공학박사 학위논문

Mobility-Aware Analytic Framework and Network Management Scheme for User-Centric Millimeter Wave Communication Systems

사용자 중심의 밀리미터파 통신 시스템을 위한 이동성 인식 분석 프레임워크 및 네트워크 관리 기법

2021년 2월

서울대학교 대학원

전기·정보공학부

최 시 영

Mobility-Aware Analytic Framework and Network Management Scheme for User-Centric Millimeter Wave Communication Systems

사용자 중심의 밀리미터파 통신 시스템을 위한 이동성 인식 분석 프레임워크 및 네트워크 관리 기법

2021년 2월

서울대학교 대학원

전기·정보공학부

최 시 영

Mobility-Aware Analytic Framework and Network Management Scheme for User-Centric Millimeter Wave Communication Systems

지도 교수 박 세 웅

이 논문을 공학박사 학위논문으로 제출함
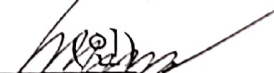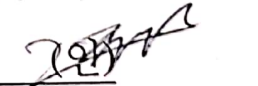
2021년 1월

서울대학교 대학원

전기·정보공학부

최 시 영

최시영의 공학박사 학위 논문을 인준함

2020년 12월

위 원 장: 김 성 철   (인)

부위원장: 박 세 웅   (인)

위   원: 심 병 효   (인)

위   원: 이 경 한   (인)

위   원: 주 창 희   (인)

Abstract

Millimeter wave (mmWave) communication enables high rate transmission, but its network performance may be degraded significantly due to blockages between the transmitter and receiver. There have been two approaches to overcome the blockage effect and enhance link reliability: multi-connectivity and ultra-dense network (UDN). Particularly, multi-connectivity under a UDN environment facilitates user-centric communication. It requires dynamic configuration of serving base station groups so that each user experiences high quality services. This dissertation studies a mathematical framework and network manament schemes for user-centric mmWave communication systems.

First, we models user mobility and mobility-aware performance in user-centric mmWave communication systems with multi-connectivity, and proposes a new analytical framework based on the stochastic geometry. To this end, we derive compact mathematical expressions for state transitions and probabilities of various events that each user experiences. Then we investigate mobility-aware performance in terms of network overhead and downlink throughput. This helps us to understand network operation in depth, and impacts of network density and multi-connection capability on the probability of handover related events. Numerical results verify the accuracy of our analysis and illustrate the correlation between mobility-aware performance and user speed.

Next, we propose user-oriented configuration rules and price based association algorithms for user-centric mmWave networks with fully/partially wired backhauls. We develop a fair association algorithm by solving the optimization problem that we formulate for mmWave UDNs. The algorithm includes an access price based per-user request decision method and a price adjustment rule for load balancing. Based on insights from the algorithm, we develop path-aware access pricing policy for mmWave

integrated access and backhaul networks. Numerical evaluations show that our proposed methods are superior to other comparative schemes. Our findings from analysis and optimization provide useful insights into the design of user-centric mmWave communication systems.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Background

Millimeter wave (mmWave) communication is a promising technology for the next generation communication system due to its potential for multi-gigabit throughput. But it has limitations because high frequency signals propagate in a short transmission range only and are vulnerable to blockages due to the physical propagation characteristics of mmWave [1]. Blockages such as buildings, forests, and people incur Non-Line-Of-Sight (NLOS) connections between the mmWave Base Station (BS) and user. This seriously aggravates the Quality-of-Experience (QoE) of users with reduced received signal strength. Therefore it is crucial to manage Line-Of-Sight (LOS) connections well, considering the blockage effect in mmWave communications.

There have been basically two approaches to overcome performance degradation due to the blockage. The first approach is multi-connectivity that supports multiple BSs to exchange data with a user at the same time, thereby increasing the chance of LOS communication due to macrodiversity [2]. It has become a key item in the 3rd Generation Partnership Project (3GPP) New Radio (NR) specifications [3]. Specifically, master BSs manage the Control Plane (CP) and secondary BSs work for data exchange in the User Plane (UP).

The second one is the Ultra-Dense Network (UDN) that increases the possibility of LOS communication. There is an advantage of providing a better communication service as the distance between the user and serving BS becomes shortened [4]. Network densification, however, causes higher handover rate and consequently heavier network overhead [5, 6]. In addition, the massive wired backhaul can result in high capital expenditures and operating costs for network operators. To address the challenges of mmWave UDN, the standardization group 3GPP have studied about UDN with wireless backhaul named as integrated access and backhaul (IAB) [7]. The objective of the study is to identify and evaluate potential solutions for efficient operation of integrated access and wireless backhaul for NR. An IAB network is composed of IAB-node and IAB-donor, where the first is BS that supports wireless access to users and wirelessly backhauls the access traffic and the latter is BS which provides user's interface to core network and wireless backhauling functionality to IAB-nodes.

Multi-connectivity and network densification lead us to rethink about an across-the-board of communication system. In a conventional communication system, there is a great deal of asymmetry between the performance between BS and UE. In the past, the computing power of BSs was superior to that of UEs, so the network composed of BSs determines a communication configuration such as modulation and coding scheme (MCS) and initial access based on cell-centric method. This is why we call the conventional mobile communication system as cellular system. Nowadays, state-of-the-art hardware architectures allow UEs to enjoy a wide variety of applications based on excellent computing power and advanced connectivity, which makes the user's requirements much more diverse. To meet various UEs' needs, cell-centric management schemes become too complicated due to many objects and constraints needed to be considered. Therefore, it is necessary to consider a user-centric communication beyond the existing cell-centric era.

1.2 Outline and Contributions

This dissertation is structured as follows. Chapter 2 introduces a mobility-aware analytic framework for user-centric mmWave communication systems. Next, We propose user-centric configuration rules and fair association algorithm for mmWave UDN in Chapter 3. In Chapter 4, we develop routing strategies and an access pricing policy for mmWave IAB network. Lastly, we conclude the work with a summary and further research direction in Chapter 5.

In Chapter 2, we conduct stochastic geometry based analysis to investigate the impact of user mobility on network performance in user-centric mmWave communication systems. To the best of our knowledge, this work is the first that uses stochastic geometry to study user mobility in blockage environments. Specifically, we propose a system model for user-centric mmWave communication systems with multi-connectivity. We also develop a state machine according to user mobility and define some events by classifying transitions in the state machine. Then, we derive compact mathematical expressions for transitions and event probabilities for various rural and urban scenarios. We pay special attention to blockages because they affect communication performance significantly in mmWave systems. Blockages decrease the BS density in a user's communication range since the LOS probability between the user and its serving BS is reduced. They increase the handover probability as the link state of the user changes rapidly with mobility. Especially in the latter case, blockages make a temporal correlation between successive observations of the link state between the BS and mobile user, which makes mobility analysis difficult. To address the challenges in mobility analysis with blockages, we derive the effective BS density and the link status correlation. From these, we obtain tractable bounds on the handover probability. We investigate mobility-aware performance in terms of network overhead and downlink throughput, improving our understanding of network operations to support user mobility. Extensive simulations verify the accuracy and usefulness of our analysis.

In Chapter 3, we propose user-oriented configuration rules and association man-

agement schemes for mmWave UDNs. We first propose user-side configuration rules for achieving user requirement and investigate impacts on traffic flow control. Then we consider a fair association based on user request through self-configuration to balance BS load across the whole network. To this end, we formulate an optimization problem, which is a mixed-integer linear programming so an NP-hard problem. We develop a fair association algorithm by applying dual approach to the primal optimization problem. The algorithm includes an access price based per-user request decision method and a price adjustment rule for load balancing. We also develop the analytic framework to investigate the outage probability. We evaluate the proposed method in terms of average resource budget, network utility, peak load, fairness, and outage. In simulations, SLM shows better performance in terms of average resource budget and network utility, while PLM shows better performance in terms of fairness and peak load. We observe that our proposed methods perform very comparably to the centralized scheme and work well in dynamic environments. We find that the fair load distribution (i.e., PLM) makes the network more resistant to outage.

We extend the work in Chapter 3 to mmWave IAB networks in Chapter 4. We investigate the geographic and pathloss models suitable for describing urban street canyon environments. Then, we propose a mmWave IAB network model and find the dominant signal path for a given transmitter and receiver in an urban street canyon. In order to develop a path selection strategy for mmWave IAB networks, we formulate a routing problem by revisiting the idea of NUM. Specifically, we newly define the link metric for Dijkstra method and the utility maximization problem for QoS guaranteed networks. We also propose a path-aware access pricing policy for user association based on path dependency of multi-hop networking. The access price for a user is sum of backhauling cost and access cost, where the former implies path-dependency and the latter include both spectral efficiency with access BS and channel dynamics. We evaluate our proposed scheme in various urban street canyon scenarios with comparing the existing schemes.

Chapter 2

Mobility-Aware Analysis of Millimeter Wave Communication Systems with Blockages

In this chapter, we propose a new analytical framework based on stochastic geometry, aiming to model user mobility and evaluate mobility-aware performance in user-centric mmWave communication systems with multi-connectivity.

2.1   Introduction

Millimeter wave (mmWave) communication is a promising technology for the next generation communication system due to its potential for multi-gigabit throughput. But it has limitations because high frequency signals propagate in a short transmission range only and are vulnerable to blockages due to the physical propagation characteristics of mmWave [1]. Blockages such as buildings, forests, and people incur Non-Line-Of-Sight (NLOS) connections between the mmWave Base Station (BS) and user. This seriously aggravates the Quality-of-Experience (QoE) of users with reduced received signal strength. Therefore it is crucial to manage Line-Of-Sight (LOS) connections well, considering the blockage effect in mmWave communications.

There have been basically two approaches to overcome performance degradation due to the blockage. The first approach is multi-connectivity that supports multiple

BSs to exchange data with a user at the same time, thereby increasing the chance of LOS communication due to macrodiversity [2]. It has become a key item in the 3rd Generation Partnership Project (3GPP) New Radio (NR) specifications [3]. Specifically, master BSs manage the Control Plane (CP) and secondary BSs work for data exchange in the User Plane (UP).

The second one is the Ultra-Dense Network (UDN) that increases the possibility of LOS communication. There is an advantage of providing a better communication service as the distance between the user and serving BS becomes shortened [4]. Employing multi-connectivity and network densification, however, causes higher handover rate and consequently heavier network overhead [5, 6]. To overcome these disadvantages, there has been a paradigm shift from cell-centric to user-centric communication, where networks serve users in a de-cellular way [8]. The user-centric UDN is a wireless network with a BS density comparable to user density. It forms a dynamic BS group with multi-connectivity, and allocates resources to each user in a flexible and seamless fashion.

To investigate the impact of user mobility on network performance in user-centric mmWave communication systems, we conduct stochastic geometry based analysis. Through analysis, we study the average behavior in various realizations of the network where nodes are placed according to a certain probability distribution [9, 10]. Our analysis captures the spatial randomness inherent in wireless networks, and can be extended to account for other uncertainty factors. In some cases, the analysis leads to closed-form expressions that help us to understand network operation and provide insightful guidelines for network design.

2.1.1 Related Work

To analyze the blockage effect and mobility-aware performance without loss of spatial randomness, the homogeneous Poisson Point Process (PPP) model has been often used since it is mathematically tractable and applicable [2, 5, 6, 11–19]. A math-

ematical framework based on random shape theory to model blockages with random sizes, orientations, and locations was proposed in [16]. However, it did not consider multi-connectivity and user mobility. The blockage effect caused by mobile blockers in mmWave systems has been studied in [12, 17]. In [12], the authors developed a prediction model for user states with mobile blockers. An analytical model was proposed in [17], combining the effects of static blockages, mobile blockers, and self blockage. Unfortunately, these studies did not consider multi-connectivity and user mobility either.

There is a previous literature on analyzing system performance in terms of coverage, handover rate, and outage probability considering blockages [2, 11, 13–15]. In [2] and [13], the authors studied LOS connection probability in a macrodiversity employed cellular system and an urban mmWave network modeled by a random grid, respectively. The authors in [11] derived a closed-form expression for the outage probability in a mmWave cellular system with macrodiversity. Note that multi-connectivity is different from macrodiversity in the perspective of the number of serving BSs. Macrodiversity allows one of LOS BSs to serve a user, while multi-connectivity enables multiple BSs to serve a user at the same time.

The coverage and handover rate were evaluated in mmWave cellular systems under the sectored beamforming model [14]. In [15], the authors investigated the downlink performance of a dense cellular network by incorporating LOS and NLOS propagation, where all BSs are elevated to the same height while users are at the ground level. These approaches did not consider the dynamics caused by user mobility. In the past, stochastic geometric approaches have been proposed to analyze user mobility and corresponding performance metrics such as handover rate and mobility-aware data rate.

Compact expressions for handover rate and mobility-aware data rate were obtained in a multi-connectivity enabled user-centric UDN [6]. In [18], the authors studied handover performance in a dense multi-tier heterogeneous network, introducing a bias factor to maximize the coverage of small cells. This work was extended to user-centric co-

operative wireless networks in [5]. The authors in [19] developed a tractable mobility-aware model for two-tier dense cellular networks under the CP/UP split architecture. However, these approaches did not consider blockages in the analysis.

2.1.2 Contributions

In this chapter, we propose a new analytical framework based on stochastic geometry, aiming to model user mobility and evaluate mobility-aware performance in user-centric mmWave communication systems with multi-connectivity. We pay special attention to blockages because they affect communication performance significantly in mmWave systems. Blockages decrease the BS density in a user's communication range since the LOS probability between the user and its serving BS is reduced. They increase the handover probability as the link state of the user changes rapidly with mobility. Especially in the latter case, blockages make a temporal correlation between successive observations of the link state between the BS and mobile user, which makes mobility analysis difficult. To the best of our knowledge, this work is the first that uses stochastic geometry to study user mobility in blockage environments. Our contributions are as follows:

- We propose a system model for user-centric mmWave communication systems with multi-connectivity. We develop a state machine according to user mobility and define some events by classifying transitions in the state machine.

- We derive compact mathematical expressions for transitions and event probabilities for various rural and urban scenarios. To address the challenges in mobility analysis with blockages, we derive the effective BS density and the link status correlation. From these, we obtain tractable bounds on the handover probability.

- We investigate mobility-aware performance in terms of network overhead and downlink throughput, improving our understanding of network operations to support user mobility. Extensive simulations verify the accuracy and usefulness
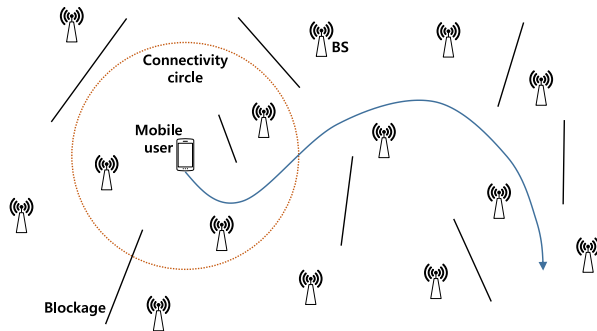
Figure 2.1. System model for User Plane (UP).

of our analysis.

The rest of this chapter is organized as follows. In Section 2.2, we describe the system model. In Section 2.3, we analyze user mobility and mobility-aware performance. In Section 2.4, we justify the correctness of our derivations through extensive simulations. Finally, Section 2.5 concludes the chapter.

2.2 System Model

In this section, we describe our model for user-centric mmWave communication systems, which is depicted in Fig. 2.1. We focus on the downlink communication and throughput, experienced by mobile users in outdoor environments.

2.2.1 Network Model

We consider a mmWave communication system with BSs and users whose locations are modeled as a homogeneous Poisson Point Process (PPP) Λ with density λ and a stationary point process, respectively. Let $\mathbf{x}_i$ denote the location of BS i, and $\mathbf{y}_j$ denote the location of user j. We consider blockages to capture LOS and NLOS properties of a wireless link. Each blockage is assumed to be a line segment, evenly located throughout the entire area [2]. Then, we describe the centers of line segments as a homogeneous PPP Ψ with density ψ. Blockage k has length l_k, uniformly distributed

9

in the range $[0, L_{\max}]$. Furthermore, lengths are assumed *i.i.d.* [2]. Blockage k has orientation θ_k, which is also an *i.i.d.* random variable and uniformly distributed in $[0, \pi]$.

Each BS has an LOS or NLOS channel for a given user. We let $\Lambda = \Lambda_L \cup \Lambda_N$, where Λ_L (Λ_N) is the PPP for BSs with LOS (NLOS) channels with the user. According to the thinning property, we note that Λ_L and Λ_N are random processes that are independent of each other [9]. Finally, we assume the CP/UP-split network architecture. That is, mmWave BSs serve users in the UP, and the others such as centimeter wave BSs manage the CP, operating in the low frequency band.

We adopt a pathloss model for wireless channels. Let $r_{i,j}$ denote the distance between BS i and user j, and $\mathcal{Q}_{i,j}$ denote its link state as either LOS or NLOS. However we observe the channel gain of NLOS links is significantly smaller than that of LOS links, i.e., about 20 dB [20]. We hence neglect NLOS links without sacrificing accuracy and express the channel gain of link (i, j) as

$$\mathcal{G}_{i,j} = \mathbb{I}(\mathcal{Q}_{i,j} = \text{LOS})C_L r_{i,j}^{-\alpha_L}, \tag{2.1}$$

where $\mathbb{I}(\cdot)$ is the indicator function, C_L is the pathloss intercept constant of LOS links, and α_L is the pathloss exponent of LOS links, respectively. When BS i transmits a signal at power P_{TX}, it arrives at user j with the strength of

$$\gamma_{i,j} = P_{TX}\mathcal{G}_{i,j}. \tag{2.2}$$

2.2.2 Connectivity Model

To model user-centric communication, we define the connectivity circle $\mathcal{C}_j$ for user j. It is a circle with radius R, centered at $\mathbf{y}_j$, and user j can communicate with BSs within the connectivity circle only. We set the radius R, considering the user's Quality-of-Service (QoS) requirements. For example, if a user runs a high rate application such as video streaming, the user's connectivity circle has a small radius to attain high SNR
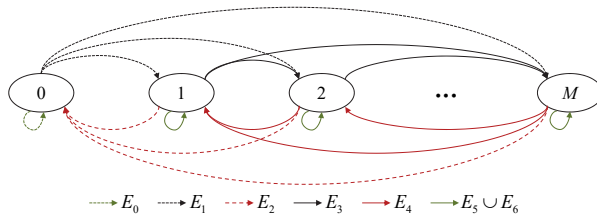
Figure 2.2. Mobility diagram of a user in the User Plane.

links.

We assume that our system supports multi-connectivity. That is, a user can connect to up to M BSs within the connectivity circle simultaneously. The maximum number of parallel connections, M, is limited by the user's capability. Even when the number of LOS BSs is larger than M, the user can be served by M BSs at the maximum. Let $\mathcal{S}_j$ denote the set of serving BSs for user j, which can be written as

$$\mathcal{S}_j = \big\{ i \,\big|\, \mathcal{Q}_{i,j} = \text{LOS}, \ \mathbf{x}_i \in \mathcal{C}_j \big\}. \tag{2.3}$$

We denote $\mathcal{S}_j$ as user j's connection table. In the initial stage, a user connects to the M nearest BSs at most. The user maintains a connection while each serving BS resides in the connectivity circle. Under this association rule, as the user moves around, serving BSs are randomly located in the connectivity circle.

2.2.3 Mobility Model

We assume that each user j moves a distance of ν_j per unit time at an arbitrary direction, and then changes its direction at random. We observe the connection tables of all users at discrete times t_n for $n = 0, 1, 2, \cdots$[1]. Depending on the mobility pattern, the distance between user j and each BS changes and consequently connection table $\mathcal{S}_j$ is updated. We describe the progress of $\mathcal{S}_j$ as a mobility diagram in the state machine of Fig. 2.2. In Section 2.3, we check that our model satisfies Markovian properties

---

[1]The network can see connection tables using uplink pilot signals such as sounding.

and global balance conditions. Each state m stands for the cardinality of a connection table, i.e. $m = |\mathcal{S}_j|$ and an arrow between two states represents the state transition between consecutive observation times. Let $p_{k,l}$ denote the transition probability from state $m^{(t_{n-1})} = k$ at t_{n-1} to state $m^{(t_n)} = l$ at t_n. We categorize the transitions $(m^{(t_{n-1})}, m^{(t_n)})$ as the following seven events, where $\mathcal{S}_j^{(t_{n-1})}$ indicates user j's connection table at time t_{n-1}.

- Out of service: $E_0 = \{(k, l) \mid k = 0, \; l = 0\}$

- Initial access: $E_1 = \{(k, l) \mid k = 0, \; \forall l \neq 0\}$

- Disconnect: $E_2 = \{(k, l) \mid \forall k \neq 0, \; l = 0\}$

- BS addition: $E_3 = \{(k, l) \mid 0 < k < l\}$

- BS removal: $E_4 = \{(k, l) \mid 0 < l < k\}$

- Handover: $E_5 = \{(k, l) \mid k = l > 0, \; \mathcal{S}_j^{(t_{n-1})} \neq \mathcal{S}_j^{(t_n)}\}$

- No BS change: $E_6 = \{(k, l) \mid k = l > 0, \; \mathcal{S}_j^{(t_{n-1})} = \mathcal{S}_j^{(t_n)}\}$.

We note that the self-looped transition probability $p_{k,k}$ ($k \neq 0$) is the sum of two probabilities. That is, $p_{k,k} = h_k + g_k$, where $h_k = \Pr\{E_5 \mid k\}$ and $g_k = \Pr\{E_6 \mid k\}$. For example, let $k = l = M$. If $x(\leq M)$ BSs in $\mathcal{S}_j^{(t_{n-1})} \cap \mathcal{S}_j^{(t_n)}$ continue to serve the user and $(M - x)$ new BSs are chosen to serve the user among $\mathcal{S}_j^{(t_n)} \setminus \mathcal{S}_j^{(t_{n-1})}$, handover occurs at $(M - x)$ BSs. In other words, handover is triggered if *any* of the M BSs leave the connectivity circle while the same number of new LOS BSs are in the circle. We emphasize that the handover event, E_5, represents the change of serving BSs while keeping the same cardinality. If the number of serving BSs changes, an event of BS addition (E_3) or removal (E_4) happens [3].

2.3 Mobility-Aware Analysis

In this section, we establish an analytical framework based on stochastic geometry and derive event occurrence probabilities for $E_0, ..., E_6$, under urban and ultra-densely deployed BSs environments.

2.3.1 Analytical Framework

We begin in a rural environment without blockage. We recall that locations of BSs can be described as a PPP in our system. Then, Slivnyak's theorem states that, in a connectivity circle centered at any point, the number of BSs is distributed as a Poisson random variable [9]. Hence,

$$\Pr\{m = k\} = \begin{cases} f(\lambda \pi R^2, k), & \text{if } k = 0, \cdots, M - 1 \\ F(\lambda \pi R^2, M), & \text{if } k = M, \end{cases} \tag{2.4}$$

where $f(a, n) = e^{-a} \frac{a^n}{n!}$ is the Probability Mass Function (PMF) of a Poisson distribution and $F(a, n) = 1 - \frac{\int_a^\infty t^{n-1} e^{-t} dt}{(n-1)!}$ is its Complementary Cumulative Density Function (CCDF) [21].

Consider two consecutive connectivity circles of user j, i.e., $\mathcal{C}_j^{(t_{n-1})}$ and $\mathcal{C}_j^{(t_n)}$, depicted in Fig. 2.3. If the moving distance is less than the circle's diameter, i.e., $\nu < 2R$, there exists an intersection between the two circles. On the other hand, if $\nu_j \geq 2R$, the two circles do not overlap at all, so the user makes handover to be served continuously. We define three regions, $\mathcal{X} = \mathcal{C}_j^{(t_{n-1})} \setminus \mathcal{C}_j^{(t_n)}$, $\mathcal{Y} = \mathcal{C}_j^{(t_n)} \setminus \mathcal{C}_j^{(t_{n-1})}$, and $\mathcal{Z} = \mathcal{C}_j^{(t_{n-1})} \cap \mathcal{C}_j^{(t_n)}$. And, let X, Y, and Z denote the areas of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. We remove the user index j and time index t_{n-1} and t_n for simplicity.

During the movement from t_{n-1} to t_n, D BSs leave the connectivity circle and A BSs join the circle newly, where D and A are Poisson random variables with means, λX and λY, respectively. Furthermore, departure process D and arrival process A are independent of each other because their associated areas do not overlap. Therefore
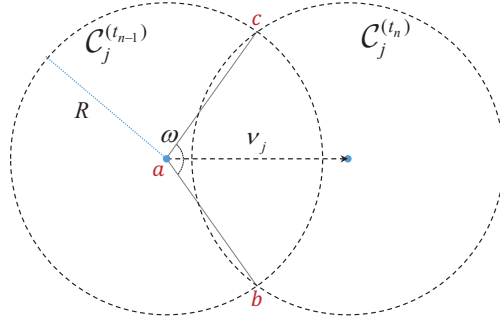
Figure 2.3. Geometry of two consecutive connectivity circles, $\mathcal{C}_j^{(t_{n-1})}$ and $\mathcal{C}_j^{(t_n)}$, incurred by the movement of user j.

we conclude that the number of BSs in the connectivity circle satisfies the Markov property. The transition probability can be written as

$$p_{k,l} = \Pr\left\{m^{(t_n)} = l \;\middle|\; m^{(t_{n-1})} = k\right\}. \tag{2.5}$$

We calculate the intersection area, denoted by Z, which will be used to derive state transition probabilities. We obtain the half of Z by subtracting the area of triangle abc from that of the circular sector abc (See Fig. 2.3). Consequently,

$$Z = 2\left(\frac{1}{2}R^2\omega - \frac{1}{2}R^2\sin\omega\right) = R^2\left(\omega - \sin\omega\right), \tag{2.6}$$

where ω is the central angle of the circular sector abc, given as $\omega = 2\cos^{-1}\left(\frac{\nu_j}{2R}\right)$.

We derive the transition probabilities of the mobility diagram in Fig. 2.2. We first set $M = \infty$, i.e., consider the unlimited connectivity case, and then limits M to one finite value. The locations of BSs follow a PPP, so the probability of having k BSs in a certain region can be obtained from the PMF of the Poisson distribution. For example, $\Pr\{N(\mathcal{X}) = k\} = f(\lambda X, k)$, where $N(\mathcal{X})$ is the number of BSs in $\mathcal{X}$. We obtain the transition probability $q_{k,l}$ from states k and l for the mobility diagram with infinite M,

i.e.,

$$q_{k,l} = \frac{\sum_{i=0}^{\min\{k,l\}} \Pr\left\{N\left(\mathcal{X}\right) = k - i, \; N\left(\mathcal{Z}\right) = i, \; N\left(\mathcal{Y}\right) = l - i\right\}}{\Pr\left\{N\left(\mathcal{C}_j^{(t_{n-1})}\right) = k\right\}}$$
$$= \frac{\sum_{i=0}^{\min\{k,l\}} f(\lambda X, k - i) \cdot f(\lambda Z, i) \cdot f(\lambda Y, l - i)}{f(\lambda \pi R^2, k)}. \tag{2.7}$$

We now aggregate the states $M, M+1, ...$ into a single state M. Then, for finite M, the transition probability $p_{k,l}$ can be written as follows.

$$p_{k,l} = \begin{cases} q_{k,l}, & \text{if } k < M, \; l < M, \\ \sum_{k=M}^{\infty} q_{k,l}, & \text{if } k = M, \; l < M, \\ \sum_{l=M}^{\infty} q_{k,l}, & \text{if } k < M, \; l = M, \\ \sum_{k=M}^{\infty} \sum_{l=M}^{\infty} q_{k,l}, & \text{otherwise,} \end{cases} \tag{2.8}$$

where the detailed expressions for $l = M$ and $k = M$ are given by

$$p_{k,M} = \sum_{l=M}^{\infty} q_{k,l} = \frac{\sum_{i=0}^{k} f(\lambda X, k - i) \cdot f(\lambda Z, i) \cdot F(\lambda Y, M - i)}{f(\lambda \pi R^2, k)}, \tag{2.9}$$

$$p_{M,l} = \sum_{k=M}^{\infty} q_{k,l} = \frac{\sum_{i=0}^{l} F(\lambda X, M - i) \cdot f(\lambda Z, i) \cdot f(\lambda Y, l - i)}{F(\lambda \pi R^2, M)}, \tag{2.10}$$

and

$$p_{M,M} = \sum_{k=M}^{\infty} \sum_{l=M}^{\infty} q_{k,l} = \frac{\sum_{i=0}^{k} F(\lambda X, M - i) \cdot f(\lambda Z, i) \cdot F(\lambda Y, M - i) + F(\lambda Z, M)}{F(\lambda \pi R^2, M)}. \tag{2.11}$$

Our mobility diagram represents a Markov chain, so we can calculate event occurrence probabilities from the current user status only, such as speed and connectivity circle radius. We can apply our model to dynamic scenarios with various user speeds and/or QoS requirements too. Such dynamics are not allowed in the analytical framework

in [6].

We now pay attention to the handover probability. To this end, we denote the self-looped transition probability at state k as $p_{k,k}$ and express this as the sum of handover probability h_k and no-change probability g_k. Then we obtain the handover probability h_k by subtracting g_k from $p_{k,k}$. For $1 \le k < M$, the condition of no-change is that all k BSs are located only in $\mathcal{Z}$, i.e., the intersection of two connectivity circles. Then, the no-change probability g_k is

$$g_k = \frac{\Pr\left\{N\left(\mathcal{X}\right)=0,\ N\left(\mathcal{Z}\right)=k,\ N\left(\mathcal{Y}\right)=0\right\}}{\Pr\left\{N\left(\mathcal{C}_j^{(t_{n-1})}\right)=k\right\}}$$
$$= \frac{f(\lambda X, 0) \cdot f(\lambda Z, k) \cdot f(\lambda Y, 0)}{f(\lambda \pi R^2, k)}. \tag{2.12}$$

Therefore, the handover probability at state k is given as

$$h_k = p_{k,k} - \frac{f(\lambda X, 0) \cdot f(\lambda Z, k) \cdot f(\lambda Y, 0)}{f(\lambda \pi R^2, k)}. \tag{2.13}$$

Especially, for state M, handover does not occur when all the serving BSs are in the intersection region. Hence, the handover probability at state M is given as

$$h_M = p_{M,M} - \underbrace{\left(\frac{Z}{\pi R^2}\right)^M}_{=g_M}, \tag{2.14}$$

where $p_{M,M}$ is given in (2.11) and g_M is derived by using a Binomial point process (BPP) [9]. In [6], the handover probability is derived, assuming that each user is connected to the nearest M BSs at the same time. This can be applied to our model by multiplying $p_{M,M}$ by $P(H_u)$ in [6].

We now calculate the event occurrence probabilities for $E_0, \cdots, E_6$, using the transition probabilities. First, we obtain $P_{E_0} = \Pr\{E_0\}$ as

$$P_{E_0} = \Pr\{m = 0\} \cdot p_{0,0} = f(\lambda(2\pi R^2 - Z), 0). \tag{2.15}$$

This represents the probability that there is no BS in the union of two connectivity circles, which is intuitively correct. Then we obtain the probabilities for events $E_1, E_2, E_3,$ and E_4 as

$$P_{E_1} = \Pr\{m = 0\} \cdot \sum_{l=1}^{M} p_{0,l}, \tag{2.16}$$

$$P_{E_2} = \sum_{k=1}^{M} \Pr\{m = k\} \cdot p_{k,0}, \tag{2.17}$$

$$P_{E_3} = \sum_{k=1}^{M-1} \Pr\{m = k\} \cdot \sum_{l=k+1}^{M} p_{k,l}, \tag{2.18}$$

and

$$P_{E_4} = \sum_{k=2}^{M} \Pr\{m = k\} \cdot \sum_{l=1}^{k-1} p_{k,l}. \tag{2.19}$$

We observe that $P_{E_1} = P_{E_2}$ and $P_{E_3} = P_{E_4}$, which is reasonable since our mobility diagram satisfies the global balance condition. That is,

$$\Pr\{m = k\} = \sum_{\forall l} \Pr\{m = l\} \cdot q_{l,k}.$$

Similarly, the probabilities for E_5 and E_6 are $P_{E_5} = \sum_{k=1}^{M} \Pr\{m = k\} \cdot h_k$ and $P_{E_6} = \sum_{k=1}^{M} \Pr\{m = k\} \cdot g_k$, respectively.

We find the event probabilities when λ goes to infinity, which describes an environment with densely deployed BSs. Note that for $S > 0$, $\lim_{\lambda \to \infty} f(\lambda S, n) = 0$ and $\lim_{\lambda \to \infty} F(\lambda S, n) = 1$ for $n = 0, 1, \cdots$. Hence, we have $\lim_{\lambda \to \infty} P_{E_i} = 0$ for $i = 0, \cdots, 4$ and $\lim_{\lambda \to \infty} (P_{E_5} + P_{E_6}) = 1$, which implies that at state M, handover incurs most of the network overhead.

We also consider the extreme case of $\nu_j \to 0$, i.e., user j rarely moves. Then $q_{k,l} = 0$ for $k \neq l$ and $q_{k,k} = 1$. Moreover, we have that $h_k = 0$ and $g_k = 1$ for $k = 1, \cdots, M$. From these, we can see that $P_{E_0} = e^{-\lambda \pi R^2}$, $P_{E_5} + P_{E_6} = 1 - e^{-\lambda \pi R^2}$, and $P_{E_i} = 0 \ \forall i = 1, \cdots, 4$ for stopped users.

On the opposite case, if user j moves at high speed, i.e., $\nu_j \to 2R$, we can rewrite the transition probability $q_{k,l}$ given in (2.7) as $f(\lambda \pi R^2, l)$. As $\lim_{\nu_j \to 2R} Z = 0$, no-change probability g_k goes to 0. Hence, handover probability $h_k = f(\lambda \pi R^2, k)$ for $k = 1, \cdots, M - 1$, and $h_M = F(\lambda \pi R^2, M)$. These results lead to $P_{E_0} = e^{-2\lambda \pi R^2}$, $P_{E_1} = P_{E_2} = e^{-\lambda \pi R^2}(1 - e^{-\lambda \pi R^2})$, and

$$
\begin{aligned}
P_{E_5} =& \frac{(M-1)! - \Gamma(M, \lambda \pi R^2)}{(M-1)!^2} + e^{-2\lambda \pi R^2}(I_0(2\lambda \pi R^2) - 1) \\
&- \frac{\lambda^2 \pi^2 R^4 e^{-2\lambda \pi R^2} {}_1F_2(1; \{M+1, M+1\}; \lambda^2 \pi^2 R^4)}{M!^2},
\end{aligned}
$$

where ${}_pF_q(a; b; z)$ is the generalized hypergeometric function and $I_n(z)$ is the modified Bessel function of the first kind [22], and $P_{E_6} = 0$. Finally, two remaining event probabilities are given as $P_{E_3} = P_{E_4} = \frac{1}{2}(1 - (P_{E_0} + 2P_{E_1} + P_{E_5}))$.

This analytic framework is suitable for rural environments, where users have a high probability of having LOS links and most BSs have no blockage. The framework holds in the CP too because blockages do not affect communication performance significantly at the low carrier frequency. The handover probability in the CP can be obtained easily by plugging $M = 1$, CP BS density, and communication range for the CP into (2.14).

2.3.2 Urban Scenario with Ultra-Densely Deployed BSs

In urban areas, many obstacles such as buildings, cars, and people affect wireless communication performance severely. So we incorporate the blockage effect into our analytic framework to obtain accurate system performance. We consider the blockage effect in two perspectives. First, blockages result in reduced BS density in the communication area because the LOS probability between the user and serving BS decreases. Second, blockages increase the handover probability because the link status between the user and serving BS can change rapidly according to the user mobility.

1) Effective BS density: We derive effective BS density while taking into account blockages. The average number of LOS BSs within the connectivity circle is given as

$$\kappa = \mathbb{E}\left[\sum_{\mathbf{x}_l \in \Lambda} \underbrace{\mathbb{I}\left(\mathbf{x}_l \in \Lambda_L\right)}_{=p(r)}\right] \overset{\text{(a)}}{=} \int_{\mathbb{R}^2} \lambda p(r) d\mathbb{R}^2, \tag{2.20}$$

where $p(r)$ is the probability that the link between the user and BS, which is at a distance r away from the user, is LOS, $\mathbb{R}^2$ stands for 2-dimensional coordinate, and (a) follows the Campbell's theorem on the PPP [9]. When blockages are placed uniformly, the probability $p(r)$ is given as

$$p(r) = e^{-\beta r}, \tag{2.21}$$

where β is a parameter depending on the density and size of blockages. When the blockages are located by the PPP with density ψ, $\beta = \psi \frac{2}{\pi} \mathbb{E}[l_k] = \frac{\psi L_{\max}}{\pi}$ (m^{-1}) [16]. Bai *et al.* showed that incorporating the height of blockages introduces only a constant scaling factor (Theorem 3 in [16]). Hence, we do not consider the height of blockages in the following analysis. We can modify our analysis to account for the height of blockages simply by adopting an appropriate scaling factor. We plug (2.21) into the truncated version of (2.20) and obtain the average number of LOS BSs in the connectivity circle as

$$\kappa(R) = 2\pi\lambda \int_0^R e^{-\beta r} r dr = 2\pi\lambda \frac{1 - e^{-\beta R}(1 + \beta R)}{\beta^2}.$$

Dividing $\kappa(R)$ by the area of the circle, we obtain the BS density within the LOS as

$$\lambda_{\text{eff}}(R) = \frac{\kappa(R)}{\pi R^2} = \lambda \cdot d(\beta R). \tag{2.22}$$

Here $d(t) = \frac{2\left(1 - e^{-t}(1+t)\right)}{t^2}$ is a monotonically decreasing function for $t > 0$. The function has a maximum of 1 at $t = 0$ and decreases to 0 as $t \to \infty$. We can see that $d(\beta R) = \frac{\lambda_{\text{eff}}(R)}{\lambda}$ represents the ratio of the number of LOS BSs to that of all the BSs

within the connectivity circle with radius R for a given β. If we replace BS density λ with (2.22) in our analytic framework, we can obtain the state, state transition, and event probabilities in urban environments.

2) Link status correlation: Knowing the effective BS density alone may not be sufficient to obtain accurate analysis results. Especially, for handover analysis, we should consider the effect of blockages on link status changes during movement. Unfortunately, it is not mathematically tractable, so we focus on a specific scenario of UDNs. Network densification is a practical solution for mmWave communications in urban areas because it helps to extend the communication coverage and provides a good QoS under blockages [2, 8].

In UDNs, a user stays in state M most of the time, i.e., connected with M BSs simultaneously. So the handover probability at state M is critical to system performance. We first answer the question how large BS density is required to provide the user with full multi-connectivity. There are more than or equal to M LOS BSs in the connectivity circle with probability $F\left(\lambda_{\mathrm{eff}}(R)\pi R^2, M\right)$. Hence the user sees M or more LOS BSs with probability δ at least if

$$F\left(\lambda_{\mathrm{eff}}(R)\pi R^2, M\right) \geq \delta. \tag{2.23}$$

Therefore, we obtain the required BS density as

$$\tilde{\lambda} \geq \frac{\beta^2 \Gamma^{-1}(M, 1-\delta)}{2\pi\left(1 - e^{-\beta R}(1 + \beta R)\right)}, \tag{2.24}$$

where $\Gamma^{-1}(m, y)$ is the inverse regularized upper incomplete gamma function, i.e., $\Gamma^{-1}(m, y)$ is the solution for x in $y = \Gamma(m, x)$. If BSs are deployed densely enough, e.g., $\delta = 0.99$, the user observes more than or equal to M LOS BSs most of the time. In this case, the mobility state diagram can be simplified to a single state M with $\Pr\{m = M\} = 1$ and $p_{M,M} = 1$.

We focus on deriving lower and upper bounds of the handover probability h_M^U,

where the superscript U represents the UDN scenario. The condition that handover does not occur is that M serving BSs are in $\mathcal{Z}$ and their link state to user j maintains the LOS in the two consecutive observations. Accordingly, we obtain the no-change probability g_M^U as

$$
\begin{aligned}
g_M^U &= \Pr\left\{\mathbf{x}_i \in \mathcal{C}_j^{(t_n)}, \mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}\middle|\, \mathbf{x}_i \in \mathcal{C}_j^{(t_{n-1})}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}, i \in \mathcal{S}_j\right\} \\
&= \Pr\left\{\mathbf{x}_i \in \mathcal{C}_j^{(t_n)}\middle|\, \mathbf{x}_i \in \mathcal{C}_j^{(t_{n-1})}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}, i \in \mathcal{S}_j\right\} \\
&\quad \times \Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}\middle|\, \mathbf{x}_i \in \mathcal{C}_j^{(t_{n-1})}, \mathbf{x}_i \in \mathcal{C}_j^{(t_n)}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}, i \in \mathcal{S}_j\right\} \\
&= \Pr\left\{\mathbf{x}_i \in \mathcal{C}_j^{(t_n)}\middle|\, \mathbf{x}_i \in \mathcal{C}_j^{(t_{n-1})}, i \in \mathcal{S}_j\right\} \cdot \Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}\middle|\, \mathbf{x}_i \in \mathcal{Z}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}, i \in \mathcal{S}_j\right\} \\
&= g_M \cdot \rho \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.25)
\end{aligned}
$$

The third equality holds because the inclusion in the connectivity circle is determined only by the location of the BS, not by the link state. In the fourth equality, we define the link state correlation as

$$
\rho \triangleq \Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}\middle|\, \mathbf{x}_i \in \mathcal{Z}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}, i \in \mathcal{S}_j\right\}. \qquad (2.26)
$$

We can see that $g_M^U \leq g_M$ since $\rho \leq 1$. Then, it is straightforward to obtain the lower bound of the handover probability as

$$
h_M^U = 1 - g_M^U \geq 1 - g_M = \underline{h_M^U}. \qquad (2.27)
$$

This result is reasonable because handover occurs more often with more BSs and blockages.

We next find the lower bound of the link status correlation ρ to obtain the upper bound of the handover probability h_M^U. To this end, we consider two extreme cases. We observe strong ρ when blockage density ψ is small and blockage size $L_{\max}$ is large, and weak ρ for large ψ and small $L_{\max}$ [2]. The link status correlation ρ has a lower

bound as follows.

$$
\begin{aligned}
\rho &= \frac{\Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}\big|\, \mathbf{x}_i \in \mathcal{Z}, i \in \mathcal{S}_j\right\}}{\Pr\left\{\mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}\right\}} \\
&\geq \Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}\big|\, \mathbf{x}_i \in \mathcal{Z}, i \in \mathcal{S}_j\right\} \\
&\geq \Pr\left\{\mathcal{Q}_{i,j}^{(t_n)} = \mathrm{LOS}, \mathcal{Q}_{i,j}^{(t_{n-1})} = \mathrm{LOS}\big|\, \mathbf{x}_i \in \mathcal{C}_j^{(t_n)}, \forall i\right\} \\
&= \begin{cases} \zeta, & \text{for strong link correlation,} \\ \zeta^2, & \text{for weak link correlation,} \end{cases} \\
&\stackrel{\text{(a)}}{=} \begin{cases} d(\beta R), & \text{for strong link correlation,} \\ d(\beta R)^2, & \text{for weak link correlation,} \end{cases}
\end{aligned}
\tag{2.28}
$$

where $\zeta \triangleq \Pr\left\{\mathcal{Q}_{i,j} = \mathrm{LOS} \mid \mathbf{x}_i \in \mathcal{C}_j, \ \forall i\right\}$ and (a) follows

$$
\zeta = \sum_{b=0}^{\infty} \int_0^R e^{-\beta r} \cdot s(r|b)\, dr \cdot e^{-\lambda \pi R^2} \cdot \frac{(\lambda \pi R^2)^b}{b!} = d(\beta R).
$$

Here, $s(r|b) = \frac{2r}{R^2}$ for $0 \leq r \leq R$ is the distribution of the distance between the user and BS when the number of BSs is b and the radius of the circle is R [17].

From (2.25) and (2.28), we obtain an upper bound of h_M^U as

$$
h_M^U \leq \overline{h_M^U} = \begin{cases} 1 - d(\beta R) \cdot g_M, & \text{for strong link correlation,} \\ 1 - d(\beta R)^2 \cdot g_M, & \text{for weak link correlation.} \end{cases}
\tag{2.29}
$$

2.3.3 Handover Analysis for Macrodiversity

We extend the proposed analytical framework when users exploit macrodiversity. As mentioned in Section 2.1, macrodiversity allows *any* of the multiple associated BSs to serve one user continuously, while multi-connectivity enables *all* associated BSs to serve one user together at the same time. This leads to different definitions of handover for macrodiversity and multi-connectivity. We assume that under macrodiversity, han-

dover decision is made when all the associated BSs should be changed at one time [23]. When a user's serving BS leaves the connectivity circle, it is served by another associated BS without handover, for example, through path switching [24]. In the following, we derive the handover probability for macrodiversity. Notice that occurrence probabilities for $E_0, \ldots, E_4$ are not different from the multi-connectivity case because such events are the same for macrodiversity and multi-connectivity.

We recall Fig. 2.3 to geographically describe handover for macrodiversity. Suppose that the user's serving BS is in $\mathcal{X}$ at time t_{n-1} and it no longer exists in the connectivity circle at t_n. The user will be served continuously if at least one associated BS is located in $\mathcal{Z}$. On the other hand, if all the associated BSs as well as the serving BS are in $\mathcal{X}$ at t_{n-1}, there will no BS to serve the user at time t_n and handover occurs. So, when the number of associated BSs is $k, \forall k = 1, \cdots, M - 1$, we can write the handover probability for macrodiversity as

$$
\begin{aligned}
\tilde{h}_k &= \frac{\Pr\left\{N\left(\mathcal{X}\right) = k,\ N\left(\mathcal{Y} \cup \mathcal{Z}\right) = 0\right\}}{\Pr\left\{N\left(\mathcal{C}_j^{(t_{n-1})}\right) = k\right\}} \\
&= \frac{f(\lambda X, k) \cdot f(\lambda(Y + Z), 0)}{f(\lambda \pi R^2, k)} = e^{-\lambda X}\left(\frac{X}{\pi R^2}\right)^k.
\end{aligned}
\tag{2.30}
$$

When the number of associated BSs is M, we obtain the handover probability by using a BPP as

$$
\tilde{h}_M = \left(\frac{X}{\pi R^2}\right)^M.
\tag{2.31}
$$

Similar to Section 2.3.2, the lower and upper bounds on handover probability for macrodiversity in urban UDNs are given as

$$
\tilde{h}_M \leq \tilde{h}_M^U \leq \overline{\tilde{h}_M^U},
\tag{2.32}
$$

where

$$\overline{\tilde{h}_M^U} = \begin{cases} 1 - d(\beta R) \cdot (1 - \tilde{h}_M), & \text{for strong link correlation,} \\ 1 - d(\beta R)^2 \cdot (1 - \tilde{h}_M), & \text{for weak link correlation.} \end{cases} \tag{2.33}$$

2.3.4 Normalized Network Overhead and Mobility-Aware Downlink Throughput with Greedy User Association

In our mobility diagram, whenever a user makes a state transition, its connection table changes. So we evaluate the network overhead incurred by user mobility, using the state transition probability. Let c_i denote the cost of event E_i for $i = 0, \cdots, 6$. We assume that c_i is the average time consumed by event E_i per unit time, as in [19]. Then we write the average network overhead at state k as

$$\begin{aligned} \eta_0 =& c_0 p_{0,0} + c_1 \sum_{l=1}^{M} p_{0,l}, \\ \eta_k =& c_2 p_{k,0} + c_3 \sum_{l=k+1}^{M} p_{k,l} + c_4 \sum_{l=1}^{k-1} p_{k,l} + c_5 h_k + c_6 g_k \\ & \forall k = 1, 2, \cdots, M-1, \\ \eta_M =& c_2 p_{M,0} + c_4 \sum_{l=1}^{M-1} p_{M,l} + c_5 h_M + c_6 g_M. \end{aligned} \tag{2.34}$$

Here, $\eta_k \in [0, 1]$ for all k's if $c_i \geq 0$ for all i's, and $\sum_{i=0}^{6} c_i = 1$ because η_k is a convex combination of $p_{k,l}$'s for $l = 0, \cdots, M$. Note that $p_{k,l} \geq 0$ for all l's and $\sum_{l=0}^{M} p_{k,l} = 1$. Finally, we define the normalized network overhead η as the average of η_k's, i.e.,

$$\eta = \sum_{k=0}^{M} \Pr\{m = k\} \cdot \eta_k. \tag{2.35}$$

In mmWave communication systems, BSs and users are considered to have directional antennas so that a connected BS-user pair can steer their beams to transmit or

receive signals successfully [14]. We assume that each BS and user have B and M RF chains, respectively. In the initial access, each user attempts to connect to M nearest BSs in the connectivity circle at most. If more than B users are connected to a BS, the BS should serve the users in Time Division Multiple Access (TDMA) manner, e.g., round robin. So, if $L_i(> B)$ users are connected to BS i, downlink user throughput decreases by the factor of $\left\lceil \frac{L_i}{B} \right\rceil$, where $\lceil \cdot \rceil$ is the ceiling function. That is, the achievable rate of link (i, j) is given as

$$\mathcal{R}_{i,j} = \frac{W}{\lceil L_i/B \rceil} \log_2 \left(1 + \frac{\gamma_{i,j}}{N_0 + I_j} \right), \tag{2.36}$$

where W is the system bandwidth, N_0 is the background noise power, and $I_j = \sum_{k:\mathbf{x}_k \in \Lambda_L \setminus \mathcal{S}_k} \gamma_{k,j}$ is received inter-cell interference at user j.

Incorporating the network overhead, we can define the mobility-aware downlink throughput of a user as

$$\mathcal{T} = (1 - h^{CP}) \sum_{k=1}^{M} \Pr\{m = k\}(1 - \eta_k)\mathcal{R}_k, \tag{2.37}$$

where $1 - h^{CP}$ is the complementary handover probability in the CP, and $\mathcal{R}_k = \sum_{i=1}^{k} \mathcal{R}_{i,j}$ is a user j's achievable rate from all the serving BSs in state k. Since the distance $r_{i,j}$ between BS i to user j has a value between $[0, R]$, the achievable rate is lower bounded as

$$\mathcal{R}_k^{\min} \geq \sum_{i=1}^{k} \frac{W}{\lceil L_i/B \rceil} \log_2 \left(1 + \frac{P_{TX} C_L R^{-\alpha_L}}{N_0 + I_j} \right). \tag{2.38}$$

Consequently, for each user, we find the lower bound of the mobility-aware downlink throughput as

$$\mathcal{T} \geq (1 - h^{CP}) \sum_{k=1}^{M} \Pr\{m = k\}(1 - \eta_k)\mathcal{R}_k^{\min}. \tag{2.39}$$

We now consider a UDN, which is a special case with $\Pr\{m = M\} = 1$. In this

Carrier freq.	C_L	α_L	$\frac{P_{TX}}{N_0}$	β
28 GHz	$10^{-6.14}$	2.0	110 dB	5.4/km

Table 2.1: Simulation parameters.

case, the normalized network overhead decreases to $\eta = \eta_M$ and the lower bound of mobility-aware downlink throughput is simplified as $\mathcal{T} \geq (1 - h^{CP})(1 - \eta_M)\mathcal{R}_M^{\min}$. If we estimate the handover probability as in (2.29), we can approximate the normalized network overhead as

$$\eta^U = c_5 h_M^U + c_6 g_M^U \approx c_5 \overline{h_M^U} + c_6 g_M. \tag{2.40}$$

Then we obtain the approximate expression for the mobility-aware downlink throughput in the UDN with blockages as

$$\mathcal{T}^U \approx (1 - h^{CP})(1 - (c_5 \overline{h_M^U} + c_6 g_M))\mathcal{R}_M^{\min}. \tag{2.41}$$

2.4 Numerical Results

In this section, we present numerical results to evaluate our analysis. We also show the mobility-aware downlink throughput and network overhead caused by user mobility. We assume that BSs are deployed densely in the CP, so every user always has at least one reliable CP connection to BSs. For example, suppose a user moves a distance of 100 m in every unit time and its connectivity circle has a radius of $R_c = 1$ km in the CP. The user experiences CP handover with probability 0.06 only, when CP BSs are deployed with $\lambda_c = 20/\text{km}^2$. We focus on UP performance without hurting CP performance. Simulation parameters are summarized in Table 2.1. We perform simulations through MATLAB and compare the results with our analysis. In simulations, we place blockages uniformly with density β and various BS density λ, depending on scenarios.

We first investigate the occurrence probabilities of the events defined in Section 2.2.3,
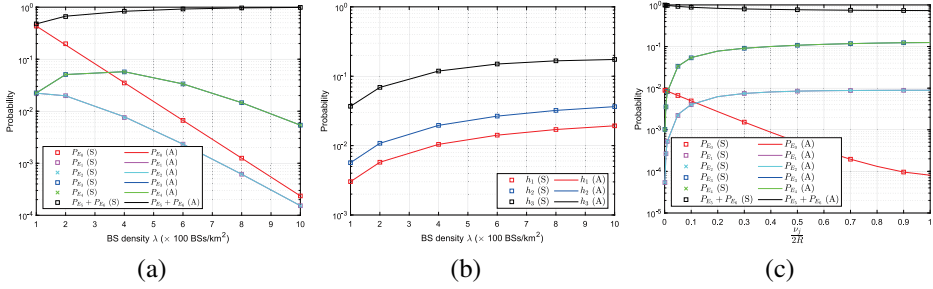
Figure 2.4. Rural scenario (A: analysis, S: simulation). (a) Event probability P_{E_i} ($i = 0, ..., 6$) versus BS density λ. $R = 50$, $\nu_j = 0.1R$, and $M = 3$. (b) Handover probability h_k ($k = 1, ..., M$) versus BS density λ. $R = 50$, $\nu_j = 0.1R$, and $M = 3$. (c) Event probability P_{E_i} ($i = 0, ..., 6$) versus normalized user speed $\nu_j/2R$. $R = 50$, $M = 3$, and $\lambda = 6 \times 10^{-4}$.

in the rural scenario. Fig. 2.4 shows that simulation results match our analysis results perfectly. In Fig. 2.4a, we present event probabilities, varying the BS density λ. We observe that the probability of $E_5 \cup E_6$ increases with the BS density, while those of the other events decrease monotonically. It can be explained since, with high BS density, both the probabilities of state M and state transition $(m^{(t_{n-1})}, m^{(t_n)}) = (M, M)$ become dominant. Fig. 2.4b shows handover probability at state k for $k = 1, 2, 3$. Mobile users experience frequent handovers in dense networks, as expected. Fig. 2.4c presents event probabilities as a function of the normalized user speed $\frac{\nu_j}{2R}$. The probabilities of E_0 and $E_5 \cup E_6$ decrease as the user speed increases. On the other hand, the probabilities of E_1, E_2, E_3 and E_4 increase then.

Furthermore, we observe that P_{E_0} is close to zero and the other probabilities converge as the user speed increases, i.e., $P_{E_1} = P_{E_2} \approx 0.009$, $P_{E_3} = P_{E_4} \approx 0.11$, and $P_{E_5} = P_{E_6} \approx 0.75$ when $\nu_j \geq R$. We find that handover events account for most of the network overhead when BSs are deployed densely. For example, $P_{E_5} + P_{E_6} \approx 0.92$ and $P_{E_i} \ll 1 (i = 0, ..., 4)$ for $\lambda = 600/\text{km}^2$ according to Fig. 2.4a.

In Fig. 2.5, we present the handover probability as a function of user speed in an urban scenario (i.e., $\lambda \geq 600$ /km^2). We also show the handover probability when each user is served by the closest M BSs as in [6]. We observe that the handover probability increases with user speed ν_j, which is reasonable since high speed users

27
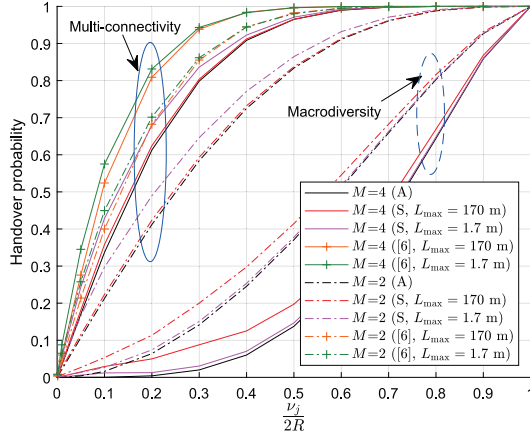
Figure 2.5. Handover probability versus normalized user speed $\nu_j/2R$ for various M, where $R = 50$ and $\lambda = 6 \times 10^{-4}$.

move away from their serving BS fast. Moreover, the handover probability increases with larger M since, with more parallel connections, a user is liable to move away from at least one of the serving BSs. It is notable that a user experiences higher handover probability when maintaining the nearest M serving BSs. That is, our association rule of sticking to the current serving BSs can lead to smaller handover probability. As mentioned in Section 2.3.3, we can use the analytic framework to calculate handover probability for macrodiversity, as shown in Fig. 2.5. For the same M, the handover probability for multi-connectivity is larger than that for macrodiversity. Handover in macrodiversity occurs when all M associated BSs leave the connectivity circle, which happens less frequently than that in multi-connectivity. So the handover probability in macrodiversity decreases with larger M. Note that the handover probabilities are 0 and 1 for $\frac{\nu_j}{2R} = 0$ and 1, respectively. This is because that $\frac{\nu_j}{2R} = 0$ indicates that user j does not move, and $\frac{\nu_j}{2R} = 1$ indicates that user j moves too fast so that its consecutive connectivity circles do not overlap (i.e., handover occurs). The gap between analysis and simulation results decreases as the average length of blockages decreases, which means that our derivation becomes accurate when $\rho \to 1$ (i.e., strong link correlation) and has a tractable lower bound for weak link correlation.

Fig. 2.6a shows the required BS density $\tilde{\lambda}$ to establish a UDN as a function of
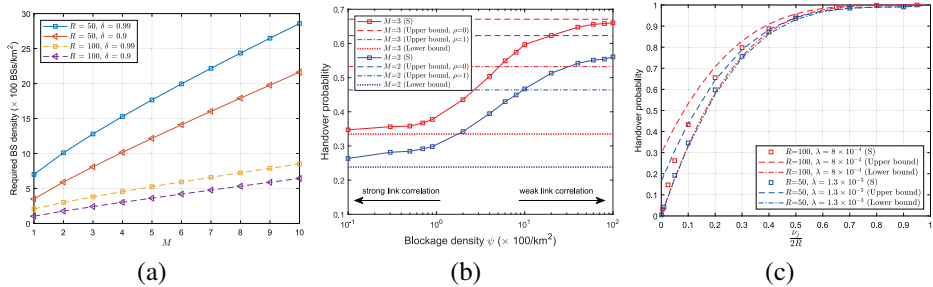
Figure 2.6. Required BS density and handover probability in an urban ultra-dense network. (a) Required BS density $\tilde{\lambda}$ versus multi-connection capability M for various R and δ. (b) Handover probability versus blockage density ψ for $R = 100$ m, $\nu_j = 0.2R$, and fixed $\beta = 5.4$/km. (c) Handover probability versus normalized user speed $\nu_j/2R$ for $\psi = 200$/km^2 and $M = 3$.

M for various R and δ. It shows that the required BS density linearly increases with M, and its slope becomes larger for smaller R. We also observe that the required BS density for $\delta = 0.99$ is larger than that for $\delta = 0.9$. The results shown in Fig. 2.6a enable network operators to calculate the BS density to provide good QoS to their subscribers. For example, mobile network operators deploy BSs with the density of at least 1,280/km^2 to guarantee that users are served by 3 BSs with probability 0.99.

For the rest of simulations (i.e., numerical results for the urban UDN), we set the BS density λ as 1,300/km^2 and 800/km^2 for $R = 50$ and $R = 100$, respectively.[2] Fig. 2.6b illustrates the upper/lower bounds on the handover probability as a function of the blockage density for a fixed β. The blockage density ψ and the maximum length of blockage $L_{\max}$ have an inverse relationship for a fixed β. Since blockages with long lengths affect a large area, the link status correlation ρ is proportional to the maximum length of blockage $L_{\max}$. We observe how much the correlation affects handover probability. The handover probability increases with the blockage density, i.e., less handover occurs in areas with strong link correlation. We argue that the upper bound for the case of $\rho = 1$ can be used for investigating whether the link correlation

---

[2]A typical criterion for UDN is the BS density of more than or equal to 1,000/km^2 [25]. Our assumption of 800/km^2 is smaller than the UDN condition but sufficient to prove the correctness of our analysis. We also present the result for $\lambda = 1,300$/km^2($\geq 1,000$/km^2). We can calculate the average inter-BS distance as $\frac{1}{2\sqrt{\lambda}}$ [9]. Hence, the average inter-BS distances for $\lambda = 1,300$ and 800 are given by 13.9 m and 17.7 m, respectively.
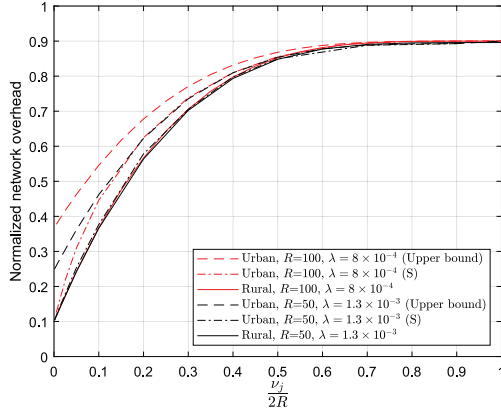
Figure 2.7. Normalized network overhead versus normalized user speed $\nu_j/2R$ for $M = 3$.

of a specific user is large or not. For example, it can be seen that the area of blockage density $\psi = 200$/km^2 with $\beta = 5.4$/km has a strong link correlation for user j with $M = 3$, $R = 100$ and $\nu_j = 20$ because the probability is bounded by the case of $\rho = 1$.

We also find that the handover probability increases with M, which is the same trend that Fig. 2.5 shows. Fig. 2.6c presents the variation of handover probability and its upper/lower bounds with respect to the normalized user speed. Our analytical upper/lower bounds provide tractable numerical results, matching simulation results well. Specifically, the upper bound for the case of $\rho = 1$ is shown because the parameters used in this simulation make the environment have a strong correlation (i.e., $\rho \to 1$). We observe that the handover probability increases with $\nu_j/2R$ and R. Specifically, the probability is above 90% when the user moves fast enough (i.e., $\nu_j \geq R$) for the cases of $R = 50$ and 100. It is intuitive that a user with high speed experiences handover frequently. Moreover, it is more likely to block the link between the user and BS in the connectivity circle with a large R. This is because a BS, which is far from the user, has more opportunities to service the user, and the LOS probability decays exponentially with the distance to the user.

We now investigate network performance, normalized network overhead, and mobility-aware downlink throughput, as a function of user speed. We assume each event has the
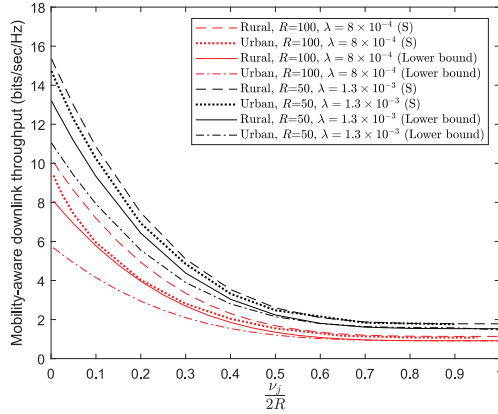
Figure 2.8. Mobility-aware downlink throughput versus normalized user speed $\nu_j/2R$ for $M = 3$.
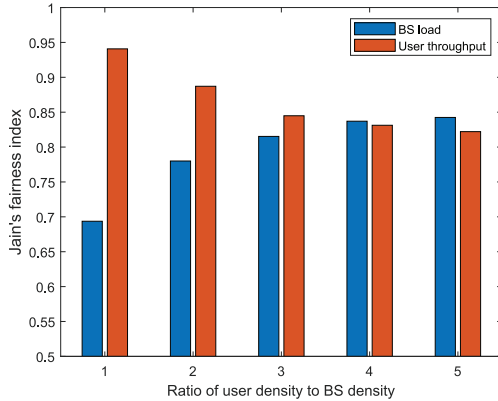
Figure 2.9. Jain's fairness indices of BS load and user throughput in the multi-user scenario, where $\lambda = 1{,}000/\text{km}^2$. Each user attempts to make greedy association with up to 3 BSs.

cost of $(c_0, c_1, c_2, c_3, c_4, c_5, c_6) = (0, 0, 0, 0, 0, 0.9, 0.1)$ because handover takes most of network overhead, so $E_0, \cdots, E_4$ are negligible to our attention. Fig. 2.7 shows the normalized network overhead for various R and λ in rural and urban scenarios. It presents that the normalized network overhead of the urban scenario is higher than that of the rural scenario. We observe that the overhead increases with R, because the handover probability is proportional to R as shown in Fig. 2.6c. Moreover, the overhead converges at $c_5 = 0.9$ when $\nu \to 2R$, otherwise at $c_6 = 0.1$ when $\nu \to 0$. It can be seen that E_5 is dominant when the user speed is very high, otherwise E_6 is dominant.
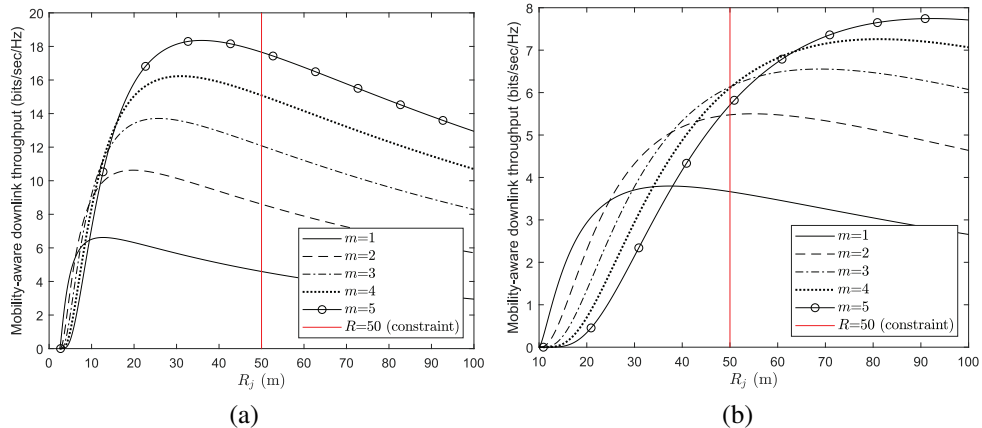
Figure 2.10. Mobility-aware downlink throughput of user j versus the radius of connectivity circle R_j. The red line shows the system constraint of $R_j \leq 50$. (a) $\nu_j = 5$. (b) $\nu_j = 20$.

For a given ν_j, the overhead with smaller R becomes higher than that with larger R (e.g., 0.58 with $R = 50$ and 0.45 with $R = 100$ for $\nu_j = 20$). Fig. 2.8 illustrates the mobility-aware downlink throughput for various R and λ in rural and urban scenarios. It shows that the throughput decreases with R, which implies that a user who requires low QoS can be served by BSs in a large connectivity circle. It is obvious that the throughput decreases with the user speed ν_j because the handover probability is proportional to ν_j. For a given ν_j, the throughput with smaller R becomes higher than that with larger R (e.g., 6.93 bps/Hz with $R = 50$ and 5.94 bps/Hz with $R = 100$ for $\nu_j = 20$).

We measure the Jain's fairness indices (JFIs) of BS load and user throughput in the multi-user scenario as shown in Fig. 2.9. In the multi-user scenario, each user makes greedy association with up to M BSs, that is, a user is connected to the M nearest LOS BSs in its connectivity circle at the same time. BS load is measured with the number of associated users at a BS. That is, we assume that the load of BS i is L_i. We use $\lambda =$1,000/km^2 and M=3 in this simulation. We observe that the greedy user association shows poor fairness performance. This is because greedy association rules do not consider user requirements. Specifically, some users with low data requirements

may exploit multi-connectivity to achieve much greater throughput than their requirements, which leads to unfair use of radio resources. Fig. 2.10 shows the mobility-aware downlink throughput as a function of R_j under various multi-connectivity scenarios of $m = 1, \ldots, 5$. In this figure, we assume that the system limits the communication range of each user to $R_j \leq 50$ (red line in the figure) and each user chooses its appropriate multi-connectivity and communication range to meet the requirements. For example, the user j selects $(m, R) = (5, 35.88)$ for $\nu_j = 5$ and $(m, R) = (3, 50)$ for $\nu_j = 20$ to maximize its throughput. From Figs. 2.9 and 2.10, we observe the need to have a fair association rule and to optimize multi-connectivity and communication ranges to take advantage of mmWave user-centric communication, which is our future work.

2.5 Summary

We evaluated the impact of user mobility on user-centric mmWave communication systems in the presence of uniform random blockages. We developed a mathematical framework based on the stochastic geometry to analyze user mobility and mobility-aware performance. Specifically, we used a connectivity circle to model the user-centric communication and a state machine to describe a user's mobility.

We derived compact expressions on the probabilities of state transitions and associated events in rural scenarios, and showed that the handover probabilities are bounded in urban scenarios. Through numerical results and computer simulations, we investigated the mobility-aware performance in terms of handover probability, network overhead, and downlink throughput, which provides deep insights for designing user-centric mmWave communication systems. We also observed that handover accounts for most of the network overhead in UDNs. Finally, we found the need of a fair association rule and joint optimization of multi-connectivity and communication range to realize the benefits of mmWave user-centric communication.

Chapter 3

Association Control for User-Centric Millimeter Wave Communication Systems

In this chapter, we introduce user-oriented configuration rules and association management schemes for user-centric mmWave communication systems.

3.1 Introduction

Millimeter wave (mmWave) communication is a promising technology for next generation communication systems that can provide multi-gigabits/sec throughput to users. But it has limitations because mmWave signals do not propagate long distance in the air due to high frequency characteristics and they are vulnerable to blockages such as buildings, forests, people, etc [1]. Blockages make wireless channels between base stations (BSs) and user equipments (UEs) as non-line-of-sight (NLOS), and seriously aggravate users' quality-of-experience (QoE) by lowering received signal strengths. So it is crucial to keep connections in line-of-sight (LOS) to mitigate the blockage effect in mmWave communications.

There are two approaches we can consider to overcome the performance degradation due to blockages: multi-connectivity and network densification. First, the multi-connectivity approach allows a user to communicate with multiple BSs at the same

time, which increases the chance of LOS connections by macrodiversity [2]. It has become a key item in the 3rd generation partnership project (3GPP) new radio (NR) specifications [3], where the master BS manages the control plane (CP) and more than one secondary BS controlled by the CP performs data exchange in the user plane (UP). Second, the network densification approach uses ultra dense network (UDN), where network operators deploy a lot of BSs in a limited area and provide high quality communication services to users within short distances from the serving BSs [4].

Multi-connectivity and network densification lead us to rethink about communication systems across-the-board. In conventional systems, there have been many asymmetries in the capabilities of BS and UE in terms of computing power, communication hardware, electrical installation, etc. The network, consisting of BSs, determines all aspects of communications such as modulation and coding scheme (MCS), transmission power, and resource allocation in a cell-centric manner. This is why we call a conventional mobile communication system a cellular system. Nowadays, UEs with state-of-the-art hardware can run a variety of applications (apps) due to their high computing power and advanced connectivity, which makes users' service requirements very diverse. To meet the various needs of UEs, cell centric management schemes become very complicated since they should incorporate various objectives and constraints into algorithm design. This motivates us to consider user-centric communications beyond existing cell-centric communication paradigms [6, 26, 27].

In mmWave communication systems, active connections in a heavy loaded BS are easily dropped because dynamic blockages cause unexpected changes in radio resource demands [4, 28]. The authors of [4, 28] proposed a multi-connectivity scheme that enables ongoing connections to reroute data through other nearby BSs, and developed its analytical model that considers dynamic blockages and bandwidth reservation. In [26], we analyzed the impact of user mobility on the network performance in user-centric mmWave communication systems, based on the stochastic geometry model. Through extensive research, we noticed that we should jointly optimize multi-

connectivity and communication ranges, to cope with dynamics of blockages and mobility. In addition, fair user association is essential for connection continuity, which facilitates resource reservation by evenly distributing available resources to all BSs.

There are previous literature on user fairness and load balancing in wireless networks [29–32]. In [29], the authors show a strong correlation between user fairness and load balancing that justifies leveraging load balancing to achieve better user fairness in our proposed schemes. The authors of [29] define the BS load as the amount of time, necessary to provide access links to all users associated with the BS. Based on the definition, they derived the main theorem that the min-max load balanced association results in max-min user fairness in multiple association cases. In [30], the authors incorporated the channel quality and specific communication characteristics of 60GHz channels into the BS load, and proposed an association scheme that ensures fair load distribution among BSs. However, they did not consider multi-connectivity and blockages in the formulation.

In this chapter, we propose UE association control schemes for user-centric mmWave communication systems, and study their impact on overall network performance. We discuss the expected impact of association strategies on network performance and evaluate them in terms of network resource budget, network utility, peak load, and load balancing. We also address the traffic load management strategies among BSs in the network. It is more challenging in mmWave systems than in conventional mobile communication systems, because mmWave channels are vulnerable to various obstacles. The contributions of this chapter are as follows.

- We formulate two optimization problems that consider resource demand of each user and the blockage effect incurred by static blockages and self-body. They aim to achieve load balancing (peak load minimization (PLM)) and maximize network resource availability (sum load minimization (SLM)), respectively.

- We solve the problems by Lagrange duality [33] and develop heuristic online algorithms for realistic scenarios. For given associations, we derive the maximum

outage probability which is defined as buffer overflow at a BS due to NLOS channels.

- We numerically evaluate the impact of our proposed load management schemes on network performance. We examine their impact on the data exchange phase where connections are intermittently blocked by dynamic obstacles.

We organize the rest of this chapter as follows. In Section 3.2, we describe the system model and formulate the global optimization problem. In Section 3.3, we develop a heuristic method based on access pricing and peak load limit control and analyze the outage probability. We investigate the performance of our proposed schemes through extensive simulations in Section 3.4. Section 3.5 concludes the chapter.

3.2 System Model

In this section, we describe our system model for user-centric mmWave communication systems.

3.2.1 Network Model

We consider a radio access network (RAN) architecture in which the BS is functionally divided into central unit (CU) and distributed unit (DU). The CU controls the operation of one or more DUs, and the DU rovides access links to UEs. In the 3GPP NR, the connection between CU and DU is called F1 interface [34, 35]. We assume one CU supervises all DUs in the considered area, and wired links instantiate F1 interfaces to provide higher capacity than access links. Fig. 3.1a presents a macroscopic view of user-centric mmWave communication systems. We denote the set of DUs as $\mathcal{N}$ and its cardinality as N. Similarly, the set of UEs is denoted as $\mathcal{K}$ and its cardinality as K. Each DU is equipped with multiple antennas to serve UEs with analog beamforming. It has multiple RF chains and receives data from more than one DU within the coverage in parallel. We assume that each UE has M RF chains. For user-centric

(a) Network perspective

(b) User perspective
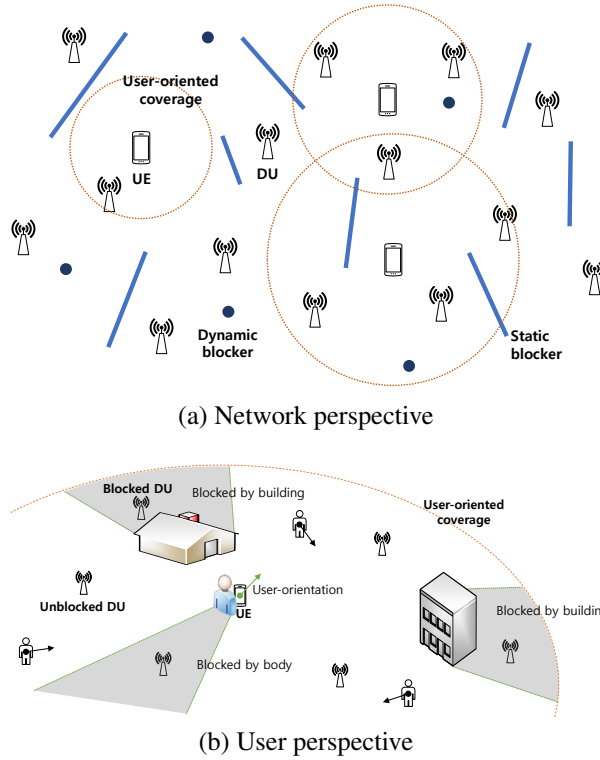
Figure 3.1. User centric mmWave communication system.

communications, each UE configures the coverage area independently according to its traffic demand [26].

We consider three types of blockage: static blockages, dynamic blockages, and self-body blocking [36], as shown in Fig. 3.1b. Permanent structures such as buildings and trees form static blockages, and mobile objects such as people and cars create dynamic blockages. We adopt a cone-shaped self-body blocking model where the user's body blocks signals within a shape of a cone with angle θ. In mmWave communications, it is preferable for UEs to be served by BSs providing LOS links because channel gains of NLOS links are significantly smaller than those of LOS links [20].

3.2.2 Channel Model and Achievable Rate

We adopt a bounded pathloss model for wireless channels. That is, the signal power is constant when the distance between DU and UE is smaller than 1, and decays exponentially with the distance. Let d_{ij} denote the distance between the i-th DU (DU_i) and the j-th UE (UE_j). NLOS links have significantly smaller channel gains compared to LOS links in mmWave communications. Therefore we neglect the NLOS links. We denote the channel gain of LOS link (i, j) as $G_{ij} = \kappa_L \cdot \min\left(1, d_{ij}^{-\alpha_L}\right)$, where κ_L and α_L are the pathloss intercept and the pathloss exponent for the LOS channel, respectively.

When DU_i transmits a signal with power P_{DU}, the signal arrives at UE_j with received strength $G_{ij} P_{DU}$. By Shannon's formula, the achievable downlink rate at link (i, j) is given as

$$r_{ij} = W \log_2 \left(1 + \frac{G_{ij} P_{DU}}{(N_0 + I_j) W}\right), \tag{3.1}$$

where W is the system bandwidth, N_0 is the background noise density, and I_j is the interference density at UE_j. However, it is well known that mmWaves propagate only in narrow beams and experience large attenuation with the distance. The interference is minimal and does not affect the link capacity substantially [20, 37]. We incorporate this property into our model. Then we simply obtain the achievable rate at link (i, j) as

$$r_{ij} = W \log_2 \left(1 + \frac{G_{ij} P_{DU}}{N_0 W}\right). \tag{3.2}$$

3.2.3 User Centric mmWave Communication Framework

We model a user-centric mmWave communication scenario as the following three stages: 1) user access allows a newly arriving UE to request an association with the network. The association request message includes contextual information such as traffic demand, multi-connectivity capability, and the set of associable DUs; 2) admission and association control makes the network (e.g., CU) decide admission and association of

the UE based on the long term channel condition in the online manner. The network periodically relocates existing associations in the offline manner to efficiently utilize radio resources networkwide; 3) data exchange allows each DU to transmit downlink data to the associated UEs with dynamic resource allocation considering short term channel conditions such as dynamic blockages.

User Access

In user-centric communications, a UE can associate with multiple DUs in its coverage according to the total traffic demand. For example, when the UE runs a high data rate application such as video streaming, it scales the coverage area small to attain high SNR or exploits multi-connectivity to achieve high throughput. The user access proceeds as follows: A newly arriving UE receives pilot signals such as system information signal in the 3GPP [38] from neighboring DUs and finds the set of associable LOS DUs. Then the UE requests an association with each candidate DU reporting its traffic demand.

We denote the uplink SNR at distance d from the UE as

$$\gamma(d) = \kappa_L \cdot \min(1, d^{-\alpha_L}) \frac{P_{UE}}{N_0 W},$$
(3.3)

where P_{UE} is the UE's transmit power[1]. Even when the channel becomes NLOS, the UE requires a minimum SNR of γ_{th} for reliable uplink transmission. Then, the UE's communication range is limited to

$$D_{\max} = \left(\frac{\kappa_N P_{UE}}{\gamma_{th} \cdot N_0 W} \right)^{\frac{1}{\alpha_N}},$$
(3.4)

where κ_N and α_N are the pathloss intercept and the pathloss exponent for NLOS channels, respectively. If the UE's coverage is defined as the circle centered at the UE's location with the radius of $D_{\max}$, the set of UE$_j$'s candidate DUs, $\mathcal{N}_j$, can be

---

[1]The 3GPP limits the UE's maximum transmit power to 23 dBm [39].

written as

$$\mathcal{N}_j = \{i \mid i \in \mathcal{N}, \ d_{ij} \leq D_{\max}, \ S_{ij} = 1\} \ \forall j \in \mathcal{K}, \qquad (3.5)$$

where $S_{ij} = 1$ if the channel (i, j) is LOS, and 0 otherwise. We denote the cardinality of $\mathcal{N}_j$ as N_j. If we denote the total traffic demand of UE_j by R_j, the maximum radio resource requirement of UE_j for DU_i is given as

$$\pi_{ij} = \begin{cases} \frac{R_j}{r_{ij}}, & \text{if } i \in \mathcal{N}_j \\ \infty, & \text{otherwise.} \end{cases} \qquad (3.6)$$

We assume that the total amount of radio resources for each DU is normalized to 1. UE_j can request associations with multiple DUs in $\mathcal{N}_j$ simultaneously, so the UE's radio resource requirement can be written as $\boldsymbol{\pi}_j = (\pi_{ij})_{i \in \mathcal{N}}$.

Admission and Association Control

We consider the scenario that all K UEs arrive at the same time to formulate a global optimization problem for admission and association control. The admission and association control based on the short term channel conditions may lead to high communication overhead and potentially disruptive connections due to channel dynamics [29]. Hence, it is reasonable to incorporate long term channel conditions such as pathloss, static blockages, and self-body blocking for association control in mmWave communications.

We define the load at DU_i incurred by UE_j as $w_{ij} \triangleq \pi_{ij}x_{ij}$, where $x_{ij} \in [0, 1]$ is the traffic split ratio of UE_j for DU_i and $\sum_i x_{ij} = 1$. Then, the total load at DU_i is given as

$$\rho_i = \sum_j w_{ij}a_{ij}, \qquad (3.7)$$

where $a_{ij} \in \{0, 1\}$ is a binary indicator whether DU_i is associated with UE_j or not.

The network should distribute the traffic load efficiently over the DUs to reduce

delays at bottleneck DUs and improve user satisfaction. We consider two load management strategies: 1) peak load minimization (PLM) and 2) sum load minimization (SLM). PLM aims at minimizing the maximum load of DUs while SLM targets maximizing the total amount of remaining radio resources in the network. We can write the objective function as

$$(1 - \tau) \max_i \rho_i + \tau \sum_i \rho_i, \tag{3.8}$$

where $\tau \in [0, 1]$. Especially, it results in PLM when $\tau = 0$, and SLM when $\tau = 1$. For $\tau \in (0, 1)$, the function represents a hybrid scheme.

Each UE i can be associated with M DUs at most, and we have two constraints, $\sum_i a_{ij} \leq M$ and $\sum_i x_{ij} a_{ij} = 1$ for $j \in \mathcal{K}$. At DUs, the radio resource requirement from each UE should be smaller than or equal to one, so we have $w_{ij} a_{ij} \leq 1 \ \forall i, j$. Further, the total load at each DU should be kept less than or equal to one to serve each associated UE at a required service level. That is, the network should admit the association requests only when $\rho_i \leq 1, \ \forall i \in \mathcal{N}$. We finally formulate the global optimization problem P_{global} for associations as follows.

$$P_{\text{global}} : \underset{(a_{ij}, x_{ij})_{\forall i,j}}{\text{minimize}} \quad (1 - \tau) \max_i \rho_i + \tau \sum_i \rho_i \tag{3.9a}$$

$$\text{subject to} \quad \rho_i \leq 1, \tag{3.9b}$$

$$w_{ij} a_{ij} \leq 1, \tag{3.9c}$$

$$\sum_i x_{ij} a_{ij} = 1, \tag{3.9d}$$

$$\sum_i a_{ij} \leq M, \tag{3.9e}$$

$$a_{ij} \in \{0, 1\}, \ x_{ij} \in [0, 1], \tag{3.9f}$$

$$i \in \mathcal{N}, \ j \in \mathcal{K}. \tag{3.9g}$$

The problem P_{global} is a mixed integer programming, whose computational complexity increases exponentially with the problem size KN. Moreover it requires global

information for construction. We notice that it is impractical to solve the problem in an online manner. However, P_{global} can be resolved in offline, i.e., the network periodically relocates existing associations to enhance the network utility or fairly distribute the network load. We also develop a heuristic online algorithm for P_{global} in the next section.

Data Exchange

During the data exchange phase, dynamic blockages can interrupt communication of established links. We assume that dynamic blocking events can be modeled by an exponential on-off process with the blocking rate Cd_{ij} and unblocking rate μ [36], where C is the dynamic blocking coefficient (see [36, *Lemma 1*]) and d_{ij} is the distance between DU_i and UE_j. Then, the blocking probability at link (i,j) due to dynamic blockers is given as

$$p_{ij} = \frac{Cd_{ij}}{Cd_{ij} + \mu}. \tag{3.10}$$

We consider two types of transmission in mmWave communications: delayed transmission and NLOS transmission. The delayed transmission assumes that the DU transmits data to the UEs on LOS links only. When the associated link becomes NLOS, the DU buffers data and transmits after the link returns to LOS, i.e., the transmission is delayed due to blockages. In the NLOS transmission, the DU transmits data to the UEs even on NLOS links accepting the reduced transmission rate.

For an arbitrary link, let r_L denote the achievable rate in the LOS state and r_N denote the rate in the NLOS state. Suppose that a DU transmits data to the UE whose demand is R for n time slots. If we use the delayed transmission, the expected amount of radio resources is $s_D = \frac{R}{r_L(1-p)}$, where p represents the blocking probability at the associated link. If we adopt the NLOS transmission, the expected amount of radio resources is $s_N = \frac{Rp}{r_L} + \frac{R(1-p)}{r_N}$. For $s_D < s_N$, the delayed transmission is more spectral efficient than the NLOS transmission. It is reasonable because in the NLOS transmission, each DU transmit data to the UEs whose achievable rates are low. In this

43

chapter, we consider the delayed transmission only.

Let $W_{ij}(t)$ be the amount of radio resources in DU_i required to serve UE_j at time slot t. We can express it as

$$W_{ij}(t) = w_{ij} + \sum_{k=1}^{t-1} p_{ij}^k (1 - p_{ij}) k w_{ij} \qquad (3.11)$$

$$= \frac{1 - t p_{ij}^t + (t-1) p_{ij}^{t+1}}{1 - p_{ij}} w_{ij}, \qquad (3.12)$$

where the first term of (3.11) is the radio resources for current incoming data transmission and the second term indicates the radio resources for buffered (delayed) data transmission. We define the average load at DU_i during T time slots as

$$\Omega_i(T) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{K_i} W_{ij}(t), \qquad (3.13)$$

where K_i is the number of associated UEs with DU_i. We assume that DUs are equipped with a sufficiently large buffer for downlink data for T time slots. Then the outage probability of DU_i can be expressed as

$$outage_i = \Pr\left\{\Omega_i(T) > 1\right\} = 1 - \Pr\left\{\Omega_i(T) \leq 1\right\}. \qquad (3.14)$$

3.3 Traffic Load Management

In this section, we develop online and offline methods for admission and association control. We then analyze the impact of the load management strategy on the performance of data exchange in terms of DU outage.

3.3.1 Optimal Association and Admission Control

PLM

For network load management, we first consider the PLM problem (i.e., $\tau = 0$) when UE_j's arrives newly. Let σ_i denote the load of DU_i just before the arrival of UE_j. Minimizing the maximum value of $(\rho_i)_i$ is equivalent to minimizing the upper bound of $(\rho_i)_i$. Then, we can express the online PLM problem as follows.

$$P_{PLM} : \underset{(a_{ij}, x_{ij})_{\forall i}}{\text{minimize}} \quad u \tag{3.15a}$$

$$\text{subject to} \quad \pi_{ij} x_{ij} a_{ij} + \sigma_i \leq u, \tag{3.15b}$$

$$\pi_{ij} x_{ij} a_{ij} + \sigma_i \leq 1, \tag{3.15c}$$

$$\pi_{ij} x_{ij} a_{ij} \leq 1, \tag{3.15d}$$

$$\sum_i x_{ij} a_{ij} = 1, \tag{3.15e}$$

$$\sum_i a_{ij} \leq M, \tag{3.15f}$$

$$a_{ij} \in \{0, 1\}, \ x_{ij} \in [0, 1], \tag{3.15g}$$

$$i \in \mathcal{N}_j. \tag{3.15h}$$

We can omit the constraint (3.15d) since it is implied by (3.15c). We notice that $u \leq 1$ if the feasible set formed by (3.15c), (3.15e), (3.15f), and (3.15g) is not empty. The problem P_{PLM} can be transformed to a mixed integer linear programming (MILP) by defining dummy variable $y_{ij} = x_{ij} a_{ij}$. We can solve the MILP problem by using the existing software packages such as CPLEX and MOSEK. However, the complexity grows rapidly with the problem size [40–42].

For now, we solve P_{PLM} by relaxing the multi-connectivity constraint (3.15f) and investigate the characteristics of its optimal solution, which will be used to develop an online algorithm for the original P_{PLM}. We write the relaxed problem P_{PLM}^{relax} as

follows.

$$P_{PLM}^{\text{relax}}: \ \underset{(x_{ij})_{\forall i}}{\text{minimize}} \quad u \tag{3.16a}$$

$$\text{subject to} \quad \pi_{ij} x_{ij} + \sigma_i \leq u, \tag{3.16b}$$

$$\sum_i x_{ij} = 1, \tag{3.16c}$$

$$x_{ij} \in [0, 1], \ i \in \mathcal{N}_j. \tag{3.16d}$$

We set the Lagrangian by dualizing (3.16b) and (3.16c) as

$$L_0(u, \mathbf{x}_j, \boldsymbol{\lambda}, \nu)$$

$$= u + \sum_{i \in \mathcal{N}_j} \lambda_i \left(\pi_{ij} x_{ij} + \sigma_i - u \right) + \nu \left(\sum_{i \in \mathcal{N}_j} x_{ij} - 1 \right)$$

$$= u \left(1 - \sum_{i \in \mathcal{N}_j} \lambda_i \right) + \nu \left(\sum_{i \in \mathcal{N}_j} x_{ij} - 1 \right) + \sum_{i \in \mathcal{N}_j} \lambda_i \left(\pi_{ij} x_{ij} + \sigma_i \right) \tag{3.17}$$

where $\mathbf{x}_j = (x_{ij})_{i \in \mathcal{N}_j}$ is the association vector, $\boldsymbol{\lambda} = (\lambda_i)_{i \in \mathcal{N}_j}$ is the Lagrange multipliers for (3.16b), and ν is the Lagrange multiplier for (3.16c).

We denote the Lagrange dual function as $g_0(\boldsymbol{\lambda}, \nu)$, which is given by

$$g_0(\boldsymbol{\lambda}, \nu) = \inf_{u, \mathbf{x}_j} L_0(u, \mathbf{x}_j, \boldsymbol{\lambda}, \nu) \tag{3.18}$$

$$= \begin{cases} \displaystyle\inf_{\mathbf{x}_j} \sum_{i \in \mathcal{N}_j} \lambda_i \left(\pi_{ij} x_{ij} + \sigma_i \right), & \text{if } \sum_{i \in \mathcal{N}_j} \lambda_i = 1 \text{ and } \sum_{i \in \mathcal{N}_j} x_{ij} = 1 \\ -\infty \text{ (unbounded below)}, & \text{otherwise,} \end{cases} \tag{3.19}$$

where the equality holds because $u \left(1 - \sum_{i \in \mathcal{N}_j} \lambda_i \right)$ is bounded below if $\left(1 - \sum_{i \in \mathcal{N}_j} \lambda_i \right) = 0$ and $\nu \left(\sum_{i \in \mathcal{N}_j} x_{ij} - 1 \right)$ is bounded below if $\left(\sum_{i \in \mathcal{N}_j} x_{ij} - 1 \right) = 0$. Therefore we

Algorithm 1 Find $x_{ij}^*(u, \boldsymbol{\lambda})$ for given u and $\boldsymbol{\lambda}$

1: Initialize $\mathbf{x}_j = \mathbf{0}$, $m = 0$, and $\epsilon = 1$
2: **repeat**
3:     Find $k = \arg\min_{i \in \mathcal{N}_j} \lambda_i \pi_{ij}$
4:     Set $x_{kj} = \min\left(1 - \epsilon, \frac{u - \sigma_k}{\pi_{kj}}\right)$ and $m = m + 1$
5:     Manipulate the DU candidate set $\mathcal{N}_j = \mathcal{N}_j \setminus \{k\}$
6:     Set $\epsilon = \epsilon - \sum_i x_{ij}$
7: **until** $\sum_i x_{ij} < 1$ and $m < M$

---

can write the Lagrange dual of P_{PLM}^{relax} as

$$P_{PLM}^{\text{dual}} : \underset{\boldsymbol{\lambda}}{\text{maximize}} \quad g_0(\boldsymbol{\lambda}) \tag{3.20a}$$

$$\text{subject to} \quad \sum_i \lambda_i = 1 \tag{3.20b}$$

$$\lambda_i \geq 0, \quad \forall i, \tag{3.20c}$$

where

$$g_0(\boldsymbol{\lambda}) = \inf_{\mathbf{x}_j \in \mathcal{X}} \sum_{i \in \mathcal{N}_j} \lambda_i \left(\pi_{ij} x_{ij} + \sigma_i\right)$$

and

$$\mathcal{X} = \left\{ \mathbf{x}_j \,\middle|\, \sum_{i \in \mathcal{N}_j} x_{ij} = 1, \, 0 \leq x_{ij} \leq \frac{u - \sigma_i}{\pi_{ij}} \right\}.$$

The dual problem is always convex, so we can solve the problem by iterative methods [33]. Specifically, P_{PLM}^{dual} can be solved via the projected subgradient method since its objective function is non-differentiable and concave [43]. The computational complexity of the subgradient method is $\mathcal{O}(1/\Delta^2)$ theoretically where, unlike the branch and bound method, Δ sets the stopping condition regardless of the problem size. However, we cannot guarantee the convergence time of the subgradient method due to its dependency on the step size.

For fixed u, P_{PLM}^{relax} becomes a convex feasibility problem since the objective function is constant. For given u, we find the optimal Lagrange multiplier $\boldsymbol{\lambda}^*(u)$ by solving

P_{PLM}^{dual}. For $\boldsymbol{\lambda}^*(u)$, a feasible $\mathbf{x}_j^*$ minimizes the Lagrangian $L_0(u, \mathbf{x}_j, \boldsymbol{\lambda}, \nu)$. Looking at (3.19), we can see that x_{ij}^* is large for DU_i with small $\lambda_i^* \pi_{ij}$. Based on this observation, we develop an online algorithm for P_{PLM} (i.e., Algorithm 1) incorporating the multi-connectivity constraint (3.15f) also. The algorithm produces the best split ratio x_{ij}^* for given u and $\boldsymbol{\lambda}$. When u is chosen small, P_{PLM}^{relax} can be infeasible or feasible but $\sum_i a_{ij} > M$ as the constraint (3.16b) is strict. If we increase u sufficiently, there exist a feasible solution with $\sum_i a_{ij} \le M$. This implies that, through controlling the peak load limit u and the access price $\boldsymbol{\lambda}$, we can make the newly arriving UE find the optimal association autonomously without exchanging information with the other UEs.

When the peak load limit u is chosen, we may determine the access price $\boldsymbol{\lambda}$ heuristically without solving the dual problem P_{PLM}^{dual}. We design the access price carefully considering both spectral efficiency and load distribution. That is, we propose a heuristic access pricing scheme for PLM as follows.

$$\tilde{\lambda}_i(\delta) = \begin{cases} \delta \frac{\pi_{ij}}{\sum_{i \in \mathcal{N}_j} \pi_{ij}} + (1 - \delta) \frac{1 - \sigma_{ij}}{\sum_{i \in \mathcal{N}_j} (1 - \sigma_{ij})}, & \text{for } i \in \mathcal{N}_j \\ \infty, & \text{otherwise,} \end{cases} \quad (3.21)$$

where the first term reflects the spectral efficiency and the second term does the current traffic load, and the two terms are combined after being weighted by $\delta \in [0, 1]$ and $1 - \delta$, respectively. Note that the weights of $\delta = 0$ and 1 represent the least-loaded-first (LLF) association and the strongest-signal-first (SSF) association, respectively. PLM minimizes the peak load by distributing the load fairly, so we can approximate the access price of PLM by choosing a proper δ in (3.21).

For given u, the network such as CU admits the UE_j's access request if $V_j(u) = \sum_{i \in \mathcal{N}_j'} \min \left(1, \frac{u - \sigma_i}{\pi_{ij}} \right) \ge 1$. This is because there remain radio resources sufficiently enough to satisfy UE_j's demand. We here define the index set of DUs with the M

Algorithm 2 Online PLM method

1: u and $\lambda_i = \tilde{\lambda}_i(\delta)$ are distributed to each DU $(i \in \mathcal{N})$
2: UE_l reports its arrival to DU_i $(i \in \mathcal{N}_l)$
3: **if** $V_l(u) \geq 1$ **then**
4:     $\mathbf{x}_j^*$ is determined by Algorithm 1
5: **else if** $V_l(1) \geq 1$ **then**
6:     **repeat**
7:         Set $\varepsilon = \min(1 - \max_i \sigma_i, \ \varepsilon + \tilde{\Delta})$
8:     **until** $V_l(u) < 1$
9:     $\mathbf{x}_j^*$ is determined by Algorithm 1
10: **else**
11:     Reject UE_l
12: **end if**
13: Update load distribution status $\sigma_i = \sigma_i + w_{ij}$ $(i \in \mathcal{N})$
---

largest available radio resource fraction, i.e.,

$$\mathcal{N}_j' = \arg \max_{i \in \mathcal{N}_j} {}_{[M]} \left(\min \left(1, \ \frac{u - \sigma_i}{\pi_{ij}} \right) \right), \tag{3.22}$$

where $\max_{[n]} \mathcal{X}$ is the operator for choosing the n largest elements in the set $\mathcal{X}$. For example, $\max_{[2]}\{2, 6, 4, 8\} = \{6, 8\}$ and $\arg\max_{[2]}\{2, 6, 4, 8\} = \{2, 4\}$. We propose to set the peak load limit as follows[2].

$$u = \min \left(1, \ \varepsilon + \max_{k \in \mathcal{N}_j} \sigma_k \right), \tag{3.23}$$

where ε is the margin for the peak load to suppress the spectral inefficiency. The margin is set small for the high network load and large for the low network load. When $V_j(1) \geq 1$, we have sufficient radio resources to accommodate UE_j. However, if the peak load limit u is chosen as $V_j(u) < 1$, the UE_j's access request is denied even though there remain radio resources to serve the user, i.e., starvation occurs. In that

---
[2]In the perspective of 3GPP NR, the peak load limit can be interpreted as the access barring. NR supports overload and access control functionality such as UE based access barring mechanisms [44, 45]. One unified access control framework specified in TS 22.261 applies to the UE for NR [46]. The network broadcasts barring control information and then the UE determines whether an access attempt is authorized based on the barring information broadcast.

Algorithm 3 Online SLM method

1: $\lambda_i = \tilde{\lambda}_i(1)$ are distributed to each DU $(i \in \mathcal{N})$
2: UE_l reports its arrival to DU_i $(i \in \mathcal{N}_l)$
3: **if** $V_l(1) \geq 1$ **then**
4:    $\mathbf{x}_j^*$ is determined by Algorithm 1 with $u = 1$ and $\lambda_i = \tilde{\lambda}_i(1)$ for $i \in \mathcal{N}$
5: **else**
6:    Reject UE_l
7: **end if**
8: Update load distribution status $\sigma_i = \sigma_i + w_{ij}$ $(i \in \mathcal{N})$
---

case, the CU increases the peak load limit. Specifically, $\varepsilon = \min(1 - \max_i \sigma_i, \ \varepsilon + \tilde{\Delta})$, where $\tilde{\Delta}$ is the increment step for the peak load limit. We present the online PLM method in Algorithm 2.

SLM

We consider the SLM problem (i.e., $\tau = 1$) when UE_j arrives newly. Then we can write the SLM online problem as

$$P_{SLM} : \underset{(a_{ij}, x_{ij})_{\forall i}}{\text{minimize}} \quad \sum_{i \in \mathcal{N}_j} (\pi_{ij} x_{ij} a_{ij} + \sigma_i) \tag{3.24a}$$

$$\text{subject to} \quad \pi_{ij} x_{ij} a_{ij} + \sigma_i \leq 1, \tag{3.24b}$$

$$\sum_i x_{ij} a_{ij} = 1, \tag{3.24c}$$

$$\sum_i a_{ij} \leq M, \tag{3.24d}$$

$$a_{ij} \in \{0, 1\}, \ x_{ij} \in [0, 1], \tag{3.24e}$$

$$i \in \mathcal{N}_j. \tag{3.24f}$$

A UE places the least burden on the network when accommodated by the nearest DU in terms of the long term channel condition. So the sum load is minimized when UE_j is associated with the m_j nearest DUs, where m_j is the minimum number of DUs for satisfying UE_j's demand. Note that the SLM is equivalent to SSF association, where

each user requests an association to the nearest DU first. Hence we can obtain the online solution by letting $u = 1$ and $\lambda_i = \tilde{\lambda}_i(1), \forall i \in \mathcal{N}$ in Algorithm 1. Algorithm 3 summarizes the online SLM method.

3.3.2 Outage Analysis

We now investigate the DU outage probability, which is given in (3.14). We first consider the average load of DU_i for T time slots (see (3.13)).

$$
\begin{aligned}
\Omega_i(T) &= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{K_i} \frac{1 - t p_{ij}^t + (t-1) p_{ij}^{t+1}}{1 - p_{ij}} w_{ij} \\
&= \sum_{j=1}^{K_i} \frac{w_{ij}}{1 - p_{ij}} \cdot \underbrace{\frac{1}{T} \sum_{t=1}^{T} \left(1 - t p_{ij}^t + (t-1) p_{ij}^{t+1} \right)}_{(*)} \\
&\overset{(a)}{\approx} \sum_{j=1}^{K_i} \frac{w_{ij}}{1 - p_{ij}} \cdot \left(1 - \frac{p_{ij}}{T} \right)
\end{aligned}
\tag{3.25}
$$

where the approximation (a) comes from $(*) \approx \left(1 - \frac{p_{ij}}{T} \right)$ when $p_{ij} \ll 1$. By plugging (3.10) into (3.25), we have

$$
\Omega_i(T) \approx \sum_{j=1}^{K_i} \left(1 + \left(1 - \frac{1}{T} \right) \frac{C}{\mu} d_{ij} \right) w_{ij}
\tag{3.26}
$$

$$
= \sum_{j=1}^{K_i} w_{ij} + \left(1 - \frac{1}{T} \right) \frac{C}{\mu} \sum_{j=1}^{K_i} w_{ij} d_{ij}
\tag{3.27}
$$

$$
\leq K_i \eta_i + \left(1 - \frac{1}{T} \right) \frac{C}{\mu} \eta_i \sum_{j=1}^{K_i} d_{ij} \triangleq \bar{\Omega}_i(T),
\tag{3.28}
$$

where $\eta_i = \max_j w_{ij}$ is the biggest weight among the associations with DU_i. The distance between UE and DU is distributed according to the probability density function $\frac{2d}{D_{\max}^2}$, $d \in [0, D_{\max}]$, whose mean and variance are $\frac{2}{3} D_{\max}$ and $\frac{1}{18} D_{\max}^2$, respec-

tively [36]. Then the mean load of DU_i is evaluated as

$$\mathbb{E}\left[\Omega_i(T)\right] \approx \sum_{j=1}^{K_i} w_{ij} + \left(1 - \frac{1}{T}\right) \frac{C}{\mu} \sum_{j=1}^{K_i} w_{ij}\mathbb{E}[d_{ij}]$$

$$= \rho_i + \left(1 - \frac{1}{T}\right) \frac{2CD_{\max}}{3\mu}\rho_i, \tag{3.29}$$

where the second term stands for the average load increment at DU_i during T slots due to dynamic blockages. The load increment by dynamic blockages is proportional to the association load ρ_i, which implies that the heavy-loaded DUs experience outages often. Hence, it is important to minimize the peak load on the network to suppress outages. Since the SSF strategy (i.e., SLM) creates heavy associations to DUs in a hot-spot area, it can lead to large fluctuations in resource demand and frequent service disruptions.

We can approximate $\bar{\Omega}_i(T)$ as a Gaussian random variable by the Central Limit Theorem [21], whose mean and variance are respectively $\left(1 + \left(1 - \frac{1}{T}\right) \frac{2CD_{\max}}{3\mu}\right) K_i\eta_i$ and $\left(1 - \frac{1}{T}\right)^2 \frac{C^2D_{\max}^2}{18\mu^2} K_i\eta_i^2$. Since $\Omega_i(T) \le \bar{\Omega}_i(T)$, the outage probability of DU_i is upper-bounded by

$$outage_i = \Pr\left\{\Omega_i(T) > 1\right\} \le \Pr\left\{\bar{\Omega}_i(T) > 1\right\}, \tag{3.30}$$

which equals

$$\Pr\left\{\bar{\Omega}_i(T) > 1\right\} = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{3T\mu}{(T-1)CD_{\max}\sqrt{K_i}\eta_i} - \frac{3T\mu\sqrt{K_i}}{(T-1)CD_{\max}} - 2\sqrt{K_i}\right)\right), \tag{3.31}$$

where $\mathrm{erf}(x) = \int_0^x e^{-t^2}dt$ is the error function. To obtain $outage_i \le 1 - \varrho$,

$$\Pr\left\{\bar{\Omega}_i(T) > 1\right\} \le 1 - \varrho \tag{3.32}$$

is sufficient. By plugging (3.31) into (3.32), we obtain the condition of the maximum

weight of DU_i, which is given by

$$\eta_i \leq \frac{3T\mu}{(T-1)CD_{\max}\sqrt{K_i}\mathrm{erf}^{-1}(2\varrho-1)+(3T\mu+2(T-1)CD_{\max})K_i}, \quad (3.33)$$

where $\mathrm{erf}^{-1}(y)$ is the inverse error function. The inequality can be applied to the peak load limit control (3.23). That is, we can make the biggest weight η_i bounded as (3.33) by limiting the peak load margin ϵ. Each DU computes the upper bound of the association weight with the current number of associations and the target outage probability. Then, the CU calculates the peak load limit to keep the outage probability below $1-\varrho$ by collecting the max association weight from DUs, i.e.,

$$\tilde{u} = \max_{k \in \mathcal{N}_j} \min\left(1, \; \sigma_k + \eta_k\right).$$

We set ε smaller than the gap between $\tilde{u}$ and the current peak load limit, i.e.,

$$\varepsilon \leq \max_{k \in \mathcal{N}_j} \min\left(1, \; \sigma_k + \eta_k\right) - \max_{k \in \mathcal{N}_j} \sigma_k. \quad (3.34)$$

3.4 Performance Evaluation

In this section, we provide numerical results to evaluate the proposed UE association methods and load management strategies.

3.4.1 Evaluation Environments

We assume that the carrier frequency is 28 GHz, the pathloss intercept is $(\kappa_L, \kappa_N) = (10^{-\frac{61.4}{10}}, 10^{-\frac{72.0}{10}})$, and the pathloss exponent is $(\alpha_L, \alpha_N) = (2.0, 2.92)$ [20, 47]. We set the transmit power of DU and UE to $P_{DU} = 20$ dBm and $P_{UE} = 23$ dBm, respectively [30]. The DU transmit power is set much lower than that of a conventional sub 6 GHz communication system (e.g., 30 dBm [44]) considering the beamforming. Reduced transmit power is beneficial to interference suppression in mmWave

communications. The background noise density and system bandwidth are $N_0 = -134$ dBm/MHz and $W = 400$ MHz, respectively. Letting the uplink SNR threshold $\gamma_{th} = 0$ dB, the maximum communication range of the UE is determined as $D_{\max} = 104.7$ m.

We place static blockages uniformly with the density $\beta = 5.4$/km [26], where the height of a static blockage is assumed to be larger than those of a DU and a UE. The orientation of a UE is generated uniformly in $(0, 2\pi)$ and the self-body blocking angle θ is set to $60°$. We assume the traffic flow control works perfectly so that the UE does not experience errors when the splitted traffic merges. Dynamic blockers are uniformly distributed with the density of 0.01 bl/m^2. We set the heights of a dynamic blocker, DU, and UE as 1.8, 5, and 1.4 m, respectively, and the speed of a mobile blocker as 1 m/s. Correspondingly, the dynamic blocking coefficient is given by $C = 7.07 \times 10^{-4}$.

We assume that each UE is equipped with three RF chains and its total traffic demand is uniformly distributed in $(0, 1000)$ Mbps. We consider the following performance metrics: 1) average resource budget (i.e., normalized unused resources) defined as $1 - \frac{\sum_{i \in \mathcal{N}} \rho_i}{N}$; 2) network utility obtained by $\frac{\sum_{j \in \mathcal{K}} R_j \cdot \mathbb{I}\left(\sum_{i \in \mathcal{N}_j} a_{ij} > 0\right)}{\sum_{i \in \mathcal{N}} \rho_i}$; 3) fairness index defined as $\frac{\left(\sum_{i \in \mathcal{N}} \rho_i\right)^2}{N \cdot \sum_{i \in \mathcal{N}} \rho_i^2}$, which becomes 1 when all DUs are equally loaded; 4) peak load $\max_i \rho_i$; and 5) outage probability.

We consider the following three schemes for performance comparison with our proposed scheme: 1) PLM.global (or SLM.global) finds the optimal PLM (or SLM) association for the existing and arriving UEs by solving P_{global} with $\tau = 0$ (or 1); 2) PLM.online (or SLM.online) finds the optimal association for the newly arriving UE without interrupting the existing associations by solving P_{PLM} (or P_{SLM}); and 3) PLM.on/offline (or SLM.on/offline) runs PLM.online (or SLM.online) when the new UE arrives and the existing associations are relocated by periodically running PLM.global (or SLM.global). We denote our proposed method by Heuristic(τ, δ), e.g., Heuristic$(1, 1)$ represents the online SLM method. Our proposed method includes the periodic association relocation with the solution of P_{global}.

(a) Average resource budget

(b) Network utility

(c) Peak load

(d) Fairness
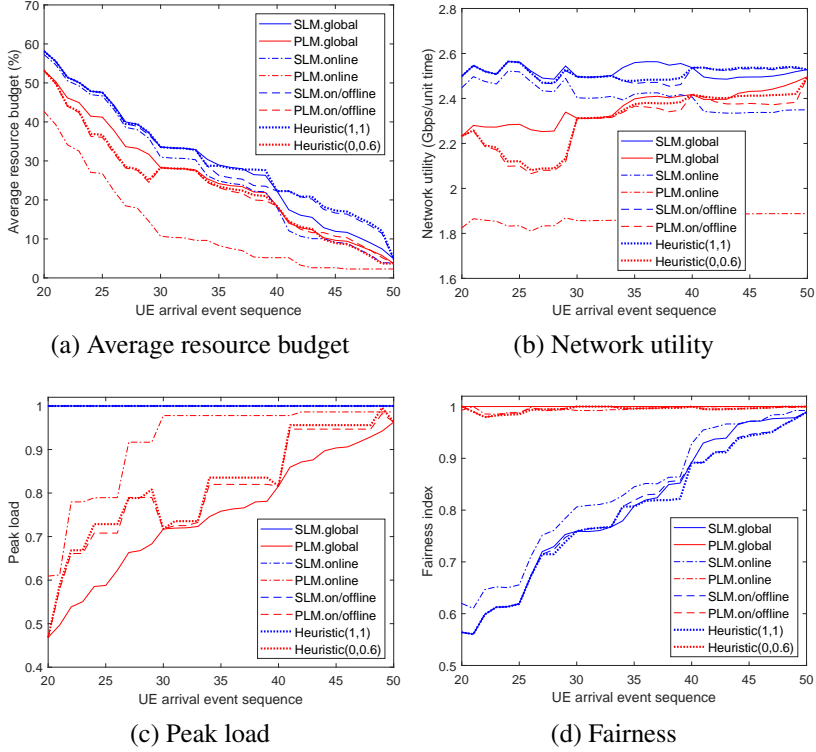
Figure 3.2. Performance evaluation vs UE arrival event sequence.

3.4.2 Performance Comparison

We first examine the online performance of our heuristic method. We assume that the DU density is 1000 DUs/km^2 and UEs sequentially arrive at the network until the network is saturated. It is difficult to get insights at a very early stage because the network resource budget is sufficient to accept any UE access request. Therefore, we observe variations of performance metrics from when the average resource budget reaches about 50%. Fig. 3.2 presents the results of performance update. SLM shows better performance in terms of average resource budget and network utility, while PLM shows better performance in terms of fairness and peak load. Specifically, for the SLM case, SLM.global and SLM.online show the best and worst performances, respectively, and SLM.on/offline shows the performance in the middle of these. The PLM

55
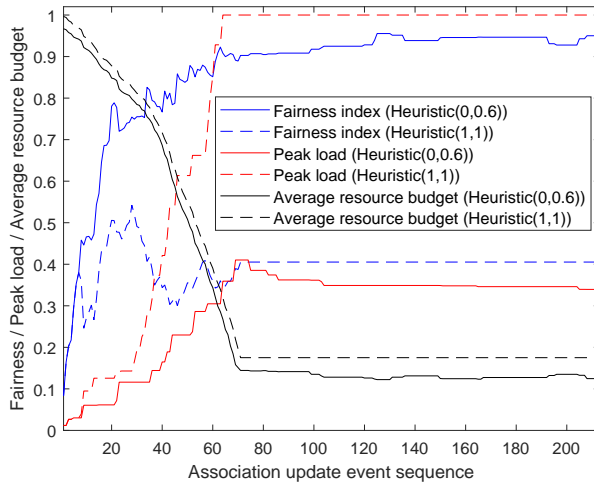
Figure 3.3. Performance change of online association algorithms over time.

case has the same tendency as the SLM case.

Fig. 3.2 shows that Heuristic(1,1) and Heuristic(0,0.6) show almost the same update trend as SLM.on/offline and PLM.on/offline, respectively. Our methods find the near optimal solutions to the online optimal problems P_{PLM} and P_{SLM}. We also observe that the performance gap narrows as the network becomes saturated. In addition, it is possible to provide association rules that achieve various purposes like PLM and SLM by adjusting τ and δ. Our methods search for associations through an intuitive access pricing scheme (3.21) rather than an iterative algorithm.

If the user mobility is dynamic, it is hard to manage associations with periodic global optimization because the problem statement can be changed before solving the problem. Fig. 3.3 reveals the online performance of our method. Here, we assume that the periodic association relocation is not included in our method. In this simulation, 10 DUs and 40 UEs are randomly distributed in 100 m $\times$ 100 m area. At the very early stage, 40 UEs arrive at the network in sequence, which is called the association phase. Each UE moves by distance v and toward φ.

We assume the random mobility model, i.e., $v \sim \text{Exp}(10)$ and $\varphi \sim \text{Uniform}(0, 2\pi)$. If UE_j's mobility affects its connection status, the UE searches for new DUs to asso-
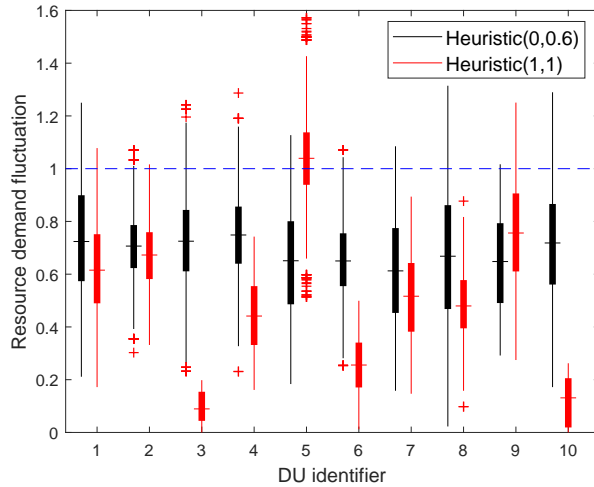
Figure 3.4. Per-DU resource demand fluctuation according to load management strategies in the data exchange stage.

ciate with $\mathsf{Heuristic}(\tau, \delta)$. We observe the performance of PLM (i.e., $\tau = 0,\ \delta = 0.6$) improves over time and outperforms the SLM (i.e., $\tau = 1, \delta = 1$). Specifically, the peak load decreases although the average resource budget changes gradually. The PLM redistributes the loads from overloaded DUs to the less loaded DUs when the UE mobility occurs. As a result, the online PLM method has the fairness index of almost one.

We evaluate the load fluctuation of each DU under the dynamic environment to study the impact of load management strategies on the data exchange stage. At the association phase, each UE makes connections with DUs according to PLM or SLM considering the long-term channel condition such as pathloss, static blockages, and self-body blocking. Then, each DU serves the associated UEs with dynamic scheduling based on fast channel variation incurred by dynamic blockers during the data exchange phase.

Fig. 3.4 shows the resource demand fluctuation of each DU, where 10 DUs and 40 UEs are uniformly distributed in the $100\,\mathrm{m} \times 100\,\mathrm{m}$ area. We observe that more than a half of the resources required by DU_5 are distributed over one, which implies that DU_5 probably goes through severe resource outages due to dynamic blockages. On the other
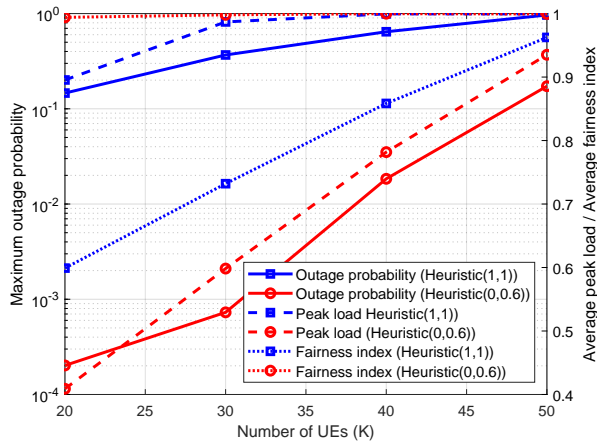
Figure 3.5. Maximum outage probability vs number of UEs.

hand, most of the resources required by the PLM strategy are distributed below one, which indicates that PLM can suppress outages. PLM reduces load variation across the network and stabilizes the data exchange phase in dynamic environments.

Fig. 3.5 presents the maximum outage probability, average peak load, and average fairness index as a function of the number of UEs. We consider a scenario that 10 DUs and K UEs are randomly distributed on the 100 m $\times$ 100 m area. In simulation, we realize the network topology 10,000 times, we run each topology for 100,000 time slots. We assume that an outage occurs when the average amount of buffered data for 10 time slots exceeds the DU's capacity. We measure the maximum outage probability as the largest outage probability among 10 DUs during data exchange for a given network realization.

Fig. 3.5 shows the probabilities obtained by averaging the results for all realizations. The peak load and fairness index results are obtained in the same manner. We observe that the outage probability increases as the number of UEs increases (i.e., the network becomes saturated). The outage performance of PLM, Heuristic(0,0.6), outperforms that of SLM, Heuristic(1,1). This is because that PLM has the characteristics of fairer load distribution than SLM. From the figure, we know that PLM adapts more flexibly to dynamic environments than SLM.

3.5 Conclusion

In this chapter, we proposed the online method for admission and association in user-centric mmWave communication systems. We considered mmWave specific characteristics such as static/dynamic blockages and self-body blocking to reflect realistic channel models. We formulated the integrated optimization problems for peak load maximization and sum load maximization whose objectives are load balancing and the maximization of radio resource budget, respectively. According to the insights obtained from solving the problems, we proposed the heuristic access pricing and peak load limit control. We also developed the analytic framework to investigate the outage probability. We evaluated the proposed method in terms of average resource budget, network utility, peak load, fairness, and outage.

Through simulation, we confirm tht SLM shows better performance than PLM in terms of average resource budget and network utility, while PLM shows better performance than SLM in terms of fairness and peak load. Our proposed methods perform very closely to the centralized scheme in overall performance and work well in dynamic environments. Finally, we found that the fair load distribution (i.e., PLM) made the network more resistant to outages. As future work, it is of our interest to design an access policy in mmWave Integrated Access and Backhaul (IAB), where mmWave BSs form a multi-hop network.

Chapter 4

Path Selection and Path-Aware Access Pricing Policy in Millimeter Wave IAB Networks

We extend the work in Chapter 3 to mmWave IAB networks in this chapter.

4.1 Introduction

Millimeter wave (mmWave) communication is a promising technology for next generation communication systems that can provide multi-gigabits/sec throughput to users. But it has limitations because mmWave signals do not propagate long distance in the air due to high frequency characteristics and they are vulnerable to blockages such as buildings, forests, people, etc [1]. Blockages make wireless channels between Base Stations (BSs) and User Equipments (UEs) as Non-Line-Of-Sight (NLOS), and seriously aggravate users' Quality-of-Experience (QoE) by lowering received signal strengths. So, in mmWave communications, it is crucial to keep connections in Line-Of-Sight (LOS) to mitigate the blockage effect.

The network densification approach uses Ultra Dense Network (UDN), where network operators deploy a lot of BSs in a limited area and provide high quality communication services to users within short distances from the serving BSs to overcome the limitation of mmWave signal propagation [4]. However, UDN involves high capital
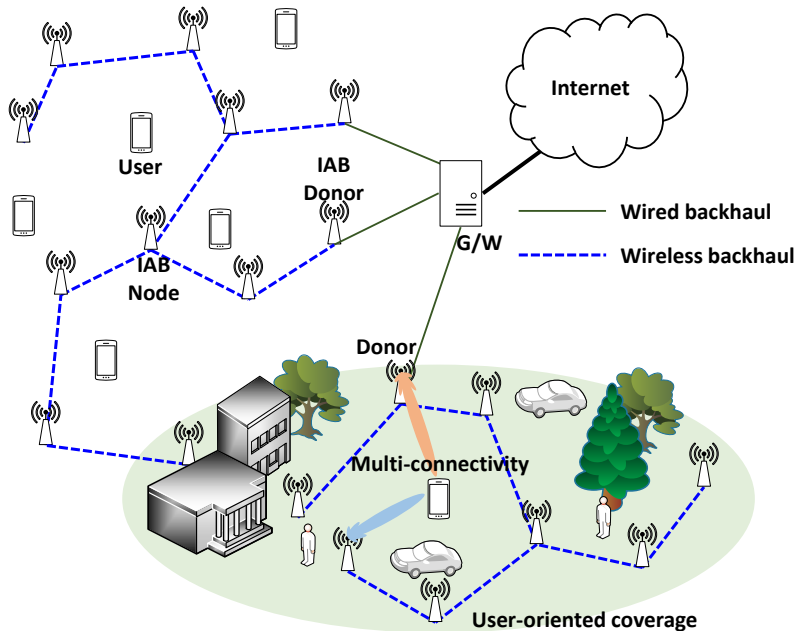
Figure 4.1. A mmWave IAB network.

and operational expenditures (CAPEX and OPEX) for network operators since high-capacity backhauling should be provided to a greater number of BSs than networks operating at lower frequencies. To address the challenges, the integrated-access-and-backhaul (IAB) framework is promising as a cost effective alternative to the fully wired backhaul. Since a mmWave network is deployed as a multi-hop network by IAB as shown in Fig. 4.1, we need to consider how to route from IAB-donor to user in mmWave IAB networks because BSs or wireless backhaul links on the path can severely degrade user QoE.

In this chapter, we investigate the geographic and pathloss models suitable for describing urban street canyon environments. Then, we propose a mmWave IAB network model and find the dominant signal path for a given transmitter and receiver in an urban street canyon. In order to develop a path selection strategy for mmWave IAB networks, we formulate a routing problem by revisiting the idea of NUM. Specifically, we newly define the link metric for Dijkstra method and the utility maximization prob-

(a) Real map	(b) Building extraction	(c) Street extraction
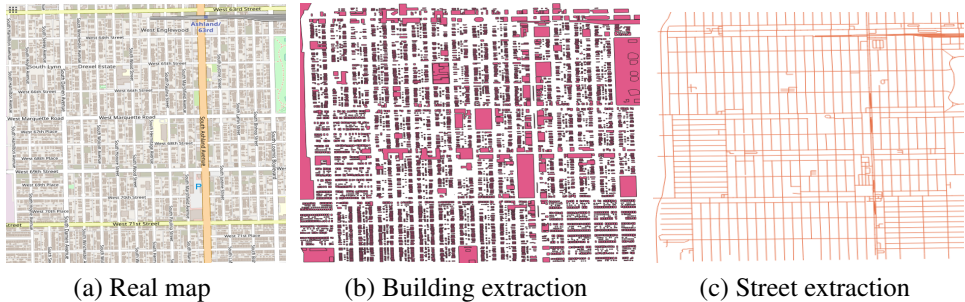
Figure 4.2. A snapshot of a region in Chicago from *OpenStreetMap* (latitude: $41.762°$N – $41.78°$N, longitude: $-87.678°$W – $-87.658°$W), with a size of 1.659×2.002 (km^2).

lem for QoS guaranteed networks. We also propose a path-aware access pricing policy for user association based on path dependency of multi-hop networking. The access price for a user is sum of backhauling cost and access cost, where the former implies path-dependency and the latter include both spectral efficiency with access BS and channel dynamics. Finally, we evaluate our proposed scheme in various urban street canyon scenarios with comparing the existing schemes.

4.2   System Model

In this section, we describe our model for user-centric mmWave IAB networks, which is depicted in Fig. 4.1. We focus on the downward communication, i.e., from the donor to UEs.

4.2.1   Geographic and Pathloss Models

To evaluate the mmWave IAB networks in an urban environment, we need to consider how to abstract geographic data from an urban area. The raw map data can be obtained by *OpenStreetMap* powered by open source software and [48, 49]. There are two kinds of geographic data abstraction method: 1) building extraction, where buildings are extracted as polygons and 2) street extraction, where outdoor streets are extracted as lines. Note that we can extract map data by using GIS tool *QGIS* [50]. Fig. 4.2 presents

a snapshot of a region in Chicago from *OpenStreetMap*, where latitude and longitude are $41.762°\mathrm{N} - 41.78°\mathrm{N}$ and $-87.678°\mathrm{W} - -87.658°\mathrm{W}$, respectively. As shown in Fig. 4.2, the building extraction is more realistic than the street extraction but the first needs to a complicated ray tracing model while the latter provides tractable model so we can setup a mathematical framework. Specifically, based on street extraction, an urban environment can be modeled as an urban street canyon. In urban street canyon scenarios, horizontal and vertical streets are generated from two independent one dimensional homogeneous PPP with identical street intensity λ_s [51].

Since the penetration through urban building walls is negligible at mmWave frequency due to the nature of mmWave signal propagation, the signal detours its way along the streets in urban canyons and changes its directions by diffractions on the buildings at intersections. Hence, we adopt the Manhattan distance based pathloss model instead of the traditional Euclidean distance based one [51]. The Manhattan distance from source to destination is composed of multiple segments, i.e., there exist one LOS segment and $M-1$ NLOS segments when the path from source to destination has $M - 1$ intersections. The pathloss is defined as follows [51]

$$
\begin{aligned}
\mathrm{PL}_{\mathrm{dB}} &= 10 \left(\alpha_L \log_{10} d_1 + \alpha_N \sum_{i=2}^{M} \log_{10} d_i \right) + (M-1)\Delta, \\
\mathrm{PL} &= d_1^{\alpha_L} \cdot \prod_{i=2}^{M} d_i^{\alpha_N} \cdot \delta^{(M-1)},
\end{aligned}
\tag{4.1}
$$

where $\alpha_L(\alpha_N)$ is the pathloss exponent of LOS(NLOS) link, d_i is length of i-th segment and $\Delta(\delta = 10^{\Delta/10})$ is corner loss in decibel(linear) scale. We assume corner losses at different corners are identical.

4.2.2 IAB Network Model

An IAB network is composed of IAB-node and IAB-donor, where the first is BS that supports wireless access to users and wirelessly backhauls the access traffic and the
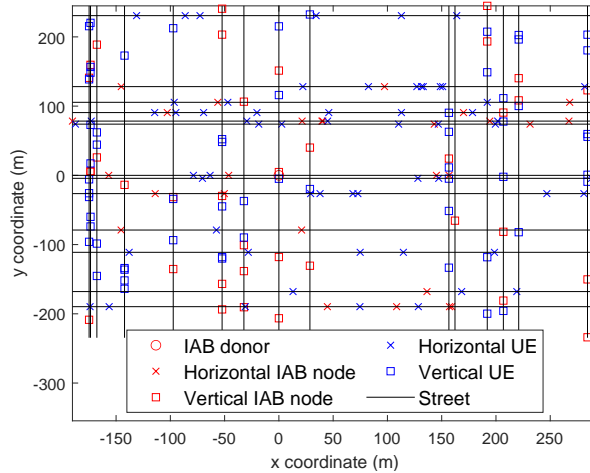
Figure 4.3. A realization of mmWave network in urban street canyon scenario with a size of $500 \times 500\,(\text{m}^2)$, $\lambda_s = 0.02/\text{m}$, $\lambda_n = 0.005/\text{m/street}$, and $\lambda_u = 0.01/\text{m/street}$.

latter is BS which provides user's interface to core network and wireless backhauling functionality to IAB-nodes. The IAB donor is connected to the core network with wired backhauling, while IAB nodes are connected to other nodes or donor with wireless backhauling as shown in Fig. 4.1. Unlike fully wired backhauling network, limited radio resources should be used for both access and backhaul, so it is necessary to consider the band operation and access-backhaul duplexing. There are two kinds of band operations in IAB networks: 1) in-band and 2) out-of-band backhaul [7]. In-band backhauling includes scenarios where access and backhaul link at least partially overlap in frequency creating half-duplexing or interference constraints, which imply that the IAB-node cannot transmit and receive simultaneously on both links. Otherwise, out-of-band scenarios are understood as not posing such constraints because access and backhaul use different frequency bands. Beamforming is an essential technique to overcome the limitations of mmWave signal propagation, where sharp beam width can divide a coverage area into multiple orthogonal spaces. Moreover, the access and backhaul beams for a traffic flow in each BS probably have opposite directions due to geo-location of BSs so we apply space-division-duplex (SDD) into access and backhaul. This means that access-backhaul full-duplexing can be achieved at each BS by
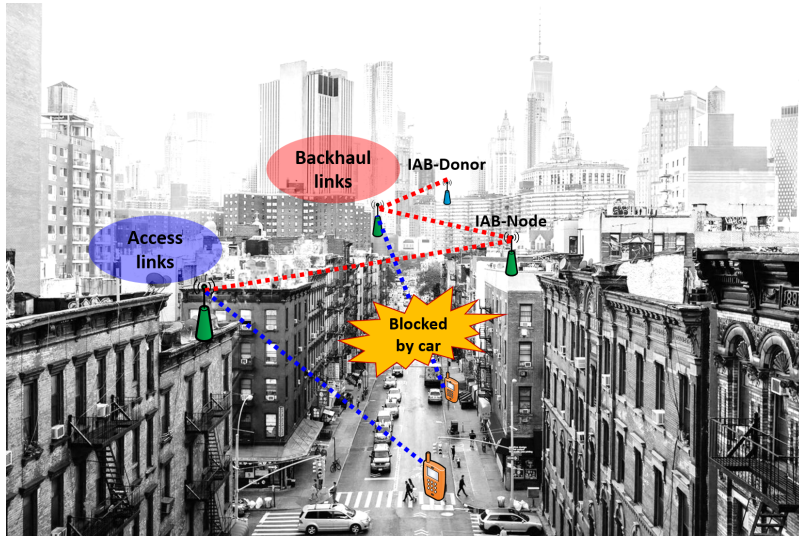
Figure 4.4. IAB communication in an urban area.

respectively allocating orthogonal space resources to access and backhaul links.

We consider a mmWave IAB network with IAB-nodes and users whose locations are modeled as one-dimensional PPP with intensities λ_n and λ_u (unit: /street). We set the coordinates of the median horizontal and vertical streets as $y = 0$ and $x = 0$, respectively. We assume that the IAB donor is located at the origin. Fig. 4.3 presents an example of a mmWave network in urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, $\lambda_n = 0.005$/m/street, and $\lambda_u = 0.01$/m/street. As shown in Fig. 4.4, the IAB-donor and IAB-nodes are installed at a high place such as the roof of a building to ensure coverage while being less affected by obstacles. With this IAB installment, we can classify the wireless links into two categories: 1) backhaul links, where the connections between different IAB-nodes or IAB-donor and IAB-node (red dashed lines in Fig. 4.4); and 2) access links, where the connections between users and IAB-donor/nodes (blue dashed lines in Fig. 4.4). Since the access link is more affected by dynamic blockers such as cars and pedestrians than the backhaul link, the backhaul link is more stable than the access one in terms of LOS-NLOS fluctuations, which can degrade the network performance by incurring outage events

(see Chapter 3). Therefore, we consider the static routing topology among the IAB-donor and nodes, otherwise, we consider the price based association between user and IAB donor/node, where the price implies link dynamics and load balancing. Since the 3GPP recommends a directed-acyclic-graph (DAG) topology not a mesh topology, we assume that an IAB routing forms a spanning-tree (ST) topology, which is a special form of DAG. To evaluate routing topology, we consider the following metrics: 1) the hop distance to the donor, i.e., height of ST topology; and 2) the number of child nodes. We adopt the sectorized antenna model and signal-to-interference-plus-noise (SINR) model from [14, 51].

4.3   Path Selection Strategies

In [52], two different distributed path selection strategies are investigated: 1) high-quality-first (HQF) selects the IAB-donor/node with the highest signal-to-noise-ratio (SNR) as a parent; and 2) wired-first (WF) selects a direct link to the IAB-donor with the best signal, even if an IAB-node with better SNR is available. The HQF approach facilitates a best quality wireless backhaul connection in the first hop, which may increase the hop distance to the IAB-donor. The WF approach may choose the donor as a parent even if the donor provide a backhaul link with poor quality, while minimizing the hop distance to the donor. Both strategies use additional contextual information related to the location of BSs to select a parent BS that is closer to the IAB donor. Therefore, when choosing a parent, each IAB-node divides the neighboring region into two half-planes identified by a line which passes through the IAB node and is perpendicular to the line passing through the IAB-node and donor. Then, the IAB-node considers for parent selection only from IAB-donor/nodes which are in the half-plane including IAB-donor, and selects that with the shortest distance or that with the minimum hops to reach the donor for HQF or WF, respectively.

The routing strategies described above, however, do not guarantee user QoE satis-

faction because the link rate is not considered when selecting a parent. As mentioned in Section 4.2.2, the backhaul links have stable quality because the IAB-donor and nodes are likely to be installed at a high area. A static routing topology is more suitable for the wireless backhaul network than a distributed path selection because the quality of wireless backhaul is likely to be stable. Open-shortest-path-first (OSPF) is a well-known routing protocol using link state routing algorithm and Dijkstra method [53]. We first design a link metric suitable for mmWave multi-hop networks to develop an OSPF-like IAB routing algorithm. To this end, we consider the network utility, which is defined as the ratio of total user throughput to total network resource usage. Let $\mathcal{N}$ and $\mathcal{K}$ be the sets of IAB-donor/nodes and users, respectively. The k-the user has minimum data rate requirement R_k. For a given routing topology, the set of IAB-donor/nodes on the route from donor to IAB-node i is denoted by $\mathcal{N}_i$. Then, we write the network utility for a given routing topology as

$$
U(\mathbf{A}, \mathcal{N}, \mathcal{K}) = \frac{\sum\limits_{k \in \mathcal{K}} R_k}{\sum\limits_{k \in \mathcal{K}} \sum\limits_{i \in \mathcal{N}} \sum\limits_{j \in \mathcal{N} \backslash \mathcal{N}_i} \frac{a_{ij} R_k}{r_{ij}}},
\tag{4.2}
$$

where $\mathbf{A} = (a_{ij})_{i,j}$ is the connected graph of the routing topology, i.e.,

$$
a_{ij} = \begin{cases} 1, & \text{if } i \text{ is the parent of } j, \\ 0, & \text{otherwise} \end{cases}
$$

and r_{ij} is the achievable rate of backhaul link (i, j).

Methodologies of maximizing total user throughput with limited network resources has been mainly studied in traditional best-effort networks, the numerator in (4.2) is correspondingly considered the objective function of the network-utility-maximization (NUM) problem. In emerging QoS guaranteed networks, however, it becomes more important to efficiently use network resources in order to provide throughput for users' QoS satisfaction in terms of NUM. Hence, minimizing the denominator in (4.2) can

be new objective of NUM for QoS guaranteed networks, i.e., the best routing topology maximizing the network utility can be obtained by solving

$$\mathbf{A}^\star = \arg\min_{\mathbf{A}} \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \mathcal{N}_i} \frac{a_{ij} R_k}{r_{ij}}. \tag{4.3}$$

The objective function of (4.3) can be rewritten as

$$\begin{aligned} f_0(\mathbf{A}) &= \sum_{k \in \mathcal{K}} R_k \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \mathcal{N}_i} \frac{a_{ij}}{r_{ij}}, \\ &= \sum_{k \in \mathcal{K}} R_k \cdot g(\mathcal{N}), \end{aligned} \tag{4.4}$$

where

$$g(\mathcal{N}) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \mathcal{N}_i} \frac{a_{ij}}{r_{ij}}$$

is the accumulated resource usage from donor to node i if user demand is 1. Then, $R_k \cdot g(\mathcal{N})$ can be interpreted as the total resource usage for backhauling the user k's downlink data. Since user demand R_k is irrespective of solving (4.3), minimizing $g(\mathcal{N})$ leads to finding the best routing topology $\mathbf{A}^\star$. Therefore, we set $t_{ij} = \frac{1}{r_{ij}}$, named as temporal efficiency, as the link metric of link (i, j). With temporal efficiency matrix $\mathbf{T} = (t_{ij})_{i,j}$, the Dijkstra method can find the optimal routing topology minimizing the total network resource usage.

We can calculate the load on IAB-node i as

$$\rho_i = \sum_{j \in \mathcal{N} \setminus \mathcal{N}_i} \sum_{k \in \mathcal{K}} \frac{a_{ij} R_k}{r_{ij}}. \tag{4.5}$$

Then, we can re-express the routing problem (4.4) with the notation $\boldsymbol{\rho} = (\rho_i)_i$ as

$$\begin{aligned} \mathbf{A}^\star &= \arg\min_{\mathbf{A}} \sum_{i \in \mathcal{N}} \rho_i \\ &= \arg\min_{\mathbf{A}} \|\boldsymbol{\rho}\|_1, \end{aligned} \tag{4.6}$$

where $\|\mathbf{x}\|_p$ is the p-norm of $\mathbf{x}$. Similar to Chapter 3, we can formulate the min-max fair problem as

$$\begin{aligned}
\mathbf{A}^\star &= \arg\min_{\mathbf{A}} \max_{i\in\mathcal{N}} \rho_i \\
&= \arg\min_{\mathbf{A}} \|\boldsymbol{\rho}\|_\infty.
\end{aligned} \tag{4.7}$$

We observe that minimizing 1-norm and infinity-norm lead to NUM and min-max fair, respectively. Finally, we formulate the p-general fairness problem as

$$\mathbf{A}^\star = \arg\min_{\mathbf{A}} \|\boldsymbol{\rho}\|_p. \tag{4.8}$$

4.4 Path-Aware Access Pricing Policy

Since the user demand cannot be decoupled from the routing problem (4.8) for $p > 1$, the $(p > 1)$-general fair routing has to solve the problem to find a new routing topology whenever a new user arrives in the network or the requirements of existing users change. In this work, we choose $p = 1$ to provide a semi-static backhaul routing topology. The access price based association for mmWave UDN given in Chapter 3 gives us an insight that network load can be balanced in a decentralized manner with a proper access pricing policy. Unlike mmWave UDN with fully wired backhauls, radio resources should be allocated for BS cooperative signaling, which increases network load, in mmWave IAB networks. Moreover, a user's access request to an IAB-node places burden on the corresponding parent IAB-donor/nodes due to path dependency, which makes analysis and optimization problem formulation difficult. So we need heuristic approaches to develop fair associations for the mmWave IAB network. In this section, we propose a path-aware access pricing policy. Let BS_i and U_k be the i-th IAB node and k-th user. Specifically, BS_1 and BS_i, $i = 2, 3, \ldots$, indicate the IAB-donor and nodes, respectively.

From the user's perspective, the route to donor appears to be a linear topology
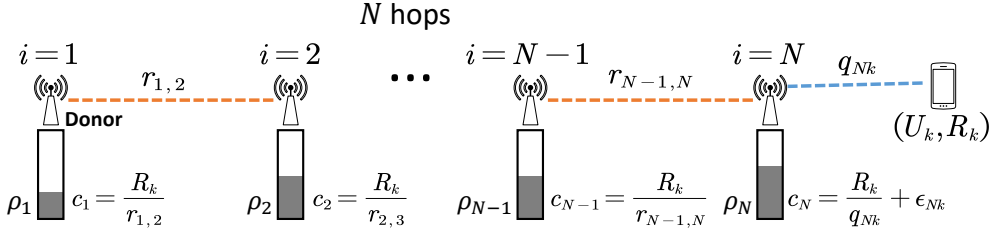
Figure 4.5. Path-aware access price model.

with N hops regardless of the routing strategy as shown in Fig. 4.5. Let c_i be the load caused by U_j on BS_i. Then, each BS load can be written as

$$c_i = \frac{R_k}{r_{i,i+1}}, \quad \forall i = 1, 2, \ldots, N - 1 \tag{4.9}$$

$$c_N = \frac{R_k}{q_{Nk}} + \epsilon_{Nk}, \tag{4.10}$$

where q_{jk} is the achievable rate of an access link (j, k) and ϵ_{jk} is the resource usage increment due to channel dynamics. The the resource usage increment due to dynamic blockers can be expressed as (see Chapter 3)

$$\epsilon_{jk} = \frac{2\lambda_B v_B(h_B - h_U)}{\mu\pi(h_{BS} - h_U)} \cdot \frac{R_k d_{jk}}{q_{jk}},$$

where λ_B is the intensity of dynamic blockers, v_B is speed of a moving blocker, (h_B, h_U, h_{BS}) is heights of blockers, users, BSs, and μ is blocking departure rate.

When the network is fully available for U_k, i.e., $1 - \rho_i \geq c_i \; \forall i$, the total network load caused by U_j can be derived by summing $c_i \; \forall i = 1, 2, \ldots, N$, i.e.,

$$C_{Nk} = \sum_{i=1}^{N} c_i = R_k \underbrace{\sum_{i=1}^{N-1} \frac{1}{r_{i,i+1}}}_{\text{Backhauling cost}} + \underbrace{\frac{R_k}{q_{Nk}} + \epsilon_{Nk}}_{\text{Access cost}}, \tag{4.11}$$

where the subscript NK means the access IAB-node identifier and user index. We note the user U_k with the data rate requirement of R_k accesses a BS with hop distance

of N to the donor is equivalent that N users with R_k access the network due to path dependency. The virtual $N - 1$ users incur backhauling cost and the actual user incurs access cost. It is intuitive that using C_{ik}, $i \in \mathcal{N}, k \in \mathcal{K}$, as access price is beneficial to both user QoE satisfaction and network management. This is because a path with the lowest C_{ik} is the solution of (4.4). Therefore, each user select a BS that provides the lowest C_{ik} among BSs in its coverage, i.e.,

$$i_k^\star = \underset{i \in \mathcal{N}_k}{\arg \min}\, C_{ik}, \tag{4.12}$$

where $\mathcal{N}_k$ is the set of BSs in U_k's coverage.

To embrace the network load distribution into access price, we modify the per-link cost c_i by adding penalty if the user demand exceeds the remaining resource amount, i.e.,

$$\tilde{c}_i = \begin{cases} c_i, & \text{if } \rho_i + c_i \leq 1 \\ c_i + \eta_i, & \text{otherwise} \end{cases} \tag{4.13}$$

where $\eta_i > 0$ is the penalty of using overloaded BS. We design the penalty η_i as the time delay caused by overloading, i.e.,

$$\eta_i = \lfloor \rho_i + c_i \rfloor. \tag{4.14}$$

From (4.4), we know that the total load induced by U_k is proportional to its demand R_k, which means that it is advantageous in terms of NUM to provide a path with small amount of accumulated resource usage to users with large demand. Then, the price for U_k to access BS_i is given by

$$\tilde{C}_{ik} = \sum_{j \in \mathcal{N}_i} \tilde{c}_i. \tag{4.15}$$

Finally, each user associates with a BS with the lowest $\tilde{C}_{ik}$ among BSs in its coverage,

Carrier freq.	α_L	α_N
28 GHz	2.0	2.92
N_0	W	P_{TX}
-114 dBm/MHz	400 MHz	30 dBm
Height of dynamic blocker	Height of DU	Height of UE
1.8 m	5 m	1.4 m

Table 4.1: Simulation parameters.

i.e.,

$$i_k^\star = \underset{i \in \mathcal{N}_k}{\arg\min} \, \tilde{C}_{ik} \tag{4.16}$$

If the network manager re-arranges user associations periodically in centralized manner, the centralized controller sorts users in the order of high demand and executes re-association according to (4.16).

4.5 Performance Evaluation

In this section, we present numerical results to evaluate various routing strategies and the proposed access pricing policy. We consider the carrier frequency of 28 GHz, where the pathloss exponent is $(\alpha_L, \alpha_N) = (2, 2.92)$ [20, 47]. We set the transmit power of BS to 30 dBm. The noise figure and system bandwidth are set to be $N_0 = -114$ dBm/MHz and $W = 400$ MHz. We assume that the data rate requirement of each user is uniformly distributed on $[0, 1000]$ Mbps, i.e., $R_k \sim \mathsf{uniform}(0, 1000)$. The simulation parameters are summarized in Table 4.1.

We first evaluate various routing scheme for a given mmWave IAB network as shown in Fig. 4.6, where streets and IAB-nodes are respectively distributed with $\lambda_s = 0.02$/m and $\lambda_n = 0.005$/m/street in urban canyon with a size of 500×500 (m^2). Here, the IAB-donor is located at the origin and assigned with the number one. IAB nodes are assigned sequentially starting with the number two. Fig. 4.7 presents Spanning-Tree (ST) routing topology generated by Dijkstra method with the proposed link metric,
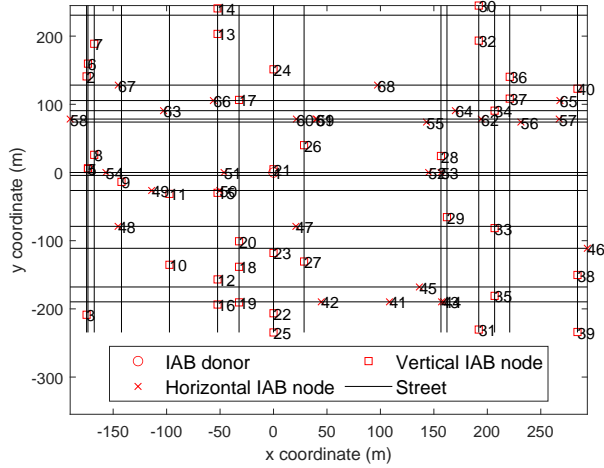
Figure 4.6. A realization of mmWave network in urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, and $\lambda_n = 0.005$/m/street.

(a) Proposed ($h = 5$)   (b) HQF ($h = 15$)   (c) WF ($h = 3$)
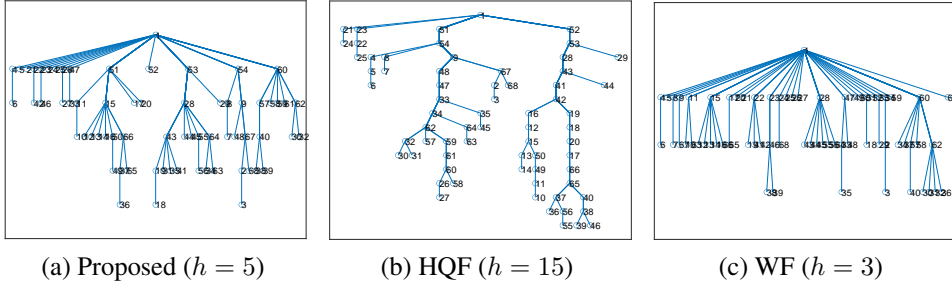
Figure 4.7. ST routing topology for the mmWave IAB network given in Fig. 4.6.

HQF, and WF. We observe that HQF makes the highest ST while WF makes the short-est ST. The proposed method creates a medium height ST. Specifically, STs generated by the proposed method, HQF, and WF have heights $h = 5$, $h = 15$, and $h = 3$, respectively, in this example. The number of child nodes tends to be opposite to the trend that height of ST shows.

Fig. 4.8 shows the statistical results of the IAB network in Fig. 4.6. In terms of the number of child nodes, it is likely to have similar performance for each routing scheme. However, it can be seen that WF has a higher probability of having many child nodes compared to other schemes. This is because the IAB-donor have many

Figure 4.8. Statistical results. (a) Number of child nodes. (b) Hop distance to IAB-donor. (c) Cumulative temporal efficiency.

Figure 4.9. Statistical results of various routing schemes in urban street canyon scenario with a size of 500×500 (m$^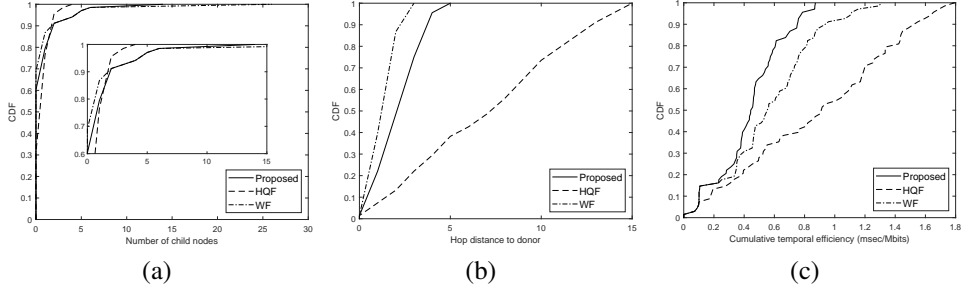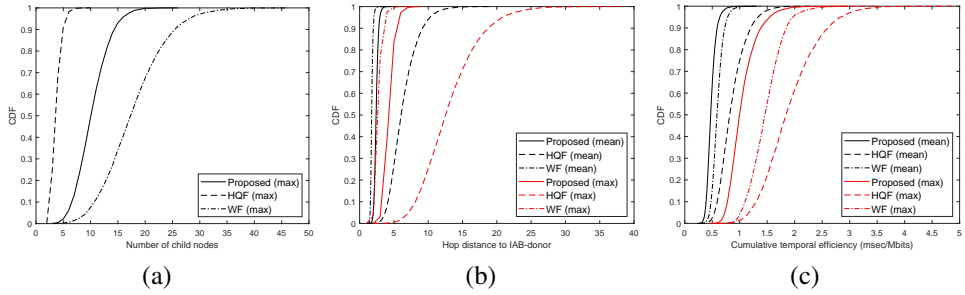2$), $\lambda_s = 0.02$/m, and $\lambda_n = 0.005$/m/street. (a) Number of child nodes. (b) Hop distance to IAB-donor. (c) Cumulative temporal efficiency.

child nodes since the IAB-nodes who can directly reach the donor connect to the IAB-donor with WF, as shown in Fig. 4.7c. Since parent nodes with a large number of child nodes can cause network disruption by creating bottlenecks, it is important to know the maximum number of child nodes in a routing topology. The result of hop distance to IAB-donor shows the same trend as the height of ST produced by the proposed method, HQF, WF. To evaluate the efficiency in terms of network utility, we define the cumulative temporal efficiency of BS_i as the summation of temporal efficiency in the path from BS_1 to BS_i. The proposed method shows the best performance than others as shown in Fig. 4.8c. HQF shows the worst performance because the hop distance is increased by selecting the parent node that provides the best channel for each node.

To evaluate various routing schemes more statistically, we generate 10,000 mmWave
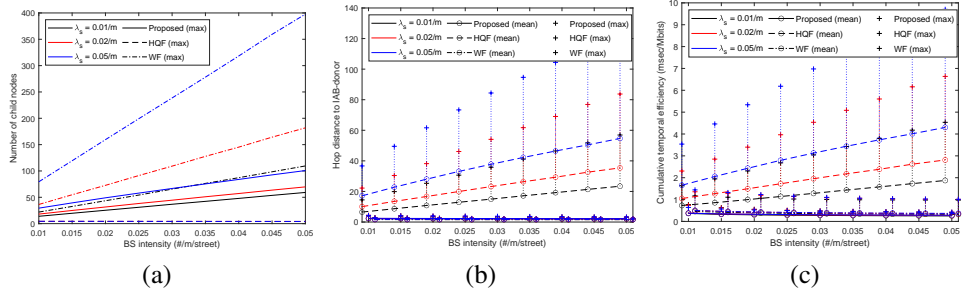
74

Figure 4.10. Performance evaluation for various street and BS intensities in urban street canyon scenario with a size of 500×500 (m^2). (a) Number of child nodes. (b) Hop distance to IAB-donor. (c) Cumulative temporal efficiency.

IAB networks and measure mean/maximum value of the number of child nodes, hop distance to IAB-donor, and cumulative temporal efficiency. We consider the urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, and $\lambda_n = 0.005$/m/street. Fig. 4.9a presents the cdf of the maxmimum number of child nodes. WF shows the largest number since the donor has too many child nodes with WF, while HQF shows the smallest number because of its criteria. In terms of hop distance to IAB-donor, the WF and the proposed method have relatively small values for both the mean and the maximum values compared to the HQF as shown in Fig. 4.9b. Fig. 4.9c shows that the proposed method has the best cumulative temporal efficiency than others.

We evaluate the routing scheme for various urban street canyon environments, i.e., $\lambda_s = \{0.01, 0.02, 0.05\}$ and $\lambda_n = \{0.01 : 0.005 : 0.05\}$, where $a : b : c$ is the sequence from a to c with interval b. It can be seen that as street and BS intensities increase, the maximum number of child nodes of WF and the proposed method increase, while that of HQF rarely increases as shown in Fig. 4.10a. Specifically, the increasing slope of the proposed method is smaller than that of WF, which implies that the proposed method creates a routing topology that causes less network congestion due to bottleneck compared to WF in dense networks. From Figs. 4.10b and 4.10c, we observe that, as street and BS intensities increase, the hop distance and cumula-
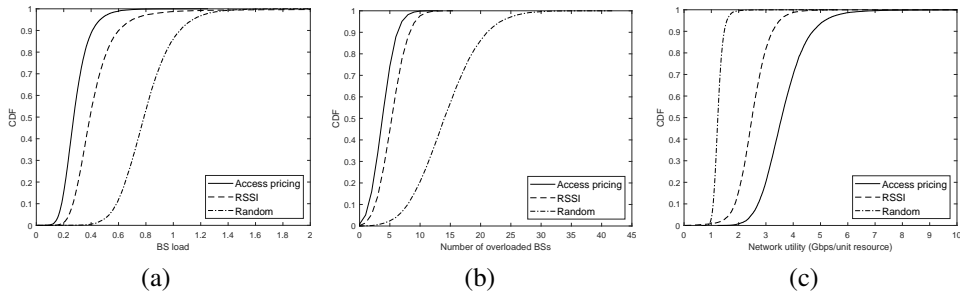
Figure 4.11. Statistical results of various association schemes in urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, $\lambda_n = 0.005$/m/street, and $\lambda_u = 0.01$/m/street. (a) BS load. (b) Number of overloaded BSs. (c) Network utility.

tive temporal efficiency of WF and the proposed method rarely change, while those of HQF increases, where the slope of HQF increases as street intensity increases. The gap between the maximum and mean value for the proposed method and WF is much smaller than that for HQF. From these results, we conclude that the proposed method provides better routing topology than HQF and WF.

Now, we evaluate the path-aware access pricing compared to following association schemes: 1) RSSI-based association, where a user connects to the BS providing the smallest pathloss; and 2) random association, where a user randomly connects to a BS among BSs in its coverage. The user-oriented coverage is defined by the set of BSs whose achievable rate to the user is equal to or larger than user's data rate requirement, i.e.,

$$\mathcal{C}_k = \{ i \mid i \in \mathcal{N},\ q_{ik} \geq R_k \} \quad \forall k \in \mathcal{K}.$$

We measure the number of overloaded BSs, BS_i is overloaded if $\rho_i > 1$, to investigate how much the association scheme causes network performance degradation. Fig. 4.11 shows the statistical results of various association schemes in urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, $\lambda_n = 0.005$/m/street, and $\lambda_u = 0.01$/m/street. The proposed path-aware access pricing shows the best performance in all aspects.

We evaluate the association scheme for various user intensity, i.e., $\lambda_u = \{0.01 :$
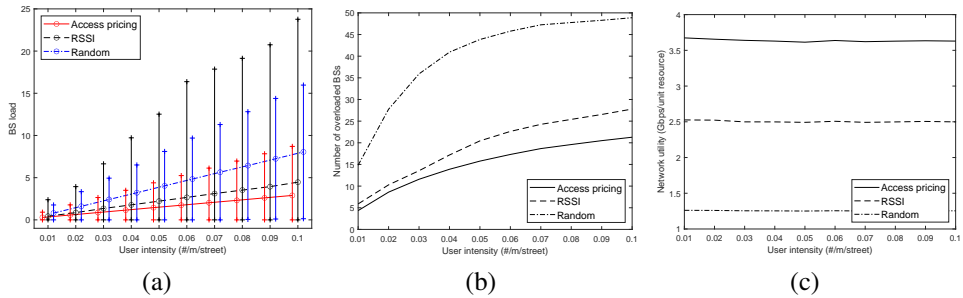
Figure 4.12. Performance evaluation for various user intensity in urban street canyon scenario with a size of 500×500 (m^2), $\lambda_s = 0.02$/m, and $\lambda_n = 0.005$/m/street. (a) BS load (+: ±1-sigma). (b) Number of overloaded BSs. (c) Network utility.
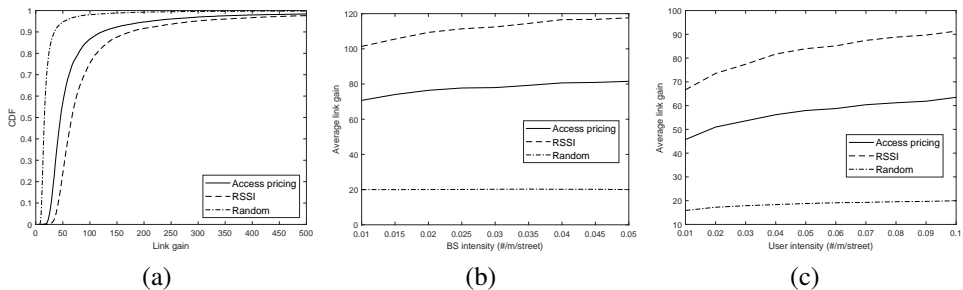
Figure 4.13. User experienced link gain performance in urban street canyon scenario with a size of 500×500 (m^2) and $\lambda_s = 0.02$/m. (a) Distribution of link gain. (b) Average link gain versus BS intensity. (c) Average link gain versus user intensity.

$0.01 : 0.1$}/m/street. Fig. 4.12a presents the BS load in terms of mean (symbol o) and variance (symbol +) as a function of user intensity. We observe that the proposed access pricing policy incurs the smallest burden on the network with small variation, which means that the proposed policy not only uses the minimal network resources, but also alleviates network congestion caused by overloading. This can be seen in Fig. 4.12b, for all user intensities, the number of overloaded BSs on the network is the lowest when the proposed policy is adopted. In terms of network utility, the proposed policy outperforms than others as shown in Fig. 4.12c. We conclude from these observations that our proposed access pricing policy provides a network-wide efficient user association.

To compare the link quality of various association schemes, we define the user experienced link gain as the ratio of associated link capacity to the user requirement. Then, we express the link gain of U_k as

$$G_k = \frac{q_{i_k^* k}}{R_k},$$

(4.17)

where i_k^* is the associated BS for U_k. Note that a link gain whose value is equal to or larger than one indicates the associated BS has sufficient capacity to serve the user. Fig. 4.13 shows the user experienced link gain results with respect to various association schemes. In these results, we omit the link gains of rejected associations (i.e., $G_k < 1$), that is, we only use the link gains of admitted associations to make the results. Fig. 4.13a illustrates the distribution of link gain for various association schemes. The RSSI based association outperforms than others because each user associates to the BS with the smallest pathloss among BSs in its coverage with the RSSI based scheme as shown in Fig. 4.13. This means that the link gain with RSSI based scheme provides upper bound. The proposed scheme achieves less link gain than RSSI based scheme but its performance is comparable to the RSSI-based scheme's. Since the proposed scheme has better performance than others in terms of load balancing as shown in Fig. 4.12, we claim that our proposed scheme is more suitable for mmWave IAB networks.

4.6 Summary

In this chapter, we first investigated the geographic and pathloss models suitable for describing urban street canyon environments. Then, we proposed a mmWave IAB network model and found the dominant signal path for a given transmitter and receiver in an urban street canyon. In order to develop a path selection strategy for mmWave IAB networks, we formulated a routing problem by revisiting the idea of NUM. In this process, we newly defined the link metric for Dijkstra method and the utility maximization

problem for QoS guaranteed networks. We also proposed a path-aware access pricing policy for user association based on path dependency of multi-hop networking. The access price for a user is sum of backhauling cost and access cost, where the former implies path-dependency and the latter include both spectral efficiency with access BS and channel dynamics. Finally, numerical evaluations presented that our method outperforms than other schemes.

Chapter 5

Conclusion

5.1 Summary

In this dissertation, we study a mathematical framework and network manament schemes for user-centric mmWave communication systems.

First, we conduct stochastic geometry based analysis to investigate the impact of user mobility on network performance in user-centric mmWave communication systems. Specifically, we propose a system model for user-centric mmWave communication systems with multi-connectivity. We also develop a state machine according to user mobility and define some events by classifying transitions in the state machine. Then, we derive compact mathematical expressions for transitions and event probabilities for various rural and urban scenarios. We pay special attention to blockages because they affect communication performance significantly in mmWave systems. Blockages decrease the BS density in a user's communication range since the LOS probability between the user and its serving BS is reduced. They increase the handover probability as the link state of the user changes rapidly with mobility. Especially in the latter case, blockages make a temporal correlation between successive observations of the link state between the BS and mobile user, which makes mobility analysis difficult. To address the challenges in mobility analysis with blockages, we derive the

effective BS density and the link status correlation. From these, we obtain tractable bounds on the handover probability. We investigate mobility-aware performance in terms of network overhead and downlink throughput, improving our understanding of network operations to support user mobility. Extensive simulations verify the accuracy and usefulness of our analysis.

Next, we propose user-oriented configuration rules and association management schemes for mmWave UDNs. We first propose user-side configuration rules for achieving user requirement and investigate impacts on traffic flow control. Then we consider a fair association based on user request through self-configuration to balance BS load across the whole network. To this end, we formulate an optimization problem, which is a mixed-integer linear programming so an NP-hard problem. We develop a fair association algorithm by applying dual approach to the primal optimization problem. The algorithm includes an access price based per-user request decision method and a price adjustment rule for load balancing. We also develop the analytic framework to investigate the outage probability. We evaluate the proposed method in terms of average resource budget, network utility, peak load, fairness, and outage. In simulations, SLM shows better performance in terms of average resource budget and network utility, while PLM shows better performance in terms of fairness and peak load. We observe that our proposed methods perform very comparably to the centralized scheme and work well in dynamic environments. We find that the fair load distribution (i.e., PLM) makes the network more resistant to outage.

Finally, We extend the work in Chapter 3 to mmWave IAB networks in Chapter 4. We investigate the geographic and pathloss models suitable for describing urban street canyon environments. Then, we propose a mmWave IAB network model and find the dominant signal path for a given transmitter and receiver in an urban street canyon. In order to develop a path selection strategy for mmWave IAB networks, we formulate a routing problem by revisiting the idea of NUM. Specifically, we newly define the link metric for Dijkstra method and the utility maximization problem for QoS guaranteed

networks. We also propose a path-aware access pricing policy for user association based on path dependency of multi-hop networking. The access price for a user is sum of backhauling cost and access cost, where the former implies path-dependency and the latter include both spectral efficiency with access BS and channel dynamics. We evaluate our proposed scheme in various urban street canyon scenarios with comparing the existing schemes.

5.2   Limitations and Future Work

In Chapter 2, the system model is established on the assumption that various factors (e.g., BS deployment, user mobility pattern, etc.) are well known random variables for tractable analysis. Moreover, we do not consider dynamic blockages that can seriously affect mmWave communications since the coupling of user and blocker mobility makes the analysis intractable. However, our sophisticated analysis can be used to statistically analyze the performance of practical mmWave networks. When developing an association control scheme in Chapter 3, we do not consider user mobility, i.e., we assume that user is stationary in order to reflect various blockage scenarios in the system model. As mentioned above, considering both various blockage scenarios and user mobility makes not only analysis intractable but also algorithm design through solving optimization problems difficult. This is because the combination of various factors creates a random process that is mathematically intractable, which may be handled by machine learning approach, specifically reinforcement learning (RL). For further work, we can model the network overhead incurred user mobility based on Chapter 2. Incorporating the network overhead into problem formulation, we can design a RL framework for mobility-aware proactive association.

To obtain insights of designing a link metric for backhaul path selection, we simplify the mmWave backhaul network and consider single donor scenario in Chapter 4. The simplified mmWave backhaul network is likely to have similar characteristics of

wired network. Based on the simple mmWave IAB model, we develop a heuristic access pricing scheme without formulating an optimization problem. In the future work, we consider additional mmWave specific features that make backhaul networks different from wired networks. For example, we take access/backhaul resource partitioning and/or duplexing into account. Considering the multiple donors scenario, furthermore, we can investigate the followings: 1) impact of donor deployment density on network performance and 2) backhaul network clustering. Finally, a sophisticated association scheme can be developed by formulating an optimization problem including exquisite mmWave IAB model.

Bibliography

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez Jr, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, no. 1, pp. 335–349, 2013.

[2] A. K. Gupta, J. G. Andrews, and R. W. Heath, "Macrodiversity in cellular networks with random blockages," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 996–1010, 2018.

[3] 3GPP TS 37.340, "E-UTRA and NR; Multi-connectivity; Stage-2," ver. 15.6.0, Sep. 2019.

[4] V. Petrov, D. Solomitckii, A. Samuylov, M. A. Lema, M. Gapeyenko, D. Moltchanov, S. Andreev, V. Naumov, K. Samouylov, M. Dohler *et al.*, "Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments," *IEEE J. Sel. Areas in Commun.*, vol. 35, no. 9, pp. 2038–2055, 2017.

[5] W. Bao and B. Liang, "Stochastic geometric analysis of handoffs in user-centric cooperative wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2016, pp. 1–9.

[6] H. Zhang and W. Huang, "Tractable mobility model for multi-Connectivity in 5G user-centric ultra-dense networks," *IEEE Access*, vol. 6, pp. 43 100–43 112, 2018.

[7] 3GPP TR 38.874, "NR; Study on Integrated Access and Backhaul," ver. 16.0.0, Dec. 2018.

[8] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78–85, 2016.

[9] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[10] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," *arXiv preprint arXiv:1604.03183*, 2016.

[11] J. Choi, "On the macro diversity with multiple BSs to mitigate blockage in millimeter-wave communications," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1653–1656, 2014.

[12] M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, M. R. Akdeniz, E. Aryafar, N. Himayat, S. Andreev, and Y. Koucheryavy, "On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10 124–10 138, 2017.

[13] K. Han, Y. Cui, Y. Wu, and K. Huang, "The connectivity of millimeter wave networks in urban environments modeled using random lattices," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3357–3372, 2018.

[14] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, 2015.

[15] I. Atzeni, J. Arnau, and M. Kountouris, "Downlink cellular network analysis with LOS/NLOS propagation and elevated base stations," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 142–156, 2017.

[16] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, 2014.

[17] I. K. Jain, R. Kumar, and S. Panwar, "Can millimeter wave cellular systems provide high reliability and low latency? an analysis of the impact of mobile blockers," *arXiv preprint arXiv:1807.04388*, 2018.

[18] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE J. Sel. Areas in Commun.*, vol. 33, no. 10, pp. 2212–2225, 2015.

[19] H. Ibrahim, H. ElSawy, U. T. Nguyen, and M.-S. Alouini, "Mobility-aware modeling and analysis of dense cellular networks with C-plane/U-plane split architecture," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4879–4894, 2016.

[20] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[21] R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes*. John Wiley & Sons, 2005.

[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Courier Corporation, 1965, vol. 55.

[23] S. Kang, S. Choi, G. Lee, and S. Bahk, "A dual-connection based handover scheme for ultra-dense millimeter-wave cellular networks," in *Proc. IEEE Globecom*, Dec. 2019.

[24] G. Lee, S. Choi, J. Kim, Y. Kim, and S. Bahk, "Prosch: Proxy aided secondary cell handover in ultra-dense mmwave network," in *Proc. IEEE WCNC*, May 2020, pp. 1–6.

[25] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 2016.

[26] S. Choi, J.-G. Choi, and S. Bahk, "Mobility-aware analysis of millimeter wave communication systems with blockages," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 5901–5912, 2020.

[27] J. Choi, S. Y. Jung, S.-L. Kim, D. M. Kim, and P. Popovski, "User-centric resource allocation with two-dimensional reverse pricing in mobile communication services," *J. Commun. Netw.*, vol. 21, no. 2, pp. 148–157, 2019.

[28] D. Moltchanov, A. Samuylov, V. Petrov, M. Gapeyenko, N. Himayat, S. Andreev, and Y. Koucheryavy, "Improving session continuity with bandwidth reservation in mmWave communications," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 105–108, 2018.

[29] Y. Bejerano, S.-J. Han, and L. Li, "Fairness and load balancing in wireless lans using association control," in *Proc. ACM MobiCom*, 2004, pp. 315–329.

[30] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 836–850, 2014.

[31] W. Wong, K. W. Chau, and S.-H. G. Chan, "Joint client association and random access control for mu-mimo wlans," *IEEE Trans. on Mobile Comput.*, 2019.

[32] S. Lin, N. Che, F. Yu, and S. Jiang, "Fairness and load balancing in sdwn using handoff-delay-based association control and load monitoring," *IEEE Access*, vol. 7, pp. 136 934–136 950, 2019.

[33] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[34] 3GPP TS 38.401, "NG-RAN; Architecture description," ver. 16.1.0, Mar. 2020.

[35] 3GPP TR 38.806, "Study of separation of NR Control Plane (CP) and User Plane (UP) for split option 2," ver. 15.0.0, Dec. 2017.

[36] I. K. Jain, R. Kumar, and S. Panwar, "The impact of mobile blockers on millimeter wave cellular systems," *IEEE J. Sel. Areas in Commun.*, vol. 37, no. 4, pp. 854–868, 2019.

[37] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, 2017.

[38] 3GPP TS 36.300, "E-UTRA and E-UTRAN; Overall description; Stage 2," ver. 16.1.0, Mar. 2020.

[39] 3GPP TS 36.101, "E-UTRA; UE radio transmission and reception," ver. 16.4.0, Dec. 2019.

[40] N. Bourgeois, R. Catellier, T. Denat, and V. T. Paschos, "Average-case complexity of a branch-and-bound algorithm for maximum independent set, under the $\mathcal{G}(n,p)$ random model," *arXiv preprint arXiv:1505.04969*, 2015.

[41] Zhang, Weixiong, "Branch-and-bound search algorithms and their computational complexity,," USC/Information Sciences Institute, Tech. Rep., May 1996.

[42] T. Denat, A. Harutyunyan, and V. T. Paschos, "Average-case complexity of a branch-and-bound algorithm for min dominating set," *arXiv preprint arXiv:1902.01874*, 2019.

[43] S. Boyd and A. Mutapcic, "Subgradient methods," *lecture notes of EE364b, Stanford University, Winter*, 2006–2007.

[44] 3GPP TS 38.300, "NR; NR and NG-RAN Overall Description; Stage 2," ver. 16.2.0, Jul. 2020.

[45] 3GPP TS 38.331, "NR; Radio Resource Control (RRC) protocol specification," ver. 16.1.0, Jul. 2020.

[46] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1," ver. 17.3.0, Jul. 2020.

[47] S.-C. Lin and I. F. Akyildiz, "Dynamic base station formation for solving NLOS problem in 5G millimeter-wave communication," in *Proc. IEEE INFOCOM*, May 2017, pp. 1–9.

[48] OpenStreetMap. [Online]. Available: http://www.openstreetmap.org/.

[49] M. Haklay and P. Weber, "Openstreetmap: user-generated street maps," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, 2008.

[50] D. QGIS, "Quantum GIS geographic information system," *Open Source Geospatial Foundation Project*, vol. 45, 2011.

[51] Y. Wang, K. Venugopal, A. F. Molisch, and R. W. Heath, "Mmwave vehicle-to-infrastructure communication: Analysis of urban microcellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7086–7100, 2018.

[52] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmwaves," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.

[53] J. Moy, "RFC2328: OSPF Version 2," Apr. 1998.

초 록

밀리미터파 통신은 고속 전송을 가능하게 하지만 송신기와 수신기 사이의 장애물로 인해 네트워크 성능이 크게 저하될 수 있다. 장애물 효과를 극복하고 링크 안정성을 향상시키는 다중 연결 및 네트워크 초고밀화 두가지 접근법이 있다. 특히 각 사용자가 고품질의 서비스를 경험할 수 있도록 서빙 기지국 그룹의 동적 구성이 필요하므로 초고밀도 네트워크 환경에서 다중 연결은 사용자 중심 통신을 용이하게 한다. 본 논문은 사용자 중심의 밀리미터파 통신 시스템을 위한 수학적 프레임워크와 네트워크 관리 체계를 연구한다. 먼저 다중 연결을 사용하여 사용자 중심의 밀리미터파 통신 시스템에서 사용자 이동성과 이동성 인식 성능 지표를 모델링하고 확률기하분석을 기반으로 하는 새로운 분석 프레임워크를 제안한다. 이를 위해 각 사용자가 경험하는 다양한 이벤트의 상태 전이 확률에 대한 수학적 표현을 도출한다. 그런 다음 네트워크 오버헤드 및 다운 링크 수율 측면에서 이동성 인식 성능을 연구한다. 이를 통해 네트워크 운영에 대한 깊이있는 이해와 네트워크 밀도 및 다중 연결 기능이 핸드 오버와 관련된 이벤트의 확률에 미치는 영향을 이해할 수 있다. 시뮬레이션 결과는 분석의 정확성을 검증하고 이동성 인식 성능과 사용자 속도 간의 상관 관계를 보여준다. 다음으로 완전 또는 부분 유선 백홀이 있는 사용자 중심 밀리미터파 네트워크를 위한 사용자 중심 구성 규칙 및 접속 가격 기반 연결 알고리즘을 제안한다. 밀리미터파 초고밀도 네트워크에 대한 최적화 문제를 해결하여 공정한 연결 알고리즘을 개발한다. 이 알고리즘에는 접속 가격 기반 사용자 별 요청 결정 방법과 로드 밸런싱을 위한 가격 조정 규칙이 포함된다. 위 알고리즘 개발을 통해 얻은 통찰력을 기반으로 밀리미터파 통합 액세스 및 백홀 네트워크를 위한 경로 인식 접속 요금 정책을 개발한다. 수치 평가에 따르면 제안된 방법이 다른 비교 기법보다 우수하다. 분석 및 최적화 결과는 사용자 중심의 밀리미터파 통신 시스템 설계에 대한 유용한 통찰력을 제공할 것 이다.

주요어: 사용자 중심 통신, 밀리미터파, 이동통신시스템, 확률기하분석, 최적화, 장애물 효과, 다중 연결, 초고밀도 네트워크, 이동성

학번: 2016-30220