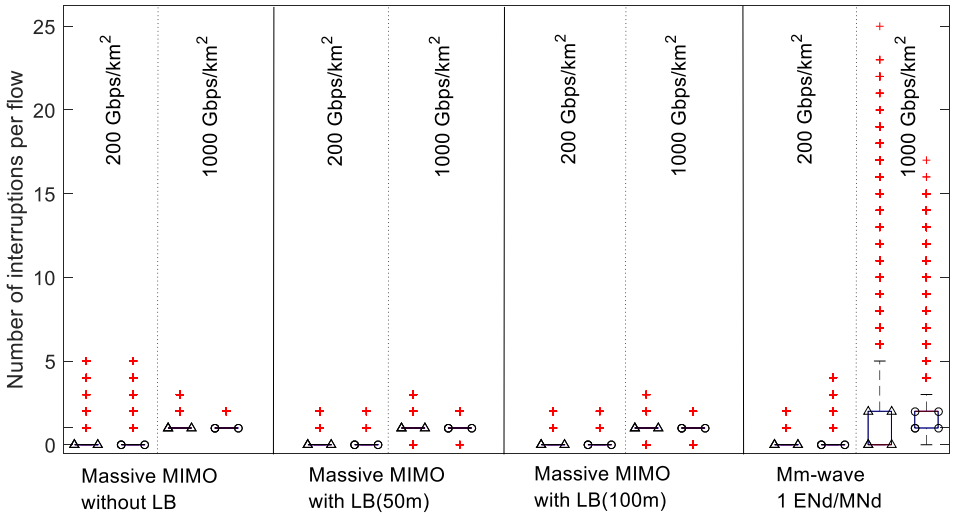

Figure 4.10: Number of interruptions (on-demand videos): - mm-wave and massive MIMO with access network load balancing (upper bound)
(For each boxplot pair: left side: perceived throughput for car and pedestrian users; right side: perceived throughput for bus users)

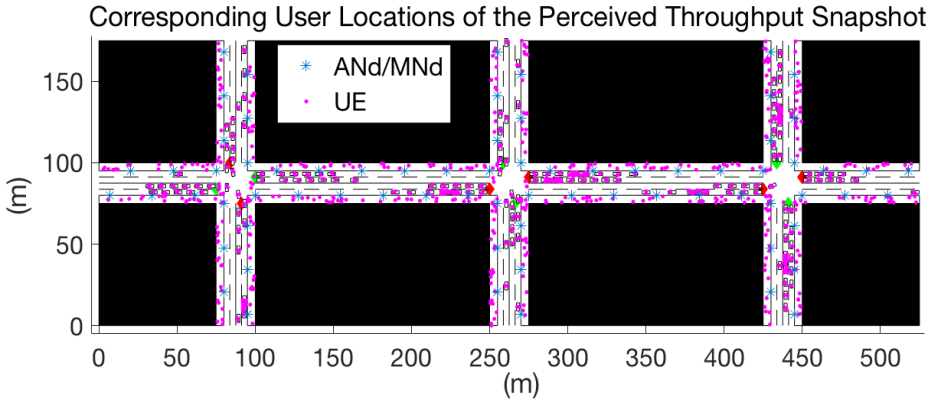

Figure 4.11: Corresponding user locations of the perceived throughput snapshot

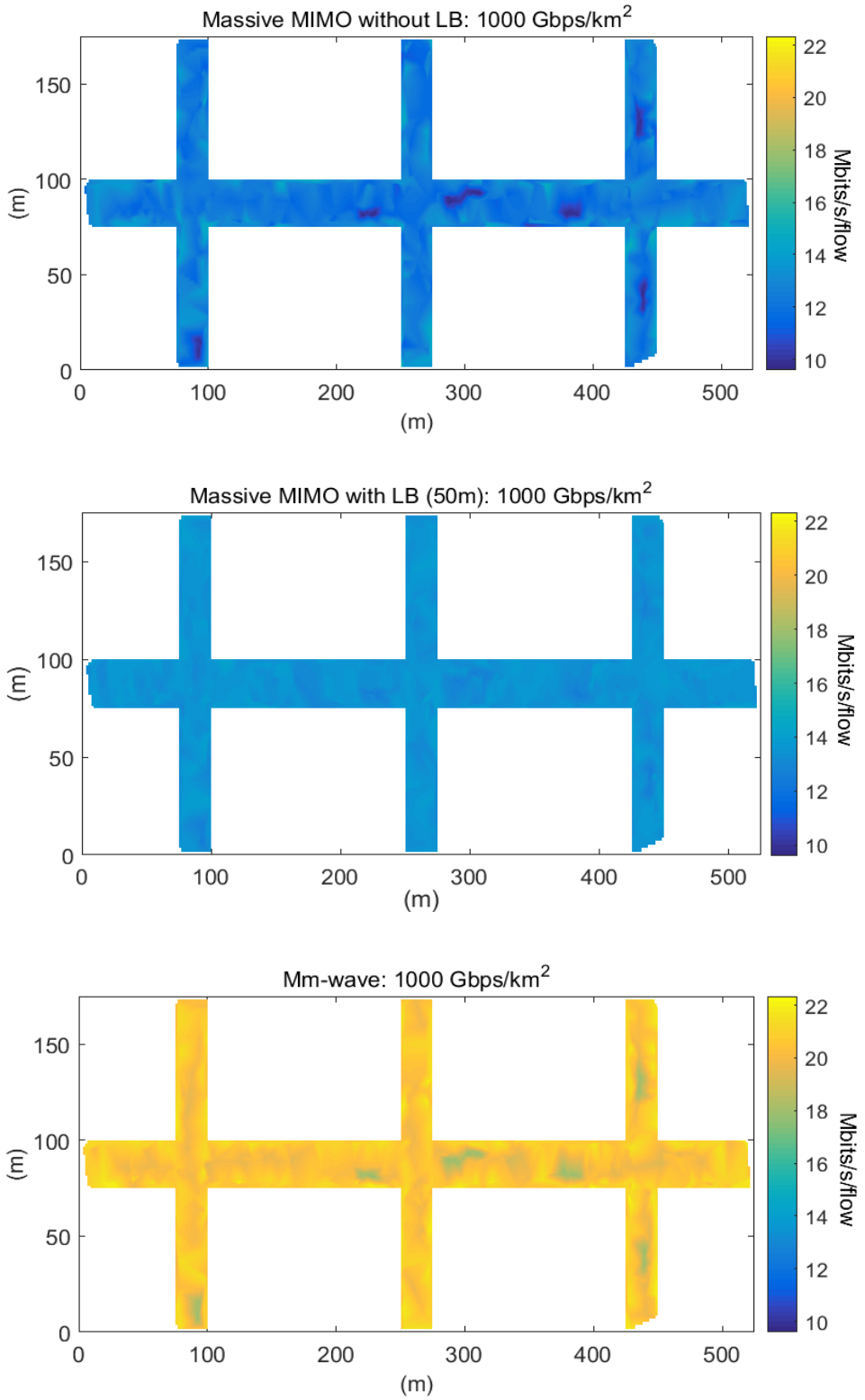

Figure 4.12: Spatial distribution of perceived throughput in Mbit/s (on-demand 4K videos, vehicles driving on the right side) - mm-wave and massive MIMO with access network load balancing (upper bound)

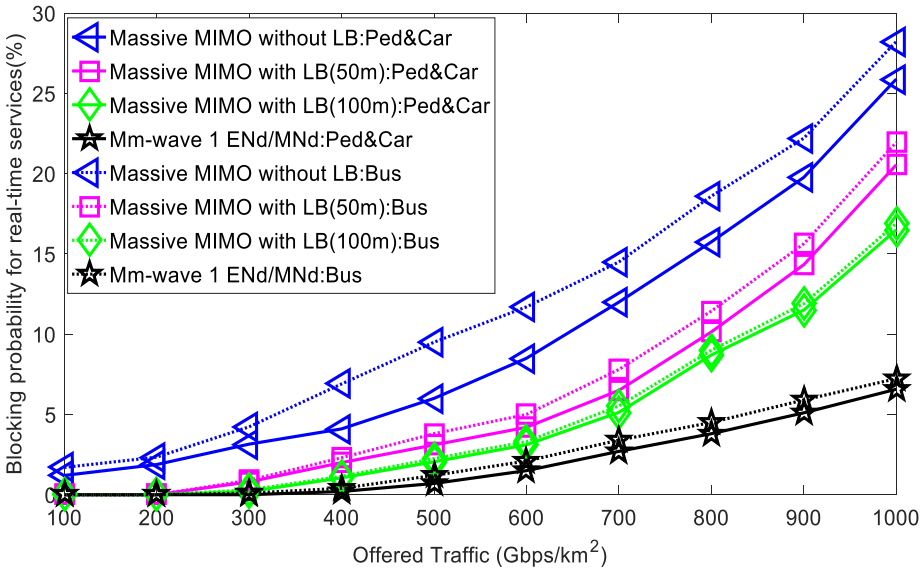

Figure 4.13: Blocking probability (real-time services) - mm-wave and massive MIMO with access network load balancing (upper bound)

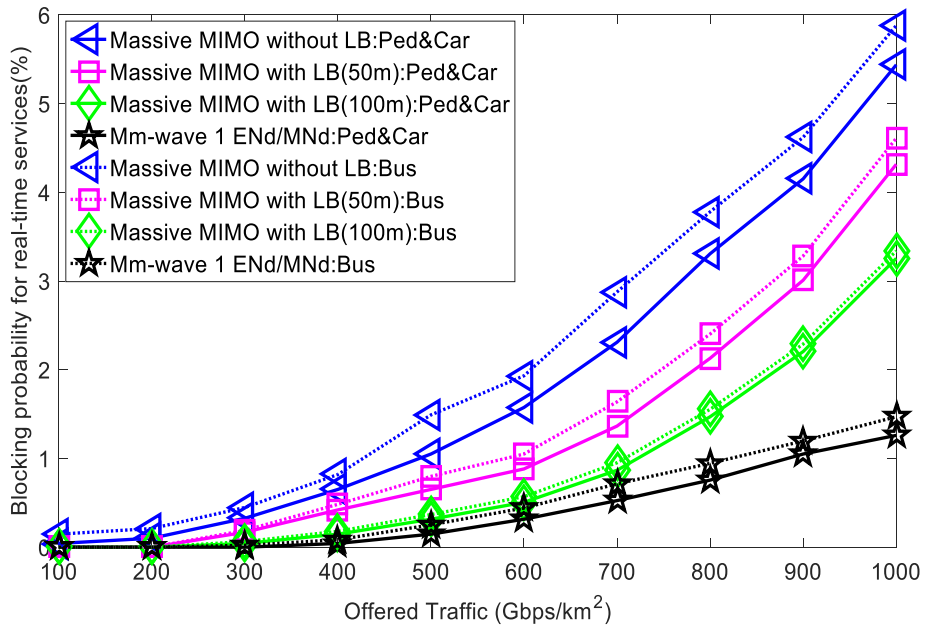

Figure 4.14: Dropping probability (real-time services) - mm-wave and massive MIMO with access network load balancing (upper bound)

4.4 Conclusion

This chapter described the methodologies used for ultra-dense outdoor small cell backhaul networks. An outdoor small cell scenario which reflects a dense urban city centre with mixed vehicular-video traffic is used as the simulation model. Two different traffic types including on-demand video streaming and real-time services were modelled in order to evaluate the capability of backhaul architectures in terms of supporting high data rate and low latency. A resource allocation scheme which prioritises the real-time traffic and allocate dynamic data rates to the on-demand videos has been introduced and will be used in the simulations in the rest of this thesis.

A moving crowd, especially bus users, results in a highly uneven traffic distribution across the small cells which may cause bottlenecks on their associated backhaul links. The upper bound simulation results demonstrated that by applying massive MIMO technology, the delivery of backhaul network is feasible at low to medium traffic loads if access network load balancing techniques is applied in the system (below 500 Gbps/km^2). By applying an access network load balancing technique, higher flexibility and therefore improved performance can be achieved. Tight interference coordination control and extensive signalling exchange are essential in order to guarantee the required QoS. In addition, the sub-6 GHz band may struggle to deliver the necessary capacity needed at high traffic loads due to insufficient frequency resources. For a congested network, mm-wave backhaul has proven to be able to significantly improve the system performance compared to massive MIMO approach using sub 6 GHz band due to its abundant spectrum.

It is worth noting that there is no definite backhaul deployment strategy which is superior to all the others when it comes to 5G. Multiple factors need to be taken into consideration in designing an appropriate backhaul solution, such as the underlying environment and infrastructure, cost, energy consumption and required capacity, etc. For example in the relative low capacity required use case introduced in Chapter 3, the need to reduce energy consumption while maintaining the coverage and QoS

should be highlighted, whereas in a highly densely populated urban city centre where thousands of users use mobile data for communication and entertainment on the go, the mm-wave backhaul option is more desirable. Therefore, mm-wave backhaul is used for the outdoor small cell network in the rest of this thesis. A number of novel backhaul architectures are proposed in the following chapters.

Chapter 5. Providing Millimetre Wave Backhaul Diversity Utilising Street Fibre Cabinets

Contents

5.1 Introduction

Using a two-tier architecture investigated in Chapter 3 and 4 can reduce the deployment costs since the number of points where fibre needs to be available is significantly less than running fibre to each small cell. However, for a dense urban

city with streets lined with high buildings and a small cell size of a few tens of metres, installing these ENds can still be cost prohibitive. For example, in Figure 4.2, assuming the propagation range for millimetre wave is 100 m to avoid weather based link outage, more than 10 ENds are needed to provide direct LOS backhaul links to the small cells in the topology shown.

The existing placement of outdoor ENds follows the similar approach as with macro sites deployments. The ENds with a fibre connection are typically mounted on tall structures such as cell towers and high buildings to increase LOS probability and avoid temporary outages caused by passing vehicles as depicted in Figure 5.1 (a). This requires installing fibre optic cables spanning the distance between a small cell access node (ANd) and a source fibre cabinet. In addition, the speed of deployment is another issue. It may take months to install the ENds on existing buildings, since such installations may be controversial especially if the route passes through private properties. In previous work on deploying multiple nodes to improve the flexibility and diversity of the backhaul, researchers usually assume several candidate links/paths are already available and focus on the link/path selection algorithms only [39] [40] [41]. The feasibility of deploying the backhaul aggregation nodes is often neglected.

One way of reducing the deployment cost as well as increasing the flexibility of the backhaul network is to consider the points where the fibre access is already available. The purpose of this chapter is to report a novel backhaul architecture delivered from street fibre cabinets designed for outdoor dense small cells, which exploits path diversity within a network to mitigate the effect of link outage as well as provide high capacity density. The ENds are co-located with street fibre cabinets to provide a flexible and low cost millimetre wave mesh backhaul network.

The remainder of the chapter is organised as follows: Section 5.2 introduces the street cabinet ENd and illustrates possible street cabinet ENd configurations along with antenna placement. In Section 5.3, an analytical framework for backhaul link outage and diversity analysis is developed. Section 5.4 explains the backhaul path management procedure both in a distributed configuration and in a centralised

configuration. In Section 5.5, a case study is discussed to illustrate the feasibility of deploying existing street cabinets to deliver backhaul. Finally, conclusions are given in Section 5.6.

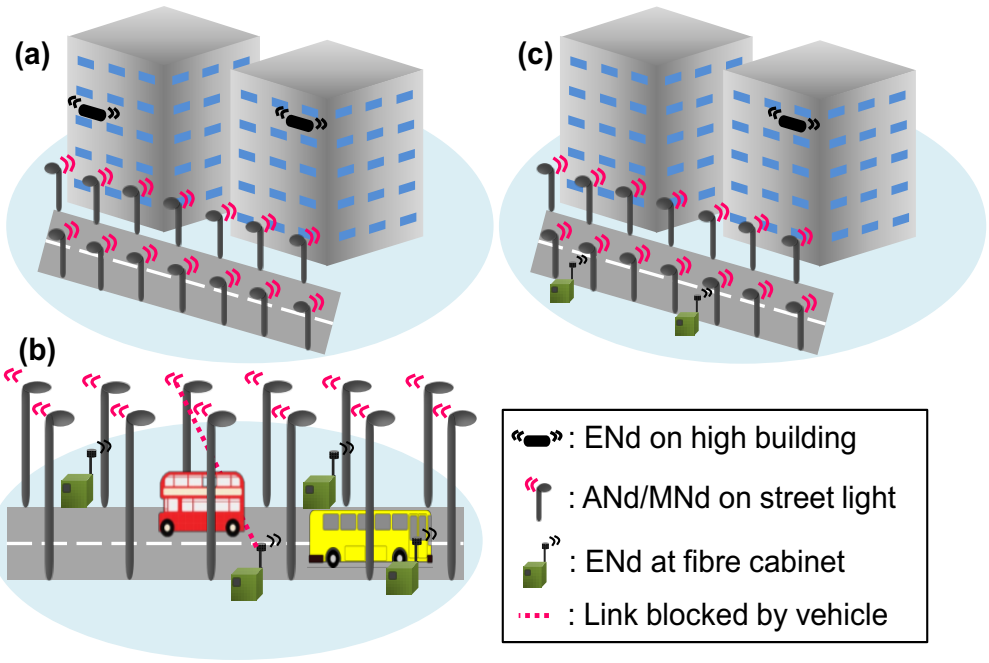

Figure 5.1: GREENBACK outdoor backhaul deployment scenarios

5.2 Street Cabinet Edge Node

Currently, operators have already installed a large number of street fibre cabinets in dense urban areas through their FTTx deployments, and these street cabinets already have fibre, power and implementation permits. Hence it is sensible to explore the possibility of using these existing street fibre cabinets to house ENds in order to overcome the need to run extra fibre and power cables between an ENd and a source cabinet. Depending on local fibre penetration level and capacity requirements, a backhaul network may be deployed exclusively by using street cabinet ENds as illustrated in Figure 5.1 (b) or a combination of existing backhaul infrastructure and street cabinet ENds as shown in Figure 5.1 (c).

The components of an ENd including RF processing unit (remote radio unit - RRU), path control/management unit, Baseband Unit (BBU, in a D-RAN configuration), and antenna array can be installed onto an existing street cabinet. The detailed path management unit will be discussed in Section 5.4.1. The street cabinet ENds can carry uplink/downlink traffic between mesh nodes and a core network using point-to-point millimetre wave links. As shown in Figure 5.2 (a) and (b), an electronically steerable antenna array (planar array for example) is attached to the exterior of the street cabinet. It is also possible to install an electronically steerable antenna array (cylindrical antenna array for example) on a pole attached to the street cabinet as shown in Figure 5.2 (c). The height of the pole depends on the outage requirement and deployment constraints.

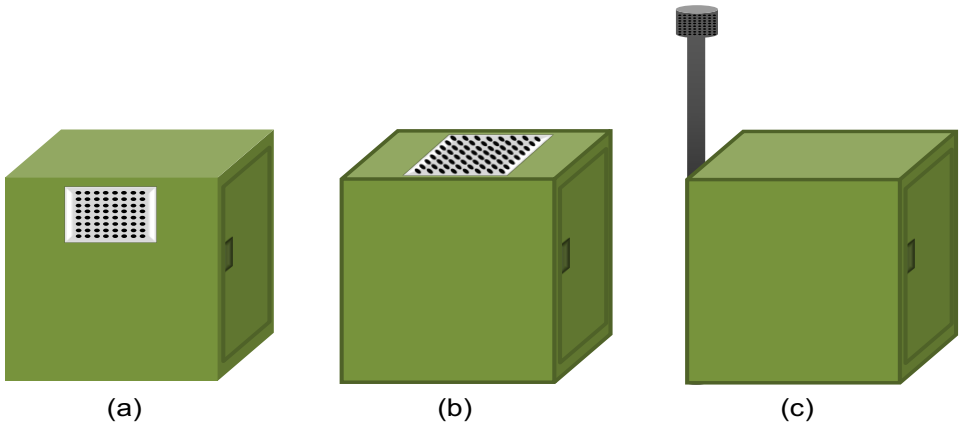

Figure 5.2: Possible antenna placements on a street cabinet

5.3 Backhaul Link Outage Analysis

5.3.1 Analytical Model

Although utilising existing street fibre cabinet ENds seems to be an extremely promising solution to deliver backhaul, the street-level millimetre wave backhaul links are more likely to be temporarily blocked by passing cars, human beings and

other obstacles compared with solutions where the ENds are installed above first storey height on building walls as illustrated in Figure 5.1 (b). Hence, it is important to examine the impact of link outage on system performance. A two-tier network model consisting of one MNd which has access to N street cabinet ENds via mm-wave backhaul links as shown in Figure 5.3 is considered. Due to the narrow beamwidth, the spatial separation of paths, and the static positions of the base station antennas, the millimetre wave links can be modelled without the need to consider the impact of mutual interference.

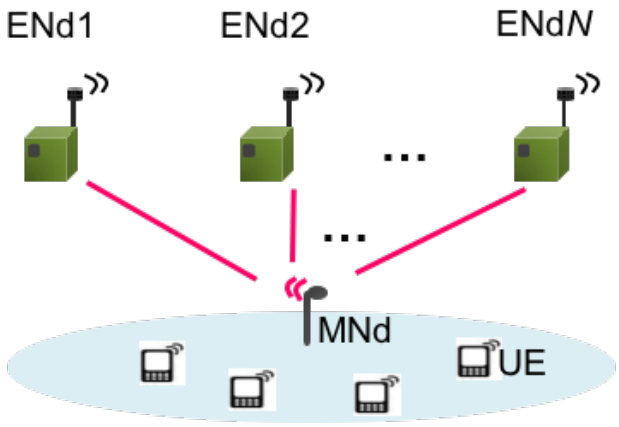

Figure 5.3: 1 MNd *N* street cabinet ENd network model

The backhaul link average outage percentage $P_o(i)$ $(i = 1, 2, \dots, N)$ here is defined as the average total amount of time of momentary outage that the i_{th} backhaul link would experience during the entire time T_s**.** For simplicity, we assume that all the links have the same average outage percentage P_o. The value can be changed without the loss of generality. Assuming a backhaul link is in outage independently of other backhaul link states, then the amount of outage backhaul links out of N follows the binomial distribution, and the probability of having m outage backhaul links for a MNd on average can be calculated by:

$$\eta_m = \binom{N}{m} P_o{}^m (1 - P_o)^{N-m}$$

$$= \frac{N!}{(N-m)!m!} P_o{}^m (1 - P_o)^{N-m} \quad (5.1)$$

It is assumed that each mm-wave link has K channels, and that the total new traffic arrival for the MNd follows a Poisson process with rate λ, and the service time is an exponential distribution with service rate μ. It is also assumed that the idle channels on the accessible backhaul links can be randomly chosen with equal probability. In this network model, all the backhaul links are independent and statistically identical. Hence, we can model the system as a M/M/c/c queue which can be represented by a one-dimensional Markov chain [98]. The blocking probability for the M/M/c/c model can be obtained from the Erlang-B formula.

In this street cabinet scenario, it is more likely that a group of traffic flows need to perform handover to another street cabinet ENd simultaneously in the event of link outage, hence modelling the enhanced arrival rate for new incoming traffic flows alone cannot capture the blocking caused by handover failures. In light of this, we classify the transmission blocking into two types: one is caused by the lack of idle channels for the new traffic; the other is caused by the handover failure for the ongoing traffic flows when a backhaul link is in outage. Given $(N - m)$ backhaul links are accessible to the MNd, then there are $K(N - m)$ channels in total. The blocking probability caused by the absence of available channels therefore can be given by:

$$P_c(m) = \frac{(\lambda/\mu)^{K(N-m)}/K(N-m)!}{\sum_{c=0}^{K(N-m)} (\lambda/\mu)^c/c!} \quad (5.2)$$

Therefore, we obtain $P_{cb}(m)$, the blocking probability for the new traffic when there are m outage backhaul links in the system:

$$P_{cb}(m) = \eta_m P_c(m)$$

$$= \frac{N!}{(N-m)!m!} {P_o}^m (1-P_o)^{N-m} \frac{(\lambda/\mu)^{K(N-m)}/K(N-m)!}{\sum_{c=0}^{K(N-m)} (\lambda/\mu)^c/c!} \quad (5.3)$$

In the situation where there is no outage link in the system, the backhaul links are chosen with equal probability. Hence, the traffic arrival rate for each backhaul link can be expressed as λ/N, and the aggregated arrival rate of m backhaul links is $\lambda m/N$. Therefore, the probability of having h ($0 \leq h \leq mK$) ongoing traffic flows (at state $p(h)$) that need to handover to other links from the m outage links can be calculated using the derivation of Erlang-B:

$$p(h) = p(0)\frac{(\lambda m/\mu N)^h}{h!}$$

$$= \frac{(\lambda m/\mu N)^h/h!}{\sum_{j=0}^{Km} (\lambda m/\mu N)^j/j!} \quad (5.4)$$

Similarly, the probability of having n ($0 \leq n \leq (N-m)K$) occupied channels on $(N-m)$ available backhaul links can be expressed as:

$$p(n) = \frac{[\lambda(N-m)/\mu N]^n/n!}{\sum_{k=0}^{K(N-m)} [\lambda(N-m)/\mu N]^k/k!} \quad (5.5)$$

The handover failure occurs when $h > (N-m)K - n$, meaning that the idle channels on the available $(N-m)$ links cannot accommodate all the handover traffic flows. Hence the probability can be calculated by:

$$P_{hb}(m) = \sum_{h=1}^{mK} \left[p(h) \sum_{n=(N-m)K-h+1}^{(N-m)K} p(n) \right] \quad (5.6)$$

Based on the above analysis, the total blocking probability can be expressed as:

$$P_b = \sum_{m=0}^{N}(P_{cb}(m) + P_{hb}(m)) \tag{5.7}$$

Note that the first term represents the new traffic blocking probability, whereas the latter is the handover failure probability for ongoing traffic flows. Substituting Equation (5.1)-(5.6) in Equation (5.7), we obtain the overall probability that a transmission is blocked P_b as:

$$P_b = \sum_{m=0}^{N}\{\frac{N!}{(N-m)!m!}{P_o}^m(1-P_o)^{N-m}\frac{(\lambda/\mu)^{K(N-m)}/K(N-m)!}{\sum_{c=0}^{K(N-m)}(\lambda/\mu)^c/c!} +$$

$$\sum_{h=1}^{mK}\left[\frac{(\lambda m/\mu N)^h/h!}{\sum_{j=0}^{Km}(\lambda m/\mu N)^j/j!}\sum_{n=(N-m)K-h+1}^{(N-m)K}\frac{[\lambda(N-m)/\mu N]^n/n!}{\sum_{k=0}^{K(N-m)}[\lambda(N-m)/\mu N]^k/k!}\right]\} \tag{5.8}$$

5.3.2 Results

The main aim of the model introduced above is to analyse the impact of high link outage and the system performance improvement by introducing path diversity to the network. In order to validate the accuracy of the analytical model, we also use MATLAB to simulate the described scenario. The parameters used in the network are listed in Table 5.1. The temporary outage probabilities are assumed to be independent between the different backhaul links, and the individual temporary outage duration is modelled as an exponential distribution with a mean duration of T_o. The outage intervals (the time interval between the end of previous outage and the next outage) are also exponentially distributed, and the inter arrival rate λ_n is calculated as:

$$\lambda_n = \frac{T_s P_o}{T_o T_s} = \frac{P_o}{T_o} \tag{5.9}$$

Note that the temporary outage probabilities for different backhaul links may not be fully independent in reality. For instance, neighbouring links may be blocked by a long vehicle. The temporary outage durations for backhaul links may also vary drastically with the time of day and between different locations. Hence, another approach would be to model the outages explicitly based on site-specific traffic and geographic data. We do not specifically model these correlations here, but further detailed modelling may be advantageous, if correlated link outage data is available.

Table 5.1: Parameter values of the analysis and simulation

Parameter	Value
Number of channels on each backhaul link: K	25
Traffic service rate: μ	0.5/s
Mean outage duration: T_o	20 s
Total simulation time: T_s	1000 s

Figure 5.4 shows the blocking probability of the system with different numbers of backhaul links accessible to a MNd. It demonstrates a good agreement between analytical and simulation results (averaged over 100 trials), which means that the analytical approach provides an effective design tool for modelling the performance of the mm-wave backhaul system with link diversity. At low traffic loads, the performance is dominated by the outage caused by obstructions, where all the backhaul links for a MNd are in outage. At medium to high traffic loads, the blocking probability increases rapidly as the traffic increases. This is a result of the blocking caused by the lack of idle channels starting to dominate. By introducing path diversity to the network, the system performance improves significantly.

Figure 5.5 demonstrates the number of backhaul links needed at different outage levels: low (10%), medium (30%) and high (50%), in order to achieve an acceptable level of performance. The blocking probability of the system with building wall ENds acts as a benchmark. Since the street level obstructions are the major concern for link outage here, the link outage of a backhaul link delivered from building wall ENds are assumed to be 0%. Fewer backhaul links are needed for a system with a lower outage level as expected. With 10% link outage, three backhaul links per MNd are required at high traffic loads as opposed to two links needed for the benchmark scenario. However, this configuration provides better performance at high traffic loads compared to the benchmark due to the path diversity introduced to the network.

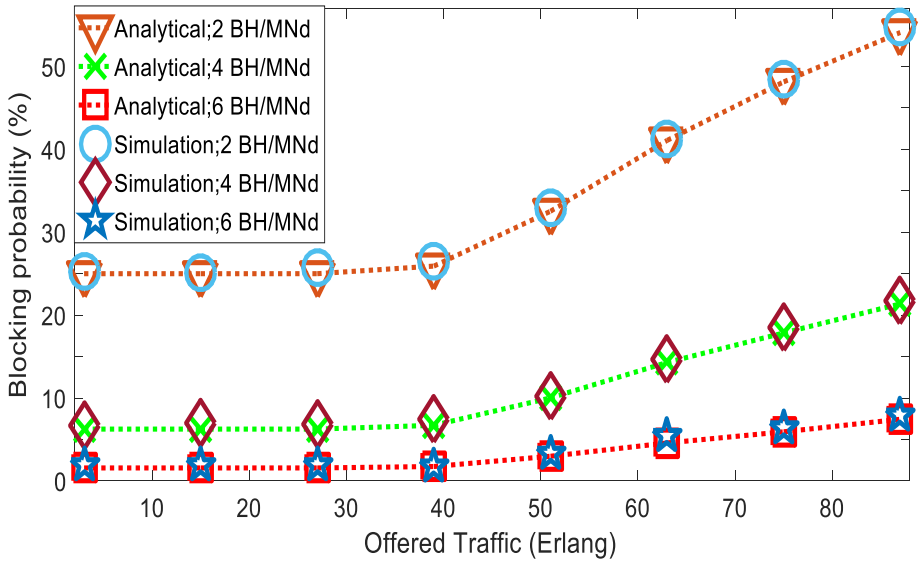

Figure 5.4: Blocking probability – analytical vs simulation (50% outage)

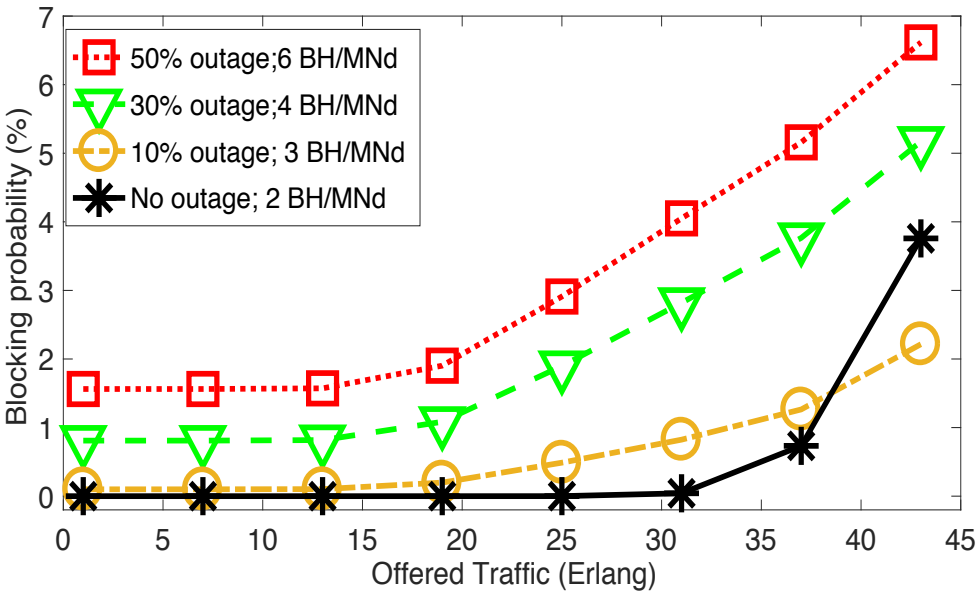

Figure 5.5: Blocking probability (analytical) - number of backhaul links needed for different outage probability

5.4 Management Procedure

As discussed above, path diversity is essential to maintain the backhaul connectivity and required QoS. Depending on the capability of the ENd devices, the backhaul/fronthaul path management module may be either distributed within each ENd or centralised at the core network level.

5.4.1 Distributed Configuration

Figure 5.6 illustrates an example of the detailed breakdown of components in a street cabinet ENd along with a distributed Path Management Unit. The MNds are interconnected with each other via the X2 interface. When another backhaul link is required in response to the current backhaul link disruption or congestion, a MNd sends a broadcast request to any ENds which is capable of receiving the request if there is no pre-stored candidate path available. Alternatively it sends directed requests to the available candidate ENds to establish a backhaul link. The QoS

parameters QoS Class Identifier (QCI) and/or Allocation and Retention Priority (ARP) can be used to indicate the required resource type, packet delay budget (ranging from 50 ms to 300 ms), packet error loss rate (ranging from 10^{-6} to 10^{-2}) and priority level. Note that the QCI can be defined by the operators. For more specific details, please see [90]. After receiving the acknowledgements from the ENds, a path selection scheme may be used to choose one or more backhaul/fronthaul links to establish based on link status, traffic loads, QoS requirements, etc. An example of the path management procedure is shown in Figure 5.7.

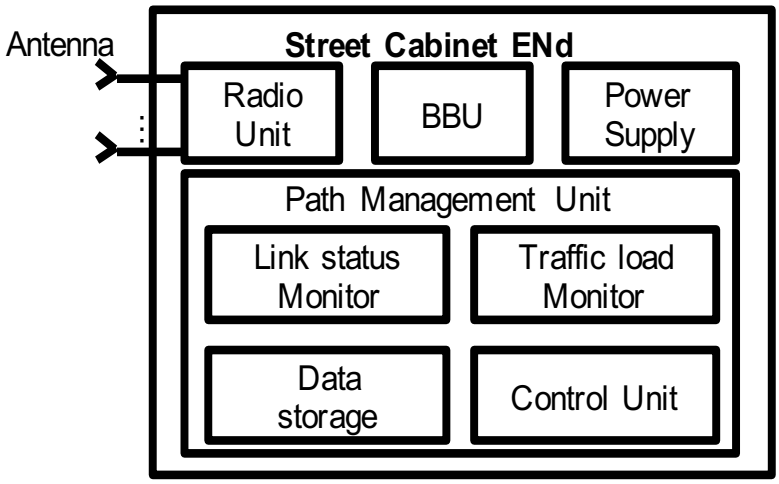

Figure 5.6: A block diagram of an example street cabinet ENd configuration with a distributed Path Management Unit

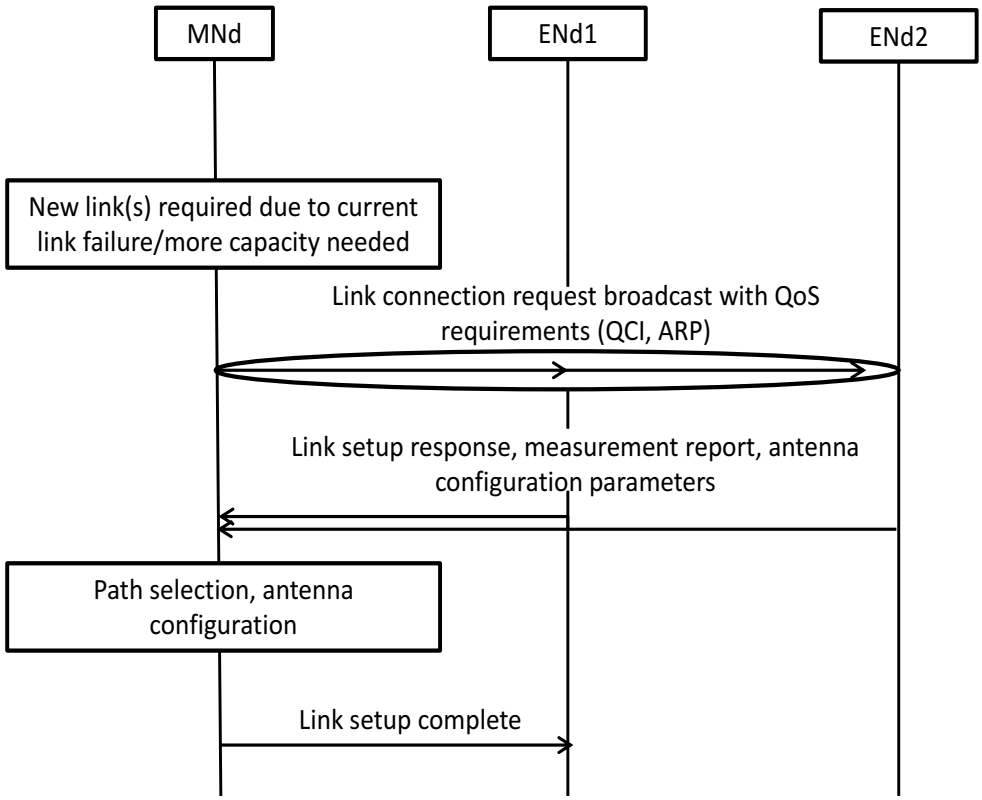

Figure 5.7: Signalling flow of an example distributed path control method

5.4.2 Centralised Configuration

The Path Management Unit can alternatively be configured within a centralised controller. The Unit can be implemented in an operator's core network. This may be realised by using SDN enabled architecture as described in Chapter 3. The SDN-enabled switches which might be placed within MNds and ENds, process the incoming traffic based on a flow table defined by the central controller. The switches also collect network utilisation statistics such as current traffic load and backhaul link status, and report them back to the central controller via a south-bound interface. The backhaul path management algorithms and resource allocation policies can be implemented by using the application program interfaces (APIs) provided by the central controller and then push the policies to the controller via a northbound interface. The controller coordinates the available backhaul links for each MNd on event basis. This approach also significantly simplifies the components of street

cabinet ENds, which further reduces the associated costs. Higher flexibility and improved resource utilisation can also be expected from this approach.

5.5 Proof-of-Concept Case Study

In Section 5.3, the analytical model to evaluate performance of different levels of link outage and path diversity is derived for regular Poisson distributed voice call traffic. This section presents the results of simulating the proposed urban street cabinet backhaul architecture in the street canyon scenario with vehicles and pedestrians using different applications described in Section 4.2 and Section 4.3.2. The simulation lasts 10,000 s for each traffic load, and the measurements are taken after 100 s when the system is relatively stable. The results are obtained by averaging 50 different simulations with randomly generated mobility traces and user traffic. In addition, the average link outage duration is assumed to be 100 s, which is the same as the red traffic light duration. The outage inter arrival rate can be calculated using Equation (5.9).

The impact of different levels of temporary link outage on system performance with single backhaul links is evaluated first. We then explore the possibility of utilising street cabinet ENds to provide backhaul redundancy for a capacity constraint network.

5.5.1 The Impact of Momentary Backhaul Link Outage on System Performance

In order to evaluate the impact of momentary link outages on system performance, we assume all the ENds are installed on the street fibre cabinets as shown in Figure 5.1 (b). Each MNd has one backhaul link connection. The outage percentage of the street cabinet links depends on the height of the antennas as well as the local environment. A lower outage percentage is possible when the antennas are placed at a greater height above ground level. On the other hand, in a dense urban city with

congested traffic, the outage might be higher. Hence, different levels of backhaul link average outage percentage P_o are tested.

In the simulation, the link outage probability values of 10%, 30% and 50% are tested respectively which are the same as the analytical analysis. A user is assumed to be able to connect to any ANds within a connecting range of 50 m (medium range as discussed in Section 4.3.3), and would initially connect to its nearest ANd which can potentially offer the best link quality. When a backhaul link experiences outage (the corresponding ANd cannot connect to the core network through a street cabinet ENd), the existing users on the link will be handed over to the next nearest ANd and its associated backhaul link. In Figure 5.8, the lines in Group 1 compare the average perceived throughput (8K on-demand videos shown, 4K on-demand videos follow the same trend) of different levels of link outage. 50% average outage performs the worst due to the high probability of the unavailable backhaul links which causes congested traffic loads on the available backhaul links. The performance is better with a lower probability of outage.

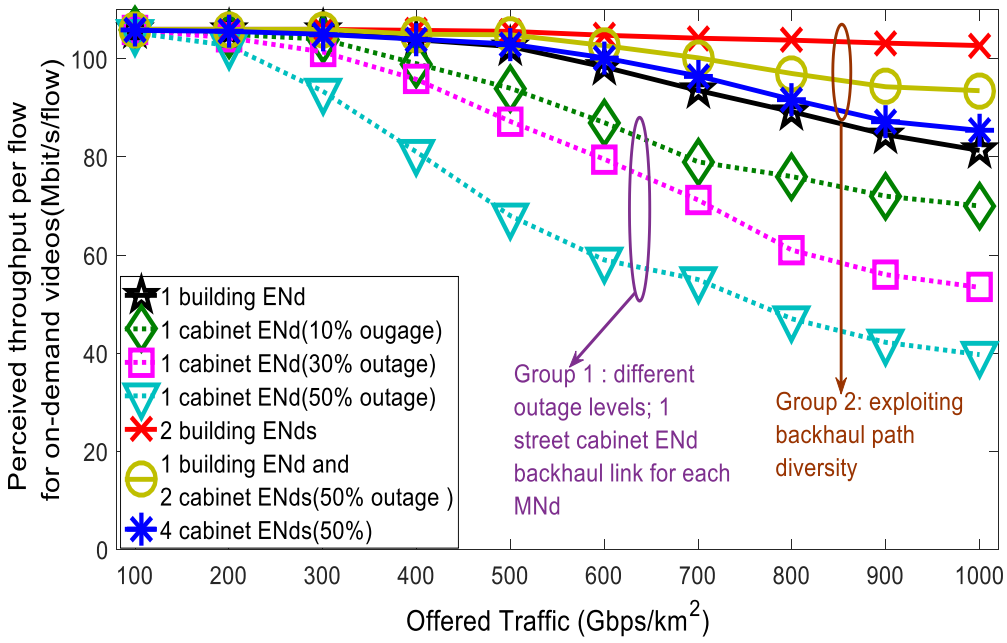

Figure 5.8: Average perceived throughput per flow – different levels of outage and redundancy (8K on-demand videos)

5.5.2 Providing Backhaul Link Diversity Using Street Cabinet ENds

Proving backhaul flexibility using street cabinet ENds provides an effective alternative to offload the traffic burden. In this subsection, we explore two possible configurations. The first one is to deploy ENds solely relying on the street cabinets as shown in Figure 5.1 (b). The other is a hybrid approach, whereby a MNd is served by a mixture of building wall and street cabinet ENds as shown in Figure 5.1 (c). Here, we assume 0% average outage for building wall ENd-MNd links, and 50% average outage for street cabinet ENd-MNd links as mentioned in Section 5.3.2.

The lines of Group 2 in Figure 5.8 compare the average perceived throughput that can be achieved by introducing redundancy into the backhaul network. The average capacity offered in these options is equivalent to the capacity provided by 2 backhaul links from building walls for each MNd. The system with 1 backhaul link and 2 backhaul links with ENds installed on building walls are the lower and upper benchmarks to evaluate the performance of the system with redundancy. Two simultaneous backhaul transmission links are allowed per MNd for the redundancy deployments, and a new flow will be allocated to the backhaul link with lower traffic load of the two.

For an all-street-cabinet configuration, 4 LoS street cabinet backhaul links are accessible for each MNd with an average outage of 50% each. This configuration has the similar average performance to the lower benchmark (1 backhaul link on building wall). However, for low to medium traffic load, the average performance loss compared to the upper benchmark is marginal (2.7% maximum), and it supports a good service level overall. Note that the encode data rate (required smooth playback rate) for 8K video is 80 Mbps.

For a hybrid configuration, 1 building wall ENd and two street cabinet ENds (with 50% average outage each) are available for each MNd and this is assumed in order to match the average capacity provided by 2 building wall ENds. As shown in Figure 5.8, it performs better than the all-street cabinet configuration (with only 1.5%

performance loss at medium traffic load) since each MNd has one reliable backhaul link on building walls, and the street cabinet ENds provides extra flexibility and capacity to the system.

Figure 5.9 – Figure 5.11 show the performance comparison of this hybrid approach versus the lower and upper benchmarks. This configuration performs significantly better than the 1 building wall ENd backhaul link per MNd approach especially for the high traffic load. Figure 5.9 shows the perceived throughput box plot. Due to the lack of backhaul diversity for the 1 building ENd configuration, the performance of pedestrian and car users vary drastically depending on their local traffic load conditions on their serving backhaul, and the bus users suffer from severe congestion overall. Introducing extra diversity into the system (hybrid approach) results in an 8.6% and 21% increase in median perceived throughput for pedestrian/car users and bus users respectively compared to the lower benchmark. Although the performance of the hybrid approach is not as good as the one supported by 2 building wall ENds per MNd configuration, it is able to provide satisfactory QoS to the end users with potentially lower costs and fewer deployment challenges (note that the smooth playback rate for 4K video is 20 Mbps).

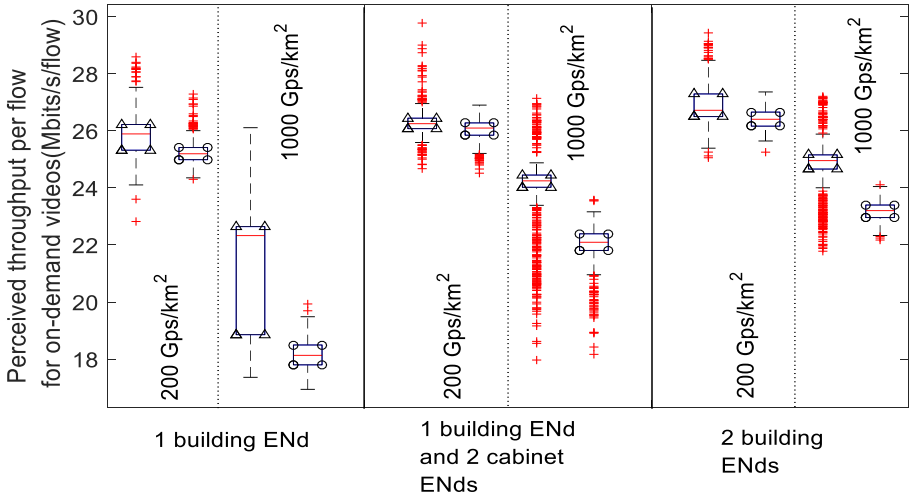

Figure 5.9: Perceived throughput (on-demand 4K videos) box plot – different redundancy levels at different offered traffic load comparison
(For each box plot pair: left side: perceived throughput for car and pedestrian users; right side: perceived throughput for bus users)

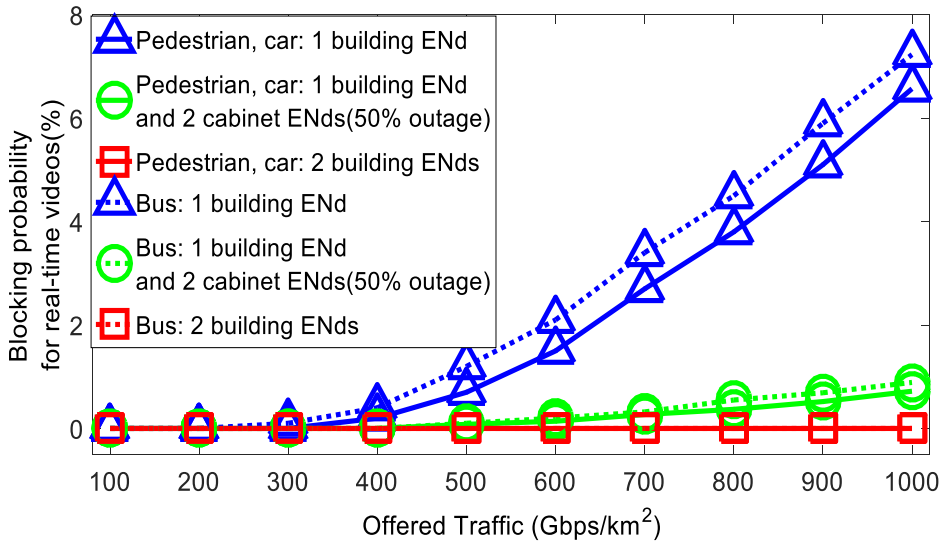

Figure 5.10: Blocking probability – different levels of backhaul redundancy (real-time videos)

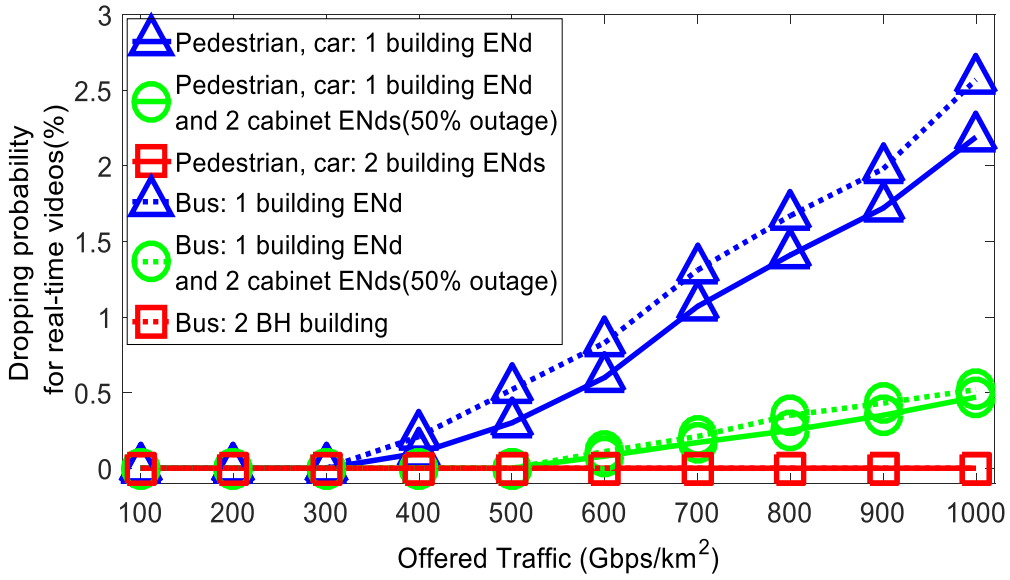

Figure 5.11: Dropping probability – different levels of backhaul redundancy (real-time videos)

5.6 Conclusion

In this chapter, a novel backhaul architecture where ENds are co-located with street fibre cabinets to provide a cost-effective and flexible millimetre wave backhaul network for outdoor dense small cells has been proposed. In the situation where installing adequate ENds with fibre connections on building walls incurs high costs and deployment challenges, the delivery of backhaul traffic directly from street cabinets with satisfactory QoS is feasible, even when there is high outage at each street cabinet. This is achieved by building sufficient path diversity within the network so that each MNd can connect to two or more street cabinet ENds to ensure high reliability of the backhaul connectivity.

An analytical framework to examine the impact of the backhaul link outage and path diversity has been presented. With 10% link outage, three backhaul links per MNd are required at high traffic loads compared to two links needed for all-building-wall deployed scenario. However, it delivers better performance at high traffic loads resulting from the extra flexibility introduced to the system. Furthermore, two

backhaul path management approaches were described – distributed configuration and centralised configuration.

Finally, a realistic simulation of a street canyon with vehicular traffic scenario showed that for an all-street-cabinet configuration, 4 street cabinet backhaul links per MNd is often sufficient, even with a conservatively high average outage of 50%. In addition, the street cabinet ENds can also be used in conjunction with building wall ENds to provide extra backhaul flexibility and capacity. The results showed that the hybrid approach with 1 building wall ENd and two street cabinet ENds per MNd achieved superior performance compared to the 1 building wall ENd per MNd configuration. It has comparable performance compared to the 2 building wall backhaul links per MNd approach, with potentially lower costs and fewer deployment challenges.

Chapter 6. Multi-hop and Topology Management for Millimetre Wave Backhaul Network

Content

6.1 Introduction

All the backhaul deployment strategies discussed in this thesis so far only consider one-hop relay: a backhaul link from an aggregation node (ENd) to a relay node (MNd). The main reason for this architecture is to reduce the system and relay node

complexity. The improvement of backhaul network flexibility is achieved by ensuring that the path diversity is created by network densification. However, in the situation where increasing the density of the infrastructure node is difficult to realise, multi-hop might be a viable solution to bringing in the extra level of flexibility, and thus improving the coverage of a mm-wave backhaul network. Here, how to design an appropriate topology management strategy is an important issue.

There is a large amount of previous work addressing various aspects of multi-hop routing algorithms, such as link quality, system capacity, hop count, and end-to-end delay [99-102]. For example in [102], a bottleneck capacity aware routing algorithm which takes link quality, interference and traffic load into account is proposed to improve the system throughput. Despite the fact that each scheme brings certain benefits, the considered scenarios are rather limited and may not be suitable for practical mm-wave multi-hop deployment due to its unique propagation characteristics. In [103] and [104], multi-hop algorithms which exploits the access point diversity are proposed to solve the blockage problem in 60 GHz band. However, only one-hop relay and overall system capacity are considered. In fact, the impact of the features of a mm-wave multi-hop backhaul on system performance and user experience remains largely unclear.

The purpose of this chapter is to investigate the dynamics of applying multi-hop technique to mm-wave backhaul when different aspects are considered in the system and, thus, to propose multi-hop schemes which account for relay node capability, end-to-end latency, QoS differentiation, and handover control.

6.2 Design Considerations on Multi-hop Millimetre Wave Backhaul

6.2.1 Full Duplex Communications

In wireless communication systems, half duplex solutions are widely adopted which means the transmission and reception of information cannot take place simultaneously. This is due to the fact that the received signal from a local transmitting node is much stronger than the arriving signal from a remote transmit antenna [105]. An attractive solution to significantly improving the network flexibility is full duplex. It allows a pair of nodes to transmit and receive information simultaneously under specific isolation conditions [105]. For example, frequency-division duplexing (FDD) has long been adopted in wireless communication systems for establishing full duplex communication links using two different frequencies. With the rapid development of advanced radio techniques, an emerging technique called in-band full duplex has been envisioned as one of the main building blocks of 5G. It allows transceivers to transmit and receive signals simultaneously over the same frequency. The resource utilisation can be significantly improved by this approach.

There are two basic topologies for full-duplex systems: bidirectional full-duplex system and relay full-duplex system as illustrated in Figure 6.1. Assume a traffic flow needs to be transmitted from source node S to destination node D. In a relay topology, if a relay node R operates in a half-duplexing mode, then it would need to receive from the source node S and forward to the destination node D alternately. With full-duplexing however, R is able to receive and forward the flow simultaneously. In a bidirectional topology, node A and node B can transmit data flows to each other simultaneously.

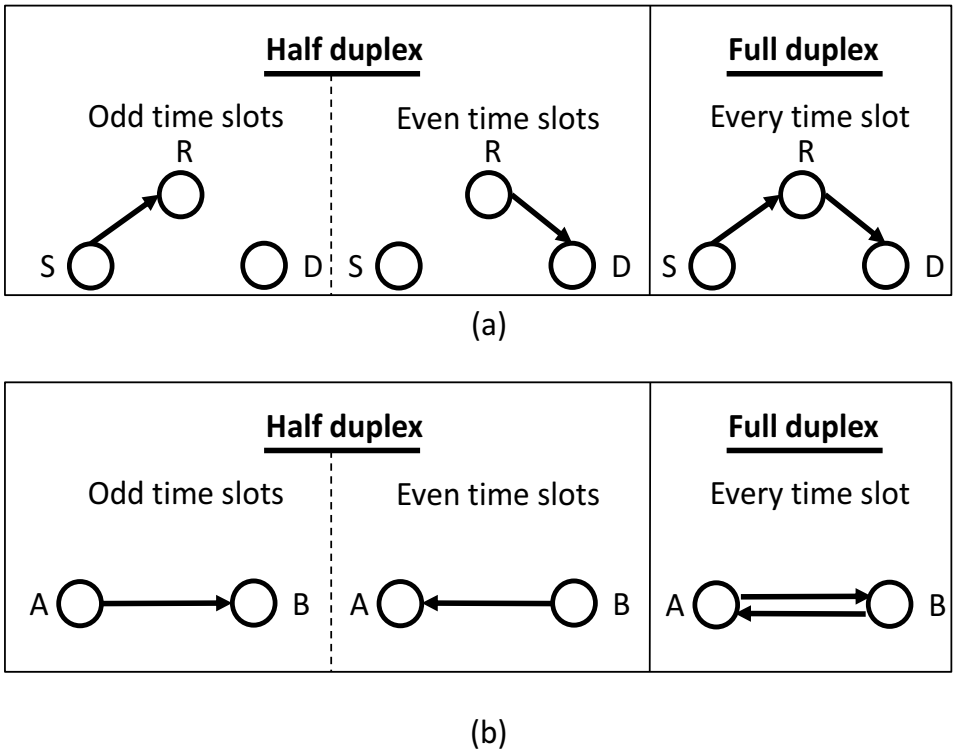

Figure 6.1: Topology of full duplex: (a) Relay topology; (b) bidirectional topology

The key challenge to design a full-duplex system is the isolation between transmission and reception. Emerging state-of-the-art techniques for self-interference cancellation include analogue and digital cancellation, and antenna separation [106] [107]. Since the beams generated by mm-wave antennas are usually highly directional, a full-duplex node with multiple directional transmission/reception links is likely to be a feasible approach to increase the spectral efficiency, assuming sufficient spatial separation of the Tx and Rx beams can be guaranteed. In Figure 6.2 for example, two simultaneous transmission/reception links are used for MNd1 and three simultaneous transmission/reception links are used for MNd2, and for each relay node, the transmit and receive antennas of a backhaul link path point to different directions (for example: MNd1→MNd2→MNd3). Note that the bidirectional full duplex is not assumed in our deployment scenario since a Rx beam is always within the Tx range for a pair of

nodes connected via a direct link (for example: MNd1↔MNd2). In this circumstance, the pair is assumed to share the resources for each direction.

In terms of the practical implementation of the advanced transceivers, the full-duplex mm-wave relay node with multiple transmission links and a bandwidth of 1GHz is still in its development phase. In addition, analogue and digital cancellation as well as the isolation provided by beamforming needs to be improved to fulfil its promises. However, rapid progress is being made in this research area and, thus it is crucial to understand its impact on system performance. Based on the mesh node's hardware capability, different numbers of simultaneous transmission/reception links using the same resource can be achieved and are compared in Section 6.4.

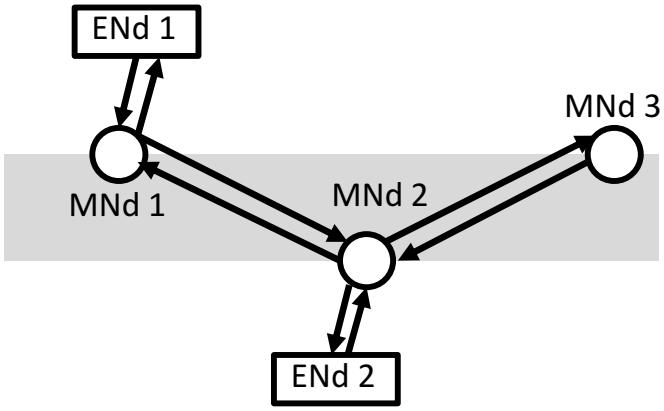

Figure 6.2: Illustration of simultaneous transmission/reception links

6.2.2 Choice of Relay Node

MNd Multi-hop

An intuitive multi-hop configuration to overcome the absence of the ENds is depicted in Figure 6.3, where a MNd can connect to an ENd with fibre connection via one or multiple MNds in a multi-hop fashion, and the MNds are not necessarily connected to their nearest peers. As discussed, sufficient spatial beam separation is

essential to achieving the relay type full duplex operation, therefore, the two MNds that carry backhaul traffic from hop to hop are assumed to be on the opposite sides of a street with a distance within 50 m as shown in Figure 6.3. Longer distances may cause the overlapping of the beams and introduce excessive signalling among MNds. This MNd multi-hop configuration requires MNds operating in full-duplex mode, hence additional complexity of the self-interference cancellation circuitry.

Wireless ENd Multi-hop

An alternative approach is shown in Figure 6.4, where a wireless ENd can be installed, and is able to connect to another ENd with fibre connectivity via one or multiple hops. Here, instead of using multi-hop links between MNds, the traffic will be aggregated at the wireless ENd, and then backhauled to/ from the wired ENds. The aggregated transmission can be directed to one or more ENds simultaneously or switched between them to improve reliability. The wireless ENds are assumed to be randomly placed on building walls between the height of a MNd (8 m) and a wired ENd (12 m) for improved spatial separation, and they can connect to another peer or ENd with fibre connection within 100 m as long as LOS is available and the beam separation can be guaranteed. No limitation on the number of simultaneous transmission links is assumed since only a small number of them are available. The rationale for this approach is that the overall backhaul deployment costs can be reduced by installing fewer wireless ENd aggregation nodes with high computational complexity and low cost MNds at every street lamp. Here, the MNds can operate in half-duplex mode receiving (downlink direction) traffic flows from the ENds, without forwarding the traffic to other peers using the same resource.

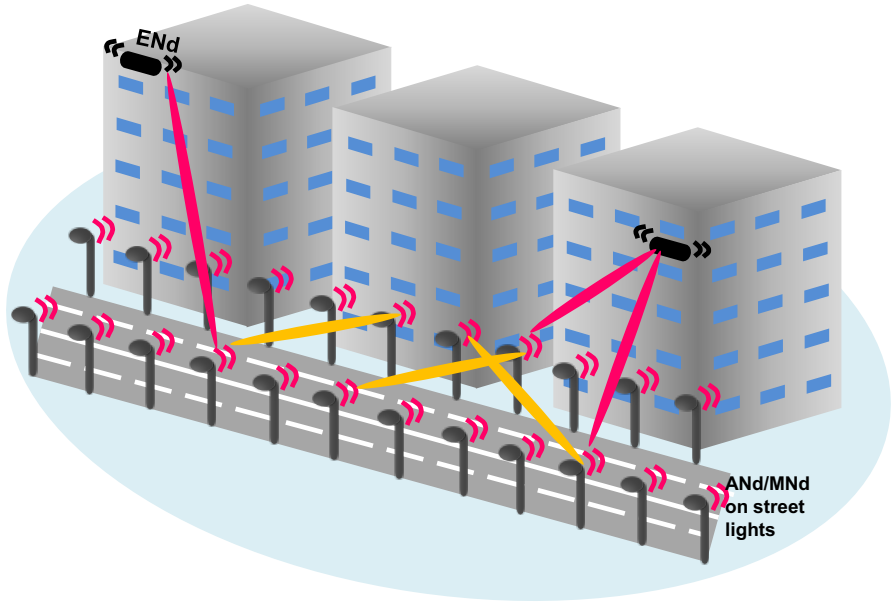

Figure 6.3: Example of using MNd as relay node

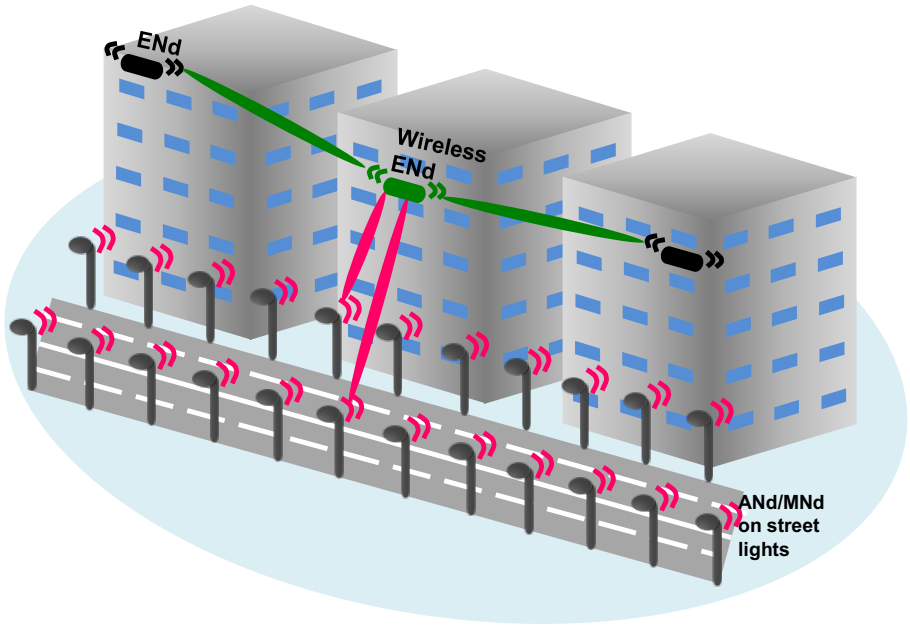

Figure 6.4: Example of using wireless ENd as relay node

6.2.3 Latency Model

Reduced latency is a key requirement of 5G. The backhaul latency depends heavily on the technology used. When multi-hop is used in the system, there is potential for latency to increase significantly, and therefore, it is important to estimate the end-to-end latency budget. For a densely deployed backhaul network, the highly directional mm-wave links can be modelled as pseudo-wired, and the end-to-end latency is mainly from the processing latency in each hop [108]. A short physical frame structure of 50 μs to 0.25 ms for mm-wave which has the control symbols located at the beginning of a frame is envisioned in [109]. This allows fast pipeline processing in the relay nodes, i.e. a relay can receive a transmission in the n_{th} sub-frame, and then forward it in the $(n+1)_{th}$ sub-frame.

Accordingly, the latency is modelled on per-hop basis in this thesis, and the end-to-end latency is the summation of the cascaded per-hop latencies as shown in Figure 6.5. There are various types of distributions have been used to model latency, such as truncated Gaussian, exponential and Gamma [110]. Here, a truncated Gaussian distribution is adopted (mean: 0.25ms; std: 0.1 ms; range: 0.05-0.45 ms). It is worth mentioning that there is no clear cut as to which distribution better describes the characteristics of the network latency since it depends on the backhaul technology used, frame structure, devices and etc. However, it is easy to adapt this per hop latency analysis approach to different models as required.

The latency for a single hop is drawn randomly from the considered distribution. Along with the number of hop information obtained from the simulations, the end-to-end cascaded latency of each flow can be calculated by:

$$\tau_i(N) = \sum_{n=1}^{N} G(\{\theta_1, \theta_2, \dots, \theta_k\}) \tag{6.1}$$

where N is the number of hops flow i passes, $G(x)$ is a random variable drawn from a considered distribution with parameters $\{\theta_1, \theta_2, \dots, \theta_k\}$.

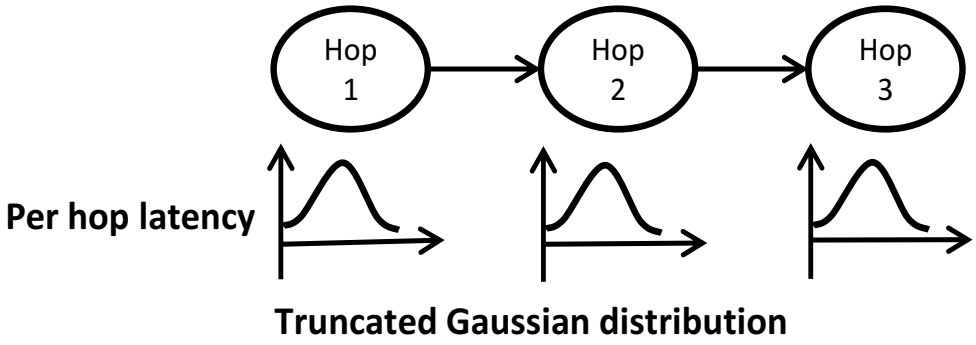

Figure 6.5: Per hop latency

6.3 Topology Management for Multi-hop Millimetre Wave Backhaul

6.3.1 Multi-hop Topology Management with QoS Control

The highly dynamic network topology in millimetre wave multi-hop networks presents major challenges to path selection and resource scheduling as it may cause unnecessary delay and lack of bandwidth for certain traffic flows. It is critical to make sure that appropriate QoS strategies are applied to different types of traffic flows. Here, a multi-hop scheme that accounts for relay node capability, end-to-end latency and QoS differentiation is proposed in this section.

Figure 6.6 shows the process of initialising the possible multi-hop path for the MNds. Each MNd is preconfigured with a list of backhaul paths to an ENd with fibre connection, and the list can be pre-programmed since the position of the nodes are practically always known. The MNds which have direct fibre ENd connections is firstly found. The MNds then search other MNds within 50 m which have direct ENd

connections and store them as the two-hop links. The search continues until a maximum of 3 multi-hop links are found to make sure the end-to-end latency is within the 1 ms 5G target.

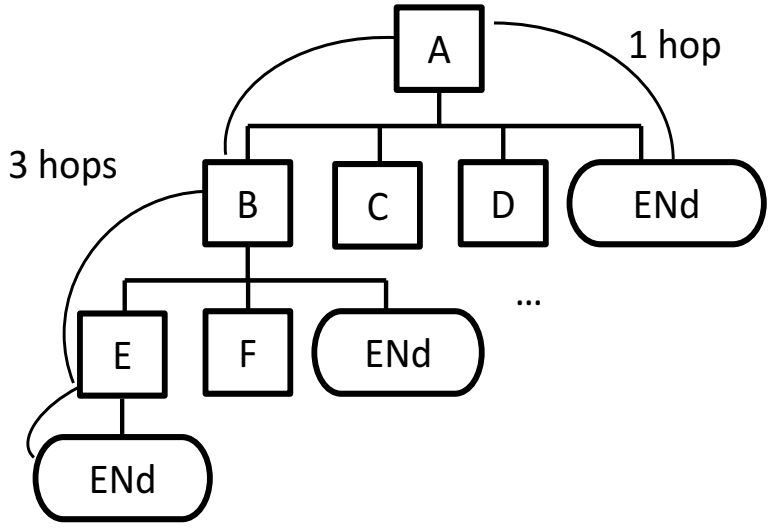

Figure 6.6: Multi-hop path initialisation

When a new/handover request which has a required capacity of r^N and a maximum number of hops h^{lim} is received at MNd i, the MNd checks the amount of available resources and number of hops on the active paths $\boldsymbol{A}_i$. Here, a maximum of two hops is used for the real-time services in order to ensure a low latency. It is assumed that the MNd maintains an up-to-date table of the capacity usage for each path to the ENds. This can be achieved by sending *RESOURCE STATUS UPDATE* message periodically among the neighbouring connected MNds. For any direct point-to-point link, the amount of resources R^{ava} which are available or can be pre-empted from the on-demand videos without compromising their QoS can be calculated by:

$$R^{ava} = R^{tot} - \sum_{m=1}^{M} r_m^{O\_ste} - \sum_{k=1}^{K} r_k^R \tag{6.2}$$

where $r_m^{O\_ste}$ $(m = 1, \dots, M)$ and r_k^R $(k = 1, \dots, K)$ are the required average data rates of the ongoing on-demand videos in steady state and real-time services on the considered link respectively. Since the real-time services have a more stringent requirement in terms of latency as well as capacity, a certain amount of capacity R^{res} can be reserved on the backhaul links for them. In this case, R^{ava} for a new on-demand video becomes:

$$R^{ava} = R^{tot} - \sum_{m=1}^{M} r_m^{O\_ste} - max\left(\sum_{k=1}^{K} r_k^R, R^{Res}\right) \tag{6.3}$$

Note that the available capacity of a path is determined by the minimum available capacity of the point-to-point links along the path, i.e.:

$$R_{ix}^{ava} = min\left\{R_{i\boldsymbol{P}_{ix}}^{ava}\right\},\ \boldsymbol{P}_{ix} \in \boldsymbol{P}_i \tag{6.4}$$

where $\boldsymbol{P}_i$ denotes the set of all possible backhaul paths for MNd i, and $\boldsymbol{P}_{ix}$ is the point-to-point links of path x for MNd i. The path management module described in Section 5.4 then searches the active path which has the maximum available capacity, and is able to meet the latency requirement, as follows:

$$l = \{l | l \in \boldsymbol{A}_i \ \wedge \ R_{il}^{ava} = \max\{R_{i\boldsymbol{A}_i}^{ava}\} \ \wedge \ h_{il} \leq h^{lim}\} \tag{6.5}$$

Path l can be chosen subject to:

$$R_{il}^{ava} \geq r^{N} \tag{6.6}$$

Otherwise, a new path from the alternative list $\boldsymbol{U}_i, \boldsymbol{U}_i \subseteq (\boldsymbol{P}_i \backslash \boldsymbol{U}_i)$ can be activated to offload the traffic, provided that the number of active links L_i on the MNd is no more than a predefined value L_{thres}. The path for the new flow can be selected based on the following condition and Equation (6.6):

$$l = \{l | l \in \boldsymbol{U}_i \ \wedge \ R_{il}^{ava} = \max\{R_{i\boldsymbol{U}_i}^{ava}\} \ \wedge \ h_{il} \leq h^{lim}\} \tag{6.7}$$

In the situation where there is no available path that is able to support the required capacity, the least loaded path of all the candidate paths $\boldsymbol{P}_i$ is chosen if the new request is an on-demand video. For a new real-time service flow, on the other hand, the amount of resources which can be pre-empted from the on-demand services on paths $\boldsymbol{P}_i$are evaluated first, and the pre-emptable resources on a point-to-point link can be obtained by:

$$R^{ava\_R} = R^{tot} - \sum_{k=1}^{K} r_k^R \tag{6.8}$$

The path set $\boldsymbol{S}_i$ which is able to accept the real-time service request can then be obtained by:

$$\boldsymbol{S}_i = \{\boldsymbol{S}_i | \boldsymbol{S}_i \subseteq \boldsymbol{P}_i \ \wedge \ R_{i\boldsymbol{S}_i}^{ava\_R} \geq r^{N} \ \wedge \ h_{i\boldsymbol{S}_i} \leq h^{lim}\} \tag{6.9}$$

Finally, the backhaul path for the real-time service is selected using the following principle:

$$l = \{l | l \in \boldsymbol{S}_i \ \wedge \ R_{il}^{ava} = max\ (R_{i\boldsymbol{S}_i}^{ava})\} \tag{6.10}$$

Algorithms 1 and 2 summarise the proposed multi-hop path selection schemes. Algorithm 1 shows the path selection scheme for a new on-demand video. Although HD on-demand videos generally require a high capacity, they have a relatively high tolerance for latency and, thus a least loaded path is chosen for them. Algorithm 2 shows the path selection scheme for a new real-time service. Since real-time services are highly sensitive to latency and usually require a fixed data rate for each flow, a limit on the maximum number of hops a real-time flow is allowed to pass through can be used. Moreover, a certain amount of capacity can be reserved for them on each backhaul link to further reduce latency and possibly blockage. Depending on whether the QoS differentiation information including capacity reservation and latency control information for real-time services can be exchanged across the network, in Section 6.4, the proposed scheme is compared with a classical capacity based routing algorithm which aims to select a path with the maximum end-to-end capacity.

Algorithm 1: Path selection for a new/handover on-demand video

Obtain possible backhaul paths for MNd i; L_{thres}; R^{Res}
if the new flow is an on-demand video **then**
  Calculate R^{ava} using Equation (6.3) and (6.4)
  Find path l within active path set $\boldsymbol{A}_i$ using Equation (6.5) and (6.6)
  if l does not exist $\wedge$ $L_i < L_{thres}$ **then**
    Find path l within alternative path set $\boldsymbol{U}_i$ Equation (6.6) and (6.7)
    if l does not exist **then**
      Select the least loaded path of all the candidate paths $\boldsymbol{P}_i$
    else
      Select path l
    end if
  else
    Select path l
  end if
end if

Algorithm 2: Path selection for a new/handover real-time service

Obtain possible backhaul paths for MNd i; L_{thres}; R^{Res}
if the new flow is a real-time service **then**
  Calculate R^{ava} using Equation (6.2) and (6.4)
  Find path l within active path set $\boldsymbol{A}_i$ using Equation (6.5) and (6.6)
  if l exists **then**
    Select path l
  elseif l does not exist $\wedge$ $L_i = L_{thres}$
    Calculate R^{ava_R} on active paths using Equation (6.8)
    $\boldsymbol{S}_i \leftarrow \{\boldsymbol{S}_i | \boldsymbol{S}_i \subseteq \boldsymbol{A}_i \quad \wedge \quad R_{i\boldsymbol{S}_i}^{ava\_R} \geq r^N \quad \wedge \quad h_{i\boldsymbol{S}_i} \leq h^{lim}\}$
    if $\boldsymbol{S}_i \neq \emptyset$ **then**
      Select path l using Equation (6.10)
    else
      Block/drop the new flow
    end if
  else
    Find path l within alternative path set $\boldsymbol{U}_i$ Equation (6.6) and (6.7)
    if l exists **then**
      Select path l
    else
      Calculate R^{ava_R} on all possible paths using Equation (6.8)
      Find path set $\boldsymbol{S}_i$ using Equation (6.9)
      if $\boldsymbol{S}_i \neq \emptyset$ **then**
        Select path l using Equation (6.10)
      else
        Block/drop the new flow
      end if
    end if
  end if
end if

6.3.2 Smallest Angle Handover Control

Several novel backhaul deployment strategies and mm-wave multi-hop schemes have been proposed so far to improve the flexibility of the backhaul network and alleviate the backhaul bottleneck and outage problems. However, the access network is also an inseparable part of completing the design of the whole network. In the context of outdoor vehicular traffic, two of the inevitable challenges for access networks need addressing are the handover control, and load balancing for small cells which has been discussed in Section 4.3. The purpose of this section is to propose a novel handover scheme which is able to dramatically reduce the handover frequency while maintaining a satisfactory level of user QoS performance. Although not the main focus of this thesis, the potential backhaul performance improvement introduced by integrating access network load balancing into the network is also explored. By distributing the traffic load among the small cells, the traffic burden on the associated backhaul links can also be reduced.

Currently, 3GPP LTE-A uses similar handover decision schemes for small cells and macro cells. These schemes perform well for handovers in conventional macro cell networks since the coverage of a macro cell is normally up to several kilometres. However, when a mobile user moves between small cells, handovers will be much more frequent even for low speed users because of the much smaller coverage of each cell. The frequent handover will cause significant signalling load to the core network entities. Hence, instead of using received signal strength as the primary criterion, the triggering of the handover decision phase should be carefully reassessed for small cell networks.

Here, a handover scheme which allows incorporating a UE's moving direction and its relative position to ANds into a standard handover procedure in order to reduce unnecessary handovers is proposed. The scheme comprises maintaining a database of the network topology, obtaining an estimate of the moving direction of a UE, and then using this information to support the handover decision. Here, since wireless backhaul networks are relatively fixed after being established, a data base of the

network topology can be easily accessible. The locations of the UEs can be obtained using global navigation satellite systems (GNSS), and then broadcasted to the nearby ANds. Alternatively, the location information can be estimated through various measurements, for example direction of arrival (DoA) and time of arrival(ToA) discussed in [111].

A decision is made to handover a UE to a target ANd such that the angle α between the UE's moving direction and the vector from the UE to the target ANd is the smallest. By selecting the target BS in this manner, the chosen ANd is more likely to lie in the same direction to where the UE is heading towards, therefore providing a relatively long serving time, and reducing unnecessary handovers. Here, it is assumed that advanced interference management techniques such as eICIC discussed in Section 2.2.2, and MIMO can be used in the access network, such that a UE can successfully connect to an ANd which is not necessarily the nearest one.

Figure 6.7 shows an example of this scenario. It is assumed that the last two (or more) location updates of the UEs can be obtained and ANds have the knowledge of the network topology information. Denote the vector from location L_{i-1} to the location of $Candidate\ AND_k$ as $v_k, (k = 1, \dots, K)$, and the UE moving vector from location L_{i-1} to location L_i as v_m. After obtaining the candidate ANd list from the UE, the serving ANd calculates and compares all the angles between $v_k, (k = 1, \dots, K)$ and v_m denoted as $\alpha_k, (k = 1, \dots, K)$, and the candidate ANd with the smallest angle α_{min} can be chosen as the target ANd. In this example, $Candidate\ AND_1$ is selected.

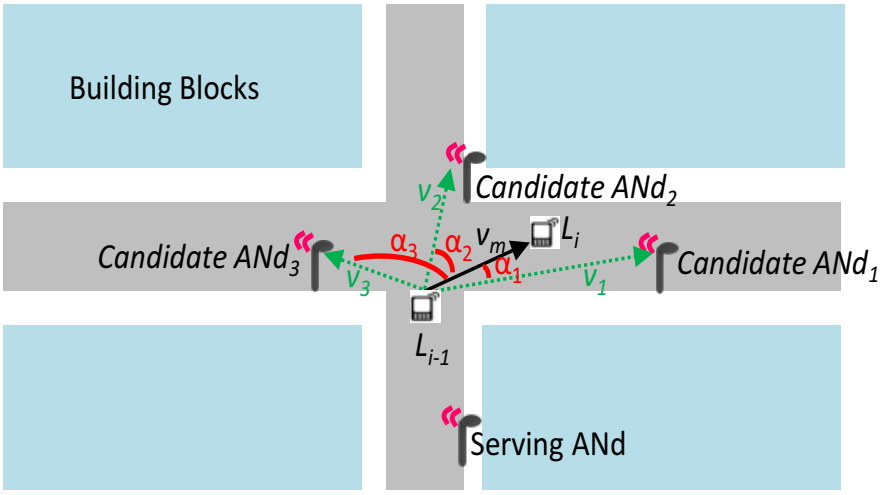

Figure 6.7: An example of choosing target BS for handover

In addition to handover frequency, the traffic load on the candidate ANds is another important factor which needs to be taken into account when designing a handover scheme due to the possible temporary hot spot created by moving crowds. Here, the smallest handover scheme can also be combined with the access network load balancing scheme discussed in Section 4.3.1.2 in order to achieve improved performance.

Figure 6.8 shows a flowchart of the handover procedure which is performed on per traffic flow basis, meaning that a UE with several ongoing traffic flows might be handed over to different ANds depending on the network conditions. This is consistent with the emerging research topic of user-centric concepts, whereby the UEs can be served by a cluster of access nodes simultaneously for improved QoS. The serving ANd configures the UE to measure the mobility level and the RSS/RSRQ of the nearby ANds, and then determine the candidate ANds for the UE. Since the access network is not the main topic in this thesis, an ANd coverage range of 50 m is used to reflect the acceptable signal quality threshold for a UE. In chapter 4, different ANd coverage range is simulated and it is shown that 50 m is a suitable distance in terms of access network load balancing as well as signalling overhead.

The smallest angle handover scheme is triggered only if there exists at least two candidate ANds, of which the traffic load along with the handover request capacity r^N is below the threshold C_{thre}. Here C_{thre} is defined as the capacity of a backhaul link, above which the selection process of an additional backhaul link path is needed. If there is a high traffic load level at each candidate node, the least loaded ANd is chosen. In this handover scheme, the handover decision is performed only when a UE moves out of the coverage of its serving ANd.

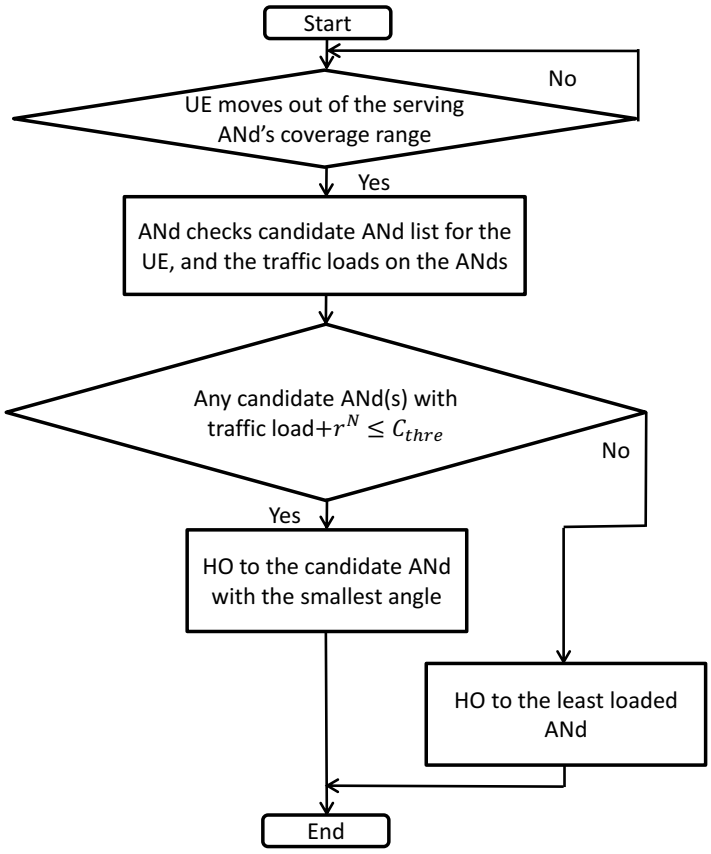

Figure 6.8: Handover procedure flow chart

6.4 Results

In this section, the results of the discussed relay node design options and proposed multi-hop topology management strategies are presented via various comparisons which isolate the factors that may affect the system performance. The underlying access network architecture, user mobility model and traffic model are described in Section 4.2 and Section 4.3.2, and if not mentioned differently, the UEs connect to their nearest ANds and no access network load balancing is assumed in order to capture the temporary hotspot effect of the outdoor vehicular traffic. In Section 6.4.1 – 6.4.3, the routing algorithm proposed in Section 6.3.1 is used, and the QoS differentiation between real-time services and on-demand videos in terms of latency is not considered.

6.4.1 Choice of Relay Node

Figure 6.9 and 6.10 compares the average perceived throughput and latency of using the two different relay node configurations discussed in Section 6.2.2: MNd multi-hop and wireless ENd multi-hop. It is assumed that the ENds are installed on building walls and there is no link outage for backhaul links. For the ease of control of the level of fibre connectivity, the percentage of MNds which have direct connection to ENds with fibre connection is defined, whereby a wide range of fibre coverage levels are chosen: 50% (low fibre coverage), 75% (medium fibre coverage), and 100% (high fibre coverage), and the MNds which have connection to ENds with fibre connection are then randomly selected. In the MNd multi-hop scenario, each MNd is assumed to be able to operate in full duplex mode and allows up to three simultaneous transmission links to achieve an appropriate trade-off between system flexibility and infrastructure deployment cost, which is discussed in Section 6.4.2. For the wireless ENd multi-hop approach, both a low density (150/km^2) and a high density (250/km^2) of wireless ENds are simulated, and the MNds operates in half-duplex mode with up to three beams pointing to the direction of the ENds.

Firstly, the performance of both schemes is affected by the level of fibre coverage, and a higher percentage of direct fibre access yield to better performance. Secondly, the MNd multi-hop scheme achieves a significantly higher throughput than that of the wireless multi-hop approach, with up to 82% throughput increase at a high traffic load of 1000 Gbps/km^2 (with 50% direct fibre access). This is due to the fact that the number of wireless ENds (150/km^2-250/km^2) is far less than the number of MNds (800/km^2), and only a limited level flexibility can be exploited, therefore the wireless ENd multi-hop links are more likely to suffer from congestion. The MNd multi-hop approach with 75% fibre coverage achieves comparable throughput compared to that of 100% fibre coverage across a wide range of traffic loads. This demonstrates that multi-hop links between MNds are highly advantageous in creating the spatial diversity which helps overcome the absence of fibre access as well as distributing the traffic load.

In contrast to the results in Figure 6.9, Figure 6.10 indicates that the wireless ENd multi-hop configuration outperforms the MNd multi-hop configuration in term of latency performance at each fibre coverage level. This is largely because the path management module selects the least loaded path which tends to have smaller number of hops in case of congestion experienced on all candidate paths, as longer links are more likely to generate bottleneck issues caused by aggregated traffic. For example, although the wireless ENd multi-hop approach with a wireless ENd density of 150/km^2 achieves the lowest latency level compared to other configurations, it results in a dramatic deterioration in average perceived throughput which cannot guarantee a smooth playback of the videos (smooth playback rate for 8K video: 80 Mbps) at medium to high traffic loads. A suitable scenario for the wireless ENd approach is where the network is largely covered by fibre, above 75% for instance. Here, an appropriate trade-off between the number of wireless ENds and the QoS provided to the UEs is likely to be found.

The plots confirm that introducing further flexibility provided by MNd multi-hop into backhaul networks makes a significantly bigger contribution to improving the QoS for the UEs, than the wireless ENd multi-hop approach. This comes with the

cost of greater investment in infrastructure. However with the rapid ongoing progress in the development of radio hardware, the infrastructure costs can be reduced dramatically. Hence, MNd multi-hop configuration is used for the rest of this chapter.

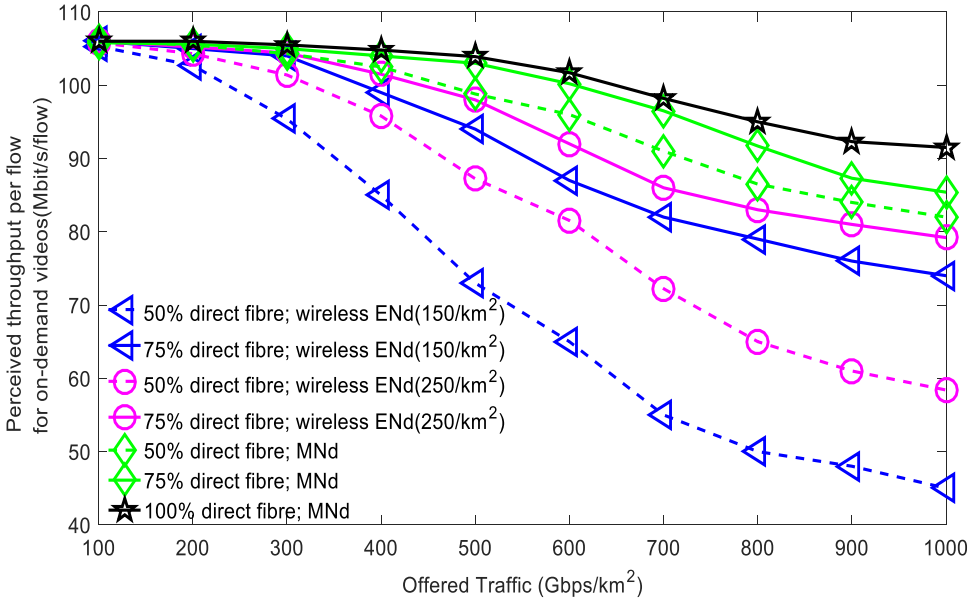

Figure 6.9: Average flow perceived throughput (8K on-demand videos) – comparison of wireless ENd multi-hop and MNd multi-hop

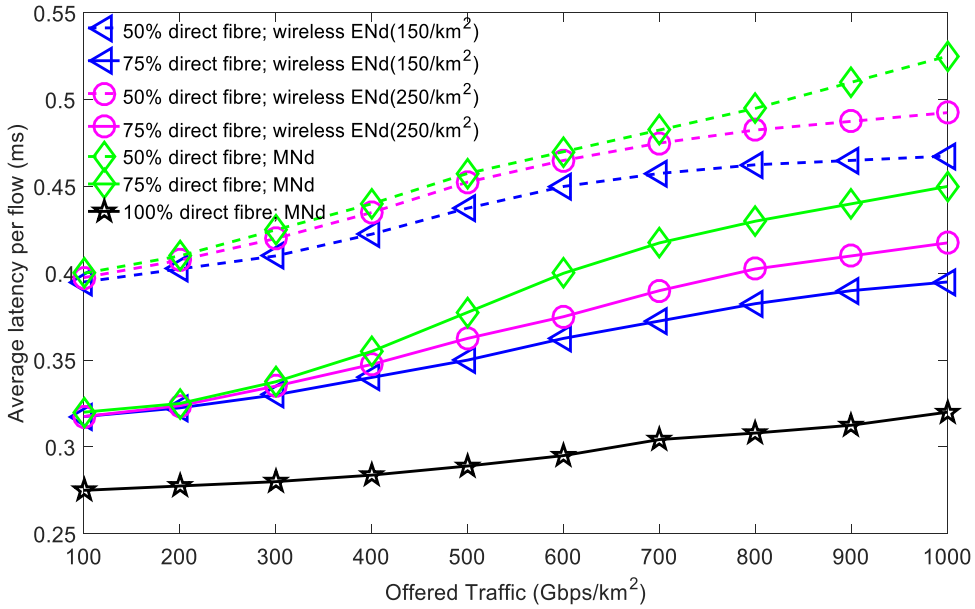

Figure 6.10: Average latency - comparison of wireless ENd multi-hop and MNd multi-hop

6.4.2 Impact of Full Duplex Capability

The utilisation of street cabinet ENds proposed in Chapter 5 is a very promising approach to providing backhaul for outdoor small cells. However so far, we have assumed each MNd has a fixed number of direct LOS backhaul links. In reality, however, street cabinets are usually randomly deployed along the streets, which means that the number of street cabinets in a certain area may not be sufficient to adequately support the local wireless backhaul traffic, especially with lower densities. The results in Section 6.4.1 indicates that multi-hop communications between MNds can be a viable solution to improve the reliability and coverage of the street cabinet backhaul network which is likely to suffer from high link outages and coverage holes. In the following sections, the dynamics of utilising street cabinet ENd and MNd multi-hop to deliver mm-wave backhaul including multi-hop topology management with QoS control as well as integrating access network aspects in terms of load balancing and handover control into the system design is investigated, but firstly, an

important detail of this type of backhaul network configuration: the appropriate choice of the number of simultaneous links enabled by the full duplex technology, is assessed in this section.

A system plan view with street fibre cabinet ENd randomly distributed along streets (density: 250/km^2) with a relatively high outage probability of 50% is depicted in Figure 6.11. Assuming each street cabinet ENd is able to generate 8 beams as discussed in Section 4.3.2, this gives approximately 2 direct access of backhaul links per MNd. Figure 6.12 - Figure 6.14 show the performance comparison of different number of simultaneous transmission links enabled by the full duplex capability (beam isolation, self-interference cancellation) of the nodes. The plots demonstrate that by allowing three simultaneous transmission links compared to two links, a higher level of flexibility is generated, which substantially improves the reliability of the backhaul network and QoS performance. However, further increasing the complexity of the nodes may significantly increase the deployment costs without bringing any notable performance improvement. For example, the performance for three simultaneous transmission links is comparable to that of unlimited number of transmission links. As shown in Figure 6.12, only a very small proportion (<5%) of the nodes use more than three simultaneous transmission links even when there is no limit on the number of transmission links that can be used. It is apparent that the trade-off between the network flexibility and deployment costs needs to be controlled by carefully considering the choice of relay nodes as well as the hardware complexity.

Although not shown, a relative low street cabinet density of 150 cabinets/km^2 and high cabinet density of 350 cabinet/km^2 with and without MNd multi-hop communication are also tested against the density of 250 cabinet/km^2 in the simulation, which yield to approximately 1 and 3 direct access of backhaul links per MNd respectively. The results confirm that the backhaul network with a higher density of street cabinet ENds performs better due to the extra capacity introduced into the network. However, a high density of 350 cabinets/km^2 without multi-hop approach does not produce any notable QoS performance improvement compared to

that of 250 cabinets/km^2 with the multi-hop approach across the traffic loads considered. Based on the above argument, a street cabinet density of 250 cabinets/km^2 and three simultaneous transmission links per MNd are assumed in the following sections.

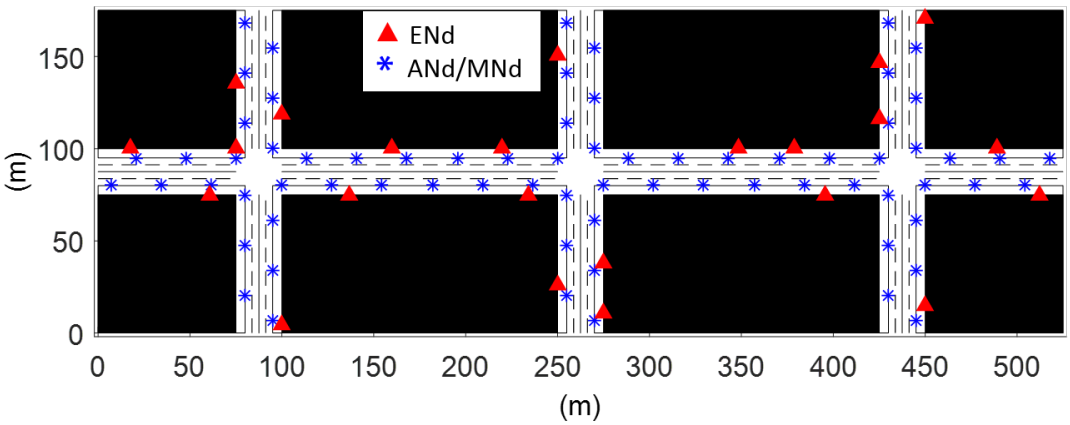

Figure 6.11: System plan view with a street cabinet density of 250/km^2

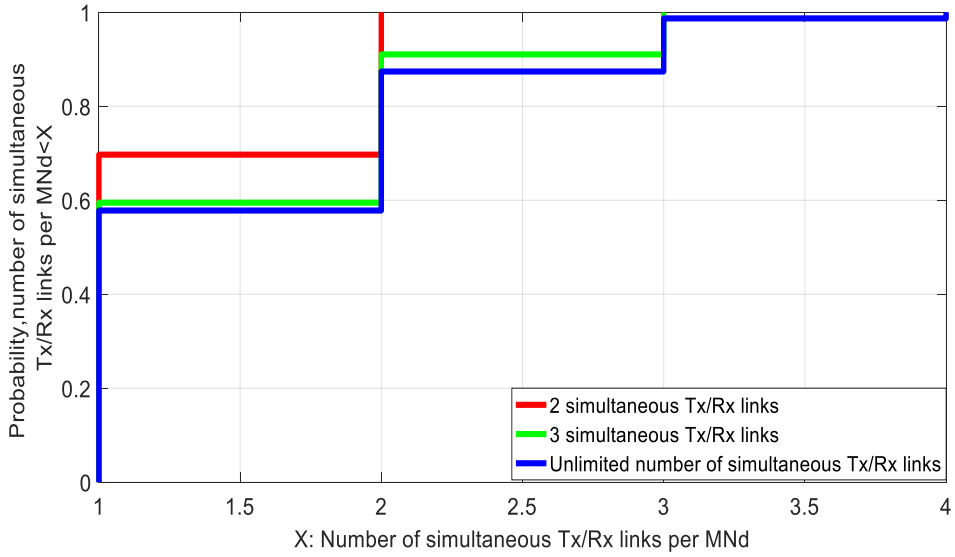

Figure 6.12: CDF – number of simultaneous transmission link per MNd

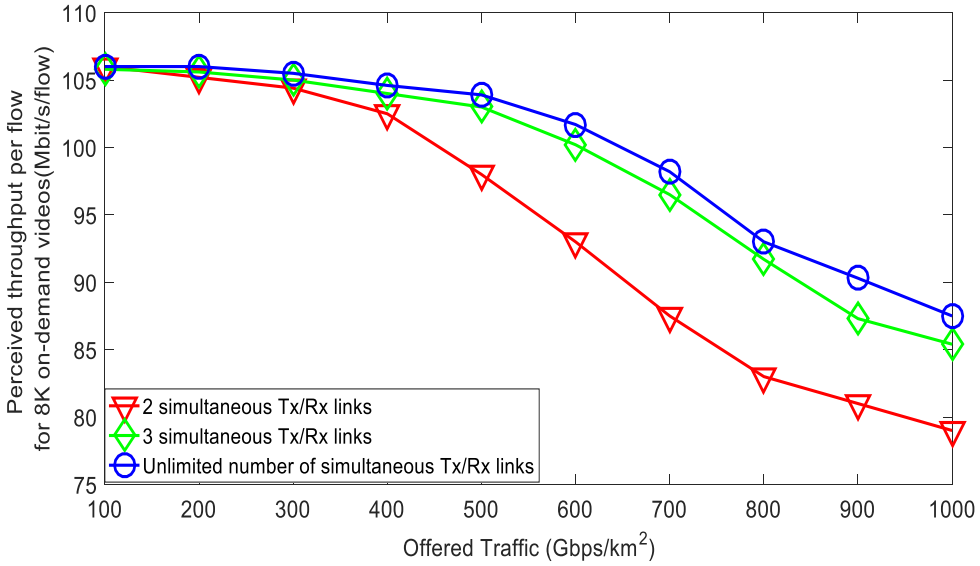

Figure 6.13: Average flow perceived throughput (8K on-demand videos) – different number of simultaneous transmission links

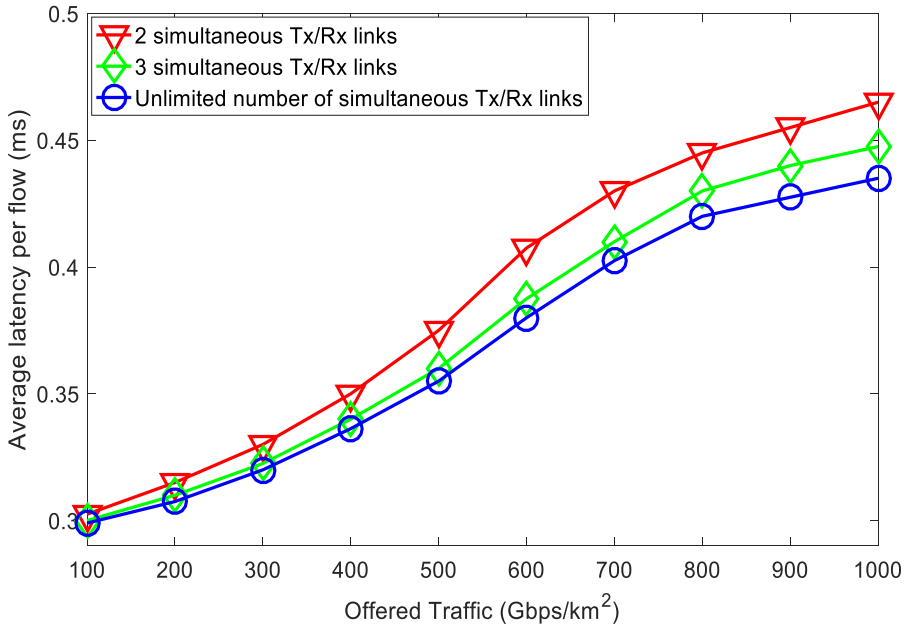

Figure 6.14: Average latency - different number of simultaneous transmission links

6.4.3 Applying Smallest Angle Handover Control and Access Network Load Balancing

The results in this section compare the performance of the smallest angle handover scheme proposed in Section 6.3.2 with a classic nearest ANd handover scheme, and show the potential QoS improvement can be achieved by introducing access node balancing. A street cabinet density of 250 cabinets/km^2 with an outage probability of 50%, and three simultaneous transmission links per MNd - MNd multi-hop communication are assumed based on the discussion in Section 6.4.2. Here, four schemes are discussed:

- *Multi-hop & smallest angle handover (HO)* – a UE is always handed over to the candidate ANd which can provide the smallest angle $\alpha_{\min}$. No access network load balancing is assumed.

- *Multi-hop & nearest ANd* – this is the benchmark scheme where by the UEs connect to their nearest ANds and no access network load balancing is assumed.

- *Multi-hop & access network load balancing (LB)* - a UE connects to the least loaded ANd within its connecting range (50m).

- *Multi-hop & access LB & smallest angle HO* – a UE connects to an ANd using the smallest angle criterion as long as the traffic loads on the candidate ANds are below the threshold discussed in Section 6.3.2. Otherwise, the UE connects to the least loaded ANd within its connecting range.

Figure 6.15 and Figure 6.16 show the comparison of the average number of handovers each flow experiences and the average perceived throughput by using different schemes. They demonstrate that the *Multi-hop & smallest angle HO* scheme dramatically reduces the number of handovers by 35% - 54% compared to the classic *Multi-hop & nearest ANd* scheme with only marginal performance deterioration. With access network load balancing, the *Multi-hop & access LB* scheme can further

improve the QoS performance due to the extra level of flexibility introduced into the system. In addition, the number of handovers is significantly less than that of the benchmark since a handover procedure is only performed when a UE moves out of the service area of its serving ANd instead of always connecting to the ANd that provides the best quality link.

Moreover, the plots show that the *Multi-hop & access LB & smallest angle HO* scheme is able to provide a good trade-off between QoS performance and handover frequency. At low to medium traffic loads (below 700 Gbps/km^2), the scheme can achieve comparable a handover frequency level compared to the *Multi-hop & smallest angle HO* scheme, which provides up to 18.7% reduction in the number of handovers compared to the *Multi-hop & access LB* scheme without noticeable performance loss. At high traffic loads, the least loaded ANd is more likely to be chosen since the network becomes more congested. In this case, the *Multi-hop & access LB & smallest angle HO* scheme exhibits similar QoS and handover frequency performance as the *Multi-hop & access LB* scheme.

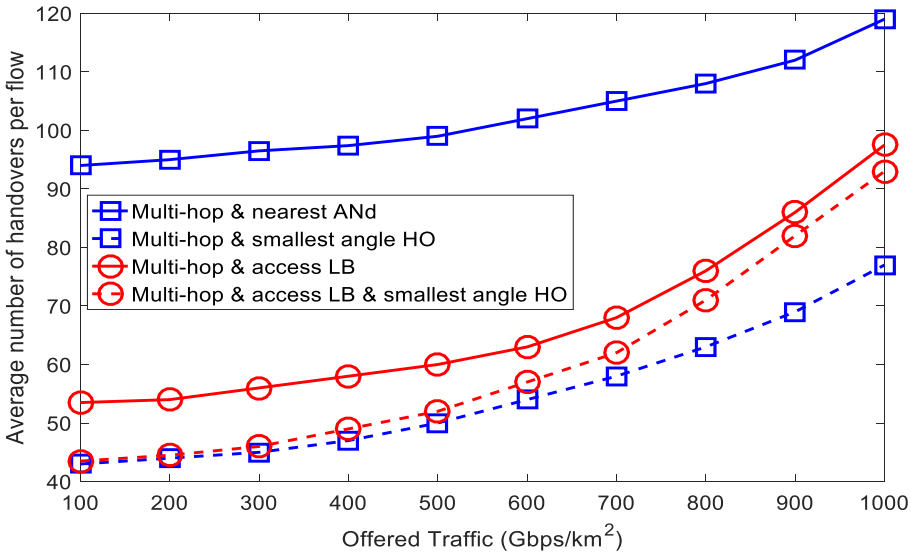

Figure 6.15: Average number of handovers - access network load balancing and handover control

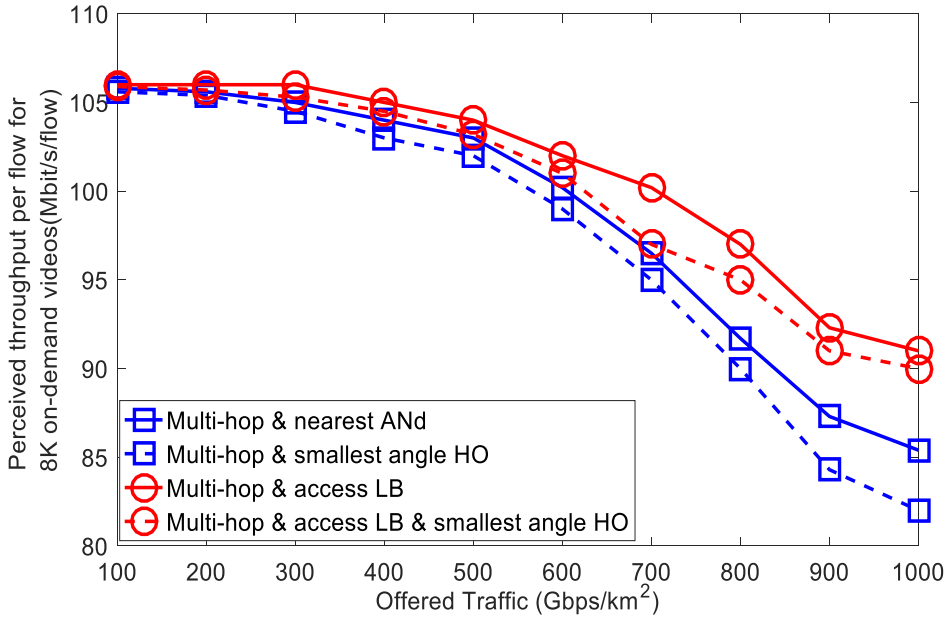

Figure 6.16: Average flow perceived throughput (8K on-demand videos) – access network load balancing and handover control

6.4.4 Topology Management with QoS Control

Figure 6.17 - Figure 6.20 compare the performance of topology management strategies with different levels of QoS control discussed in Section 6.3.1. Here, the same underlying deployment scenario, whereby a street cabinet density of 250 cabinets/km^2 with an outage probability of 50%, three simultaneous transmission links per MNd and MNd multi-hop communication are assumed. Depending on whether the QoS differentiation information including capacity reservation and latency control information for real-time services can be exchanged across the network, a certain amount of resources can be reserved and the maximum number of hops can be applied for a real-time service. Here, three different capacity reservation levels for real-time services are compared: 0% where only latency control for the real-time services are considered, medium (25%) and high (50%) level. If the QoS differentiation information is not available, the candidate path with the maximum end-to-end capacity is chosen.

The plots show that all the schemes can provide satisfactory performance in terms of perceived throughput (smooth playback rate for 8K video: 80 Mbps), blocking probability and dropping probability. With the latency control scheme, the path management module chooses the path with a lower number of hops for the real-time traffic, and therefore the latency for real-time traffic is lower than the on-demand traffic compared to the maximum capacity scheme. By allowing reserved capacity for real-time traffic on the backhaul links, further latency reduction can be achieved for the real-time services, ranging from 4% - 21% compared to the maximum capacity scheme. However, there is a trade-off between the latency of real-time traffic and the performance (latency and perceived throughput) of on-demand traffic. A higher percentage capacity reservation for real-time traffic results in performance deterioration for on-demand traffic. This is due to the fact that the on-demand traffic is more likely to use the path with higher number of hops (higher latency) and has less available capacity to use (lower perceived throughput). Even so, the perceived throughputs for the reservation schemes are higher than the required smooth playback data rate for an 8K on-demand videos. Hence, it is possible to trade off the on-demand performance for lower real-time latency.

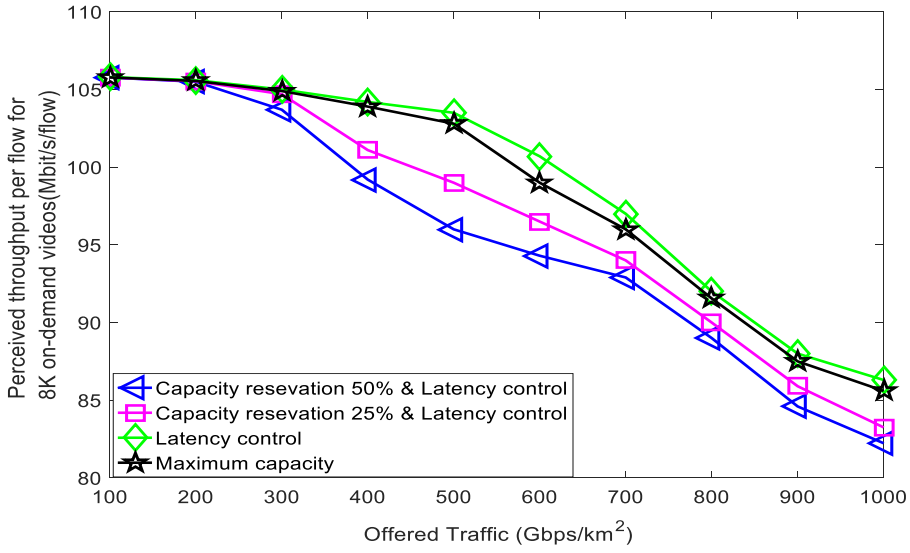

Figure 6.17: Average flow perceived throughput (8K on-demand videos) – different multi-hop topology management schemes with QoS control

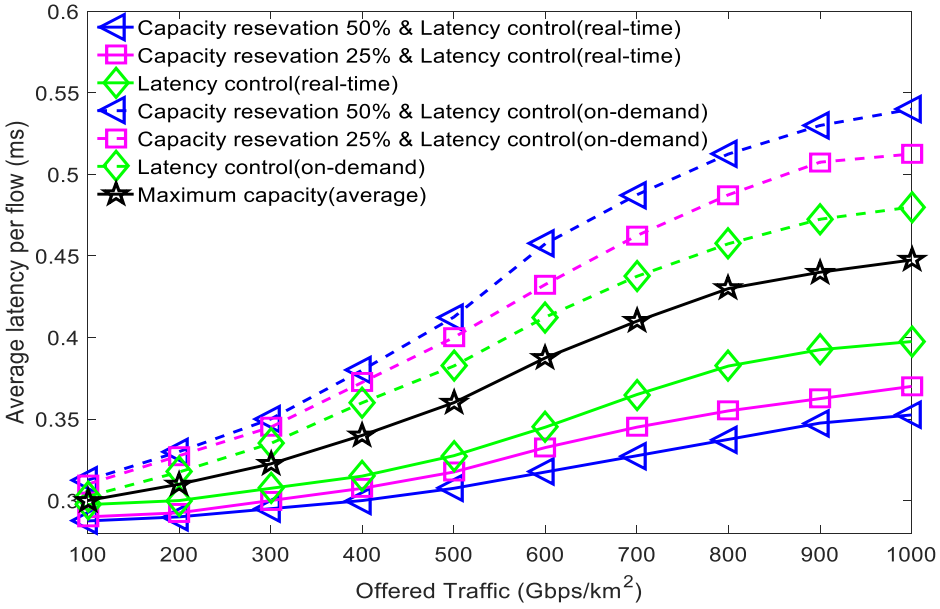

Figure 6.18: Average latency - different multi-hop topology management schemes with QoS control

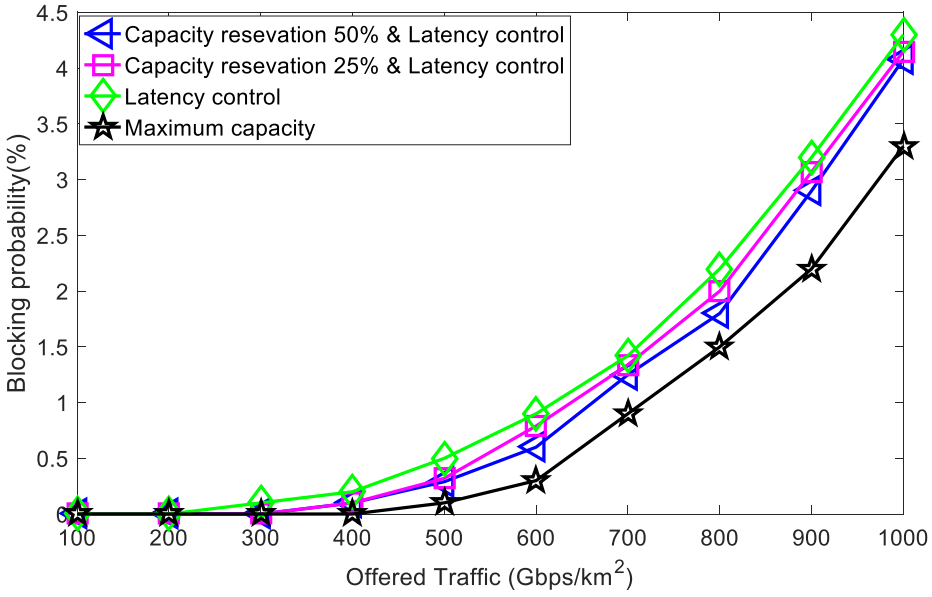

Figure 6.19: Blocking probability (real-time services) - different multi-hop topology management schemes with QoS control

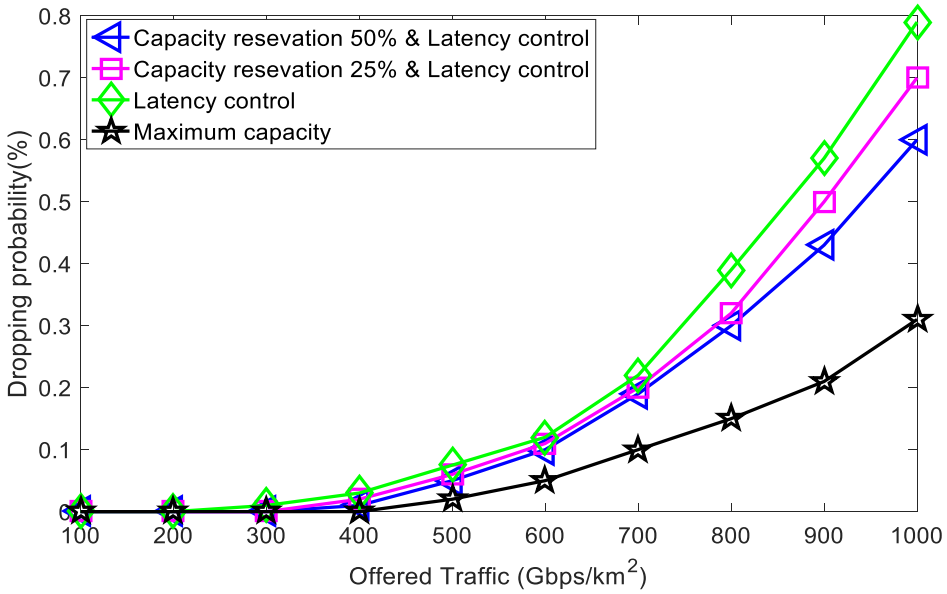

Figure 6.20: Dropping probability (real-time services) - different multi-hop topology management schemes with QoS control

6.5 Conclusion

In this chapter, various aspects of deploying multi-hop mm-wave backhaul network for outdoor small cells have been investigated, and it has become evident that one of the most desired 5G backhaul features is flexibility. The flexibility can be achieved in a number of ways. Firstly, the multi-hop configurations proposed are effective in overcoming the lack of fibre access as well as distributing traffic loads. In particular, the path diversity created by using multi-hop between MNds makes a more significant contribution to improving the system performance in the dynamic vehicular scenario considered, with up to 82% throughput increase than that of the wireless ENd multi-hop approach at a high traffic load of 1000 Gbps/km^2. Full-duplex capability is another key feature for future mm-wave backhaul mesh networks enabling multi-flow relaying at the MNds, which significantly improve the flexibility of the system. The results have demonstrated that for a mm-wave backhaul network with street cabinet density of 250 cabinets/km^2, three simultaneous transmission

links per MNd are sufficient to provide satisfactory QoS for the UEs. Another critical factor of a backhaul network design discussed is the application of appropriate QoS strategies to support different applications such as bandwidth/latency constraint real-time videos. The proposed QoS control schemes can effectively reduce the latency for real-time services, and by reserving capacity for real-time traffic on backhaul links, lower latency can be achieved (4% - 21%).

In terms of combining access network design for outdoor small cells with vehicular traffic, a smallest angle handover algorithm has been proposed which is able to dramatically reduce the number of handovers by 35% - 54% compared to the classic connecting to the nearest ANd scheme without compromising the QoS of the UEs. Moreover, by using access network load balancing, better performance can be achieved. This is due to the fact that the traffic from congested hotspots can be spatially distributed to neighbouring areas, and therefore offload the traffic burden on the backhaul links.

Chapter 7. Conclusions and Future Work

Contents

7.1 Conclusions

This thesis has highlighted that a high level of flexibility has become a fundamental requirement in order to provide a reliable and robust backhaul network which can accommodate various use cases. The work presented in this thesis has therefore focused on designing and evaluating novel wireless backhaul architectures that serves a dense 5G radio access network, and has been guided by the hypothesis mentioned in Chapter 1:

"Introducing and exploiting an appropriate level of flexibility in wireless backhaul networks can improve network utilisation and user experience, enabling improvements in capacity, reliability and latency."

First, a dynamic two-tier SDN architecture and the concept of NFV has been employed in backhaul networks to facilitate network management tasks and on-demand resource provision in Chapter 3. Two use cases were presented of the proposed SDN architecture to address two increasingly popular research area in wireless backhaul network: multi-tenancy dynamic resource sharing with QoS guarantee and energy-aware topology management for backhaul networks. I t has demonstrated that a higher level of flexibility in terms of resource allocation can be achieved with SDN and NFV technologies. The proposed architecture also enables

exploiting the backhaul diversity within the network, and thus delivers efficient resource utilisation and high energy savings.

Next, the thesis has presented wireless backhaul architectures that serve ultra-dense outdoor small cells installed on street-level fixtures. The upper bound performance that a massive MIMO backhaul network can provide has been investigated in Chapter 4. An access network load balancing scheme has been applied to the system in order to alleviate the bottleneck problem caused by the uneven traffic loads. However, at high traffic loads, the massive MIMO backhaul cannot provide sufficient capacity to meet the requirement due to the lack of frequency resources. Mm-wave backhaul on the other hand, offering abundant spectrum and therefore multi-gigabit data rates can significantly improve system performance.

Mm-wave backhaul links provided by ENds installed on building walls can provide reliable backhaul links and satisfactory QoS. However, in the situation where installing adequate ENds with fibre connections on building walls incurs high costs and deployment challenges, utilising existing street fibre cabinets is an extremely promising solution to provide backhaul, when link diversity is used, to overcome individual high link outages. Moreover, the street level backhaul links can provide extra flexibility to the system. It has been shown in Chapter 5 that backhaul networks can even rely solely on the street cabinet ENds, assuming sufficient cabinet ENd redundancy can be built into the system.

In order to further increase the flexibility and coverage of the street fibre cabinet backhaul network, multi-hop backhaul configurations have been proposed in Chapter 6. The path diversity created by using multi-hop between MNds has been shown to be highly advantageous in overcoming the negative impact of high link outages. Furthermore, topology management algorithms which aim to reduce the handover frequency and achieve appropriate QoS for different traffic types including bandwidth/latency constraint real-time videos are proposed.

The proposed backhaul architectures and schemes are evaluated through large scale simulations, and in some cases theoretical analysis, shown significant improvements

in accommodating various use cases, resource utilisation and reliability in wireless backhaul networks when a higher level of flexibility can be exploited.

7.2 Novel Contributions

This thesis has proposed novel backhaul architecture options and evaluated various aspects of designing backhaul solutions that serves a dense 5G radio access network. These highlight the need to design flexible backhaul networks which can provide tailored services for different deployment scenarios that support a diverse range of applications with different performance requirements. The details of the original contributions are given in this section.

Providing Backhaul Connectivity Utilising Street Fibre Cabinets

A novel backhaul architecture where Edge Nodes (ENds) are co-located with street fibre cabinets to provide a cost-effective and robust millimetre wave backhaul network for outdoor dense small cells has been proposed. The access network is backhauled via wireless Mesh Nodes (MNd) linked directly to one or more ENds located on existing street fibre cabinets This approach can dramatically reduce the costs and deployment time compared to a more conventional approach whereby the ENds are typically mounted on high buildings, for example the BuNGEE backhaul architecture [34].

Exploiting Backhaul Flexibility in Multi-hop Backhaul Networks

Two novel multi-hop mm-wave backhaul architectures have been proposed: MNd multi-hop; and wireless ENd multi-hop, where instead of using multi-hop links between MNds, the traffic is aggregated at a wireless ENd, and then backhauled to/ from the wired ENds. The latter approach can reduce the overall backhaul deployment costs by installing fewer wireless ENd aggregation nodes with high computational complexity and low cost MNds at every street lamp. This is motivated by the fact that although a large amount of previous work has addressed various aspects of multi-hop routing algorithms, such as link quality, system capacity, hop

count, and end-to-end delay, the impact of different mm-wave relay node design options on system performance and user experience remains largely unclear [99] [101].

Angle Based Handover Control

A smallest angle handover scheme which is able to dramatically reduce the handover frequency while maintaining a satisfactory level of user QoS performance has been proposed. Instead of using received signal strength as the primary criterion, another parameter angle α which considers a UE's moving direction and the relative positions from ANds to the UE, is incorporated in a standard handover procedure. This approach can significantly reduce the number of handovers compared to the state-of-art signal strength based handover scheme [20].

Backhaul Topology Management with QoS Control

A backhaul topology management scheme aiming to apply appropriate QoS strategies for different types of traffic flows has been proposed for mm-wave backhaul networks. The scheme takes into account the availability, capacity and the number of hops of the candidate backhaul paths in a highly dynamic network topology. The proposed method can effectively reduce the latency for real-time services, and by allowing reserved capacity for real-time traffic on backhaul links, lower latency can be achieved compared to a traditional maximum end-to-end capacity aware scheme.

Wireless Backhaul Deployments for Ultra-Dense Outdoor Small Cells

In order to achieve high flexibility and capacity for future outdoor ultra-dense small cells while accommodating various use cases and deployment constraints, a mm-wave backhaul architecture has been proposed. It is compared with a massive MIMO backhaul approach whereby the upper bound capacity and access network load balancing are considered. Although a large amount of research on mm-wave and massive MIMO networks focuses on the hardware and physical layer design including antenna array implementation, beamforming and modulation schemes, the

system level evaluation of the backhaul solutions for ultra-dense small cells using the mentioned two key technologies has not been investigated [32] [41]. For a congested network, mm-wave backhaul is shown to be able to significantly improve the system performance compared to a massive MIMO approach using sub-6 GHz band due to its abundant spectrum.

SDN for Wireless Backhaul Network Management

A novel dynamic two-tier SDN architecture has been employed in backhaul network to facilitate complex network management tasks including multi-tenancy resource sharing and energy-aware topology management. The proposed architecture can deliver efficient resource utilisation, and QoS guarantees with marginal QoS compromises compared to an all centralised framework which requires on-the-fly control information exchange [65]. The trade-off between scalability and performance of the system can be controlled by adjusting the time resolution.

7.3 Future Work

This section presents recommendations for future work, mainly built on the work of this thesis. The potential related applications and the extension of the ideas proposed in the thesis are discussed.

Energy Efficient Backhauling

Although energy-efficient topology management strategies have been investigated on a large scale in wireless access network, relatively little attention has been paid to backhaul energy consumption management aspects. In Chapter 3, a backhaul link selection scheme is proposed for BuNGee architecture in order to minimise the backhaul network energy consumption. There, the base station energy consumption models described in EARTH project was used [47]. However for mm-wave backhaul architectures, the energy efficiency aspect has not been studied. To our knowledge, the research on energy consumption models for mm-wave base stations is still in its infancy, especially with the rapid ongoing progress in the development of radio

hardware. Hence it would be interesting to assess the benefits of switching part of the backhaul network off or switching them to adaptive sleep mode, and apply backhaul link scheduling schemes to evaluate the energy efficiency.

Machine Learning Techniques for Backhaul Topology Management

The complexity of future mobile networks presents new challenges for network management. Although centralised controlled network can offer high resource utilisation, it is also worth considering distributed solutions for flexibility and scalability reasons. For the highly dynamic network environment with vehicular traffic considered in Chapter 4, machine learning techniques including reinforcement learning and transfer learning can be used to improve performance. For example, the historical information of the mobility patterns can be used to predict the user movement and traffic load conditions for routing the backhaul network. Another possible method is a hybrid approach, which can possibly utilise both a centralised control, such as SDN to guide the learning process, and a localised distributed learning scheme to reduce the unnecessary signalling. Transfer learning is an effective approach to improve the localized learning process of target cognitive agent.

In-band wireless backhauling

In this thesis, the access network and backhaul network are assumed to use different frequency band. Another emerging research topic of in-band wireless backhaul can be studied, whereby the access and backhaul links operates in the same frequency band. When the in-band wireless backhaul approach is used for small cells, the joint management of the resources and interference for the access and backhaul networks are crucial.

Millimetre Wave Link Outage Caused by Moving Obstacles

The temporary link outage of mm-wave connections caused by moving obstacles are assumed to be independent between different links. However, the outage probabilities for adjacent backhaul links may not be fully independent in reality. For instance, neighbouring links may be blocked by a long vehicle. The temporary

outage durations for backhaul links may also vary drastically with the time of day and between different locations. Hence, another approach which can be explored would be to model the correlated temporary link outages either theoretically or explicitly based on site-specific traffic and geographic data.

Millimetre Wave Massive MIMO

The upper bound performance of sub-6 GHz massive MIMO backhaul has been evaluated in Chapter 4. Recently, the combination of mm-wave and massive MIMO has drawn an increasing interest. In addition to increased capacity and spectral efficiency, such systems also have compact dimensions. Signal processing issues including channel estimation, precoding need to be investigated. Moreover, innovative antenna architectures will be required to handle the wide channels.

Glossary

ANd	Access Node
ABS	Almost Blank Subframe
ARP	Allocation and Retention Priority
API	Application Program Interface
BBU	Baseband Unit
BS	Base Station
CA	Carrier Aggregation
CAPEX	Capital Expenditure
CB	Coordinated Beamforming
CoMP	Coordinated Multipoint
CS	Coordinated Scheduling
C-RAN	Cloud-RAN
CRE	Cell Range Expansion
D2D	Device-to-Device
D-RAN	Distributed-RAN
eICIC	enhanced Inter-cell Interference Coordination
eNB	evolved Node B
ENd	Edge Node
FSO	Free Space Optics
FTP	File Transfer Traffic
FTTx	Fibre to the X
GBR	Guaranteed Bit Rate

HetNet	Heterogeneous Network
JP	Joint Processing
KPI	Key Performance Indicators
LOS	Line-of-Sight
MIMO	Multiple Input Multiple Output
M2M	Machine to Machine
NFV	Network Function Virtualisation
NLOS	Non-Line-of-Sight
MNd	Mesh Node
MS	Mobile Station
OPEX	Operating Expense
QCI	QoS Class Identifiers
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RRH	Remote Radio Head
RSRQ	Reference Signal Received Quality
RSRP	Reference Signal Received Power
RSSI	Received Signal Strength Indicator
SDN	Software Defined Networking
SINR	Signal-to-Interference plus Noise Ratio
TL	Transfer Learning
TVWS	Television White Space
UE	User Equipment
UHD	Ultra-High Definition

List of References

[1] *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*, ITU-R M.2083-0, Geneva, Sep. 2015.

[2] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," in *IEEE Communications Magazine,* vol. 52, no. 2, pp. 82-89, Feb. 2014.

[3] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," in *IEEE Communications Magazine,* vol. 52, no. 2, pp. 90-96, Feb. 2014.

[4] Cisco, "Cisco visual networking index: global mobile data traffic forecast update, 2016–2021", *Cisco White Paper*, Feb. 2017.

[5] A. Rao, A. Legout, Y.-s. Lim, D. Towsley, C. Barakat, and W. Dabbous, "Network characteristics of video streaming traffic," in *ACM Proceedings of the Seventh Conference on emerging Networking Experiments and Technologies*, pp. 25:1-25:12, Tokyo, Japan, Dec. 2011.

[6] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5G systems," in *IEEE Communications Magazine,* vol. 53, no. 9, pp. 10-17, Sep. 2015.

[7] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," in *IEEE Communications Magazine,* vol. 52, no. 5, pp. 26-35, May 2014.

[8] T. Wen and P. Zhu, "5G: A technology vision," *Huawei Technologies Co. Ltd. White Paper,* 2013.

[9] NGMN Alliance. (Feb. 2015). *NGMN 5G White Paper*. [Online]. Available: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf. [Accessed: Feb. 1, 2017].

[10] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," in *IEEE Journal on Selected Areas in Communications,* vol. 30, no. 3, pp. 497-508, Apr. 2012.

[11] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," in *IEEE/ACM transactions on networking,* vol. 22, no. 1, pp. 137-150, 2014.

[12] A. Liu, V. K. N. Lau, L. Ruan, J. Chen, and D. Xiao, "Hierarchical radio resource optimization for heterogeneous networks with enhanced inter-cell interference coordination (eICIC)," in *IEEE Transactions on Signal Processing,* vol. 62, no. 7, pp. 1684-1693, Apr. 2014.

[13] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," in *IEEE Communications Magazine,* vol. 50, no. 2, pp. 148-155, Feb. 2012.

[14] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," in *IEEE Transactions on Wireless Communications,* vol. 12, no. 6, pp. 2762-2775, June 2013.

[15] *Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects (Release 11)*. 3GPP TR 36.819 V11.2.0, Sep. 2013.

[16] Y. Chenyang, H. Shengqian, H. Xueying, and A. F. Molisch, "How do we design CoMP to achieve its promised potential?," in *IEEE Wireless Communications,* vol. 20, no. 1, pp. 67-74, Feb. 2013.

[17] T. Biermann, L. Scalia, C. Changsoon, W. Kellerer, and H. Karl, "How backhaul networks influence the feasibility of coordinated multipoint in cellular networks [Accepted From Open Call]," in *IEEE Communications Magazine,* vol. 51, no. 8, pp. 168-176, Aug. 2013.

[18] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," in *IEEE Journal on Selected Areas in Communications,* vol. 28, no. 9, pp. 1479-1489, Dec 2010.

[19] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," in *IEEE Transactions on Wireless Communications,* vol. 12, no. 1, pp. 248-257, Jan 2013.

[20] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober and L. Hanzo, "User association in 5G networks: A survey and an outlook," in *IEEE Communications Surveys & Tutorials,* vol. 18, no. 2, pp. 1018-1044, Secondquarter 2016.

[21] H. Beyranvand, W. Lim, M. Maier, C. Verikoukis, and J. A. Salehi, "Backhaul-aware user association in FiWi enhanced LTE-A heterogeneous networks," in *IEEE Transactions on Wireless Communications,* vol. 14, no. 6, pp. 2992-3003, June 2015.

[22] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," in *IEEE Transactions on Wireless Communications,* vol. 15, no. 5, pp. 3251-3268, May 2016.

[23] *Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility Enhancements in Heterogeneous Networks*. 3GPP TR 36.839 V11.1.0, Sep. 2012.

[24] S. Sadr and R. S. Adve, "Handoff rate and coverage analysis in multi-tier heterogeneous networks," in *IEEE Transactions on Wireless Communications,* vol. 14, no. 5, pp. 2626-2638, May 2015.

[25] A. Ulvan, R. Bestak, and M. Ulvan, "Handover procedure and decision strategy in LTE-based femtocell network," in *Telecommunication systems,* pp. 1-16, Apr. 2013.

[26] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: challenges and research advances," in *IEEE Network,* vol. 28, no. 6, pp. 6-11, 2014.

[27] O. Tipmongkolsilp, S. Zaghloul, and A. Jukan, "The evolution of cellular backhaul technologies: current issues and future trends," in *IEEE Communications Surveys & Tutorials,* vol. 13, no. 1, pp. 97-113, 2011.

[28] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," in *IEEE Communications Surveys & Tutorials,* vol. 16, no. 4, pp. 2231-2258, 2014.

[29] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," in *IEEE Wireless Communications,* vol. 22, no. 5, pp. 22-31, 2015.

[30] M. Fitch, M. Nekovee, S. Kawade, K. Briggs, and R. MacKenzie, "Wireless service provision in TV white space with cognitive radio technology: A telecom operator's perspective and experience," in *IEEE Communications Magazine,* vol. 49, no. 3, 2011.

[31] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-advanced*. John Wiley & Sons, 2011.

[32] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine,* vol. 52, no. 2, pp. 186-195, 2014.

[33] B. Li, D. Zhu, and P. Liang, "Small cell in-band wireless backhaul in massive MIMO systems: A cooperation of next-generation techniques," in *IEEE Transactions on Wireless Communications,* vol. 14, no. 12, pp. 7057-7069, 2015.

[34] M. Goldhamer, Z. e. Gross, N. Haklai, and O. Marinchenco, "BuNGee Deliverable D1. 2: Baseline BuNGee Architecture," *Beyond Next Generation Mobile Networks Project* Tech. Rep. V1.0, Nov. 2010.

[35] *Attenuation by atmospheric gases*, Recommendation ITU-R P.676-10, Geneva, 2013.

[36] *Characteristics of precipitation for propagation modelling*, Recommendation ITU-R P.837-6, Geneva, 2012.

[37] *Specific attenuation model for rain for use in prediction methods*, Recommendation ITU-R P.838-3, , Geneva, 2005.

[38] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," in *IEEE Transactions on Communications,* vol. 63, no. 9, pp. 3029-3056, 2015.

[39] A. A. Artemenko and R. O. Maslennikov, "System and method of relay communication with electronic beam adjustment," Patent US9391688 B2, July 2016.

[40] J. A. Louberg, P. A. Johnson, and E. Korevaar, "Wireless communication system," Patent US7769347 B2, 2010.

[41] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," in *Wireless Networks,* vol. 21, no. 8, pp. 2657-2676, 2015.

[42] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks," in *IEEE Transactions on Communications,* vol. 61, no. 10, pp. 4391-4403, 2013.

[43] S. Bloom, E. Korevaar, J. Schuster, and H. Willebrand, "Understanding the performance of free-space optics [Invited]," in *Journal of optical Networking,* vol. 2, no. 6, pp. 178-200, 2003.

[44] I. I. Kim and E. J. Korevaar, "Availability of free-space optics (FSO) and hybrid FSO/RF systems," in *ITCom 2001: International Symposium on the Convergence of IT and Communications*, International Society for Optics and Photonics, Denver, CO, United States, Nov. 2001, pp. 84-95.

[45] K. Fang-Chun, F. A. Zdarsky, J. Lessmann, and S. Schmid, "Cost-Efficient Wireless Mobile Backhaul Topologies: An Analytical Study," in *2010 IEEE Global Telecommunications Conference (GLOBECOM)*, Miami, FL, 2010, pp. 1-5.

[46] R. Nadiv and T. Naveh, "Wireless backhaul topologies: Analyzing backhaul topology strategies," *Ceragon White Paper,* 2010.

[47] G. Auer, O. Blume, V. Giannini, I. Godor, M. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, and W. Wajda, "D2. 3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," in *INFSOICT-247733 EARTH (Energy Aware Radio and NeTwork TecHnologies), Tech. Rep,* 2010.

[48] D. Grace, J. Chen, T. Jiang, and P. D. Mitchell, "Using cognitive radio to deliver "Green" communications," in *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Hannover, 2009, pp. 1-6

[49] S. Rehan and D. Grace, "Combined green resource and topology management for beyond next generation mobile broadband systems," in *2013 International Conference on Computing, Networking and Communications (ICNC)*, San Diego, CA, 2013, pp. 242-246.

[50] S. Tombaz, P. Monti, F. Farias, M. Fiorani, L. Wosinska, and J. Zander, "Is Backhaul Becoming a Bottleneck for Green Wireless Access Networks?," in *2014 IEEE International Conference on Communications (ICC)*, Sydney, NSW, 2014, pp. 4029-4035.

[51] A. Checko *et al.*, "Cloud RAN for mobile networks—a technology overview," in *IEEE Communications surveys & tutorials,* vol. 17, no. 1, pp. 405-426, 2015.

[52] A. Checko, H. L. Christiansen, and M. S. Berger, "Evaluation of energy and cost savings in mobile Cloud RAN," in *Proc. OPNETWORK 2013,* 2013, pp. 1-7.

[53] A. Checko, H. Holm, and H. Christiansen, "Optimizing small cell deployment by the use of C-RANs," in *European Wireless 2014; 20th European Wireless Conference*, Barcelona, Spain, 2014, pp. 1-6.

[54] NGMN Alliance. (Jan. 2013). *Suggestions on Potential Solutions to C-RAN Version 4.0*. [Online]. Available: http://www.ngmn-ic.info/uploads/media/NGMN_CRAN_Suggestions_on_Potential_Solutions_to_CRAN.pdf. [Accessed: Feb. 20, 2017].

[55] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella and D. Wübben, "Cloud technologies for flexible 5G radio access networks," in *IEEE Communications Magazine,* vol. 52, no. 5, pp. 68-76, May 2014.

[56] W. Anjing, M. Iyer, R. Dutta, G. N. Rouskas, and I. Baldine, "Network Virtualization: Technologies, Perspectives, and Frontiers," in *Journal of Lightwave Technology* 31.4*,* vol. 31, no. 4, pp. 523-537, 2013.

[57] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358-380, Firstquarter 2015.

[58] D.-E. Meddour, T. Rasheed, and Y. Gourhant, "On the role of infrastructure sharing for mobile network operators in emerging markets," in *Computer Networks,* vol. 55, no. 7, pp. 1576-1591, 2011.

[59] W. Xin, P. Krishnamurthy, and D. Tipper, "Wireless network virtualization," in *2013 International Conference on Computing, Networking and Communications (ICNC)*, San Diego, CA, 2013, pp. 818-822.

[60] S. Baucke and C. Görg, "D-3.1. 1 Virtualisation Approach: Concept," *FP7 4WARD Project Tech. Rep*, 2009.

[61] H. Wen, P. K. Tiwary, and T. Le-Ngoc, *Wireless Virtualization*. Springer, 2013.

[62] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, "A Novel LTE Wireless Virtualization Framework," in I*nternational Conference on Mobile Networks and Management*, Springer, Berlin, Heidelberg, 2010, pp. 245-257.

[63] R. Kokku, R. Mahindra, Z. Honghai, and S. Rangarajan, "NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks," in *IEEE/ACM Transactions on Networking,* vol. 20, no. 5, pp. 1333-1346, 2012.

[64] X. Costa-Perez, J. Swetina, G. Tao, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," in *IEEE Communications Magazine,* vol. 51, no. 7, pp. 27-35, 2013.

[65] K. Hyojoon and N. Feamster, "Improving network management with software defined networking," in *IEEE Communications Magazine,* vol. 51, no. 2, pp. 114-119, 2013.

[66] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A Survey on Software-Defined Networking," in *IEEE Communications Surveys & Tutorials,* vol. 17, no. 1, pp. 27-51, 2015.

[67] ONF Fundation, "Software-defined networking: The new norm for networks," *ONF White Paper*, 2012.

[68] C. J. Bernardos, A. de la Oliva, P. Serrano, A. Banchs, L. M. Contreras, H. Jin and J. C. Zuniga, "An architecture for software defined wireless networking," in *IEEE Wireless Communications,* vol. 21, no. 3, pp. 52-61, 2014.

[69] P. Spapis, K. Chatzikokolakis, N. Alonistioti, and A. Kaloxylos, "Using SDN as a key enabler for co-primary spectrum sharing," in *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, Chania, 2014, pp. 366-371.

[70] N. Cvijetic, A. Tanaka, K. Kanonakis, and T. Wang, "SDN-controlled topology-reconfigurable optical mobile fronthaul architecture for bidirectional CoMP and low latency inter-cell D2D in the 5G mobile era," in *Optics Express,* vol. 22, no. 17, pp. 20809-15, Aug 2014.

[71] R. Ahmed and R. Boutaba, "Design considerations for managing wide area software defined networks," in *IEEE Communications Magazine,* vol. 52, no. 7, pp. 116-123, 2014.

[72] S. H. Yeganeh, A. Tootoonchian, and Y. Ganjali, "On scalability of software-defined networking," in *IEEE Communications Magazine,* vol. 51, no. 2, pp. 136-141, 2013.

[73] S. Hassas Yeganeh and Y. Ganjali, "Kandoo: a framework for efficient and scalable offloading of control applications," in *ACM Proceedings of the first workshop on Hot topics in software defined networks*, 2012, pp. 19-24.

[74] N. Khan, C. Oestges, F. Mani, and O. Renaudin, "BuNGee Deliverable: D2.1: Final BuNGee Channel Models," *Beyond Next Generation Mobile Networks Project*, Tech. Rep., 2011.

[75] IST-WINNER II, "WINNER II Public Deliverable: WINNER II channel models," *IST-WINNER II*, Tech. Rep., 2007.

[76] A. Papadogiannis and A. G. Burr, "Multi-beam assisted MIMO - a novel approach to fixed beamforming," in *011 Future Network & Mobile Summit*, Warsaw, 2011, pp. 1-8.

[77] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects*. 3GPP TR 36.814 V9.0.0, Mar. 2010.

[78] T. Bohn, "D4. 1: Most promising tracks of green radio technologies," *INFSO-ICT-247733 EARTH,* Tech. Rep., 2010.

[79] E. Voytko, N. Blum, M. Harshish, O. Marinchenco, A. Tipograff, R. Barda, R. Iluz, O. Sporn, "BuNGee Deliverable: D4.3 BuNGee System Integration Report," *Beyond Next Generation Mobile Networks Project*, Tech. Rep., 2012.

[80] Y. Bian and D. Rao, "Small cells big opportunities," *Global Business Consulting. Huawei Technologies Co., Ltd,* 2014.

[81] F. J. Ros, J. A. Martinez, and P. M. Ruiz, "A survey on modeling and simulation of vehicular networks: Communications, mobility, and tools," in *Computer Communications,* vol. 43, pp. 1-15, 2014.

[82] A. Mahajan, N. Potnis, K. Gopalan, and A. Wang, "Urban mobility models for vanets," in *2nd IEEE International Workshop on Next Generation Wireless Networks*, Dec. 2006.

[83] A. M. Mohammed and A. F. Agamy, "A survey on the common network traffic sources models," in *International Journal of Computer Networks,* vol. 3, no. 2, 2011.

[84] K. Wang, X. Li, H. Ji, and X. Du, "Modeling and optimizing the LTE discontinuous reception mechanism under self-similar traffic," in *IEEE Transactions on Vehicular Technology,* vol. 65, no. 7, pp. 5595-5610, 2016.

[85] I. B. Aban, M. M. Meerschaert, and A. K. Panorska, "Parameter estimation for the truncated Pareto distribution," in *Journal of the American Statistical Association,* vol. 101, no. 473, pp. 270-277, 2006.

[86] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[87] YouTube. (2017, May). *Live encoder settings, bitrates and resolutions* [Online]. Available: https://support.google.com/youtube/answer/2853702?hl=en-GB. [Accessed: June 1, 2017]

[88] Netflix. (2017, May). *Internet Connection Speed Recommendations* [Online]. Available: https://help.netflix.com/en/node/306. [Accessed: June 1, 2017].

[89] Huawei, "5G system requirements and scenarios - Project Dense Outdoor Radio Access (DORA),*" Huawei Technologies Co. Ltd.*, Sweden, Tech. Rep. V1.0, 29 July 2014.

[90] *Technical Specification Group Services and System Aspects; Policy and Charging Control Architecture (Release 14)*. 3GPP TS 23.203 V14.3.0, Mar. 2017.

[91] P. Harris,W. B. Hasan, S. Malkowsky, J. Vieira, S. Zhang, M. Beach, L. Liu, E. Mellios, A. Nix, S. Armour, A. Doufexi, K. Nieman and N. Kundargi, "Serving 22 Users in Real-Time with a 128-Antenna Massive MIMO Testbed," in *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, Dallas, TX, 2016, pp. 266-272.

[92] J. Wells, "Faster than fiber: The future of multi-G/s wireless," in *IEEE Microwave Magazine,* vol. 10, no. 3, pp. 104-112, 2009.

[93] *Radio frequency channel arrangements for fixed service systems operating in the bands 71–76 GHz and 81–86 GHz*. ECC Recommendation(05)07, Lugano, 2013.

[94] A. Mozharovskiy, A. Artemenko, A. Sevastyanov, V. Ssorin, and R. Maslennikov, "Beam-steerable integrated lens antenna with waveguide feeding system for 71–76/81–86 GHz point-to-point applications," in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, Davos, 2016, pp. 1-5.

[95] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mm wave 5G cellular system," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, 2014, pp. 377-381.

[96] MiWaveS, "Deliverable D 4.5 Antenna technologies for mmW access and backhaul communications," *Beyond 2020 Heterogeneous Wireless Nework with Millimeter-wave small-cell access and backhauling,* Tech. Rep., Sep. 2016.

[97] M. Giovanni and L. Frecassetti, "E-Band and V-Band-Survey on status of worldwide regulation," *ETSI White Paper*, France, 2015.

[98] L. Kleinrock. *Queueing Systems, Volume I: Theory*. New York: Wiley, 1975.

[99] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Wireless Networks,* vol. 11, no. 4, pp. 419-434, 2005.

[100] R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks," in *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 133-144, 2004.

[101] H. Li, Y. Cheng, C. Zhou, and W. Zhuang, "Routing metrics for minimizing end-to-end delay in multiradio multichannel wireless networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 24, no. 11, pp. 2293-2303, 2013.

[102] T. Liu and W. Liao, "Capacity-aware routing in multi-channel multi-rate wireless mesh networks," in *IEEE International Conference on Communications (ICC)*, Istanbul, 2006, vol. 5, pp. 1971-1976.

[103] X. Zhang, S. Zhou, X. Wang, Z. Niu, X. Lin, D. Zhu and M. Lei, "Improving network throughput in 60GHz WLANs via multi-AP diversity," in *2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, June 2012, pp. 4803-4807.

[104] S. Singh, F. Ziliotto, U. Madhow, E. Belding, and M. Rodwell, "Blockage and directivity in 60 GHz wireless personal area networks: from cross-layer model to multihop MAC design," in *IEEE Journal on Selected Areas in Communications,* vol. 27, no. 8, pp. 1400-1413, Oct. 2009.

[105] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *ACM SIGCOMM Computer Communication Review,* vol. 43, no. 4, pp. 375-386, 2013.

[106] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: self-interference cancellation, protocol design, and relay selection," in *IEEE Communications Magazine,* vol. 53, no. 5, pp. 128-137, 2015.

[107] S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti and P. Levis, "Applications of self-interference cancellation in 5G and beyond," in *IEEE Communications Magazine,* vol. 52, no. 2, pp. 114-121, 2014.

[108] R. Mudumbai, S. K. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *IEEE INFOCOM 2009*, Rio de Janeiro, 2009, pp. 2871-2875.

[109] M. Dohler and T. Nakamura, *5G mobile and wireless communications technology*. Cambridge University Press, 2016.

[110] G. Zhang, T. Q. S. Quek, M. Kountouris, A. Huang, and H. Shan, "Fundamentals of heterogeneous backhaul design—Analysis and optimization," in *IEEE Transactions on Communications,* vol. 64, no. 2, pp. 876-889, 2016.

[111] A. Hakkarainen, J. Werner, M. Costa, K. Leppanen, and M. Valkama, "High-efficiency device localization in 5G ultra-dense networks: Prospects and enabling technologies," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Boston, MA, Sep. 2015.