

TOWARDS ROBUST AND RELIABLE COMMUNICATION FOR MILLIMETER WAVE
NETWORKS

By

Masoud Zarifneshat

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science – Doctor of Philosophy

2022

ABSTRACT

TOWARDS ROBUST AND RELIABLE COMMUNICATION FOR MILLIMETER WAVE NETWORKS

By

Masoud Zarifneshat

The future generations of wireless networks benefit significantly from millimeter wave technology (mmW) with frequencies ranging from about 30 GHz to 300 GHz. Specifically, the fifth generation of wireless networks has already implemented the mmW technology and the capacity requirements defined in 6G will also benefit from the mmW spectrum. Despite the attractions of the mmW technology, the mmW spectrum has some inherent propagation properties that introduce challenges.

The first is that free space pathloss in mmW is more severe than that in the sub 6 GHz band. To make the mmW signal travel farther, communication systems need to use phased array antennas to concentrate the signal power to a limited direction in space at each given time. Directional communication can incur high overhead on the system because it needs to probe the space for finding signal paths. To have efficient communication in the mmW spectrum, the transmitter and the receiver should align their beams on strong signal paths which is a high overhead task.

The second is a low diffraction of the mmW spectrum. The low diffraction causes almost any object including the human body to easily block the mmW signal degrading the mmW link quality. Avoiding and recovering from the blockage in the mmW communications, especially in dynamic environments, is particularly challenging because of the fast changes of the mmW channel.

Due to the unique characteristics of the mmW propagation, the traditional user association methods perform poorly in the mmW spectrum. Therefore, we propose user association methods that consider the inherent propagation characteristics of the mmW signal. We first propose a method that collects the history of blockage incidents throughout the network and exploits the historical blockage incidents to associate user equipment to the base station with lower blockage possibility. The simulation results show that our proposed algorithm performs better in terms of improving the quality of the links and blockage rate in the network.

User association based only on one objective may deteriorate other objectives. Therefore, we formulate a biobjective optimization problem to consider two objectives of load balance and blockage possibility in the network. We conduct Lagrangian dual analysis to decrease time complexity. The results show that our solution to the biobjective optimization problem has a better outcome compared to optimizing each objective alone.

After we investigate the user association problem, we further look into the problem of maintaining a robust link between a transmitter and a receiver. The directional propagation of the mmW signal creates the opportunity to exploit multipath for a robust link. The main reasons for the link quality degradation are blockage and link movement. We devise a learning-based prediction framework to classify link blockage and link movement efficiently and quickly using diffraction values for taking appropriate mitigating actions. The simulations show that the prediction framework can predict blockage with close to 90% accuracy. The prediction framework will eliminate the need for time-consuming methods to discriminate between link movement and link blockage.

After detecting the reason for the link degradation, the system needs to do the beam alignment on the updated mmW signal paths. The beam alignment on the signal paths is a high overhead task. We propose using signaling in another frequency band to discover the paths surrounding a receiver working in the mmW spectrum. In this way, the receiver does not have to do an expensive beam scan in the mmW band. Our experiments with off-the-shelf devices show that we can use a non-mmW frequency band's paths to align the beams in mmW frequency.

In this dissertation, we provide solutions to the fundamental problems in mmW communication. We propose a user association method that is designed for mmW networks considering challenges of mmW signal. A closed-form solution for a biobjective optimization problem to optimize both blockage and load balance of the network is also provided. Moreover, we show that we can efficiently use the out-of-band signal to exploit multipath created in mmW communication. The future research direction includes investigating the methods proposed in this dissertation to solve some of the classic problems in the wireless networks that exist in the mmW spectrum.

Copyright by
MASOUD ZARIFNESHAT
2022

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to Dr. Li Xiao my advisor for her academic support throughout my Ph.D. career. I was lucky to benefit from her experience and knowledge in my research during the time I spent as a Ph.D. student. Despite the difficulties and challenges on the way, Dr. Xiao's advice was always helpful for me to see the problems from a new angle. I learned a lot from my advisor and thank her for the wisdom she shared with me. I would also extend my gratitude to my graduate committee Dr. Eric Torng, Dr. Matt Mutka, and Dr. Chichia Chiu.

I would like to thank my family for their unconditional support through the years I have been studying here far away from them. I owe my mother and father for their unconditional love and support. My mother will always be missed but she has an especial place in my heart and I know I still receive her blessings. I especially thank my wonderful wife, Sepide for her support throughout my career.

I would like to appreciate the people in the ELANS lab for their support and input. I would like to thank Dr. Chin-Jung Liu for his help during my Ph.D. with different research papers and also his help with the professional aspect of my career. I thank Dr. Chen Qiu for his help with advising about careers that I can pursue. I would like to thank Dr. Alireza Ameli and Dr. Mohammad Moazzami for their support. I appreciate my co-authors Dr. Proteek Roy and Dr. Jiliang Tang for their valuable and direct input in my research. I also thank Dr. Yunhao Liu and Dr. Qiben Yan and their research groups for holding group seminars that contributed to my better understanding of the art.

I extend my appreciation to all the faculty members and instructors in the Department of Computer Science and Engineering for enriching my knowledge in the general area of computer science. I would like to thank Mr. Mark McCullen for his support in my professional career.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ALGORITHMS	xiv
CHAPTER 1 INTRODUCTION	1
1.1 Blockage Mitigation in mmW Networks	4
1.2 Biobjective Optimization for Load and Blockage	5
1.3 Blockage Prediction Using Learning Approaches	6
1.4 Out-of-Band Beam Adaptation	7
1.5 Summary	8
CHAPTER 2 BLOCKAGE MITIGATION IN MILLIMETER WAVE NETWORKS	9
2.1 Related Work	10
2.2 System Model and C-RAN Primer	11
2.2.1 System Model	11
2.2.2 C-RAN Primer	15
2.3 User Association for Blockage Mitigation	17
2.3.1 Blockage Detector Component (BD)	17
2.3.2 Blockage Point Collector Component (BPC)	19
2.3.3 Device Association Component (DA)	20
2.3.4 Time Complexity and Overhead Analysis	22
2.4 Simulations	25
2.4.1 Simulation Parameters and Metrics	25
2.4.2 Simulation Results and Plots	27
2.5 Summary	32
CHAPTER 3 BIOBJECTIVE APPROACH FOR OPTIMIZING BLOCKAGE AND LOAD BALANCING OF THE NETWORK	33
3.1 Related Work	34
3.2 System Model and Objective Functions	36
3.2.1 System Model	37
3.2.2 Objective Functions	37
3.2.2.1 Load Balance	38
3.2.2.2 Blockage Score	40
3.2.3 Scalarization Methods	41
3.3 Biobjective Optimization Problem	44
3.3.1 Formulation and Solution of ASF Scalarization Method	45
3.3.1.1 ASF Generated Single Objective Problem	45
3.3.1.2 Solution to ASF Generated Single Objective Problem via Dual Analysis	46

3.3.2	Formulation and Solution of NC Scalarization Method	51
3.3.2.1	Solution to NC Generated Single Objective Problem via Dual Analysis	52
3.3.3	Formulation and Solution of WS Scalarization Method	55
3.3.4	Semi-Distributed Algorithm for Dual Analysis	56
3.4	Proof of Duality Gaps	57
3.4.1	Duality Gap of ASF	59
3.4.2	Duality Gap of NC	62
3.4.3	Duality Gap of WS	65
3.5	Evaluations	66
3.5.1	Parameters and Metrics	66
3.5.2	Results and Discussion	68
3.6	Summary	75
CHAPTER 4 LEARNING-BASED LINK BLOCKAGE PREDICTION		77
4.1	Related Work	79
4.2	System Primer and Design	82
4.2.1	Diffraction in Millimeter-Wave Frequency	82
4.2.2	Measuring Diffraction Value on a Multipath Component	84
4.2.3	Discovering the Multipath Components Around Receiver	89
4.2.4	Learning and Classification Methods	91
4.2.4.1	Learning Modalities	91
4.2.4.2	Training Methods	92
4.3	Feature Vector Generation	95
4.3.1	Feature Generation Based on Diffraction	95
4.3.2	Burst Feature	97
4.3.3	Time Complexity Analysis	101
4.4	Simulation Results	101
4.4.1	Simulation Framework	102
4.4.2	Results	103
4.4.2.1	Results for different settings and parameters	110
4.5	Detecting the Link Blockage Using Signal Patterns	113
4.5.1	Channel State Information Reconstruction	114
4.5.2	Link Blockage Detection Based on Signal Patterns	115
4.5.3	Data Collection	117
4.5.4	Experimental Results	118
4.6	Summary	121
CHAPTER 5 BEAM ADAPTATION USING OUT-OF-BAND SIGNAL FOR ROBUST MILLIMETER WAVE COMMUNICATION		123
5.1	Related Work	125
5.2	A Primer on the Direction of Arrival and the mmW Channel	127
5.2.1	The Direction of Arrival in the Sensing Band	127
5.2.2	A Primer on mmW Channel and Beam Pattern	131
5.2.3	Beam Alignment in IEEE802.11ad	133

5.2.4	Beam Adjustment Using Out-of-Band Measurements	133
5.3	Out-of-Band Path Discovery	136
5.4	Experimental Evaluations	138
5.4.1	Experimental Hardware Platform	138
5.4.2	Evaluation Setup	139
5.4.3	Methodology of the Communication Performance Assessment	139
5.4.4	Experiments Results	140
5.4.4.1	Signal Spatial Spectrum	140
5.4.4.2	Communication Performance Assessment	144
5.5	Summary	153
CHAPTER 6 CONCLUSION AND FUTURE WORK		154
6.1	Conclusion	154
6.1.1	Blockage Score and Load Balancing Joint Optimization	154
6.1.2	Link-Aware mmW Communication Improvement	155
6.1.3	Contribution	157
6.2	Future Research Directions	157
6.2.1	Path Discovery for User Association	158
6.2.2	Out-of-Band Method for Classic Problems in mmW	158
BIBLIOGRAPHY		161

LIST OF TABLES

Table 3.1	Effect of blocking object count on blockage and blockage score	75
Table 4.1	Effect of relative permittivity value on prediction accuracy (%)	112
Table 4.2	Prediction accuracy (%) vs. different mobility models	113
Table 4.3	Time delays for training and prediction (ms)	121
Table 5.1	Statistics of the Orientation Experiment	152

LIST OF FIGURES

Figure 2.1	The concept of communication disk and human blockage in mmW network . . .	12
Figure 2.2	An example network with UEs, BSs and blocking objects	13
Figure 2.3	Different angle values for different parts of antenna beams of transmitter and receiver	14
Figure 2.4	A setting of blockage points spread in network for neighborhood functionality illustration	20
Figure 2.5	The grids affected by a new blockage point are filled with color	25
Figure 2.6	Average Blockage per blocking object	27
Figure 2.7	Average Handover per UE	28
Figure 2.8	False Positive Ratio	29
Figure 2.9	False Negative Ratio	29
Figure 2.10	Maximum Link Utility	30
Figure 2.11	The Average Rate	30
Figure 2.12	Average Blockage per UE for Exponential Delta Function	31
Figure 2.13	Average Blockage per UE for Inverted Delta Function	32
Figure 3.1	Optimal point targeting for corresponding reference direction	42
Figure 3.2	Average Blockage Count	68
Figure 3.3	Average Blockage Count Comparison Gurobi vs. Subgradient	69
Figure 3.4	Average Handover Count	70
Figure 3.5	False Positive Ratio	71
Figure 3.6	False Negative Ratio	71
Figure 3.7	Comparison of different parameters for all user association methods	72

Figure 3.8	Cumulative Distribution Function for SINR	72
Figure 3.9	Average Rate for Different User Association Algorithms	73
Figure 3.10	Average Maximum Load for Different User Association Algorithms	73
Figure 3.11	Average Maximum Load for Different Subproblem Sizes	74
Figure 3.12	SINR Comparison for Normal Constraint Method	74
Figure 4.1	Flowchart of training and prediction phases	94
Figure 4.2	Sectors are drawn around receiver and features associated with them	96
Figure 4.3	Example of EMA value change in a series of samples	98
Figure 4.4	The layout of the simulation area	103
Figure 4.5	Accuracy of cascade classifier trained for each receiver	104
Figure 4.6	Link blockage vs. non-link blockage accuracy of cascade classifier trained for each receiver	104
Figure 4.7	Accuracy of the K-Means clustering trained for each receiver	105
Figure 4.8	Accuracy of cascade classifier trained with no burst feature	106
Figure 4.9	blockage vs. non-blockage accuracy of cascade classifier trained with no burst feature	107
Figure 4.10	Accuracy of cascade classifier trained with burst feature	107
Figure 4.11	blockage vs. non-blockage accuracy of cascade classifier trained with burst feature	108
Figure 4.12	Effect of training set percentage on accuracy without online boosting and the burst feature	109
Figure 4.13	Effect of training set percentage on accuracy of blockage and non-blockage without online boosting and the burst feature	109
Figure 4.14	Effect of the training set percentage on accuracy without online boosting and with the burst feature	110

Figure 4.15	Effect of training set percentage on accuracy of blockage and non-blockage without online boosting and with the burst feature	110
Figure 4.16	The location experiment setting	111
Figure 4.17	Measured metrics for location experiments	111
Figure 4.18	Signal pattern for an unblocked link	116
Figure 4.19	Signal pattern for a blocked link	116
Figure 4.20	Average prediction accuracy in the bedroom	119
Figure 4.21	Average prediction accuracy in the living room	120
Figure 5.1	The paths formed around the receiver	125
Figure 5.2	The incidence angle of signal and the antenna array	128
Figure 5.3	The general process of sector-level sweep	134
Figure 5.4	A flowchart of out-of-band path discovery	136
Figure 5.5	The spatial spectrum of the signals in the office cubicle	140
Figure 5.6	The spatial spectrum of the signals in the bedroom	141
Figure 5.7	The spatial spectrum of the signals in the living room	142
Figure 5.8	The spatial spectrum of the signals in the living room with ESPRIT	143
Figure 5.9	The spatial spectrum of the signals in the outdoor environment	143
Figure 5.10	The topology of transmitter and receiver	144
Figure 5.11	CDF over all subcarriers N	145
Figure 5.12	Statistics over all subcarriers N	145
Figure 5.13	CDF over subcarriers 1 to 5 (ESPRIT) N	146
Figure 5.14	CDF over subcarriers 1 to 5 (MUSIC) N	147
Figure 5.15	CDF over all subcarriers NW	147

Figure 5.16 Statistics over all subcarriers NW	148
Figure 5.17 CDF over all subcarriers E	148
Figure 5.18 CDF over all subcarriers W	149
Figure 5.19 CDF of performance of the DoA estimation methods	149
Figure 5.20 Statistics of performance of the DoA estimation methods	150
Figure 5.21 CDF of performance of the beam training methods (North)	151
Figure 5.22 CDF of performance of the beam training methods (North West)	151
Figure 5.23 CDF of performance of the beam training methods (West)	152
Figure 5.24 Statistics of performance of different subcarriers	153

LIST OF ALGORITHMS

Algorithm 2.1	The algorithm of BD component	22
Algorithm 2.2	The algorithm of Handover component	22
Algorithm 3.1	Semi-distributed Algorithm for User Association (SDA)	56
Algorithm 4.1	Training algorithm of the classifier	99
Algorithm 4.2	Prediction algorithm by the classifier	99
Algorithm 5.1	The general procedure of Out-of-Band Path Discovery	138

CHAPTER 1

INTRODUCTION

With the rapid increase demand of wireless data due to the emerging applications, it is essential to go towards less populated frequency bands. Many emerging applications require high bandwidth, and the users demand a high degree of freedom in mobility while using those applications. The applications include the virtual reality interface for gaming and social networking, high-definition video streaming, and live video broadcast. To solve the problem of high wireless traffic in low-frequency bands different technologies like cognitive radio have been introduced but they cannot provide the bandwidths required for the data-hungry applications. We can also see an increasing number of mobile devices that run those applications over the time. For example, Ericsson Mobility Report [1] predicts each subscriber will have over 40 GB monthly mobile traffic by 2026. It also predicts that in 2026 the new generations of wireless networks will carry more than half of the world's smart-phone traffic.

The current frequency spectrum that is being used (sub 6 GHz) is very populated and unable to provide the required bandwidth for the performance enhancements stated in the fifth generation (5G) and the sixth generation (6G) of the wireless networks. Therefore, the applications need a much wider frequency spectrum that has the required bandwidth. One of the main enabling technologies of the future generations of wireless networks is the millimeter wave (mmW) technology that works in an enormous frequency band between 30 GHz-300 GHz of the electromagnetic spectrum [2] and is being used in the already-implemented 5G. In addition to 5G, 6G is also going to benefit from the mmW technology [3]. With the wireless traffic demands and the capacity goals that are defined in 5G and 6G, the mmW spectrum plays an important role in providing the capacities that are being met in 5G and those that are planned in 6G.

The research community calls this wide frequency band the millimeter wave spectrum because the wavelength of the signal is of the order of millimeters. The mmW spectrum has been researched for many years. Despite its long history, because of the inherent challenges of mmW propagation,

communication is sparse in this spectrum and this band has not attracted much attention from the communication research community up until the last two decades. However, in recent decades with the advances in communication electronics and integrated circuits, the mmW spectrum has started to gain more attention from the communication research community.

Despite the vast bandwidth that the mmW spectrum provides and its initial implementation in 5G, communication in this spectrum still inherently faces challenging problems due to the high frequency of the mmW signal. The mmW signal's propagation characteristics incur unique challenges to commercialize the mmW technology [4]. One of these challenges is the huge propagation loss of mmW technology due to rain, atmospheric, and molecular absorption [4]. The free space pathloss in the mmW spectrum is more severe than in the sub 6 GHz spectrum. Moreover, the common objects found in urban areas like building walls, glasses, foliage, and office cubicles can deteriorate the penetration loss of mmW [5]. Experiments show that for locations farther than 200 meters, the receiver could not detect any signals if the transmitter transmits with powers between 15 and 30 dBm [2]. The link breakage is likely due to environmental obstructions that block all paths (either via reflection or scattering). This is called outage in mmW technology [2].

Since the pathloss in mmW is significant, it is necessary to concentrate the signal power of a transmitter to a limited direction in space to make the signal travel farther. Therefore, the mmW signal propagates directionally as opposed to the sub 6 GHz signal that propagates omnidirectionally. To make the directional propagation of the mmW signal happen, the transmitter and the receiver use a phased array antenna to direct the signal to a specific direction in the space through beamforming. To have efficient communication, both the transmitter and the receiver should align their beams towards each other. When any of the nodes move, depending on the environment the transmitter and the receiver may require a beam alignment. The beam alignment is a challenging and high overhead task, especially in dynamic environments where there are many humans or other blocking objects, and User Equipment (UE) are on the move. Both nodes, UE and base station (BS), should keep their beams aligned while there is movement or blockage which is called beam tracking. Both

the node movement and the blockages make the mmW link fragile and it is necessary to tackle this challenge.

We look into the user association problem in mmW networks. The user association is performed when the UE needs to be connected to another BS. Due to the unique characteristics of the mmW propagation, the traditional user association methods perform poorly in mmW networks. To support reliable and efficient user association in mmW networks, we study effective solutions that considers mmW's unique propagation characteristics, such as signal directionality and low diffraction of the mmW signal for various blocking objects.

We further look into the problem for maintaining a robust link between a pair of nodes, or between a UE and a BS when the link is blocked or moved. Such solutions do not try to associate the UE to another BS but exploit the multiple mmW signal paths that are formed around the UE to maintain a robust link between the UE and the current BS. For example, when pedestrians block a link on a sidewalk, the solution is to find an alternate signal path, such as a path reflected off a surface, to quickly resolve such a transient situation. If the UE cannot find an alternate signal path, the UE may need to look for another BS to connect. Moreover, there are situations where the UE is stationary and does not need to change its BS. In such scenarios, the link is maintained by exploiting the multipath created by directional propagation of the mmW signal.

In the following, we describe the major areas of focus in this dissertation to give an initial exposure to the content of this dissertation. We provide a brief description of the types of problems that we define in this dissertation. Then, we depict a general view of the solution for those problems for a better understanding of future chapters.

We organize the rest of this chapter as follows. We provide an introductory explanation of different problems and their solutions in this chapter. In Section 1.1, we discuss the problem of blockage in mmW systems along with a proposed solution. Section 1.2 discusses a biobjective approach to jointly optimize the load balance and the blockage of the network. We use the approach introduced in Section 1.1 for the blockage objective function. Section 1.3 introduces the problem of detection of blockage and link movement in mmW networks and proposes a learning approach

to detect them efficiently. Finally, Section 1.4 introduces an out-of-band method to perform beamforming on the fly for low overhead beam tracking in mmW networks.

1.1 Blockage Mitigation in mmW Networks

Blockage in the mmW networks is a serious problem that we need to address efficiently to achieve robust communication. Based on outdoor experiments that researchers have done, the human body can block the mmW fatally [6]. The human blockage is also true in indoor settings [7]. Since the body mass is mainly composed of water and water is a good blocker of mm-Wave, the human body can attenuate up to 31.2 dB of mmW's beam strength if positioned close enough to the transmitter in the way of the signal [7]. The authors in [6] and [7] suggest three methods to mitigate the negative effects of human blockage, including switching the victim UE's antenna to another reflection path, dilating (widening) the beam, and connecting to another BS.

The reflected signal in the None-Line-of-Sight (NLOS) state does not always provide the network with desired performance [4]. The strength of the reflected signal also depends on the power of the original signal and the surface reflection loss. Moreover, the signal reflection method is more effective in indoor environments [7]. In the event of a human blockage, the beam dilation can help maintain the link connectivity but only if the original link SINR is high and when the blockage is far enough from the transmitter/receiver. In either case, the link rate degrades significantly [7].

We can induce that choosing a better BS in case of blockage is more effective in terms of the network performance than the other two methods in some specific cases. This problem falls into the field of User Association in cellular networks. The problem is how to associate UE to different BSs in a way that the user association improves a metric. Common metrics that the literature has considered for this purpose are SINR, rate of the links, load balance of the network, outage/coverage probability, spectrum efficiency, and fairness. UEs are associated to BSs to improve those metrics throughout the network individually or jointly [8]. The user association is more suitable for environments with limited number of blockages. This is mainly due to the overhead of the handover. Excessive blockages can result in increased number of handovers which

may decrease the gain of having the user association to combat the link blockage.

We propose a protocol to mitigate the severity of human-originated blockage in mmW cellular networks using previous blockage incidents that the network has experienced. We exploit the previous blockages to make more educated guesses of what is going to be the next best BS for a UE in terms of minimum blockage chance. This history of blockage incidents in the network acts as a heuristic for choosing the best BS for a UE to minimize the number of blockages experienced from the current time forward. For this purpose, the centralized entity of Cloud Radio Access Network (C-RAN) does all the coordination for keeping the history and associating UEs to the BSs. We elaborate on the method in Chapter 2.

1.2 Biobjective Optimization for Load and Blockage

As stated in Section 1.1, one of the main methods to combat the blockage problem in mmW is the user association to a different BS. The existing user association algorithms for reducing blockage and increasing link availability in mmW networks do not consider load balancing base stations. The blockage reduction algorithms aim to minimize the blockage effect which may result in overloading some BSs due to having more UE associated to those BSs with links that the solution considers to have a lower chance of getting blocked by blocking objects. Therefore, if the UEs associate themselves to BSs only according to the blockage objective, the UEs will overload some BSs while others remain underutilized, which may lead to performance degradation [9]. In an unbalanced network, some BSs may have extra capacity to accept new UEs but are less attractive for UEs to connect because the algorithm assigns a high blockage chance to those BSs. In this case, since the blockage chance of a link is only an estimation of the future blockage incidents based on the previous blockages, the actual blockage may not happen for a BS whose blockage chance is estimated to be high and consequently becomes underutilized. In general, the blockage reduction algorithms divert UEs to BSs with less blockage chance, which may result in overloaded or underutilized BSs and degrade network performance.

This problem calls for a more holistic view of blockage control and load balancing. Improving

the performance based on one metric results in deterioration of the other, a framework that accounts for both blockage and load balancing at the same time is needed to improve the overall performance.

To address this problem, we consider both blockage reduction and load balance in our user association algorithm design and formulate a biobjective combinatorial optimization problem to optimize user association in mmW cellular networks according to the two objectives. We can adjust the involvement of each of the objectives when we need. We provide a detailed discussion of this method in Chapter 3.

1.3 Blockage Prediction Using Learning Approaches

In case of intermittent blockages in a highly dynamic environment, it may not be efficient to associate the UE to another BS to escape a blockage. Instead, the link quality may be improved by finding an alternative path between the BS and the UE. To improve the communication throughput of the mmW link, it is important to detect the blockage quickly and efficiently and minimize the overhead of finding another beam for communication. The primary use case of mmW technology in 5G and 6G is cellular networks. People use cellular networks in outdoor environments with highly dynamic mobile objects, including transceivers and blocking objects. Beam misalignment between a transmitter and a receiver can happen when at least one of them is moving. The experiments conducted by authors in [7] show that variations in RSSI occur when the link is moving.

As mentioned in Section 1.1, because of the high frequency and lower diffraction value compared to sub 30 GHz spectrum, in the mmW spectrum, the signal cannot go around the objects on its way or penetrate them efficiently [10]. Therefore, the blocking objects can break mmW links fatally. We call this link quality drop *link blockage*. RSSI variation is a sign of the blocking effect of blocking objects on mmW link during a time window. The authors in [7] show that when a blocking object blocks the mmW link, the algorithm can use the variations in RSSI to detect the blockage. Another cause of the link quality drop is *link movement*. The link movement happens when either of the nodes move and the nodes need to adjust their beams. The initial beam misalignment causes the link quality to drop.

It is essential to discriminate the link quality drop due to link blockage or link movement. Because they need their own treatment. Applying the solution of one problem to the other may make the situation worse. Detecting a link blockage or a link movement requires 30 ms according to [7]. This delay can have a negative impact on quality of experience. We propose a blockage classification method to use diffraction to enable the nodes to predict the blockage without the need of high-overhead signal strength recording. Chapter 4 discusses the problem and the solution.

1.4 Out-of-Band Beam Adaptation

Section 1.3 discusses that it is essential to discriminate between the blockage and the link movement for better link recovery operations. When the prediction is done, the solution reacts to the source of the link degradation accordingly to improve link quality. In both cases of link degradation, the UE needs to find the signal paths that are updated due to the blockage or the movement of either node (UE or BS).

One method of aligning the beam on the strongest signal path is using the mmW signal. This method has been used extensively and has its own overhead on a communication system. The directional communication in mmW makes the communication system scan the environment which is a high overhead task. In some cases, it is not possible to communicate through LOS and NLOS paths should be used. If a signal with different carrier frequency than mmW is used, the path discovery is out-of-band.

The existing out-of-band path discovery methods focus on finding the single line-of-sight path. Many of the methods use specialized hardware to achieve their goal. Our approach looks for all the possible paths of the mmW band in the out-of-band signal. We implement and design it on commercial off-the-shelf devices. It is very challenging to obtain accurate information for multiple path discovery using the limited hardware of commercial devices. We discuss a full account of the problem and the respective solution in Chapter 5.

1.5 Summary

In this dissertation, we first investigate the user association problem in mmW networks with its inherent propagation characteristics. We propose a history-based blockage mitigation method that uses blockage history to make blockage-aware user association in mmW networks. We show that the proposed method has better performance than other common criteria for user association. We also show that using only one criterion for user association is not enough to achieve high performance in mmW networks. Therefore, we jointly optimize load balance along with the blockage in the user association process. We show that considering both blockage and load balance in the user association process makes the network balanced while reducing the number of blockages.

Moreover, we look into the problem of maintaining a robust link between two nodes. We show that using the diffraction values is a reliable method to predict the blockage in mmW networks. We also show that with a limited number of data points we can achieve high accuracy in blockage prediction.

To maintain a robust link in case of a blockage, we propose to exploit the multipath in the mmW spectrum by using the signal from a non-mmW frequency to discover mmW signal paths in the receiver. We use such paths to avoid a time-consuming beam scan in mmW frequency. Our results show that it is possible to have out-of-band beam adaptation without the high-overhead in-band beam scan on commodity devices. Through experiments we are able to show that the performance of the adapted beam is comparable to the in-band scanned beam.

CHAPTER 2

BLOCKAGE MITIGATION IN MILLIMETER WAVE NETWORKS

In this chapter, we study the impact of blockage in mmW networks and present our proposed method to mitigate the blockage for user association in such networks. Due to the propagation characteristics of the mmW signal, the blocking objects in an environment can easily block the signal. In wireless communications, the user association is a method to select a BS for a UE. There are various criteria for selecting a BS in non-mmW networks such as better SINR, higher rate or lower power consumption. However, such criteria do not work well when applied to the mmW networks. Therefore, we need to define a new criterion that considers the specific propagation characteristics of the mmW signal. According to the organization that we discuss in Chapter 1, the first step towards the robust mmW communication is to lower the blockage by circumventing it using associating the UE to a BS with the lowest probability of blockage.

In urban areas, there are highly populated spots that have more frequent human activities. Therefore, there is a higher probability that a link between a UE and a BS in such areas gets blocked. The proposed algorithm aims at decreasing the blockage rate by detecting BSs that have better beam penetration to the highly populated areas and selecting those BSs for UEs requesting data communication. The proposed algorithm also adapts itself to the change of location of highly blocked areas through time. Since the spectrum shortage is severe in populated urban areas, it is a valid assumption that we use the proposed algorithm in such networks.

We organize the rest of this chapter as follows. In Section 2.1, we discuss the papers in the field of user association in both sub 6 GHz and mmW spectrum. We dedicate Section 2.2 to providing system preliminaries and Cloud Radio Access Network (C-RAN) primer. In Section 2.3, we introduce the proposed algorithm and elaborate on it. Finally, we present the system evaluation experiment results in section 2.4.

2.1 Related Work

The goal of user association is to improve or optimize different metrics by associating UEs to appropriate BSs.

Dhillon et al. [11] have developed a closed-form formulation for coverage and outage probability in a multi-tier heterogeneous cellular network (HCN). Jo et al. [12] have involved the BS association bias in formulating the outage probability for a UE in a tier or the whole network. The user association realizes the balancing process. The solution gives Each UE-BS pair an activity fraction to maximize the utility of the whole network. The authors directly formulated the user association problem as a convex network utility maximization problem in [13]. The algorithm centrally performs the optimization process.

Bai et al. [14] have formulated SINR and rate in mmW networks. The association criterion is the minimum path loss in their work. The traditional methods that the literature proposed for sub 3 GHz networks cannot be used directly to mmW user association problem [4, 15]. In [16], the authors addressed the problem of user association as a multi-assignment optimization problem. They used auction algorithms to maximize the weighted throughput of all BSs. Xu et al. [17], proposed a distributed solution to the dynamic resource allocation problem in mmW networks as an optimization problem and proposed a distributed solution. Authors in [18] defined the user association problem in the context of load balancing and fairness as an optimization problem. The authors defined a dual Lagrangian problem and they solved it. A possible direction in harnessing the power of mmW is to use this technology in current cellular networks in the form of multi-band HCNs. Sakaguchi et al. [19] have proposed an architecture for mmW small cells in HCN that C-RAN controls the process.

The fundamental difference between the proposed method and the methods reviewed in this section is that our proposed method uses the previous incidents of blockage to determine the current blockage possibility of the UE to each BS. At the time of our work, we were not aware of any papers concerning user association has the memory to use the history of link quality fluctuations.

2.2 System Model and C-RAN Primer

In this section, we first discuss the specifics of the system we propose. Due to the nature of cellular mobility, handover is frequent, and due to the nature of mmW, beamforming is inevitable. Collaboration between BSs is critical to realizing the above tasks. C-RAN is one of the most promising candidates and the research community expects it to be popular in the future. Several C-RAN projects have been initiated in many organizations such as Next Generation Mobile Networks (NGMN) [20]. Our proposed algorithm can be regarded as a function added to C-RAN architecture.

2.2.1 System Model

In our system, we have N BSs, M UEs, and P blocking objects. The number of blocking objects is the same as the number of UEs as each UE is assumed a human being carries it which is a blocking object itself. The set of BSs is defined as $\mathcal{N} = \{1, 2, \dots, N\}$ and the set of the UEs is defined as $\mathcal{M} = \{1, 2, \dots, M\}$. For each BS_i , the set of UEs that is served by the BS is M_i . For each UE_j , the set of all BSs that is in the UE's communication range is \mathcal{N}_j . We assume the communication range of UE and BS to be equal to R meters. We call the parameter the communication disk. We assume each BS is mounted on a post with h meters from the ground. Each BS_i is equipped with a_i number of antennas meaning the BS_i can serve up to a_i UEs at a time. On a BS, the controller assigns the beam on each antenna with a different channel. We assume the antennas are able to rotate all 360 degrees or in the interval, $[-\pi, +\pi]$ in radian. Figure 2.1 shows the concept of communication disk and the human blockage and UEs. As shown in Figure 2.1, we illustrate the relative positions of the BS, the blocking object, and the UE in a three-dimensional (3D) space. Since the space of the illustration is a space that consists of x and y dimensions which is called 2-Dimensional (2D) space, it is difficult to convey the exact 3D shape of the set of elements in the network.

We also depict an example network in Figure 2.2. The figure shows a street that is equipped with BSs that are trying to serve UEs despite the human blockages. We adapt the SINR and rate equations from [14]. In this chapter, there is a formulation for each of the SINR and consequently,

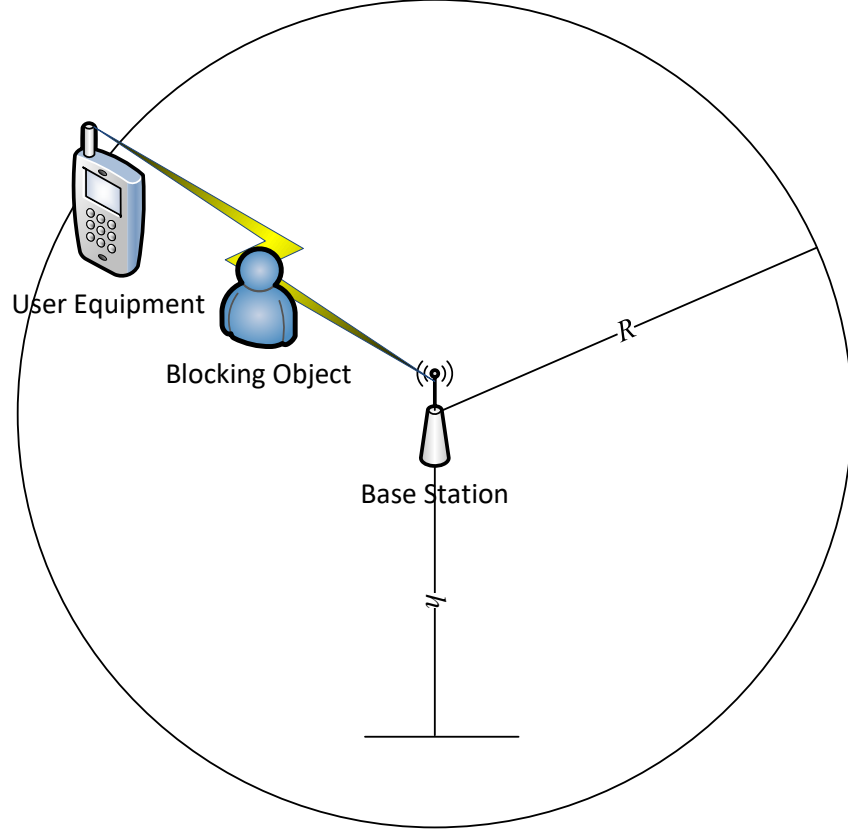


Figure 2.1 The concept of communication disk and human blockage in mmW network

the rate is based on the characteristics of mmW technology. The difference of what we calculate from the original SINR/rate formulations comes from the fact that in simulations, we have the actual values we do not have to deal with SINR/rate values as random variables. We only use the SINR/rate values for system evaluation purposes.

In [14], the path loss model introduced is a hybrid one in which we have two models for LOS and NLOS paths, respectively. $L(F)$ is the path loss of a link with length F and authors in [14] formulated it as,

$$L(F) = I(p(F))C_L F^{-\alpha_L} + (1 - I(p(F)))C_N F^{-\alpha_N} \quad (2.1)$$

In Equation (2.1), $I(x)$ is the Bernoulli random variable with parameter x , α_L is the LOS path loss exponent, α_N is the NLOS path loss exponent, and C_L and C_N are intercepts of the LOS and NLOS path loss formulas. In our system, we just use either the first term or the second term of (2.1) depending on whether there is a blockage in a link or not.

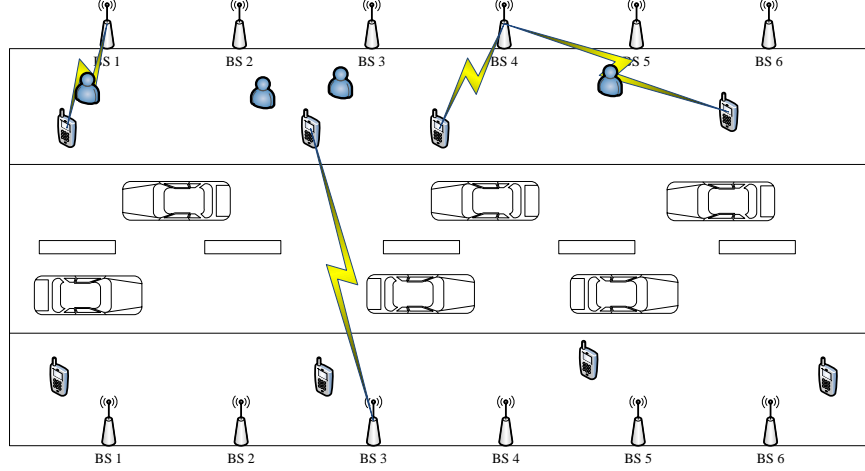


Figure 2.2 An example network with UEs, BSs and blocking objects

For the antenna gain, the authors in [14] bring the antenna directivity into attention since in mmW technology the directivity of the beam determines most of the antenna gains. In the model used in [14], B is the directivity gain of the main lobe and b is the directivity gain of the side lobe. θ is the beamwidth of the main lobe in radian and ϕ is the angle from the boresight direction. Then, we denote the directivity of an antenna as $G_{(B,b,\theta)}(\phi)$. To compute directivity gain, the authors in [14] give the formulation as

$$D_l = G_{B_t, b_t, \theta_t}(\phi_t^l) \cdot G_{B_r, b_r, \theta_r}(\phi_r^l) \quad (2.2)$$

where t and r denote parameters for transmitter and receiver, respectively. We compute the directivity gain for a BS and its associated UE as $D_0 = B_r B_t$. D_l is the link directivity of all interfering links for link 0. The authors in [14] also define the directivity gain as a random variable but we compute the real directivity gains for each link. In their work, they assume the angles of transmitter and receiver antenna beam have a uniform distribution. For each link, we compute the different portions of the main lobe and side lobes of the two ends of the link that align with each other.

We consider four different states for directivity gain, 1) when both the transmitter and the receiver completely align their main lobes, the directivity gain is $B_r B_t$, 2) when the receiver aligns the whole width of the main lobe with just side lobe of the transmitter, the directivity gain is $B_r b_t$, 3) when the receiver aligns its side lobe with all main lobe of the transmitter, the directivity is $b_r B_t$

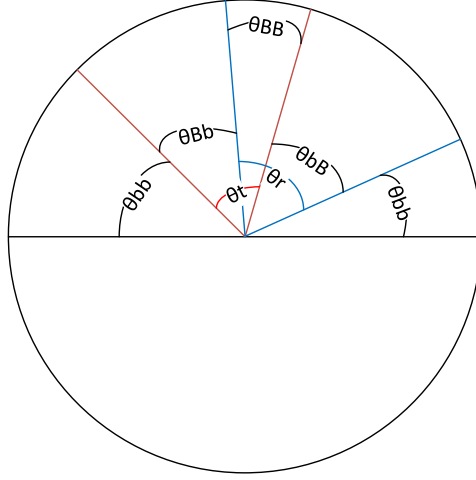


Figure 2.3 Different angle values for different parts of antenna beams of transmitter and receiver

and 4) when both the transmitter and the receiver completely align their side lobes, directivity gain is $b_r b_t$. These values are special ones that only happen under certain conditions. But in reality, there are many cases that transmitter-receiver main/side lobe alignment is partial.

To compute those gains, in [14], they used the probabilities (uniform distribution of boresight angles of transmitter and receiver). Instead, we use the actual angles of the transmitter and the receiver. For each link, we compute the different portions of the main lobe and side lobes of the two ends of the link that align with each other. The result of the computation can be the perfect portions like mentioned above. That way, we compute the directivity gains as stated above. But for proportional cases, the proportion is multiplied by each case. Figure 2.3 shows how different portions are sized according to the antenna orientation of transmitter and receiver. In Figure 3, the sector with red lines is the beam of the transmitter and the sector with blue lines is the beam of the receiver. For grayscale recognition, θ_t is the beamwidth of the transmitter and θ_r is receiver beamwidth. For the sake of simplicity, we invert the direction of the antenna of one node for the transmitter or the receiver so both sides have the same direction. θ_{BB} is the angle of beam portion that both transmitter and receiver main lobes align on each other. θ_{Bb} is the angle of the main lobe of transmitter and side lobe of receiver aligned, θ_{bB} is the angle of side lobe of transmitter and the main lobe of receiver aligned and θ_{bb} is the angle of side lobe of transmitter and side lobe of receiver aligned. We compute these alignment angles and inverting one side based on bisectors of

main lobe sectors of transmitter and receiver. We compute the directivity gain for a link connecting a UE and its serving BS k using (2.3),

$$D_k = \frac{\theta_{BB}}{2\pi} B^2 + Bb \left(\frac{\theta_{Bb}}{2\pi} + \frac{\theta_{bB}}{2\pi} \right) + \frac{\theta_{bb}}{2\pi} b^2 \quad (2.3)$$

In Equation (2.3), the value D_k is the directivity gain between transmitter and the receiver. We normalize the portions of different angles by dividing each angle by 2π .

The Authors in [14] formulate the SINR for a link connecting a UE_e and its serving BS_k as Equation (2.4),

$$SINR_{e,k} = \frac{h_k D_k L(F_k)}{\sigma^2 + \sum_{l \in (\mathcal{N}_e - k)} h_l D_l L(F_l)} \quad (2.4)$$

In Equation (2.4), h_k is the small scale fading for the link between BS_k and the UE_e associated to it, σ^2 is the thermal noise power of the propagation environment. We consider the small scale fading in Equation (2.4) independent Nakagami fading for each link according to [14]. We neglect frequency selectivity in fading as research shows delay spread is generally small [21] and we can minimize the impact of frequency selective fading by techniques like orthogonal frequency-division multiplexing or frequency domain equalization [14]. We compute the SINR value as signal power divided by noise plus interference power. The interference power is composed of the integration of power effect of antenna direction of UE_e and the other BSs that are in the communication range of e except e 's serving BS_k . We formulate the rate of the link between UE_e and its serving BS_k as the following equation,

$$\Gamma = W \log_2(1 + SINR_{e,k}) \quad (2.5)$$

where W is the bandwidth of the system and is assigned to a UE in Hz.

The material provided in this section gives a more detailed view of the system on which the proposed algorithm has been implemented. In the next section, we survey C-RAN technology.

2.2.2 C-RAN Primer

In this section, we provide a primer of Cloud Radio Access Network (C-RAN) technology. The research community has proposed this technology to equip the modern wireless system with a

centralized entity for better and integrated control. Since in the proposed algorithm, the algorithm records and processes the history of blockage and makes it available to the UE's that need the blockage information, it is more efficient to give the responsibility of collecting, processing, and distribution of the blockage history to the centralized C-RAN.

As the volume of wireless traffic increases over the years, it is important to keep the costs low. These costs include the total cost of ownership (TCO), cost of sites of equipment, and the cost of power [20]. Moreover, the complexity of future radio access networks makes it difficult to maintain or upgrade networks to new standards like 5G while different networks with different technologies co-exist. China Mobile Research Institute proposed Centralized, Collaborative, Cloud, and Clean RAN (C-RAN) to address these problems [22].

C-RAN centralized different processing resources to form a single resource pool to allocate them dynamically on demand. This helps in lowering interference and energy consumption and increasing resource utilization. This technology is not only usable in current wireless networks but also essential in future 5G standard technologies like Large Scale Antenna Systems (LSAS), full-duplex and ultra-dense networks [20] mainly because of its centralized nature.

C-RAN architecture is composed of three main components: Base-band Unit (BBU) pool, Remote Radio Unit (RRU) networks, and transport networks. A BBU pool is located at a centralized site and consists of time-varying sets of soft BBU nodes. A soft BBU is a BBU instance in a traditional network where processing resources and capabilities are dynamically allocated and reconfigured on real-time conditions [20]. The RRU networks are the same as the traditional networks. They provide the UEs the wireless coverage. Transport networks are networks that connect BBU pools to RRU networks. Depending on different configurations the connecting medium can differ from fiber-optic to microwave transmission.

The network can use the C-RAN technology to implement the proposed algorithm in this paper. We use the C-RAN as a coordinator that can store the distributed data of blockage incidents and do the processing and then use the processed data as information to make BS association blockage aware. We use C-RAN to manage Het-Net that is a possible usage point for mmW technology.

As depicted in Figure 1 in [20], the General Purpose Processor Platform can be used to do the processing of the algorithm. It includes running the required functions and storing the necessary data. Then the solution makes the user association decision based on information received from RRU of UEs and the current blockage map of the network in C-RAN.

Considering the primary information provided in this section, we are now ready to discuss our proposed algorithm of User Association for blockage mitigation.

2.3 User Association for Blockage Mitigation

In this section, we introduce and elaborate on our proposed algorithm of user association for blockage mitigation in mmW networks.

The main principle that justifies the proposed algorithm is that the places that have more common blockage incidents have a higher chance for the blocking objects to block UEs in the future. It means if we can record the points in the coverage area that more blockage incidents have occurred in the past, it is more likely that UEs that move to those areas will experience blockage than other areas. Based on this observation, we propose an algorithm that associates UEs to BSs that the blocking objects are less likely to block the UEs in the future. In urban areas like the entry of department stores or subway stations, we can easily see that if blocking objects block a group of UEs; it is more likely for other UEs that will be present in those locations in the future to have the blockage. A higher blockage score for a BS from a specific geographical location means there has been a considerable number of blockages so far and heuristically, and there will be in the future. The algorithm has three main components that we discuss in different subsections.

2.3.1 Blockage Detector Component (BD)

The first component is the blockage detector (BD) in the UEs. The BD is responsible for detecting the blockage incidents and sending the blockage information to the associated BS in a blockage information packet (BIP). Upon receiving the BIP, the BS forwards the BIP to the C-RAN for further processing.

The BIP contains three information fields. The fields include blocked UE identifier (BUE), location of blocked UE at the time of the blockage (served as blockage point), movement specification of UE at the time of blockages like direction and speed. We can use these optional fields for other mining operations performed on blockage information to further improve the algorithm.

For UE location-dependent information, we can say that since the manufacturers of today's smart devices equip them with GPS modules it is possible to obtain the location and consequently speed and movement direction of the UE. The only problem of using the location of each UE at the time of blockage is the privacy issue. We can resolve that by storing BUEs as coded information anonymously. This way C-RAN has BIP, but in the field of BUE, an identifier is stored. The blockage mitigation algorithm uses this identifier exclusively.

Based on the measurements on RSSI variations in a limited amount of time, in [7] the authors have done measurements to find a pattern for each blockage and UE motion. The human blockage has a more regular pattern of RSSI variation that the algorithm can exploit to distinguish between the blockage and UE movement. For this reason, they defined a measurement window of RSSI. A window is classified as blockage if $N_{\partial^2 RSSI}$ is less than threshold Θ_{block} and as motion, if $N_{\partial^2 RSSI}$ is more than an upper threshold Θ_{motion} . Then the confidence of classification is accumulated over time until it is over 95% or exceeds a detection latency bound. If $RSSI(t)$ denotes RSSI at time t , and T is the window length, then the authors in [7] calculated the second-order statistics of RSSI values at the receiver side over a time window of T as follows:

$$\partial^2 RSSI(t, T) = \frac{\partial^2}{\partial^2 t} [RSSI(t), \dots, RSSI(t + T)] \text{ and } N_{\partial^2 RSSI} = Var\{\partial^2 RSSI(t, T)\} \quad (2.6)$$

It means the receiver computes the variance of a window of RSSI measurements and the receiver compares it against blockage and motion thresholds. In [7], the threshold values for blockage and motion classification are set empirically to 0.002 and 0.2, respectively.

2.3.2 Blockage Point Collector Component (BPC)

The second component is the blockage point collector (BPC), which resides in the C-RAN. The BPC is responsible for collecting the blockage point information from the UEs all over the network and storing them in an informative way. The algorithm in BPC receives the BIPs from BSs and extracts the information. The CRAN then stores the blockage information to use them as a database for computing blockage scores.

Blockage score is a quantity that BPC computes to determine the blockage likelihood of a link between a UE at a certain location trying to connect to a specific BS. We compute the blockage score of a link between BS_i and UE_j as follows:

$$\gamma_{ij} = \sum_{p \in NH(i,j)} \Delta(p, j) \quad (2.7)$$

where p is a blockage point recorded in the past by its BIP, i is the index of BS and j is the index of UE. $NH(i, j)$ is a set of blockage points that happened in the past between BS_i and UEs in the neighborhood of the current location of UE_j . We define the neighborhood as the distance between UE_j , and the blockage point is not more than a threshold called $d_{effective}$. This quantity is the maximum distance of a blockage point that would affect the blockage likelihood of a link. It is logical to select $d_{effective} = R$. Since the communication range of a UE and BS is not more than R , so there is no point in selecting $d_{effective}$ more than R .

$\Delta(p, j)$ is a function of distance between blockage point p and current location of UE_j . We define two different functions Δ_1 and Δ_2 . The distance function defines the effect intensity of different blockage points on a certain link. If the function tends to return a low value, the function may lead to the false positive association, a BS that may be unsafe is falsely considered safe. If the function tends to return a high value, the function may cause false negatives, the algorithm falsely considers a safe BS unsafe. We propose two flavors of the function delta, inverted distance or exponential function of distance.

$$\Delta_1(p, j) = \frac{1}{\sqrt{(x_p - x_j)^2 + (y_p - y_j)^2}} \quad (2.8a)$$

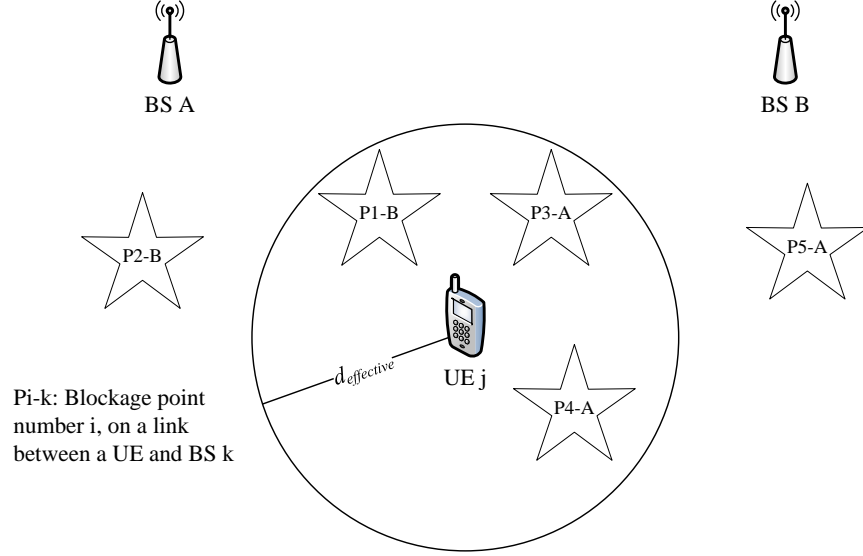


Figure 2.4 A setting of blockage points spread in network for neighborhood functionality illustration

$$\Delta_2(p, j) = e^{-\sqrt{(x_p - x_j)^2 + (y_p - y_j)^2}} \quad (2.8b)$$

In Equation (2.8), (x_p, y_p) is the coordinate of blockage point p in 2D space, (x_j, y_j) is the coordinate of current location of UE_j . We can imply from (2.8) that if the distance of UE_j from a blockage point is more, the resulting blockage score for UE_j would be lower. This comparison makes sense as we do not expect a blocking object blocks a link if the link had a safe distance from previous blockage points. Figure 2.4 shows a simple network setting to illustrate the subset of blockage points that $NH(i, j)$ selects. In Figure 2.4, the set of neighborhood of UE_j for BS_A is $NH(A, j) = P3, P4$ and for BS_B is $NH(B, j) = P1$. Blockage points $P2$ and $P5$ do not belong to any neighborhoods of UE_j since they are not in UE_j 's $d_{effective}$ range.

2.3.3 Device Association Component (DA)

We call the third component of the system the Device Association (DA). The responsibility of DA is to use the blockage scores produced by BPC to select the best BS according to their blockage scores.

When BS_i forwards a BIP from UE_j to the C-RAN, the BPC computes the blockage score γ_{ij} of the link between BS_i and UE_j . Therefore, the output of BPC is a vector $\vec{\lambda}_j$. This would be

a vector of the blockage scores of UE_j if the UE were going to be associated to different BSs in its range, R . Therefore, for each of the BSs in \mathcal{N}_j , the algorithm should compute the vector. We compute each element of $\vec{\lambda}_j$ as follows,

$$\vec{\lambda}_j(i) = \gamma_{ij}, \quad \forall i \in \mathcal{N}_j \quad (2.9)$$

In Equation (2.9), DA collects the blockage scores of UE_j concerning BSs that are in its communication range R , which generates vector $\vec{\lambda}_j$.

Up to this point, DA has the required information to decide associating UE_j to one of the BSs in \mathcal{N}_j . Our method associates UE_j to a BS with minimum blockage score in $\vec{\lambda}_j$. Since the capacity of a BS is limited and there is a limited number of antennas, it is possible that the BS with minimum blockage score cannot serve UE_j . In this case, the method selects the BS with the next minimum blockage score. This process continues until UE is associated to a BS that has the capacity to serve the new UE and has the lowest blockage score. The best BS to serve UE_j called α_j and our proposed algorithm chooses it by the following formulation,

$$\alpha_j = \arg \min_{i \in \mathcal{N}_j} \vec{\lambda}_j(i) \cdot x_i \quad (2.10a)$$

$$x_i = \begin{cases} 1, & |u_i| \leq a_i \\ cost_{max}, & \text{otherwise} \end{cases} \quad (2.10b)$$

Where x_i detects if BS_i has the capacity to accept UE_j as one of its clients. Since the blockage scores are numbers between 0 and 1, the blockage score γ_{ij} cannot be more than the number of blockage points in $NH(i, j)$. Therefore, we should choose $cost_{max}$ in a way to maximize the term $\vec{\lambda}_j(i) \cdot x_i$, so that the algorithm does not choose BS_i if BS_i does not have the capacity to serve UE_j .

The traditional method looks for a distance from BS that the RSSI is less than a threshold. We consider that distance as the borderline of the communication disk of the BS. The handover takes place at the borderline. Since in our algorithm minimizing the blockage incidents is of high priority, the algorithm triggers the handover procedure whenever the blockage score of the link between UE_j and its current associate BS is more than the link between UE_j and other BSs in \mathcal{N}_j .

Algorithm 2.1 The algorithm of BD component

- 1: Sense RSSI and compute the RSSI window;
- 2: Compute $N_{\partial^2 RSSI}$ according to (2.6);
- 3: Compare to thresholds to classify window as blockage
- 4: Send the BIP to associated BS;

Algorithm 2.2 The algorithm of Handover component

- 1: **if** *handover needed* **then**
- 2: Compute γ_{ij} using (2.7);
- 3: Compute $\tilde{\lambda}_j(i)$ using (2.9);
- 4: Find BS with best blockage score using (2.10);
- 5: Do the handover;
- 6: **end if**

If that is the case, the algorithm hands over UE_j to BS with the lowest blockage score according to Equation (2.10).

The method mentioned above has one major drawback. The UEs need to report their locations to their BS, and the BSs need to relay that information to C-RAN to compute the blockage score vectors. This algorithm incurs a certain amount of overhead. However, the proposed algorithm effectively reduces the chance of UEs being blocked. Although we can apply the proposed algorithm to the handover task, the proposed algorithm is appropriate for the user association task, not the handover for the reason mentioned. Therefore, it is logical to use the traditional handover methods to detect the need for handover. When the need is evident, then our proposed algorithm can play its role to compute the best BS to hand over to.

The Algorithms 2.1 and 2.2 generalized the whole process of blockage point registration and using blockage information for selecting the best BS as the next BS for a UE. This algorithm has two parts. The first part implements the BD algorithm and runs on the UE the second part implements the BPC and DA components which the algorithm uses them in the handover detection procedure.

2.3.4 Time Complexity and Overhead Analysis

The BD component needs to measure RSSI and record RSSI T times. Therefore, the complexity of this component is of order $O(T)$ with T being the RSSI window size. BD also sends BIP once for

each blockage that occurs, meaning for each $O(T)$ computation there is $O(1)$ overhead for sending BIP. However, for the total number of BIPs sent by a UE, we can say the time complexity is not of constant order and we need to address it. If the maximum number of BIPs sent by UE_j while connected to BS_i is H , then the total BIP transfer in the network for all of the UEs sending to all of the BSs is of order $O(NMH)$.

In the BPC component, for computing blockage score γ_{ij} in Equation (2.7) we need to know how many $\Delta(p, j)$ summations we do. This means that we need to know the size of set $NH(i, j)$, the number of blockage points that occurred in $d_{effective}$ range of the current location of the UE_j while UEs in those points connected to BS_i . In order to have a good estimate of the size of $NH(i, j)$, we assume the number of new blockage points in each time instance compared to the previous one is bounded. We call this bound c . This quantity says that the number of new blockage points added to $NH(i, j)$, $\forall i \in \mathcal{N}, j \in \mathcal{M}$ cannot go more than c . This assumption is valid since there is a limited number of blockages in different time instances added. One can find this limit through an experiment at different hours of a day and different days of a week. This is equivalent to the recursive equation below,

$$n(t) = n(t - 1) + c, n(0) = 0 \quad (2.11a)$$

$$n(t) = ct \quad (2.11b)$$

As shown in Equation (2.11), the recursive function's closed-form solution is the function $n(t) = ct$. Variable t is the time instance number t . Therefore, the order of number of elements of set $NH(i, j)$ is $O(ct)$ for time instance number t . There is no specific packet overhead for BPC component since BPC computes the blockage scores based on the current location of UE_j .

In the DA component, Equation (2.10) has the time complexity related to the size of the set \mathcal{N}_j . We can estimate this quantity by the density of BS in the implementation field. If the area of the implementation field is A , there are N BSs deployed in this area; thus, the density of BS in this field is $\frac{N}{A}$. Therefore, the expected number of elements of \mathcal{N}_j is equal to $\pi \times R^2 \times \frac{N}{A}$ and the time complexity of DA component is equal to $O(\frac{N}{A} \times R^2)$. As we can see, the time complexity has square relation with the radius of the blockage effect disk. It means if the radius R is halved, the

time complexity becomes a quarter of the original. The only information transfer overhead for the DA component is to decide for the best BS and inform the BS and UE of the decision. Therefore, the time complexity to transfer data should be of order $O(1)$.

According to Equation (2.11), $n(t) = ct$ can serve as a cap for the total number of the blockage incidents that happened in the communication disk of BS_i that are in the effective blockage disk of UE_j . According to Equation (2.7), to compute the blockage score of each UE that has a different current location than the previous one, we should reconsider all the points in $NH(i, j)$. This process leaves a big overhead on the system as a slight change in location of UE_j obligates BPC to redo all the computations. We propose a modification to our algorithm to enhance the algorithm in terms of computation complexity and compromising reasonable prediction accuracy.

In our proposed modification, we define fixed points in the implementation field. We call these points *anchor points*. The implementation field is divided into grids of the same size virtually. It means BPC can compute from the location of a UE in which virtual grid the UE lands. We define the anchor points as the center of each grid. When a new BIP is received, the algorithm computes the blockage scores for each anchor point. In other words, the algorithm considers all the points in the same grid equally. Therefore, in BPC instead of computing the blockage score from scratch for each new location of each UE, the blockage score of each UE is the blockage score of the grid the UE is located. The modification enhances the time complexity of the BPC in Equation (2.11) to $O(1)$. This is because the BPC only needs to find the blockage score for UE_j and BS_i from a table according to the location of UE_j . In the originally proposed algorithm, BPC needs a place to hold all BIPs. However, since in the modified version BPC stores the BIPs in the form of blockage scores for grids, there is no need for that storage.

The BPC computes the blockage score in an online manner as it receives the new BIPs, and there is no need for storing BIPs. The new BIP updates the blockage score of the grids that the distance of the blockage point to those anchor points is not more than $d_{effective}$. Figure 2.5 shows the update of the blockage score of grids within $d_{effective}$ range.

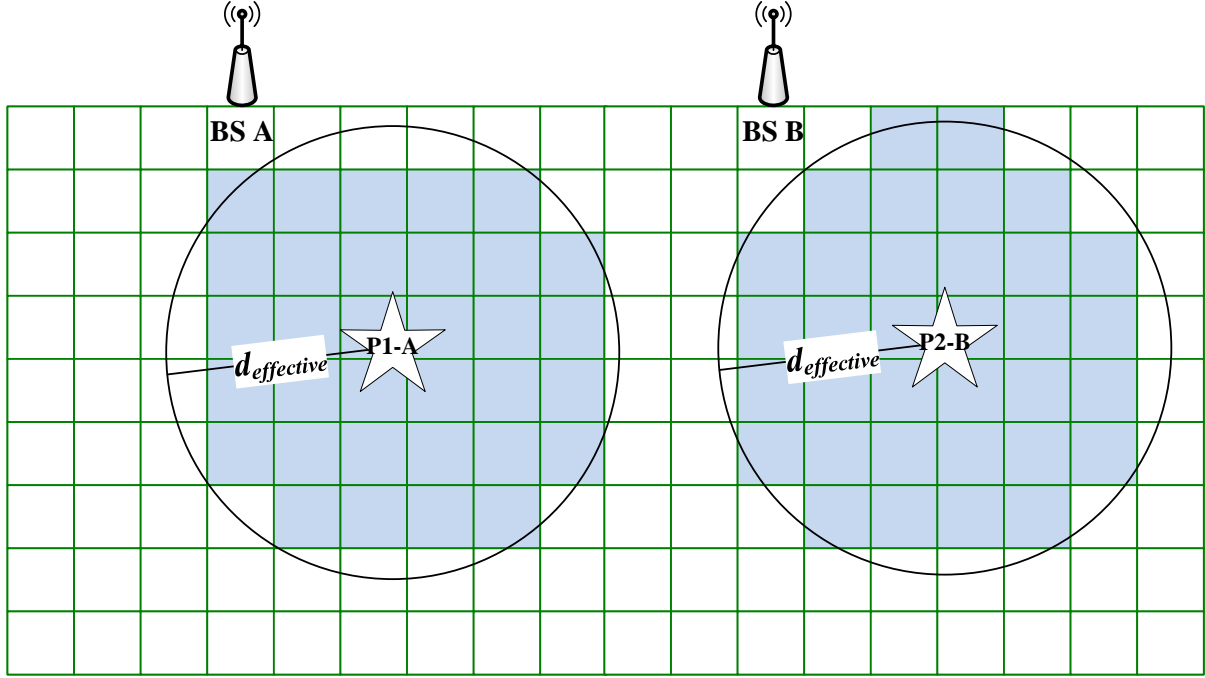


Figure 2.5 The grids affected by a new blockage point are filled with color

2.4 Simulations

In this section, we present the simulation results of the proposed algorithm and another state-of-the-art algorithm for solving the user association problem in mmW networks.

2.4.1 Simulation Parameters and Metrics

The tool used in this paper is iNet framework 2.6 for simulating the node mobility on top of the OMNeT++ 4.6 framework to control the whole simulation process.

The simulation area is a rectangle of 100×1000 square meters. There are 100 BSs, 250 blocking objects, and 200 UEs roaming around the simulation area similar to the map depicted in Figure 2.2. As shown in Figure 2.2, the blocking objects considered in the simulations are human body blockage and vehicles moving in the street since the user association algorithm can connect a UE to a BS on the other side of the street. Therefore, we need to account for the blocking objects in the street (i.e., vehicles). We consider the mobile nodes including nodes carrying UEs and blocking objects to be humans plus the vehicles that we assume to have the same amount of signal absorption

as human beings. The communication range of each BS and UE is set to 100 meters. The height of the post that each BS is installed h is set to 3 meters. Parameter W in Equation (2.5) is set to 1200 MHz [18].

The algorithm detects the blockage incidents based on the experiments done in [7] as stated in Section 2.3.1. Considering each event happens on the scale of 0.1 seconds, each blockage detection window takes three events. The BD records the RSSI in this window to detect if there has been a blockage or not. There is a module in the simulator that knows the location of the objects moving in the area. The module is responsible for collecting blockage data and computing the blockage scores.

According to [14], the path loss model is different for NLOS and LOS paths. The authors of [14] also have used Nakagami distribution for small-scale fading. We experiment with six user association algorithms, including original and modified versions. The first one is our proposed algorithm of computing the blockage scores for each link. We call this algorithm Blockage Point. The second one is the modification that we make to the Blockage Point by using grids and anchor points. We call the algorithm Blockage Point Grid. The third algorithm [18] addresses the problem of user association with the purpose of load balancing. We call this algorithm Load Balancing. The fourth one is our modification to Load Balancing to see how our proposed algorithm can affect the performance of other user association algorithms. The fifth algorithm associates a user to the closest BS. We call this Closest BS. The last algorithm does the association based on maximum SINR. The user is associated to a BS with maximum SINR. We call this Max SINR.

We use our heuristic of blockage scores in computing the rate and consequently, the utility that the algorithm uses in the Load Balancing. In this way, when the blockage score of a link is high, the SINR and consequently the rate of the link decreases. According to the definition in [18] of utility, the utility is equal to the division of demand by rate. When the rate decreases, the utility increases. Therefore, the change of the UE to be associated to the BS decreases because of the load balancing target. We call this algorithm Load Balancing with Heuristic.

There are four metrics defined in this paper to size the performance of the proposed algorithms

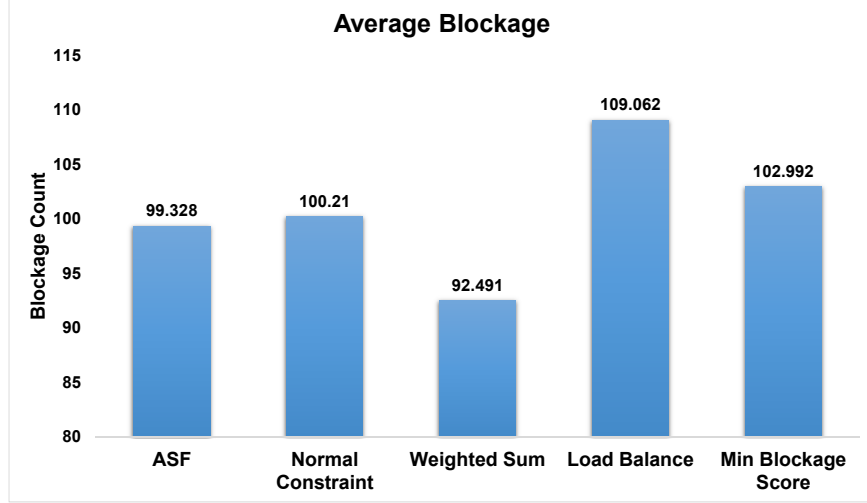


Figure 2.6 Average Blockage per blocking object

and compare the performance with the Load Balancing. We provide the metrics in the following.

Average Blockage per blocking object is the total number of blockages that occurred in the network divided by the number of blocking objects. The average Handover per UE is total handovers in the network divided by the number of UEs. We implement both Rate and SINR according to formulations defined in [14]. According to [18], we compute the link utility, and the algorithm records the maximum of that for a link at each time as Maximum Link Utility.

There are two other metrics defined, namely False Positive Ratio and False Negative Ratio. In order to define these two metrics, we need to define the optimal behavior in this context. UE_1 is associated to BS_A if there is a blockage between UE_1 and BS_A , UE_1 is handed over to BS_B . If there is no blockage in BS_B for UE_1 , we say the handover is optimal. If UE_1 has no blockage with BS_A , it is not handed over to BS_B . If UE_1 has blockage with BS_A and BS_B it stays in BS_A . Positive and negative is defined as if the UE is handed over to another BS or not, respectively.

2.4.2 Simulation Results and Plots

In this section, we discuss the simulation results of the algorithms and different metrics for comparing them. In the first part, the Δ function used is exponential as in (2.8b).

Figure 2.6 shows the Average Blockage per blocking object for different algorithms. As we can

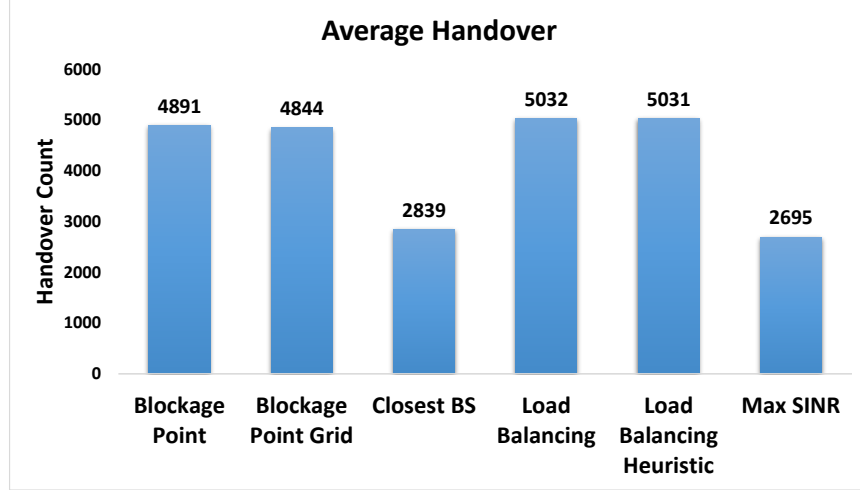


Figure 2.7 Average Handover per UE

see in Figure 2.6, our proposed algorithm, Blockage Point, and its modification Blockage Point Grid have considerably reduced the number of blockages in the network compared to the Load Balancing, Closest BS, and Max SINR algorithms. The reduction is at least 21%. They also outperform Closest BS and Max SINR algorithms. The difference between Blockage Point and Blockage Point Grid is not much, and Blockage Point Grid shows a little merit over Blockage Point. Therefore, modifying the Blockage Point to Blockage Point Grid can be a good compromise in the time complexity-accuracy trade-off. As we can see, both Load Balancing algorithms show more blockages experienced by UEs compared to our algorithm. One primary reason for that is the focus of Load Balancing on balancing UE load among all BSs in range, hence the name. Figure 2.6 also shows that implementing our blockage score system in other optimization problems may not lead to better results.

In Figure 2.7, we plot the Average Handover per UE. As we can see, handover count for Load Balancing and Load Balancing with Heuristic is more than of Blockage Point and Blockage Point Grid. The difference is not much, however. The Closest BS and Max SINR algorithms have the lowest handover counts almost half of others.

We plot False Positive Ratio and False Negative Ratio metrics in Figures 2.8 and 2.9, respectively. As we show in Figure 2.8, False Positive Ratio for Blockage Point and Blockage Point Grid is less

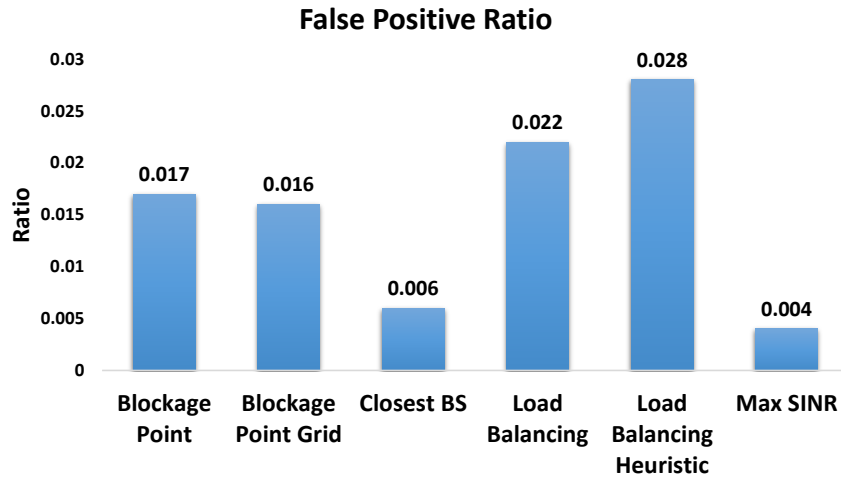


Figure 2.8 False Positive Ratio

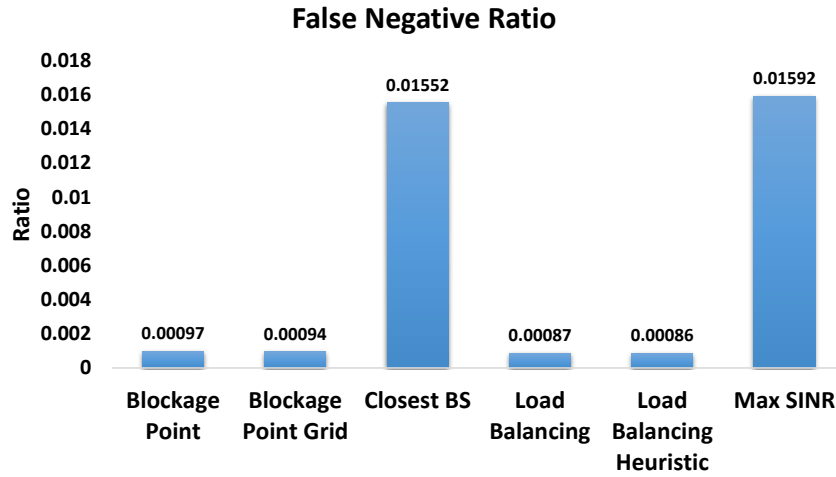


Figure 2.9 False Negative Ratio

than Load Balancing algorithms. Closest BS and Max SINR have the least values, and the reason is the way False Positive Ratio is defined. For the False Negative ratio, all algorithms except Closest BS and Max SINR are very close to zero. We can also find the reason for the higher False Negative Ratio in those two algorithms in the metric definition. Load Balancing algorithms perform better in terms of False Negative Ratio than our proposed algorithms by a minimal margin.

Figure 2.10 compares the maximum link utility in the network for Blockage point and Load Balancing. As we can see, Blockage Point has its 100% link utility reached immediately after starting the simulation, but the Load Balancing maintains the link utility between 40% and 70%

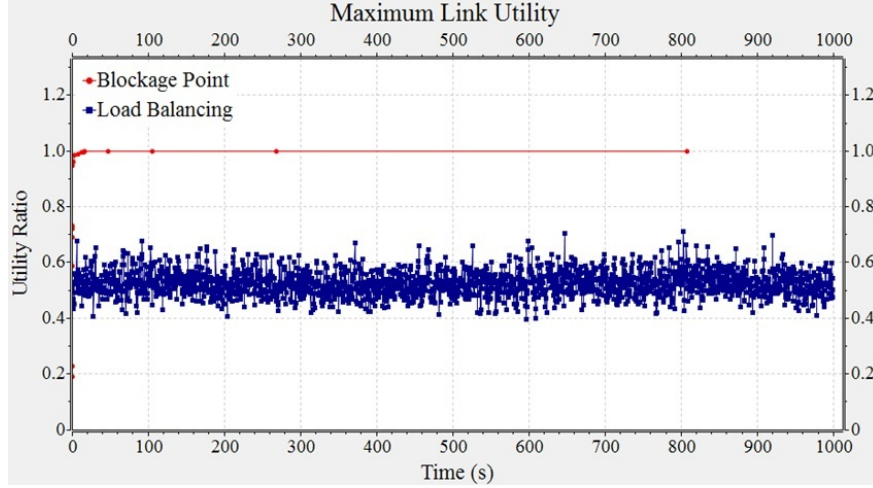


Figure 2.10 Maximum Link Utility

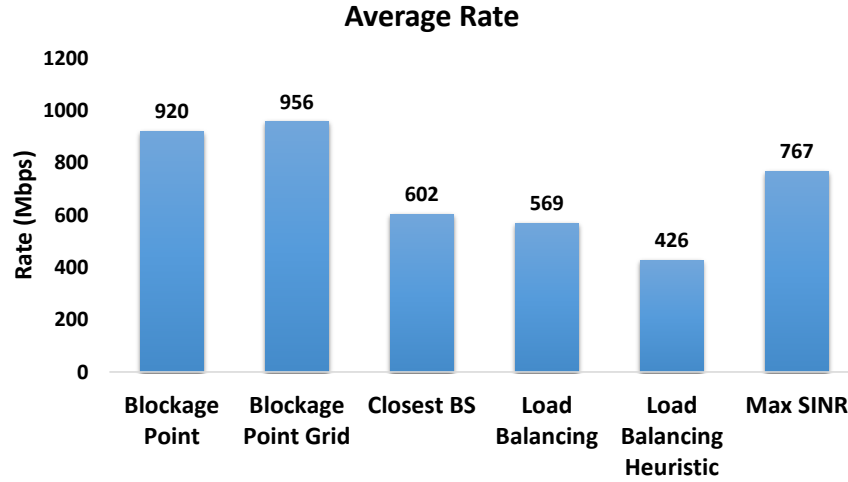


Figure 2.11 The Average Rate

for all the time. We expect this behavior from Load Balancing as it is specially designed for the purpose.

Figure 2.11 compares average rate of all algorithms. It is evident that in Blockage Point, the number of links with higher rates is more than the Load Balancing. As we expect from the blockage count figure, the rate of our proposed algorithms outperform other algorithms. The Load Balancing algorithms have worse performance than Closest BS and Max SINR algorithms. We expect these results from the average blockage counts.

We also study the effect of choosing different Δ functions (defined in Section 2.3.2) and the

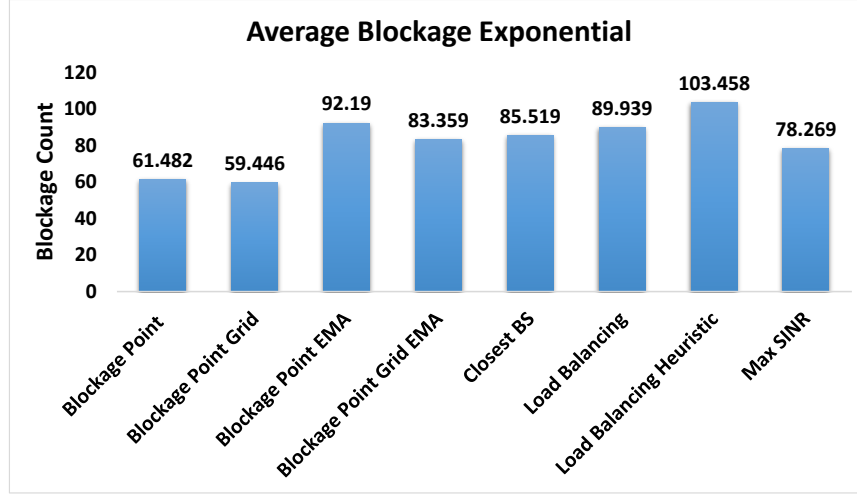


Figure 2.12 Average Blockage per UE for Exponential Delta Function

inclusion of timeliness on the performance of our proposed algorithm.

So far, in this chapter, there was no mechanism to deal with the aging of blockage data. That is, older blockage incidents are dealt with as new ones. For having an aging process of blockage data, we use the Exponential Moving Average (EMA) for blockage data to emphasize newer blockage data.

Figure 2.12 shows the average blockage for different algorithms when the algorithm uses exponential Δ function. As we can see in Figure 2.12, except for Blockage Point EMA, our algorithms still outperform all other algorithms except Max SINR. The Blockage Point Grid EMA has a slightly better performance than Blockage Point EMA. Compared to the case without having EMA, more blockage count at about 33% increase for Blockage Point. The Blockage Point Grid shows a 28% increase when using the EMA algorithm to add timeliness to the algorithm. It seems that, unlike Blockage Point Grid, Blockage Point is more sensitive to have older data with the same weight as the newer data. Therefore, we can say that Blockage Point Grid has more timeliness than Blockage Point when the Δ function is exponential.

Figure 2.13 shows the Average Blockage per blocking object for different algorithms with inverted function as Δ function. As we show in Figure 2.13, when the algorithm uses the inverted distance function in (2.8a) as Δ function instead of using exponential distance function in (2.8b),

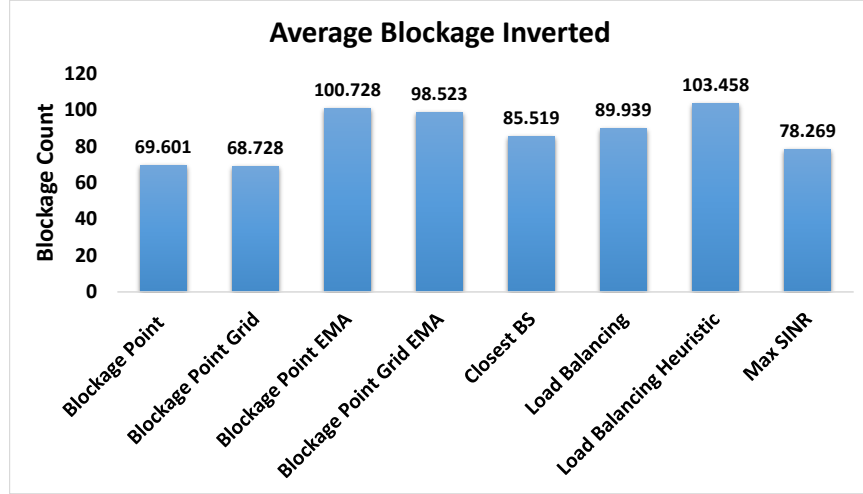


Figure 2.13 Average Blockage per UE for Inverted Delta Function

it worsens the average blockage count. Both EMA modifications to our algorithms have worse performance than other algorithms. In the case of the Load Balancing Heuristic, both EMA algorithms perform better. Comparing our algorithms in two Figures 2.12 and 2.13 shows that exponential function is a better choice for Δ function.

2.5 Summary

In this chapter, we propose a user association method especially designed for mmW networks. The method considers the inherent propagation characteristics of the mmW signal. We use the blockage history to associate a UE to a BS that has a lower chance of blockage. We show that our proposed method of user association can perform well in terms of decreasing the blockages and improving the other important network metrics. We compare the performance of our proposed method with other user association methods. The results show that our proposed method outperforms the compared methods in different metrics.

CHAPTER 3

BIOBJECTIVE APPROACH FOR OPTIMIZING BLOCKAGE AND LOAD BALANCING OF THE NETWORK

In the previous chapter, we discuss the problem of blockage in mmW networks and present the solution to this problem for user association. If the user association method only considers the blockage, some BSs would become more favorable than others and it may make the network unbalanced with degraded performance. In this chapter, we consider both blockage reduction and load balance in the user association algorithm design and formulate a biobjective combinatorial optimization problem to optimize user association in mmW cellular networks according to the two objectives. We then transform the biobjective problem into a single objective problem using three scalarization methods. For all of the scalarization methods, we solve both primal and dual problems. The challenge of the problem is formulating the optimization problem as a single objective problem from the biobjective problem and capturing the true relation between two objectives. The objectives are discrepant in the way their parameters are collected and influence the performance of the network. The involvement of each of the objectives can be adjusted when needed. In this way, the solver can prioritize an objective over the other.

The main points in this chapter include defining an optimization problem and providing the solution to the optimization problem. The solution to the optimization problem will give an opportunity to the network to adjust its performance policy via user association with respect to the objective functions. We provide a subgradient solution to different single objective problems of the biobjective problem, which has polynomial time complexity compared to the primal solver that has exponential time complexity.

We summarize the contributions as follows:

1. We define a biobjective optimization problem to improve load balance and minimize the blockage score in mmW networks.

2. We develop the solution for the optimization problem that is generated by each scalarization method with polynomial time complexity.
3. We provide the proof of the upper bound for the duality gap between the dual and primal solutions.

We organize the rest of this chapter as follows. Section 3.1 reviews the literature on the topic of user association in wireless networks. Section 3.2 describes the system model and different parameters. We formulate the optimization problems in Section 3.3. The proof to duality gaps provided in the chapter comes in Section 3.4. We present the performance of the solutions provided in this chapter in Section 3.5.

3.1 Related Work

This section discusses the literature of user association problems in wireless networks on joint optimization/improvement domains. Liu et al. [8] have proposed a taxonomy for user association methods in wireless networks. According to the methodology used in that paper, we review some of the related works.

The user association can be done to improve or optimize the outage/coverage probability. Dhillon et al. [11] have proposed a framework for computing the coverage probability in multi-tier HetNets. Cheung et al. [23] have introduced a model to compute success probability in each tier, considering different spectrum allocation policies.

The authors used the combinatorial approach for user association in [24]. The authors formulated a joint optimization problem of user association and channel allocation decision between macrocells and small cells in a HetNet. As a follow-up, Ghimire et al. [25] have proposed a framework to analyze the performance of HetNets by optimizing resource allocation, transmission coordination, and user association to maximize data rate throughput. The authors discuss joint optimization of base station sleep mode, user association, and subcarrier allocation to maximize total power consumption in [26].

In the mmW spectrum area, Xu et al. [17] have discussed the problem of joint optimization of user association and relaying traffic to other clients. The objective is to maximize the total network throughput. In [18], the authors defined an optimization problem to minimize the maximum load across all base stations to optimize user association. Sakaguchi et al. [27] have proposed a user association method in mmW networks that considers the supported available rate and number of users in each cell. For energy harvesting networks, [28] have formulated an optimization problem to maximize network utility while each base station's energy consumption does not exceed the harvested amount. Alizadeh et al. [29] have proposed a method for load balancing with consideration of blockage in millimeter-wave networks. However, they do not consider the blockage independently and detect it as rate degradation instead. This assumption leads to false blockage detection because the mean movement is also a source of rate degradation. Semiari et al. [30] have proposed one-to-many matching solutions for the association problem in hybrid millimeter-wave and microwave networks. They also do not consider the blockage as an independent objective. They do not consider the joint optimization of load and blockage in millimeter-wave networks.

There are many other user association methods in Heterogeneous Networks (HetNet) [31, 32, 33] that consider user association with having different performance metrics in mind. Authors in [34, 35] also studied Energy-efficient user association.

The blockage mitigation and optimizing blockage in mmW networks have been studied to some extent. Bai et al. [14] have talked about the blockage analysis in the mmW network. They provided a model considering static blockages in the network. However, they did not study the effect of human blockage more dynamically. Some papers discuss modeling the human body as a blocking object to mmW networks [36, 37]. Although they discuss the blockage in the mmW spectrum in its dynamic format, they only provide a model in terms of signal outcome in the blocking object's presence. They do not consider the load balancing of the network nor give a solution to avoid blockage in the future. Choi [38] has proposed a model on blockage probability in a network with multiple BSs. The authors used the diversity to mitigate the effect of blockage on the performance of the network. They provided a closed-form expression for the outage probability concerning the

transmission power. However, they do not provide a method to associate users to the BSs and solve the association problem. They also do not discuss the blockage with other critical objectives. They mention the power, but the power is not a parameter that can be changed much neither on BS nor UE.

For the choice of blockage model, there are different schemes to be considered. Niknam et al. [39] have proposed a geometrical blockage model in mmW networks. They use geometry to calculate a shadow region created by a blockage. In 3D space, Han et al. [40] have proposed a model to capture the blockage by a human body. They involve both dynamic human body blockage and static blockage from the objects found around the human body. Blockage exists in terahertz communications and the authors quantized the blockage based on the probability of having line-of-sight [41] and the authors investigated the effect of blockage on quality of experiment for wireless virtual reality in [42]. Although the papers mentioned propose a blockage model in high-frequency communication, the solutions in these papers are not able to address the problem that this chapter investigates, which is user association. The existing work utilizes geometric principles for estimating the shadow region of the blockage, but they lack a quantitative measure that we need to identify the next best base station for a blocked mobile user to associate.

All of the discussed literature share the fact that they try to optimize one or more variables based on single objectives. The factor that distinguishes our work from the ones discussed here is that in our work, we want to optimize user association in mmW networks considering two objectives at the same time. Since the input variables are binary, the problem we are considering is a biobjective combinatorial optimization problem. To the best of our knowledge, the authors did not consider the joint consideration of load balancing and blockage avoidance in the same optimization problem in any of their work.

3.2 System Model and Objective Functions

In this section, we discuss the specifics of the system that we consider for our proposed method. Then, we introduce the two objective functions, load balance and blockage score. We use the

objective functions to optimize the user association variable. We also briefly introduce three scalarization methods used to transform the 2-dimensional objective vector to a scalar.

3.2.1 System Model

In our system, we have N BSs, M UE, and P blocking objects. The set of BSs is defined as $\mathcal{N} = \{1, 2, \dots, N\}$ and we define the set of the UEs as $\mathcal{M} = \{1, 2, \dots, M\}$. The value of parameter P is not known in the problem formulation and consequently in the solution. The reason is that the network does not know the value for P . Therefore, the network should be able to work without knowing P . It can be variable as the blocking objects are on the move so that they can come and go. The solution is adaptable to the number of blockages that are relevant to P .

We define M_i as the set of UEs that are served by BS_i . Analogously, we define N_j as the set of all BSs in the UE j 's communication range. We assume the communication range of UE and BS to be equal to R meters. The communication range is a cut-off value for signal propagation. The signal is negligible when it travels farther than R . However, the signal propagation and its quality in different locations are a function of many factors. We use a model for signal propagation [14] that the authors designed for mmW signal propagation.

We assume each BS to be mounted on a post with h meters from the ground. Each BS_i is equipped with a_i number of antennas meaning the BS_i can serve up to a_i UEs at a given time. The channels of each beam of these antennas are different from other antennas in a BS. We assume the antennas to be able to rotate all 360° degrees or in the interval, $[-\pi, +\pi]$ in radian.

3.2.2 Objective Functions

We want to optimize user association based on two objectives. The first one is the *Load Balance* (t) objective, which we define as the maximum total utility of one base station across the BS $i \in \mathcal{N}$. The second objective is *Blockage Score* (B) that we define as the total blockage score of all links in the network. In the following two subsections, we discuss these two objective functions in more detail.

3.2.2.1 Load Balance

For the first objective, we use the load balancing optimization problem definition and solution provided by Athanasiou et al [18]. The authors have defined the channel utilization between BS i and UE j as $\beta_{ij} = \frac{Q_j}{R_{ij}}$ where R_{ij} is the rate of the link between BS i and UE j and Q_j is the demanded data rate of UE j . We define the rate as,

$$R_{i,j} = W \log_2(1 + \text{SINR}_{i,j}) \quad (3.1)$$

In (3.1), W is the bandwidth.

$$\text{SINR}_{i,j} = \frac{h D_{ij} L(F_{ij})}{\sigma^2 + \sum_{k \in \{\mathcal{N}_j - i\}} h D_{kj} L(F_{kj})} \quad (3.2)$$

The SINR [14] works with the loss due to LOS and Non-Line NLOS signals. L is the loss based on these two types of signal. F_{ij} is the distance between BS i and UE j . h is small scale fading, σ^2 is thermal noise, and, D_{ij} is the directivity gain between BS i and UE j . The index k goes through the interfering BSs. These BSs are the ones that are in the communication range of UE j but UE j is not connected to them.

We define the user association variable that we want to optimize as,

$$x_{ij} = \begin{cases} 1, & \text{if UE } j \text{ associated to BS } i \\ 0, & \text{Otherwise} \end{cases} \quad (3.3)$$

We formulate the combinatorial optimization problem for optimizing user association variable x as,

$$\text{minimize } t \quad (3.4a)$$

$$\text{subject to } \sum_{j \in \mathcal{M}_i} \beta_{ij} x_{ij} \leq t, \quad \forall i \in \mathcal{N} \quad (3.4b)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.4c)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.4d)$$

In Equation (3.4), constraint (3.4a) makes sure that the maximum utility of a BS does not exceed variable t . Therefore, if the solver minimizes t , it minimizes the maximum BS utility as well. Constraints (3.4c) and (3.4d) make sure that the solution to this optimization problem is feasible i.e., UE_j is associated with only one BS and a UE is associated with a BS or not with no in-between states.

The authors of [18] have used Lagrangian dual analysis to solve optimization problem in (3.4). The Lagrangian of optimization problem (3.4) is,

$$L(t, \mathbf{x}, \boldsymbol{\lambda}) = t \left(1 - \sum_{i \in \mathcal{N}} \lambda_i \right) + \sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}_j} \beta_{ij} \cdot \lambda_i \cdot x_{ij} \quad (3.5)$$

In Equation (3.5), there are N Lagrangian variables, one for each inequality in (3.4b). The Lagrangian dual function is,

$$g(\boldsymbol{\lambda}) = \inf_{\substack{t \in \mathbb{R} \\ \mathbf{x} \in \mathcal{X}}} L(t, \mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} \sum_{j \in \mathcal{M}} g_j(\boldsymbol{\lambda}), & \sum_{i \in \mathcal{N}} \lambda_i = 1 \\ -\infty, & \text{Otherwise} \end{cases} \quad (3.6)$$

In Equation (3.6), \mathcal{X} is a set of x_j vectors for all $j \in \mathcal{M}$ that conform constraints (3.4c) and (3.4d). $g_j(\boldsymbol{\lambda})$ is,

$$\text{minimize } \sum_{i \in \mathcal{N}_j} \beta_{ij} \cdot \lambda_i \cdot x_{ij} \quad (3.7)$$

Subject to $x_j \in \mathcal{X}_j$

The Lagrangian dual problem is,

$$\text{maximize } g(\boldsymbol{\lambda}) = \sum_{j \in \mathcal{M}} g_j(\boldsymbol{\lambda}) \quad (3.8a)$$

$$\text{subject to } \sum_{i \in \mathcal{N}} \lambda_i = 1 \quad (3.8b)$$

$$\lambda_i \geq 0, \quad \forall i \in \mathcal{N} \quad (3.8c)$$

The algorithm then solves the optimization problem (3.8) via the iterative subgradient method. The algorithm chooses the best solution for optimization problem (3.4) from solutions resulting from each iteration of the subgradient method. Throughout this chapter, we refer to the load balance objective as the first objective.

3.2.2.2 Blockage Score

For the second objective, we use the idea introduced in Chapter 2. We use the blockage model to capture the randomness that is inherent in mobile networks. The randomness is especially significant in outdoor mobile networks, which is our proposed user association model's primary target. The blockage score model we use has the user association structure, which is an integer optimization problem. This structure fits into our proposed biobjective optimization design. The blockage score model conveniently provides our model with a measure of how probable a blockage is by capturing the history of past blockage incidents in the non-deterministic and dynamic outdoor environment. While this blockage model uses geometric principles to calculate the shadow region of a blocking object like the other blockage models, we also design it to optimize user association which its counterparts are not capable of doing. The other blockage models that we survey in Section 3.1 provide the probability measure for the blockage. However, they do not use the model to enhance user association. We design the model in a way to reduce the blockage through user association.

The algorithm computes the blockage score objective for the whole network. The algorithm generates the objective value by adding the blockage score for all the links in the network. This includes the links that exist between the UEs that are associated with the BSs. As discussed in Chapter 2, BPC computes the blockage score of one link between a UE and a BS based on the previous blockage incidences that occurred in the UE neighborhood in past. The BPC computes the blockage score using the distance function from the UE location and the history of the UEs' locations that had blockages in the past. The blockage score is a temporal and spatial parameter. It is spatial in the sense that it contains the blockage report of the previous blockage incidents in the area in which a UE is trying to connect to a BS. It is temporal because the value of the blockage score ages. They use an Exponential Moving Average (EMA) on the blockage score to decrease over time if there are no new blockages reported recently. The solutions that will be developed for the optimization problem assume that the blockage score objective is constant while the solution forms and implements the user association matrix. The solution's distributed nature makes this

assumption possible since UE performs each user association. We completely implement the time variability of the blockage score objective for the evaluation of our solution.

In the provided model, we do not use the number of the blocking objects (P). In reality, the network and the solvers do not know the value of P . Therefore, the potential solution should be independent of P , so that the solution is practical in a real system. What matters to the solvers is the historical blockage data that BPC gathers from across network in the form of the blockage score. P affects the historical blockage data.

3.2.3 Scalarization Methods

In a biobjective optimization problem, we call a solution Pareto-optimal or Pareto-efficient if there exists no other feasible solution that dominates it. A solution A dominates other solution B if it is the same or better in one objective and strictly better in the other objective. By solving the problem, we get a Pareto-front (PF) or trade-off surface (concave or convex) of optimal values. Any preference-based multi-objective optimization problem can be solved with either apriori or a-posteriori method. Among apriori methods, researchers widely use utility-based method, lexicographic method, goal programming [43]. Since we don't assume any prior preference of objectives, we only discuss a-posteriori methods. Posteriori methods find a good distribution of Pareto-optimal solutions and give them to the decision-maker (DM) to pick one. Solvers generate Pareto-efficient solutions one by one in an iterative manner. Literature proposes various scalarization methods to combine multiple objectives in such a way that it targets specific optimal solutions of the front in each subproblem. From literature, some of the well-known scalarization methods are Achievement Scalarization Function (ASF) [44], Normal Constraint (NC) [45] and Weighted Sum (WS) [46].

WS approach combines two objectives into a single one by giving weights to the objectives. Then we use a single objective optimizer to solve the converted single objective problem. WS approach has the drawback of finding only the convex part of the whole Pareto-front. For example, in Figure 3.1, the WS approach could only find one of the two extreme points (A or B). But it can be very efficient in finding the convex part of PF. We give the formulation of WS below. Here m is

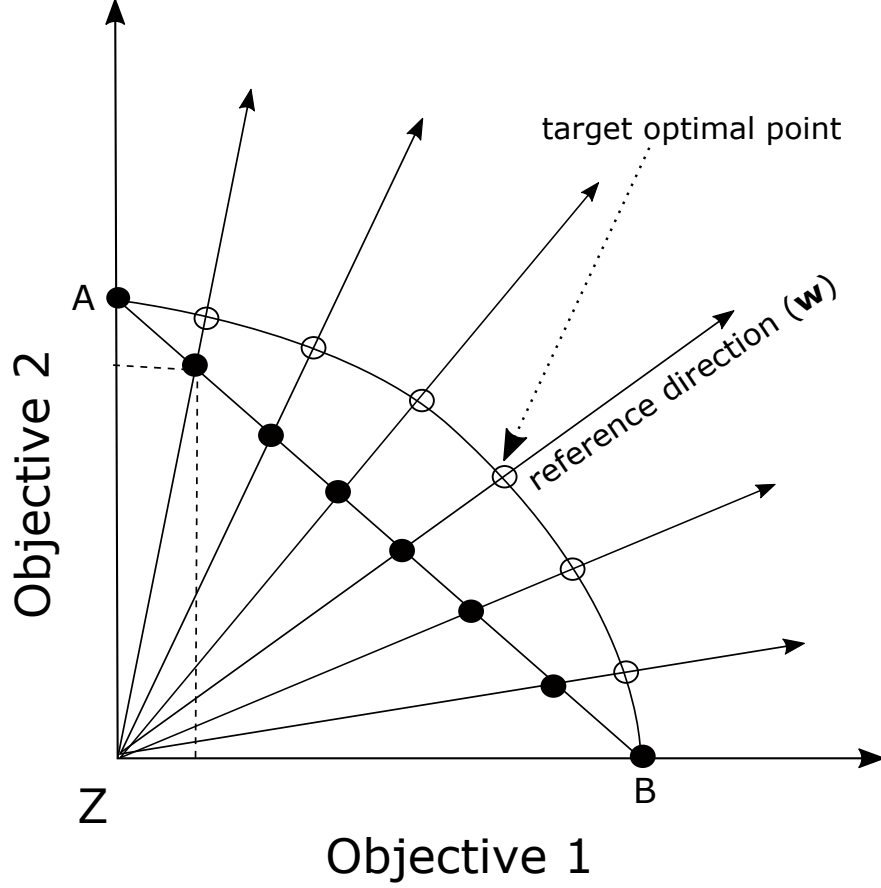


Figure 3.1 Optimal point targeting for corresponding reference direction

the number of objectives, \mathbf{w} is the weight vector, $\mathbf{F} = (f_1, \dots, f_m)$ is objective function.

$$WS(\mathbf{x}) = \sum_{i=1}^m w_i \times f_i(\mathbf{x}) \quad (3.9)$$

In Achievement Scalarizing Function (ASF) method[47], the problem gives one ideal minimum point \mathbf{z} and a set of well-distributed reference directions (see Figure 3.1). Unlike WS, the reference directions \mathbf{w} can easily specify the target optimal point of PF. We optimize the following equation.

$$ASF(\mathbf{x}) = \max_{i=1}^m \frac{f_i(\mathbf{x}) - z_i}{w_i} \quad (3.10)$$

The normal constraint method (NC)[45] optimizes two objectives separately as a single objective optimization to get utopian points $\bar{\mu}_1$ and $\bar{\mu}_2$. The solver uses these points to normalize the objectives. The solver uses converted objective values $\tilde{\mathbf{F}} = (\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_m)$ in the following equation. It uses parallel reference vectors to generate subproblems. After solving each subproblem,

it makes the corresponding region infeasible.

$$\begin{aligned} \text{NC}(\mathbf{x}) &= \max_{i=1}^m \left(\tilde{f}_i(\mathbf{x}) - (w_i - \frac{1}{m}) \right) \\ \text{such that } \mathbf{v}_i \cdot (\tilde{\mathbf{F}}(\mathbf{x}) - \mathbf{w}) &\leq 0, \quad \forall i \in \{1, 2, \dots, m\} \\ \text{where } \mathbf{v}_i &= \bar{\mu}_j - \bar{\mu}_i, \quad i \neq j \end{aligned} \quad (3.11)$$

We use more than one scalarization method to see how different scalarization methods perform in solving the defined problem. Since different methods have different approaches to finding the Pareto front in optimization problems, we are interested to see the performance of different scalarization methods in solving the problem. We see the different performance of the scalarization methods in Figure 3.7.

The intuition behind choosing the mentioned scalarization methods comes from the nature of the biobjective optimization problem. The defined optimization problem is a combinatorial type and needs special methods for converting an objective vector to a scalar. The NC method generates evenly spaced solutions on a Pareto frontier [45]. Due to the solution space's unknown characteristics in our proposed optimization problem, we conclude that using the NC scalarization method helps us find an optimal solution due to the method's unbiased search strategy. On the other hand, our optimization problem has a clear ideal objective value vector that can be useful for searching for the solution set using the ASF method as this method is sensitive to the ideal point selection [48]. The WS method does not change the feasible set and boils down to the single objective version of the multi-objective problem [49]. This property of the WS scalarization method is beneficial for combinatorial problems.

To generate the weight vectors, we have used Das and Dennis method[50]. This method generates equally spaced weight vectors in any number of dimensions. Note that, the scalarization methods can potentially generate non-Pareto-optimal solutions if there exists no optimal solution with the corresponding weight or reference direction[50]. Thus we need to use a Pareto filter to remove the dominated solutions from the set.

3.3 Biobjective Optimization Problem

In this section, we first formulate the biobjective optimization problem. We formulate the single objective optimization problem of biobjective optimization problem (3.12) by the scalarization methods introduced in Section 3.2.3. Then, we solve the transformed single objective problem by appropriate methods. We borrow the general approach used in [18] to solve the single-objective problem generated by the three scalarization methods via dual analysis.

To optimize user association based on both load balancing and blockage score objectives, we formulate the optimization problem as,

$$\text{minimize } f(t, B) \quad (3.12a)$$

$$\text{subject to } \sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t, \quad \forall i \in \mathcal{N} \quad (3.12b)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.12c)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.12d)$$

In Equation (3.12), function f is a general function that converts the vector of objectives to one scalar for optimization, and the objective B is the total blockage score in the network. The problem formulated in (3.12) is a Biobjective Combinatorial Optimization (BOCO) problem. The user association problem is combinatorial. The algorithms for solving these problems have exponential time complexity [18]. We transform the problem (3.12) into single objective problem by using three scalarization methods introduced in Section 3.2.3. Then we solve the resulting primal single-objective problems. We also provide the Lagrangian dual solutions to the single objective problems with lower time complexity than the exponential primal solution.

The constraint (3.12c) says that all UEs must connect to exactly one BS. It may seem like a strong constraint. However, if a solution works for $\sum_{i \in \mathcal{N}_j} x_{ij} = 1$, it will work for $\sum_{i \in \mathcal{N}_j} x_{ij} \leq 1$. Since the solution works on each UE independently, it is possible to define a threshold for the signal received by the UE. If it falls below that threshold, the algorithm does not perform the association for that UE.

The purpose of defining this optimization problem is to optimize (minimize) both blockage probability and maximum load on the BSs in the network. We pursue this purpose by selecting BSs for different UEs based on the objective functions. In other words, solving this optimization problem makes the network more balanced while decreasing the blockages that can happen in the network. The blockage score objective B captures the blockage probability of a UE if it is associated with a certain BS. We would like to select a BS for a UE that minimizes such probability. Since it is a joint optimization problem, the solution selects a BS that decreases both blockage probability and maximum BS load at the same time.

3.3.1 Formulation and Solution of ASF Scalarization Method

This section will formulate the single-objective optimization problem of the biobjective problem defined in (3.12) generated by the ASF scalarization method. Then, we use dual analysis to solve the single-objective optimization problem for the ASF scalarization method.

3.3.1.1 ASF Generated Single Objective Problem

ASF method needs an ideal minimum point z . The ideal point in our defined problem (3.12) is $z = (0, 0)$ for objective vector (t, B) . Since both objectives of maximum BS utilization and total blockage score cannot get negative values, thus assuming the ideal point as zero for both objectives is logical. The ideal point may or may not be reachable. We formulate the ASF generated single objective problem as,

$$\begin{aligned}
& \text{minimize} \quad \max\left(\frac{t}{w_0}, \frac{B}{w_1}\right) \\
& \text{subject to} \quad \sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t, & \forall i \in \mathcal{N} \\
& \quad \quad \quad \sum_{i \in \mathcal{N}_j} x_{ij} = 1, & \forall j \in \mathcal{M} \\
& \quad \quad \quad x_{ij} \in \{0, 1\}, & \forall j \in \mathcal{M}, i \in \mathcal{N}_j
\end{aligned} \tag{3.13}$$

In problem (3.13), B is total blockage score, $\mathbf{w} = (w_0, w_1)$ is the weight vector. Since there are two objectives the size of weight vector is 2. We can reformulate the problem in (3.13) to linear as authors in [18] did to formulate min-max problem. We can reformulate the problem as,

$$\text{minimize } S \quad (3.14a)$$

$$\text{subject to } \frac{t}{w_0} \leq S \quad (3.14b)$$

$$\frac{B}{w_1} \leq S \quad (3.14c)$$

$$\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t \quad \forall i \in \mathcal{N} \quad (3.14d)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.14e)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.14f)$$

In problem (3.14), is a linear combinatorial optimization problem. The objective is the maximum of two terms of original ASF formulation. We solve the primal problem in (3.14) by a linear program solver to get to its optimum solution.

3.3.1.2 Solution to ASF Generated Single Objective Problem via Dual Analysis

In this section, we elaborate on a solution to ASF that generated a single optimization problem formulated in Equation (3.14). The problem in (3.14) is a mixed-integer linear program (MILP) and its complexity is proved to be NP-hard [18]. To have a distributed low complexity method of solving the biobjective problem defined in Equation (3.12), we discuss the Lagrangian dual analysis of a single objective optimization problem generated by the ASF scalarization method.

First, we eliminate variable t from Equation (3.14) and replacing B with its equivalent. In order to do that, we can remove the constraint (3.14b) by using it in constraint (3.14d) and replacing variable B with its value defined in [51], making the resulting optimization problem as,

$$\text{minimize } S \quad (3.15a)$$

$$\text{subject to } \frac{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} \gamma_{ij} \cdot x_{ij}}{w_1} \leq S \quad (3.15b)$$

$$\frac{\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij}}{w_0} \leq S \quad \forall i \in \mathcal{N} \quad (3.15c)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.15d)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.15e)$$

We solve Equation (3.15) as the optimization problem for ASF scalarization method.

In order to obtain Lagrangian of an optimization problem, the objective function and the constraints of the primal problem go together in another optimization problem. Then, we derive the Lagrangian dual by computing infimum of Lagrangian with respect to all variables other than Lagrangian multipliers. We formulate the Lagrangian of problem in Equation (3.15) using only objective function (3.15a) and first two constraints (3.15b) and (3.15c) as,

$$L(S, \mathbf{x}, \boldsymbol{\lambda}) = S \cdot \left(1 - \lambda_{N+1} - \sum_{i \in \mathcal{N}} \lambda_i\right) + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} x_{ij} \cdot \left(\frac{\gamma_{ij}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{w_0} \cdot \lambda_i\right) \quad (3.16)$$

In Equation (3.16), we gather the Lagrangian multipliers in vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{N+1})$. There are N Lagrangian multipliers, one for each constraint in (3.15c) and another one for constraint (3.15b). We also swap the index of nested sum operators according to equivalence $\{(i, j) \mid i \in \mathcal{N}, j \in \mathcal{M}_i\} \equiv \{(n, m) \mid m \in \mathcal{M}, n \in \mathcal{N}_m\}$ used in [18]. We also borrowed from [18] the definition of a vector space that input \mathbf{x} satisfies constraints (3.15d) and (3.15e). The vector space \mathcal{X} is Cartesian product of $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_M$. We define each \mathcal{X}_j as,

$$\mathcal{X}_j = \{\mathbf{x}_j = (x_{ij})_{i \in \mathcal{N}_j} \mid \sum_{i \in \mathcal{N}_j} x_{ij} = 1, x_{ij} \in \{0, 1\}, i \in \mathcal{N}_j\} \quad (3.17)$$

According to Equation (3.17), each \mathcal{X}_j is j^{th} column in matrix \mathbf{x} that has a single one and all other entries are zeros. The single one can be in any N entries of the column j . Now, we are ready to define the Lagrangian dual problem. We define it as,

$$g(\boldsymbol{\lambda}) = \inf_{\substack{S \in \mathbb{R} \\ \mathbf{x} \in \mathcal{X}}} L(S, \mathbf{x}, \boldsymbol{\lambda}) \quad (3.18a)$$

$$= \begin{cases} \inf_{\mathbf{x} \in \mathcal{X}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} x_{ij} \cdot \left(\frac{\gamma_{ij}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{w_0} \cdot \lambda_i\right), & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.18b)$$

$$= \begin{cases} \sum_{j \in \mathcal{M}} \inf_{\mathbf{x}_j \in \mathcal{X}_j} \sum_{i \in \mathcal{N}_j} x_{ij} \cdot \left(\frac{\gamma_{ij}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{w_0} \cdot \lambda_i \right), & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.18c)$$

$$= \begin{cases} \sum_{j \in \mathcal{M}} g_j(\lambda), & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.18d)$$

We formulate the infimum of Lagrangian in Equation (3.16) with respect to two variables S and \mathbf{x} . The infimum of the term containing S with respect to S in Equation (3.16) is $-\infty$. Therefore, it is unbounded. In order to have a bounded S term, we need to make the coefficient of S , zero in Equation (3.16). That is made possible by having coefficient of S equal to zero as condition in dual problem. The vector space we define in (3.17) helps us to deduce Equation (3.18c). In Equation (3.18c), we compute the infimum over all \mathbf{x}_j members of the set \mathcal{X}_j . We formulate the Lagrangian dual problem as,

$$\text{maximize } g(\lambda) = \sum_{j \in \mathcal{M}} x_{ij}^* \cdot \left(\frac{\gamma_{ij}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{w_0} \cdot \lambda_i \right) \quad (3.19a)$$

$$\text{subject to } \sum_{i=1}^{N+1} \lambda_i = 1 \quad (3.19b)$$

$$\lambda_i \geq 0 \quad (3.19c)$$

In Equation (3.19a), we compute x_{ij}^* using the following formula,

$$x_{ij}^* = \begin{cases} 1, & i = \arg \min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{kj}}{w_0} \cdot \lambda_k \right) \\ 0, & \text{otherwise} \end{cases} \quad (3.20)$$

The fact that the infimum in $g_j(\lambda)$ is over the set \mathcal{X}_j results Equation (3.20). Therefore, the entries of x_{ij}^* are all zeros except the i^{th} entry that has minimum $\frac{\gamma_{ij}}{w_1} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{w_0} \cdot \lambda_i$ for UE_j .

The problem in (3.19), is a concave function. It is also non-smooth, leading to the use of the subgradient method because it is a non-differentiable function. Since the dual problem has constraint (3.19b), we use the projected version of the subgradient method. In this version, to ensure

that the solver chooses the variable values from the constraint polygon, we generate a Euclidean projection of the current point on the constraint polygon as the next solution to the optimization problem. There are $N + 1$ variables of λ_i that each will have a subgradient. Therefore, subgradient of $-g$ (since subgradient is for minimization problems) is a vector with $N + 1$ elements. The subgradient of $-g$ in a feasible λ is

$$u_i = \begin{cases} -\sum_{j \in \mathcal{M}_i} \frac{\beta_{ij}}{w_0} x_{ij}^*, & i \in \mathcal{N} \\ -\sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}_j} \frac{\gamma_{ij}}{w_1} x_{ij}^*, & i = N + 1 \end{cases} \quad (3.21)$$

Then we compute the next value of vector λ by,

$$\lambda^{(k+1)} = P(\lambda^{(k)} - \alpha_k \cdot u^{(k)}) \quad (3.22)$$

In Equation (3.22), $P(\lambda^{(k)} - \alpha_k \cdot u^{(k)})$ is Euclidean projection of point $\lambda^{(k)} - \alpha_k \cdot u^{(k)}$ in $N + 1$ dimensional space into simplex defined by $\{\lambda \mid \sum_{i=1}^{N+1} \lambda_i = 1\}$. The parameter α_k is a step size variable in k^{th} iteration.

The subgradient algorithm works iteratively. In each iteration (k), it considers vector $\lambda^{(k-1)}$ and computes a new vector $\lambda^{(k)}$ for current iteration. Then, we use vector $\lambda^{(k)}$ to compute x_{ij}^* via Equation (3.36). Then, we use the value of x_{ij}^* to compute the subgradient vector. This cycle continues for a constant number of times defined by iteration number of subgradient algorithm. As stated in Equation (3.20), the algorithm selects only one i that minimizes the term. This leads to satisfying constraints (3.15d) and (3.15e). The simplex projection satisfies the constraint in Lagrangian dual problem. Therefore, we always have a feasible solution for the problem.

The subgradient method does not find the optimal solution but can find a near-optimal solution within reasonable time complexity. To compute the time complexity of the subgradient algorithm, we analyze all the steps mentioned earlier. The computation of x^* matrix is of order $O(M.N)$. Computing subgradient vector also costs $O(M.N)$. There are different algorithms to compute the Euclidean projection on simplex. The one we use has a time complexity of order $O(N \log(N))$. If we call the iteration number of the subgradient method K , the subgradient method's total time

complexity is $O(K.M.N)$. The worst order is $O(\max(K, M, N)^3)$, which is of polynomial time complexity compared to exponential time complexity for methods that solve the primal problem. This difference is significant for systems with limited processing resources and real-time response expectations.

We formulate the duality gap for the solution of ASF scalarization method as,

$$p^* - d^* \leq (N + 2) \left(\max \left(\max_{j \in \mathcal{M}, i \in \mathcal{N}_j} \gamma_{ij}, \max_{j \in \mathcal{M}, i \in \mathcal{N}_j} \beta_{ij} \right) + \max_{j \in \mathcal{M}} \left(\min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{w_1} + \frac{\beta_{kj}}{w_0} \right) \right) \right) \quad (3.23)$$

In inequality (3.23), p^* and d^* are the optimum solutions to primal and dual optimization problems, respectively. In this inequality, the value that determines the maximum value of right hand side of the inequality is $\max_{j \in \mathcal{M}, i \in \mathcal{N}_j} \gamma_{ij}$. According to [51], this term equals the maximum blockage score in the whole network. As stated in [51], the BPC sums the blockage score over time using the exponential moving average (EMA) to account for the aging of the blockage data. It means that if the blockage data gets old and no new blockages happen, the blockage score does not increase, or it even decreases. The non-increasing property of blockage score under such conditions leads to a bounded duality gap given the fact that β_{ij} is bounded to 1.

On the other hand, if the blockages happen frequently, the upper bound for the duality gap increases over time. One approach can be using a bound for blockage score to keep it from growing excessively. This approach is useful for making less biased decisions for selecting a base station because growing the blockage score of a base station to a great value does not necessarily make the other base stations that have a little less blockage score more attractive. They all have high blockage scores. A simple implementation may involve choosing a value and the maximum blockage score. The BPC does not record the blockage incidents leading to an increase of such maximum value, and the respective base station is not associable. If the algorithm defines such bound on blockage score, the duality gap is bounded.

3.3.2 Formulation and Solution of NC Scalarization Method

In this section, we formulate and solve the linear single objective optimization problem that we generate by the NC scalarization method.

Normal Constraint (NC) method requires two Utopian points to do the scalarization. These two points are the optimum values for all objectives when solved separately. The NC is the same as ASF method, but the difference is in how they draw the reference vectors. In NC, unlike ASF, the weight vector generating method draws the reference vectors in parallel. We show the reformulated version for NC generated single objective problem as,

$$\text{minimize } \max \left(NF_1 - (w_0 - 0.5), NF_2 - (w_1 - 0.5) \right) \quad (3.24a)$$

$$\text{subject to } \sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t \quad \forall i \in \mathcal{N} \quad (3.24b)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.24c)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.24d)$$

$$\text{In which } NF_1 = \frac{t - f_1(X_l)}{f_1(X_r) - f_1(X_l)} = \frac{t - f_1(X_l)}{C_1} \quad (3.24e)$$

$$NF_2 = \frac{B - f_2(X_r)}{f_2(X_l) - f_2(X_r)} = \frac{B - f_2(X_r)}{C_2} \quad (3.24f)$$

In problem (3.24), NF_1 and NF_2 are normalized objective values for first and second objectives. X_l is the optimal solution to the optimization problem that minimizes maximum load t in the network. Similarly, X_r is the optimal solution to optimization problem that minimizes total blockage score B . $f_1(x)$ is the objective value for maximum load of solution x and $f_2(x)$ is the objective value for total blockage score B . The values in \max function are normalized objective values. Like ASF formulation (3.14), we want to reformulate problem (3.24) to make it a linear optimization problem. We formulate the result as,

$$\text{minimize } S \quad (3.25a)$$

$$\text{subject to } NF_1 - (w_0 - 0.5) \leq S \quad (3.25b)$$

$$NF_2 - (w_1 - 0.5) \leq S \quad (3.25c)$$

$$\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t, \quad \forall i \in \mathcal{N} \quad (3.25d)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.25e)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.25f)$$

The problem in (3.25) is a linear optimization problem that LP solvers can solve.

3.3.2.1 Solution to NC Generated Single Objective Problem via Dual Analysis

The solution to the NC-generated single objective problem uses the same methodology as the ASF scalarization method.

In order to get to the linear optimization problem, we need to eliminate the t variable and replace NF_1 and NF_2 parameters with extreme constants as in Equations (3.24e) and (3.24f). We also replace parameter B by its equivalent according to [51]. We do the variable elimination through using constraint (3.25b) in constraint (3.25d). Therefore, we eliminate variable t and reformulate the problem as,

$$\text{minimize } S \quad (3.26a)$$

$$\text{subject to } \frac{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} \gamma_{ij} \cdot x_{ij}}{C_2} - \frac{f_2(X_r)}{C_2} - w_1 + 0.5 \leq S \quad (3.26b)$$

$$\frac{\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij}}{C_1} - \frac{f_1(X_l)}{C_1} - w_0 + 0.5 \leq S, \quad \forall i \in \mathcal{N} \quad (3.26c)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.26d)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.26e)$$

We compute the Lagrangian of optimization problem in (3.26), like ASF that we provide in Section 3.3.1.2. We use objective function (3.26a) along with constraints (3.26b) and (3.26c). Therefore,

we formulate the Lagrangian as,

$$L(S, \mathbf{x}, \lambda) = S \cdot \left(1 - \lambda_{N+1} - \sum_{i \in \mathcal{N}} \lambda_i\right) + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} x_{ij} \cdot \left(\frac{\gamma_{ij}}{C_2} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{C_1} \cdot \lambda_i\right) - \left(\frac{f_2(X_r)}{C_2} + w_1 - 0.5\right) \cdot \lambda_{N+1} - \left(\frac{f_1(X_l)}{C_1} + w_0 - 0.5\right) \cdot \sum_{i \in \mathcal{N}} \lambda_i \quad (3.27)$$

From Lagrangian in (3.27), we can deduce Lagrangian dual optimization problem for (3.26). We formulate the Lagrangian dual as,

$$g(\lambda) = \inf_{\substack{S \in \mathbb{R} \\ \mathbf{x} \in \mathcal{X}}} L(S, \mathbf{x}, \lambda) \quad (3.28a)$$

$$= \begin{cases} \inf_{\mathbf{x} \in \mathcal{X}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} x_{ij} \cdot \left(\frac{\gamma_{ij}}{C_2} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{C_1} \cdot \lambda_i\right) - \left(\frac{f_2(X_r)}{C_2} + w_1 - 0.5\right) \cdot \lambda_{N+1} \\ - \left(\frac{f_1(X_l)}{C_1} + w_0 - 0.5\right) \cdot \sum_{i \in \mathcal{N}} \lambda_i, & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.28b)$$

$$= \begin{cases} \sum_{j \in \mathcal{M}} \inf_{\mathbf{x}_j \in \mathcal{X}_j} \sum_{i \in \mathcal{N}_j} x_{ij} \cdot \left(\frac{\gamma_{ij}}{C_2} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{C_1} \cdot \lambda_i\right) - \left(\frac{f_2(X_r)}{C_2} + w_1 - 0.5\right) \cdot \lambda_{N+1} \\ - \left(\frac{f_1(X_l)}{C_1} + w_0 - 0.5\right) \cdot \sum_{i \in \mathcal{N}} \lambda_i, & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.28c)$$

$$= \begin{cases} \sum_{j \in \mathcal{M}} g_j(\lambda), & \sum_{i=1}^{N+1} \lambda_i = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (3.28d)$$

In (3.28b) and (3.28c), the whole term before each colon is the term for the first case and the term after colon is the condition. Similar to the ASF solution, we get different formulations for Lagrangian dual.

The Lagrangian dual optimization problem is,

$$\begin{aligned}
& \text{maximize } g(\lambda) = \\
& \sum_{j \in \mathcal{M}} x_{ij}^* \left(\frac{\gamma_{ij}}{C_2} \cdot \lambda_{N+1} + \frac{\beta_{ij}}{C_1} \cdot \lambda_i \right) - \left(\frac{f_2(X_r)}{C_2} + w_1 - 0.5 \right) \cdot \lambda_{N+1} - \left(\frac{f_1(X_l)}{C_1} + w_0 - 0.5 \right) \cdot \sum_{i \in \mathcal{N}} \lambda_i \\
& \text{subject to } \sum_{i=1}^{N+1} \lambda_i = 1 \\
& \lambda_i \geq 0
\end{aligned} \tag{3.29}$$

In Equation (3.29), we compute x_{ij}^* as,

$$x_{ij}^* = \begin{cases} 1, & i = \arg \min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{C_2} \cdot \lambda_{N+1} + \frac{\beta_{kj}}{C_1} \cdot \lambda_k \right) \\ 0, & \text{otherwise} \end{cases} \tag{3.30}$$

We formulate the subgradient of optimization problem in (3.29) as,

$$u_i = \begin{cases} - \sum_{j \in \mathcal{M}_i} \frac{\beta_{ij}}{C_1} \cdot x_{ij}^* + \frac{f_1(X_l)}{C_1} + w_0 - 0.5, & i \in \mathcal{N} \\ - \sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}_j} \frac{\gamma_{ij}}{C_2} \cdot x_{ij}^* + \frac{f_2(X_r)}{C_2} + w_1 - 0.5, & i = N+1 \end{cases} \tag{3.31}$$

By using the subgradient in Equation (3.31) and subgradient method it is possible to solve the optimization problem in (3.29). The general approach is the same as described in Section 3.3.1.2.

The duality gap inequality for NC scalarization method is as follows,

$$p^* - d^* \leq (N+2) \left(\max \left(\max_{j \in \mathcal{M}, i \in \mathcal{N}_j} (\gamma_{ij}) + CT_2, \max_{j \in \mathcal{M}, i \in \mathcal{N}_j} (\beta_{ij}) + CT_1 \right) + \max_{j \in \mathcal{M}} \left(\min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{C_2} + \frac{\beta_{kj}}{C_1} \right) \right) \right) \tag{3.32}$$

where

$$\begin{aligned}
CT_1 &= -\frac{f_2(X_r)}{C_2} - w_1 + 0.5 \\
CT_2 &= -\frac{f_1(X_l)}{C_1} - w_0 + 0.5
\end{aligned} \tag{3.33}$$

We can use the same explanation of bounded duality gap as inequality in (3.23) for inequality in (3.32).

3.3.3 Formulation and Solution of WS Scalarization Method

Weighted sum method multiplies objective values by a weight vector and then sums over the results.

For our specific problem, we formulate weighted sum scalarized optimization problem as,

$$\text{minimize } w_0.t + w_1.B \quad (3.34a)$$

$$\text{subject to } \sum_{j \in \mathcal{M}_i} \beta_{ij}.x_{ij} \leq t, \quad \forall i \in \mathcal{N} \quad (3.34b)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.34c)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.34d)$$

The objective function in the weighted sum method is a linear function, so there is no need for reformulation of problem (3.34). We formulate the Lagrangian dual problem as,

$$\text{maximize } g(\lambda) = \sum_{j \in \mathcal{M}} x_{ij}^*.(w_1.\gamma_{ij} + \lambda_i.\beta_{ij}) \quad (3.35a)$$

$$\text{subject to } \sum_{i \in \mathcal{N}} \lambda_i = w_0 \quad (3.35b)$$

$$\lambda_i \geq 0, i \in \mathcal{N} \quad (3.35c)$$

In Equation (3.35a), we compute x_{ij}^* using the following formula,

$$x_{ij}^* = \begin{cases} 1, & i = \arg \min_{k \in \mathcal{N}_j} (w_1.\gamma_{kj} + \lambda_k.\beta_{kj}) \\ 0, & \text{otherwise} \end{cases} \quad (3.36)$$

For subgradient method we compute the subgradient of optimization problem in (3.35). The subgradient of $-g$ in a feasible λ is $\mathbf{u} = (u_j)_{j \in \mathcal{N}}$ and $u_i = -\sum_{j \in \mathcal{M}_i} x_{ij}^*.\beta_{ij}$. Then we compute the next value of vector λ by recursive subgradient method formulation in (3.22).

In Equation (3.22), $P(\lambda^{(k)} - \alpha_k.\mathbf{u}^{(k)})$ is Euclidean projection of point $\lambda^{(k)} - \alpha_k.\mathbf{u}^{(k)}$ in N dimensional space into simplex defined by $\{\lambda \mid \sum_{i \in \mathcal{N}} \lambda_i = w_0\}$.

We formulate the duality gap for the solution provided for WS scalarization method as,

$$p^* - d^* \leq (N + 1) \left(\max_{j \in \mathcal{M}, i \in \mathcal{N}_j} \beta_{ij} + \max_{j \in \mathcal{M}} \beta_{njj} \right) \quad (3.37)$$

Algorithm 3.1 Semi-distributed Algorithm for User Association (SDA)

- 1: Initialize: Central {CRAN distributes the parameters to BSs}
- 2: Initialize: Distributed {Each BS_i distributes $\lambda^{(k)}$ at time k and γ_{ij} to UEs $j \in \mathcal{M}_i$ }
- 3: UE_j computes its associated BS_i by Equations (3.20), (3.30) and (3.36)
- 4: UE_j only signals its associated BS_i and BS_i compute subgradient according to subgradient equations for each scalarization method
- 5: $\lambda^{(k+1)}$ computed by communication between BS s for the next step
- 6: if stopping criterion not met go to 3

where,

$$n_j = \arg \min_{k \in \mathcal{N}_j} \beta_{kj} \quad (3.38)$$

As we can see in inequality (3.37), the right-hand side of the inequality does not depend on the blockage score. This independence means that the duality gap is bounded regardless of the assumption of having maximum value for blockage score.

3.3.4 Semi-Distributed Algorithm for Dual Analysis

In this section, we introduce a semi-distributed algorithm used to do the computations of the solution to the dual problem discussed in previous sections. We need to compute the subgradient of the dual problem and the step in the subgradient method to solve the dual problem for three single objective problems generated by all three scalarization methods in Equation (3.22). The evaluation of this equation requires the evaluation of subgradients of three scalarization methods. It also needs to compute x_{ij}^* for the scalarization methods.

The algorithm is a semi-distributed one as there are some parts of it needs a centralized entity like CRAN to run them [20]. The rest of the operations can be done in a distributed manner between BSs and UEs. We elaborate on the semi-distributed algorithm for user association (SDA) in Algorithm 3.1. In Algorithm 3.1, there are two different initializations, central and distributed. In central initialization, CRAN is responsible for distributing the general parameters to BSs. These parameters include weight vectors for all scalarization methods and λ_{N+1} for ASF and NC scalarization methods and C_1 and C_2 for NC scalarization method. An entity that has a holistic view of the network can only compute the general parameters. That is why CRAN is responsible for it.

According to this algorithm, a change in the number of BS or UE should not incur a considerable load on the distributed process. For the central process, the CRAN can handle the change properly as it is designed to do so.

As we can see in Algorithm 3.1, the algorithm can perform the different tasks for implementing the solutions to different biobjective optimization problems in a distributed manner. The solution to the optimization problem tries to associate UEs to BSs suitable to improve the objective values. The objective functions push the UEs to BSs with less chance of being blocked while maintaining the network as balanced as possible. The different user association methods that we evaluate are user association policies (UAP) generated by solutions to different optimization problems. Since the solution has a distributed nature, it is feasible to have the solution running in higher frequencies and have a negligible overhead on the system. When the UE goes out of the reach of a BS, the algorithm can start the solution to find another BS for the UE based on the objectives defined.

3.4 Proof of Duality Gaps

In this section, we present the proof of the duality gaps provided in the chapter. For duality inequality of each scalarization method, we provide the required proofs and then the proof for duality inequality.

Before we discuss the duality gap of the scalarization methods, we need to present a proposition that provides a general framework for the duality gap of a certain class of optimization problems. In the following, we present Proposition 1 adopted from [18].

Proposition 1 *Consider a possibly nonconvex problem in Equation (3.39).*

$$\text{minimize } \sum_{j \in \mathcal{J}} f_j(\mathbf{y}_j) \quad (3.39a)$$

$$\text{subject to } \mathbf{y}_j \in \mathcal{Y}_j, \quad j \in \mathcal{J} \quad (3.39b)$$

$$\sum_{j \in \mathcal{J}} \mathbf{h}_j(\mathbf{y}_j) \leq \mathbf{b} \quad (3.39c)$$

where the variables are $\mathbf{y}_j \in \mathbb{R}^{y_j}$. The problem parameters are $\mathcal{J} = \{1, 2, \dots, J\}$, \mathbf{b} is a vector in \mathbb{R}^Q , \mathcal{Y}_j is a subset of \mathbb{R}^{y_j} and $f_j : \text{conv}(\mathcal{Y}_j) \rightarrow \mathbb{R}$ and $\mathbf{h}_j : \text{conv}(\mathcal{Y}_j) \rightarrow \mathbb{R}^Q$ are functions that are

defined on convex hull of \mathcal{Y}_j . There are assumptions that hold for the optimization problem (3.39).

Assumption 1 There exists at least one feasible solution for problem (3.39).

Assumption 2 For each j , the subset of \mathbb{R}^{y_j+Q+1}

$$\{(\mathbf{y}_j, \mathbf{h}_j(\mathbf{y}_j), f_j(\mathbf{y}_j)) | \mathbf{y}_j \in \mathcal{Y}_j\} \quad (3.40)$$

is compact.

Assumption 3 For each j , with any vector $\tilde{\mathbf{y}}$ in $\text{conv}(\mathcal{Y}_j)$, there exists $\mathbf{y} \in \mathcal{Y}_j$ such that $\mathbf{h}_j(\mathbf{y}) \leq \tilde{\mathbf{h}}_j(\tilde{\mathbf{y}})$, where $\tilde{\mathbf{h}}_j : \text{conv}(\mathcal{Y}_j) \rightarrow \mathbb{R}^Q$ is convexified version of \mathbf{h}_j on $\text{conv}(\mathcal{Y}_j)$ and \leq means component-wise inequality. To be specific, for all $\tilde{\mathbf{y}} \in \text{conv}(\mathcal{Y}_j)$ we have,

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \inf \left\{ \sum_{k=1}^{y_j+1} \alpha^k \mathbf{h}_j(\mathbf{y}^k) \middle| \tilde{\mathbf{y}} = \sum_{k=1}^{y_j+1} \alpha^k \mathbf{y}^k, \mathbf{y}^k \in \mathcal{Y}_j, \sum_{k=1}^{y_j+1} \alpha^k = 1, \alpha^k \geq 0 \right\} \quad (3.41)$$

We define the dual problem of problem (3.39) as,

$$\begin{aligned} & \text{maximize } d(\mathbf{v}) = \inf_{\substack{\mathbf{y}_j \in \mathcal{Y}_j \\ j \in \mathcal{J}}} \left\{ \sum_{j \in \mathcal{J}} [f_j(\mathbf{y}_j) + \mathbf{v}^T \mathbf{h}_j(\mathbf{y}_j)] - \mathbf{v}^T \mathbf{b} \right\} \\ & \text{subject to } \mathbf{v} \geq \mathbf{0} \end{aligned} \quad (3.42)$$

The Lagrangian multipliers are $\mathbf{v} = (v_1, \dots, v_Q) \in \mathbb{R}^Q$. Then, the duality gap of the problem in (3.39) is

$$P^* - D^* \leq (Q+1) \max_{j \in \mathcal{J}} \rho_j \quad (3.43)$$

In Equation (3.43), P^* is the optimal value for the problem (3.39), D^* is the optimal value for the dual problem in (3.42) and ρ_j is a nonnegative scalar such that,

$$\rho_j \leq \sup_{\mathbf{y}_i \in \mathcal{Y}_j} f_j(\mathbf{y}_j) - \inf_{\mathbf{y}_i \in \mathcal{Y}_j} f_j(\mathbf{y}_j) \quad (3.44)$$

Relying Proposition 1, for each optimization problem generated by each scalarization function, we reformulate the optimization problem to an equivalent problem so that it fits in general form provided in problem (3.39). Then we show that the reformulated problem is compatible with all assumptions mentioned. Then, we formulate the duality gap of each optimization problem generated by each scalarization method based on Equation (3.43).

3.4.1 Duality Gap of ASF

In this section, we elaborate the duality gap inequality presented in the chapter. The ASF generated optimization problem that we prepare for the Lagrangian dual analysis is as follows.

$$\begin{aligned}
& \text{minimize } S \\
& \text{subject to } \frac{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} \gamma_{ij} \cdot x_{ij}}{w_1} \leq S \\
& \quad \frac{\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij}}{w_0} \leq S \quad \forall i \in \mathcal{N} \\
& \quad \sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \\
& \quad x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j
\end{aligned} \tag{3.45}$$

We rewrite the problem formulated in (3.45) in a way to conform with the formulation of the problem in (3.43). We reformulate problem (3.45) as follows.

$$\begin{aligned}
& \text{minimize } \sum_{j \in \mathcal{M}} S_j \\
& \text{subject to } \frac{\sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}} \bar{\gamma}_{ij} \cdot x_{ij}}{w_1} \leq \sum_{j \in \mathcal{M}} S_j \\
& \quad \frac{\sum_{j \in \mathcal{M}} \bar{\beta}_{ij} \cdot x_{ij}}{w_0} \leq \sum_{j \in \mathcal{M}} S_j \quad \forall i \in \mathcal{N} \\
& \quad \sum_{i \in \mathcal{N}} \delta_{ij} \cdot x_{ij} = 1, \quad \forall j \in \mathcal{M} \\
& \quad x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N} \\
& \quad 0 \leq S_j \leq S^{\max} + \eta_j \cdot x_{n_j j}
\end{aligned} \tag{3.46}$$

In problem defined in (3.46), the variables are $\mathbf{S} = (S_1, \dots, S_M)$. We define the variables in Equation (3.47).

$$\bar{\beta}_{ij} = \begin{cases} \beta_{ij}, & i \in \mathcal{N}, j \in \mathcal{M}_i \\ 0, & \text{otherwise} \end{cases} \tag{3.47a}$$

$$\bar{\gamma}_{ij} = \begin{cases} \gamma_{ij}, & i \in \mathcal{N}, j \in \mathcal{M}_i \\ 0, & \text{otherwise} \end{cases} \quad (3.47b)$$

$$\delta_{ij} = \begin{cases} 1, & j \in \mathcal{M}, i \in \mathcal{N}_j \\ 0, & \text{otherwise} \end{cases} \quad (3.47c)$$

$$S^{max} = \max(\gamma, \beta) \quad (3.47d)$$

$$\gamma = \max_{j \in \mathcal{M}, i \in \mathcal{N}} \bar{\gamma}_{ij} \quad (3.47e)$$

$$\beta = \max_{i \in \mathcal{N}, j \in \mathcal{M}} \bar{\beta}_{ij} \quad (3.47f)$$

$$\eta_j = \min_{k \in \mathcal{N}_j} \left(\frac{\bar{\gamma}_{kj}}{w_1} + \frac{\bar{\beta}_{kj}}{w_0} \right) \quad (3.47g)$$

$$n_j = \arg \min_{k \in \mathcal{N}_j} \left(\frac{\bar{\gamma}_{kj}}{w_1} + \frac{\bar{\beta}_{kj}}{w_0} \right) \quad (3.47h)$$

We can show easily that problem in (3.46) is the same as the problem in (3.45). Therefore, if P^* is the optimal value of problem (3.46) and p^* is the optimal value of problem (3.45), then we have

$$P^* = p^* \quad (3.48)$$

This modified version of problem (3.46) is based on (3.39) which we compare its different parts as follows.

1. $\mathcal{J} = \mathcal{M}, J = M$
2. $\mathbf{y}_j = (Z_j, S_j) \in \mathbb{R}^{y_j}, Z_j = (x_{ij})_{i \in \mathcal{N}}, y_j = N + 1$
3. $f_j(\mathbf{y}_j) = S_j$
4. $\mathcal{Y}_j = \{((x_{ij})_{i \in \mathcal{N}}, S_j) \mid \sum_{i \in \mathcal{N}} \delta_{ij} \cdot x_{ij} = 1, x_{ij} \in \{0, 1\}, i \in \mathcal{N}, S_j \in [0, S^{max} + \eta_j \cdot x_{n_j j}]\}$
5. $\mathbf{h}_j(\mathbf{y}_j) = ((\frac{\bar{\beta}_{1j}}{w_0} \cdot x_{1j} - S_j), \dots, (\frac{\bar{\beta}_{Nj}}{w_0} \cdot x_{Nj} - S_j), (\frac{\sum_{i \in \mathcal{N}} \bar{\gamma}_{ij} \cdot x_{ij}}{w_1} - S_j)) \in \mathbb{R}^Q, Q = N + 1$
6. $\mathbf{b} = \mathbf{0}$

We are ready to show that the assumptions in Proposition 1. The Assumptions 1 and 2 hold for the problem (3.46). We can show it easily. For Assumption 3, we need to make some more elaboration. According to definition of $\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}})$ in Equation (3.41) we can write it as follows.

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \sum_{k=1}^{y_j+1} \alpha^k \mathbf{h}_j(\mathbf{y}^k) \quad (3.49)$$

After plugging the parameters of problem (3.46) in Equation (3.49) we have,

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \sum_k \alpha^k \left(\left(\frac{\bar{\beta}_{1j}}{w_0} x_{1j}^k - S_j^k \right), \dots, \left(\frac{\bar{\beta}_{njj}}{w_0} x_{njj}^k - S_j^k \right), \dots, \left(\frac{\bar{\beta}_{Nj}}{w_0} x_{Nj}^k - S_j^k \right), \left(\frac{\bar{\gamma}_{njj}}{w_1} x_{njj}^k - S_j^k \right) \right) \quad (3.50a)$$

$$= \sum_k \left(\left(\frac{\bar{\beta}_{1j}}{w_0} \alpha^k x_{1j}^k - \alpha^k S_j^k \right), \dots, \left(\frac{\bar{\beta}_{njj}}{w_0} \alpha^k x_{njj}^k - \alpha^k S_j^k \right), \dots, \left(\frac{\bar{\beta}_{Nj}}{w_0} \alpha^k x_{Nj}^k - \alpha^k S_j^k \right), \left(\frac{\bar{\gamma}_{njj}}{w_1} \alpha^k x_{njj}^k - \alpha^k S_j^k \right) \right) \quad (3.50b)$$

$$= \left(\frac{\bar{\beta}_{1j}}{w_0} \cdot \sum_k (\alpha^k x_{1j}^k) - \sum_k (\alpha^k S_j^k), \dots, \frac{\bar{\beta}_{njj}}{w_0} \cdot \sum_k (\alpha^k x_{njj}^k) - \sum_k (\alpha^k S_j^k), \dots, \frac{\bar{\beta}_{Nj}}{w_0} \cdot \sum_k (\alpha^k x_{Nj}^k) - \sum_k (\alpha^k S_j^k), \frac{\bar{\gamma}_{njj}}{w_1} \cdot \sum_k (\alpha^k x_{njj}^k) - \sum_k (\alpha^k S_j^k) \right) \quad (3.50c)$$

$$\geq \left(\frac{\bar{\beta}_{1j}}{w_0} \cdot \sum_k (\alpha^k x_{1j}^k) - \sum_k (\alpha^k S_j^k), \dots, \frac{\bar{\beta}_{njj}}{w_0} \cdot \sum_k (\alpha^k x_{njj}^k) - \sum_k (\alpha^k (S_c^{max} + \eta_j x_{njj})), \dots, \frac{\bar{\beta}_{Nj}}{w_0} \cdot \sum_k (\alpha^k x_{Nj}^k) - \sum_k (\alpha^k S_j^k), \frac{\bar{\gamma}_{njj}}{w_1} \cdot \sum_k (\alpha^k x_{njj}^k) - \sum_k (\alpha^k S_j^k) \right) \quad (3.50d)$$

$$\geq \left(\frac{\bar{\beta}_{1j}}{w_0} \cdot 0 - S_c^{max}, \dots, \frac{\bar{\beta}_{njj}}{w_0} \cdot 0 - S_c^{max}, \dots, \frac{\bar{\beta}_{Nj}}{w_0} \cdot 0 - S_c^{max}, \frac{\bar{\gamma}_{njj}}{w_1} \cdot 0 - S_c^{max} \right) \quad (3.50e)$$

$$= \left(-S_c^{max}, \dots, -\frac{\bar{\gamma}_{njj}}{w_1} - S_c^{max}, \dots, -S_c^{max}, -\frac{\bar{\beta}_{njj}}{w_0} - S_c^{max} \right) = \mathbf{h}_j(\mathbf{y}) \quad (3.50f)$$

In Equation (3.50), $S_c^{max} = S^{max} + \eta_j$ and $\mathbf{y} = (0, \dots, 1, 0, \dots, 0, S_c^{max})$ is a feasible solution, thus Assumption 3 holds.

According to (3.51), the dual optimal value for problem (3.46) is equal to dual optimal value of problem (3.45).

$$D^* = P_{relax}^* = P_{relax}^* = d^* \quad (3.51)$$

where D^* is dual optimal value for problem in (3.46) and P_{relax}^* is the optimal solution of a LP relaxation of problem (3.46) and p_{relax}^* is optimal value for LP relaxation of problem (3.45) and d^* is dual optimal value for problem (3.45). The authors proved the first equality in (3.51) in [18] for their problem. Using the same process to prove it for our problem is trivial. Second equality comes from the fact that both problems are equivalent and therefore, their LP relaxations have the same optimal value. The authors again proved the last equality in [18] and we can use it for our problem as well.

According to Equations (3.48), (3.51) and (3.43) we have

$$p^* - d^* \leq (N+2) \max_{j \in \mathcal{M}} \rho_j \leq (N+2) \max_{j \in \mathcal{M}} (S^{max} + \eta_j) = (N+2) \left(\max(\gamma, \beta) + \max_{j \in \mathcal{M}} \left(\min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{w_1} + \frac{\beta_{kj}}{w_0} \right) \right) \right) \quad (3.52)$$

3.4.2 Duality Gap of NC

In this section we discuss the duality gap analysis for optimization problem that we generate by Normal Constraint (NC) scalarization method. We use Proposition 1 and its assumptions. We formulate the optimization problem generated by NC scalarization method as,

$$\text{minimize } S \quad (3.53a)$$

$$\text{subject to } \frac{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}_i} \gamma_{ij} \cdot x_{ij}}{C_2} - \frac{f_2(X_r)}{C_2} - w_1 + 0.5 \leq S \quad (3.53b)$$

$$\frac{\sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij}}{C_1} - \frac{f_1(X_l)}{C_1} - w_0 + 0.5 \leq S, \quad \forall i \in \mathcal{N} \quad (3.53c)$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \quad (3.53d)$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.53e)$$

We formulate the modified version of the problem (3.53) based on the general form in problem (3.39) as,

$$\begin{aligned}
& \text{minimize } \sum_{j \in \mathcal{M}} S_j \\
& \text{subject to } \frac{\sum_{j \in \mathcal{M}} \sum_{i \in \mathcal{N}} \bar{\gamma}_{ij} \cdot x_{ij}}{C_2} + CT_2 \leq \sum_{j \in \mathcal{M}} S_j \\
& \quad \frac{\sum_{j \in \mathcal{M}} \bar{\beta}_{ij} \cdot x_{ij}}{C_1} + CT_1 \leq \sum_{j \in \mathcal{M}} S_j \quad \forall i \in \mathcal{N} \quad (3.54) \\
& \quad \sum_{i \in \mathcal{N}} \delta_{ij} \cdot x_{ij} = 1, \quad \forall j \in \mathcal{M} \\
& \quad x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N} \\
& \quad 0 \leq S_j \leq S^{max} + \eta_j \cdot x_{n_j j}
\end{aligned}$$

In problem defined in (3.54), the variables are $S = (S_1, \dots, S_M)$. We define the variables in Equation (3.55).

$$\bar{\beta}_{ij} = \begin{cases} \beta_{ij}, & i \in \mathcal{N}, j \in \mathcal{M}_i \\ 0, & \text{otherwise} \end{cases} \quad (3.55a)$$

$$\bar{\gamma}_{ij} = \begin{cases} \gamma_{ij}, & i \in \mathcal{N}, j \in \mathcal{M}_i \\ 0, & \text{otherwise} \end{cases} \quad (3.55b)$$

$$\delta_{ij} = \begin{cases} 1, & j \in \mathcal{M}, i \in \mathcal{N}_j \\ 0, & \text{otherwise} \end{cases} \quad (3.55c)$$

$$CT_1 = -\frac{f_2(X_r)}{C_2} - w_1 + 0.5 \quad (3.55d)$$

$$CT_2 = -\frac{f_1(X_l)}{C_1} - w_0 + 0.5 \quad (3.55e)$$

$$S^{max} = \max(\gamma + CT_2, \beta + CT_1) \quad (3.55f)$$

$$\gamma = \max_{j \in \mathcal{M}, i \in \mathcal{N}} \bar{\gamma}_{ij} \quad (3.55g)$$

$$\beta = \max_{i \in \mathcal{N}, j \in \mathcal{M}} \bar{\beta}_{ij} \quad (3.55h)$$

$$\eta_j = \min_{k \in \mathcal{N}_j} \left(\frac{\bar{\gamma}_{kj}}{C_2} + \frac{\bar{\beta}_{kj}}{C_1} \right) \quad (3.55i)$$

$$n_j = \arg \min_{k \in \mathcal{N}_j} \left(\frac{\bar{\gamma}_{kj}}{C_2} + \frac{\bar{\beta}_{kj}}{C_1} \right) \quad (3.55j)$$

Like ASF, we can show that problem (3.54) is the same as the original NC problem (3.53). Therefore we have the equality (3.48) for NC as well as ASF. This modified version of problem (3.54) is based on (3.39) which we define its different parts as follows.

$$1. \mathcal{J} = \mathcal{M}, J = M$$

$$2. \mathbf{y}_j = (Z_j, S_j) \in \mathbb{R}^{y_j}, Z_j = (x_{ij})_{i \in \mathcal{N}}, y_j = N + 1$$

$$3. f_j(\mathbf{y}_j) = S_j$$

$$4. \mathcal{Y}_j = \left\{ ((x_{ij})_{i \in \mathcal{N}}, S_j) \mid \sum_{i \in \mathcal{N}} \delta_{ij} \cdot x_{ij} = 1, x_{ij} \in \{0, 1\}, i \in \mathcal{N}, S_j \in [0, S^{max} + \eta_j \cdot x_{n_j j}] \right\}$$

$$5. \mathbf{h}_j(\mathbf{y}_j) = \left(\left(\frac{\bar{\beta}_{1j}}{C_1} \cdot x_{1j} + \frac{CT_2}{M} - S_j \right), \dots, \left(\frac{\bar{\beta}_{Nj}}{C_1} \cdot x_{Nj} + \frac{CT_2}{M} - S_j \right), \left(\frac{\sum_{i \in \mathcal{N}} \bar{\gamma}_{ij} \cdot x_{ij}}{C_2} + \frac{CT_1}{M} - S_j \right) \right) \in \mathbb{R}^Q, \\ Q = N + 1$$

$$6. \mathbf{b} = \mathbf{0}$$

The validity of Assumptions 1 and 2 of Proposition 1 are straightforward to show for problem (3.54). For Assumption 3, we have the following series of equalities and inequalities in an approach similar to ASF.

$$\tilde{\mathbf{h}}_j(\tilde{\mathbf{y}}) = \sum_k \alpha^k \left(\left(\frac{\bar{\beta}_{1j}}{C_1} \cdot x_{1j}^k + \frac{CT_1}{M} - S_j^k \right), \dots, \left(\frac{\bar{\beta}_{n_j j}}{C_1} \cdot x_{n_j j}^k + \frac{CT_1}{M} - S_j^k \right), \right. \\ \left. \dots, \left(\frac{\bar{\beta}_{Nj}}{C_1} \cdot x_{Nj}^k + \frac{CT_1}{M} - S_j^k \right), \left(\frac{\bar{\gamma}_{n_j j}}{C_2} \cdot x_{n_j j}^k + \frac{CT_2}{M} - S_j^k \right) \right) \quad (3.56a)$$

$$= \sum_k \left(\left(\frac{\bar{\beta}_{1j}}{C_1} \cdot \alpha^k \cdot x_{1j}^k + \alpha^k \cdot \frac{CT_1}{M} - \alpha^k \cdot S_j^k \right), \dots, \left(\frac{\bar{\beta}_{n_j j}}{C_1} \cdot \alpha^k \cdot x_{n_j j}^k + \alpha^k \cdot \frac{CT_1}{M} - \alpha^k \cdot S_j^k \right), \dots, \right. \\ \left. \left(\frac{\bar{\beta}_{Nj}}{C_1} \cdot \alpha^k \cdot x_{Nj}^k + \alpha^k \cdot \frac{CT_1}{M} - \alpha^k \cdot S_j^k \right), \left(\frac{\bar{\gamma}_{n_j j}}{C_2} \cdot \alpha^k \cdot x_{n_j j}^k + \alpha^k \cdot \frac{CT_2}{M} - \alpha^k \cdot S_j^k \right) \right) \quad (3.56b)$$

$$= \left(\frac{\bar{\beta}_{1j}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{1j}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k), \dots, \frac{\bar{\beta}_{n_j j}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{n_j j}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k), \dots, \right. \\ \left. \frac{\bar{\beta}_{Nj}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{Nj}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k), \frac{\bar{\gamma}_{n_j j}}{C_2} \cdot \sum_k (\alpha^k \cdot x_{n_j j}^k) + \frac{CT_2}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k) \right) \quad (3.56c)$$

$$\begin{aligned}
&\geq \left(\frac{\bar{\beta}_{1j}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{1j}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k), \dots, \right. \\
&\quad \frac{\bar{\beta}_{n_j j}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{n_j j}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot (S_c^{max} + \eta_j \cdot x_{n_j j}^k)) \\
&\quad \left. , \dots, \frac{\bar{\beta}_{Nj}}{C_1} \cdot \sum_k (\alpha^k \cdot x_{Nj}^k) + \frac{CT_1}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k), \right.
\end{aligned} \tag{3.56d}$$

$$\begin{aligned}
&\quad \frac{\bar{\gamma}_{n_j j}}{C_2} \cdot \sum_k (\alpha^k \cdot x_{n_j j}^k) + \frac{CT_2}{M} \cdot \sum_k (\alpha^k) - \sum_k (\alpha^k \cdot S_j^k) \Big) \\
&\geq \left(\frac{\bar{\beta}_{1j}}{C_1} \cdot 0 + \frac{CT_1}{M} - S_c^{max}, \dots, \frac{\bar{\beta}_{n_j j}}{C_1} + \frac{CT_1}{M} - S_c^{max}, \dots, \frac{\bar{\beta}_{Nj}}{C_1} \cdot 0 + \frac{CT_1}{M} - S_c^{max}, \frac{\bar{\gamma}_{n_j j}}{C_2} + \frac{CT_2}{M} - S_c^{max} \right)
\end{aligned} \tag{3.56e}$$

$$= \left(\frac{CT_1}{M} - S_c^{max}, \dots, \frac{CT_1}{M} - \frac{\bar{\gamma}_{n_j j}}{C_2} - S_c^{max}, \dots, \frac{CT_1}{M} - S_c^{max}, \frac{CT_2}{M} - \frac{\bar{\beta}_{n_j j}}{C_1} - S_c^{max} \right) = \mathbf{h}_j(\mathbf{y}) \tag{3.56f}$$

In Equation (3.56), $S_c^{max} = S_c^{max} + \eta_j$ and $\mathbf{y} = (0, \dots, 1, 0, \dots, 0, S_c^{max})$ is a feasible solution, thus Assumption 3 holds.

According to (3.51), the dual optimal value for problem (3.54) is equal to dual optimal value of problem (3.53). We can make the same justification as ASF for the explanation for equality (3.51) in case of NC.

According to Equations (3.48), (3.51) and (3.43) we have

$$\begin{aligned}
p^* - d^* &\leq (N+2) \max_{j \in \mathcal{M}} \rho_j \leq (N+2) \max_{j \in \mathcal{M}} (S_c^{max} + \eta_j) \\
&= (N+2) \left(\max(\gamma + CT_2, \beta + CT_1) + \max_{j \in \mathcal{M}} \left(\min_{k \in \mathcal{N}_j} \left(\frac{\gamma_{kj}}{C_2} + \frac{\beta_{kj}}{C_1} \right) \right) \right)
\end{aligned} \tag{3.57}$$

3.4.3 Duality Gap of WS

We discuss the duality gap of Weighted Sum (WS) scalarization method in this section. This analysis is very close to the proof that the authors provided in [18]. In order to avoid repetition, we just mention the difference between WS and the duality gap that the authors proved in [18]. We formulate the WS generated optimization problem as,

$$\text{minimize } w_0 \cdot t + w_1 \cdot B \tag{3.58a}$$

$$\text{subject to } \sum_{j \in \mathcal{M}_i} \beta_{ij} \cdot x_{ij} \leq t, \quad \forall i \in \mathcal{N} \tag{3.58b}$$

$$\sum_{i \in \mathcal{N}_j} x_{ij} = 1, \quad \forall j \in \mathcal{M} \tag{3.58c}$$

$$x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{M}, i \in \mathcal{N}_j \quad (3.58d)$$

The objective function of the problem that conforms to problem (3.39) is the only difference of WS optimization problem with problem defined in [18] and formulated as,

$$\text{minimize} \quad \sum_{j \in \mathcal{M}} \left(w_0 \cdot t_j + w_1 \cdot \sum_{i \in \mathcal{N}} \tilde{\gamma}_{ij} \cdot x_{ij} \right) \quad (3.59)$$

Therefore, the objective function of modified WS problem that is comparable to problem (3.39) is as follows,

$$f_j(\mathbf{y}_j) = w_0 \cdot t_j + w_1 \cdot \sum_{i \in \mathcal{N}} \tilde{\gamma}_{ij} \cdot x_{ij} \quad (3.60)$$

According to proof provided in [18], the duality gap for WS scalarization method is,

$$p^* - d^* \leq (N+1) \max_{j \in \mathcal{M}} \rho_j \leq (N+1) \max_{j \in \mathcal{M}} (t_j^{max} + \beta_{n_j j}) = (N+1) \left(\max_{j \in \mathcal{M}, i \in \mathcal{N}_j} \beta_{ij} + \max_{j \in \mathcal{M}} \beta_{n_j j} \right) \quad (3.61)$$

3.5 Evaluations

This section discusses the simulation results of different scalarization methods discussed in the previous section based on two different objective functions defined. We also use a primal problem solver called Gurobi [52]. We use Version 7.0.1 of the Gurobi solver in our simulations. To compute the β_{ij} parameter, we use the SINR model introduced in [14].

3.5.1 Parameters and Metrics

We use OMNet++ [53] network simulator to simulate the environment. The simulation area is a rectangle of 100×500 square meters. There are $M = 100$ user equipments, $N = 50$ base stations and $P = 130$ blocking objects. To simulate a street in a city, we design two sidewalks that the UEs and the blocking objects move. The BSs are installed every 50 meters. We fix their positions. We use the Gauss-Markov mobility model [54] to move the mobile nodes (UEs and blocking objects) in the network.

The user association methods studied in this simulation section are Achievement Scalarizing Function (ASF), Normal Constraint (NC), and Weighted Sum (WS) methods. We also solve the optimization problem for each of the load balance and blockage score objective functions separately. We use two different solvers for solving different methods. We use a subgradient algorithm to solve the dual optimization problem.

We use this algorithm to solve ASF, NC, and WS generated single objective problems and single-objective optimization problems for load balance objective function. We also use a primal problem solver to solve all single objective problems at hand. It includes single objective scalarized optimization problems and the load balance and blockage score optimization problems separately. Gurobi is the name of the solver, and we show it in figures.

We use several metrics to measure and compare the performance of our biobjective optimization approach to single-objective approaches. The first metric is *Average Blockage per Blocking Object*. The BPC records the blockage each time it occurs in the network. The blockage count is cumulative over time. Then, the total number of blockages in the network is divided by the number of blocking objects. To detect blockage in the simulation environment, we assume that each dynamic blocking object, a human, is a cylinder with a height of 1.8 meters and a 0.4-meter width. Each BS is installed on a 3-meter post. The field-of-view of the BS is a sector of about 30° . Using this information, we calculate the shadowed area behind each blocking object. If a UE falls in these shadowed areas, we count that as one blockage.

The second metric is *Average Handover per User Equipment*. A counter increments when a handover occurs. We divide the total by M . This metric is cumulative over time, as well. We also compute rate and SINR for all active links in the network. We use the formulation in [14]. In this formulation, the SINR for a link in the mmW network is a function of antenna orientation, path loss model, independent Nakagami fading model for small-scale fading, and thermal noise of propagation environment. We also use the objective values of load balance and total blockage score for all user association methods we discuss.

The values for the parameters are absolute quantities. All of them are the results of simulations with equal duration. Therefore, the results show the differences in the performance of the algorithms under the same conditions. Thus, we use the values of the parameters to fairly compare the performance of the algorithms.

There are two metrics named *False Positive* and *False Negative*. We define them based on the optimal behavior of handover. We provide those plots to show how the different user association methods are aligned with the defined optimal behavior. For the implementation of the optimal behavior, we access the parameters of the mobility generator based on the mobility model used to mobilize both the UE and blocking objects. The algorithm performs the association and handover based on how far in time an impending blockage would be. The UE is associated to another BS if a blockage happens on the current BS. The algorithm chooses the destination BS based on the least chance of blockage based on the movement of the UE and the blocking

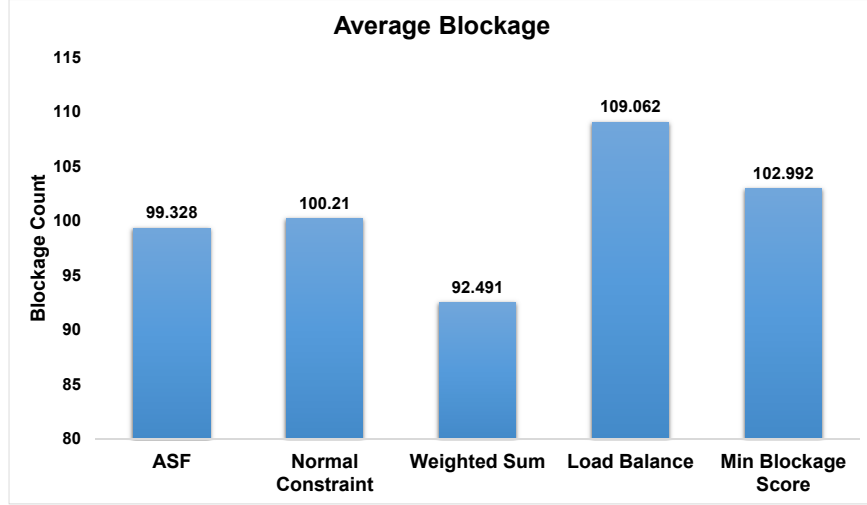


Figure 3.2 Average Blockage Count

objects. This information is not available in the real world and the simulator knows them because it generates the movement of the objects.

The simulator knows the optimal BS for a UE to connect. We define false-positive and false-negative ratios based on whether the algorithm connects a UE to an optimal BS or not. If based on the optimal behavior, the UE should be connected to a BS, but it does not, it counts as one false negative.

On the other hand, if a UE should stay in the current BS, but the algorithm connects it to another BS, it counts as one false positive. The algorithm detects the negativity based on the handover event. We define the optimal behavior based on the blockage. Since the blockage objective is heuristic, it is possible that after handover to a BS, the UE experiences a blockage. Therefore, this handover is not optimal. In general, the optimal behavior is when performing a handover does not cause blockage.

3.5.2 Results and Discussion

In the following, we depict figures for different metrics we discuss.

All of the scalarization methods have 40 subproblems. It means that we generate 40 different weight vectors and then solved the single optimization problem generated by each of the scalarization methods using each weight vector. Unless specifically mentioned, we make the comparison between load balance and minimum blockage score and scalarization methods with a subgradient solution of the scalarization method with 40 subproblems. Unless stated otherwise, the solver for all methods is subgradient. Whenever there is

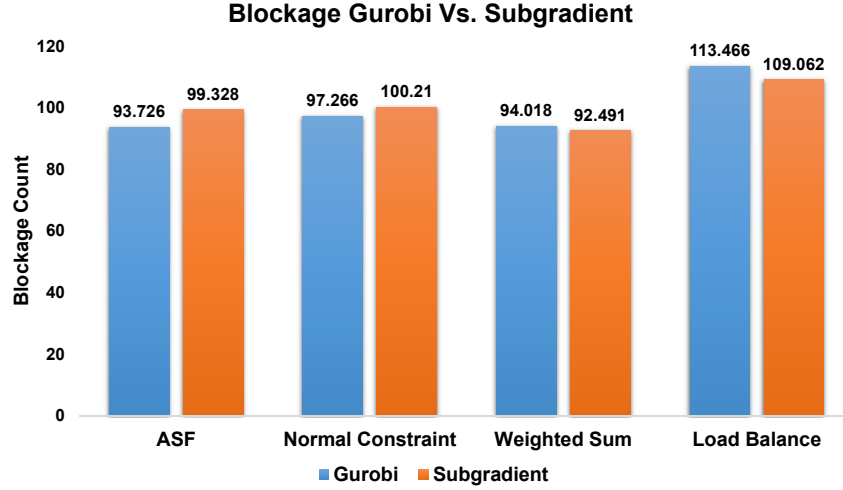


Figure 3.3 Average Blockage Count Comparison Gurobi vs. Subgradient

a comparison involving primal solver, the number of the subproblem is 40. For the load balance method, we use the subgradient method used in [18].

As depicted in Figure 3.2, the three scalarization methods have a lower average blockage score than each load balance and minimum blockage score methods. Among the scalarization methods, the weighted sum has the best performance, while ASF and normal constraints are close to each other. The normal constraint has the worst performance than the other two scalarization methods. Since blockage score is a heuristic, minimizing that does not guarantee the best performance. Therefore, the number of blockages for the Min Blockage Score is not the minimum one. The heuristic nature of the blockage score model is responsible for higher blockage rates than other methods. The combination of minimum blockage score policy and load balance policy results in a better average blockage value. The better performance is owed to the fact that having a balanced network leads to a balanced distribution of users, which can act as blocking objects. If we only use the blockage score minimization as the user association policy, the network gets unbalanced, and it causes the concentration of users in certain areas and leads to more blockages.

As expected, the optimal method has the best blockage score among the association methods thanks to its knowledge of the mobilities.

In Figure 3.3, we make the comparison between all scalarization methods and load balance method solved by subgradient and Gurobi solvers. As we can see in this figure, the subgradient solver has a slightly better performance than Gurobi solver for weighted sum and load balance methods.

We show the average handover count in Figure 3.4. As shown in this figure, the handover count of

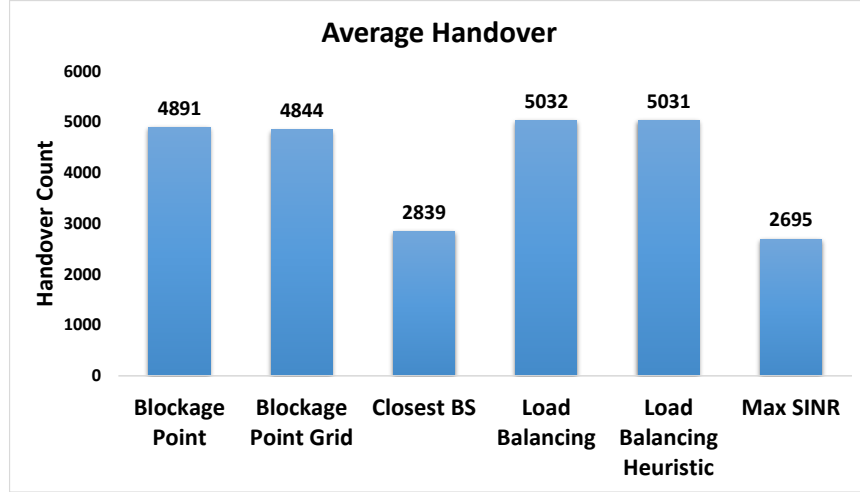


Figure 3.4 Average Handover Count

scalarization methods is nearly the same but more than the minimum blockage score method. Load balance has the worst performance in this metric. The minimum blockage score, on the other hand, has the best performance. The optimal method decreases the handover count significantly and the Min Blockage Score is the closest method to the optimal one with a 43% increase in handover count over the optimal method. As we can see, the scalarization methods' results are between blockage score and load balance methods. We expect this because the scalarization methods receive something from both objective functions. The load balance method's handover is higher mainly because the blockage score is more stable with the EMA method and calculation of blockage score according to a reference point in the simulation area.

We illustrate the false-positive ratio in Figure 3.5. The false-positive counts for the three scalarization methods are less than the separate objectives. Among separate objectives, load balance has a better ratio. We see almost the same pattern for the false-negative ratio in Figure 3.6. Within scalarization methods, the weighted sum has the best performance of all. For both figures, the patterns of difference for the FPR and the FNR are the same. It shows that the scalarization methods cause the UEs to be connected to the BSs with longer uninterrupted periods. It is mainly because of the reconciliation of the two objectives by these methods. Moreover, the scalarization methods tame the blockage score uncertainty.

Figure 3.7 shows a radar plot comparing the parameters for all of the user association methods. We convert the values for each side to normalized values for better comparison. For each parameter, the smaller value is better. As shown in Figure 3.7, the Load Balance method has overall better performance than the other methods, but it is a single objective method. Among the biobjective methods, we can see that the

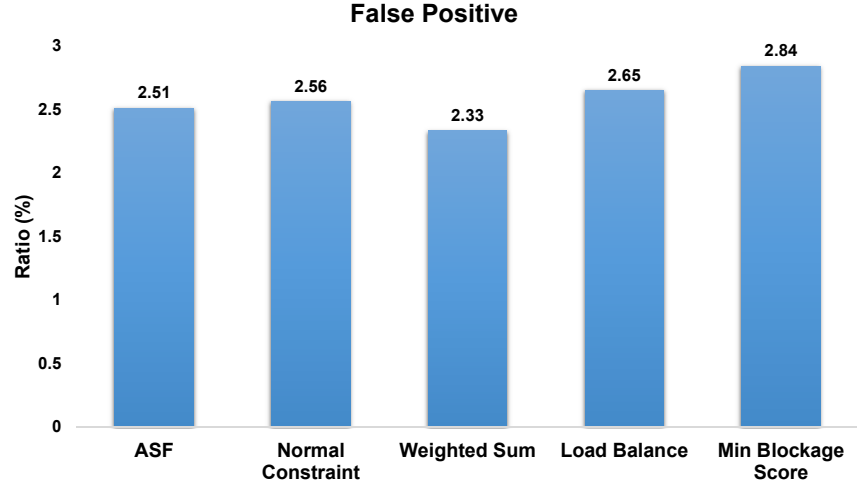


Figure 3.5 False Positive Ratio

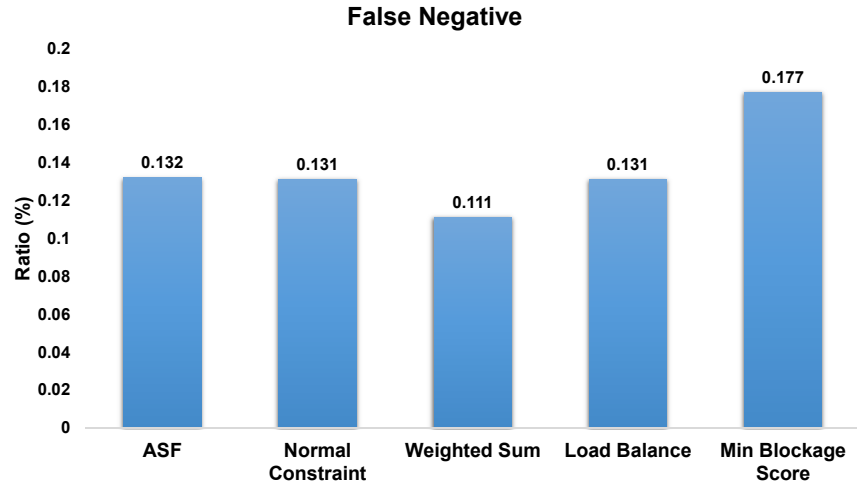


Figure 3.6 False Negative Ratio

Weighted Sum method performs the best. The solution to the WS scalarization method generates a user association policy (UAP) that results in the best overall network performance. We illustrate various network metrics in Figure 3.7 that shows the better performance of such UAP.

We depict the CDF of SINR in Figure 3.8. It is evident that the normal constraint solved with Gurobi solver has links with high SINR with higher frequency than two objective functions separately. Minimum blockage score and normal constraint solved with subgradient solver are the same in terms of SINR. Load balance has the worst performance.

We present the column diagram of the average rate for different user association algorithms in Figure 3.9. In this figure, we include the result for both solvers. However, for the minimum blockage score method,

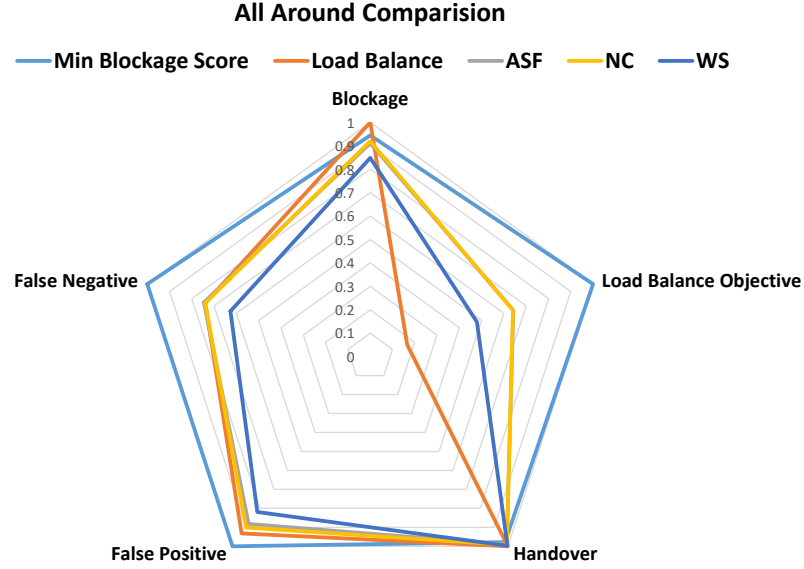


Figure 3.7 Comparison of different parameters for all user association methods

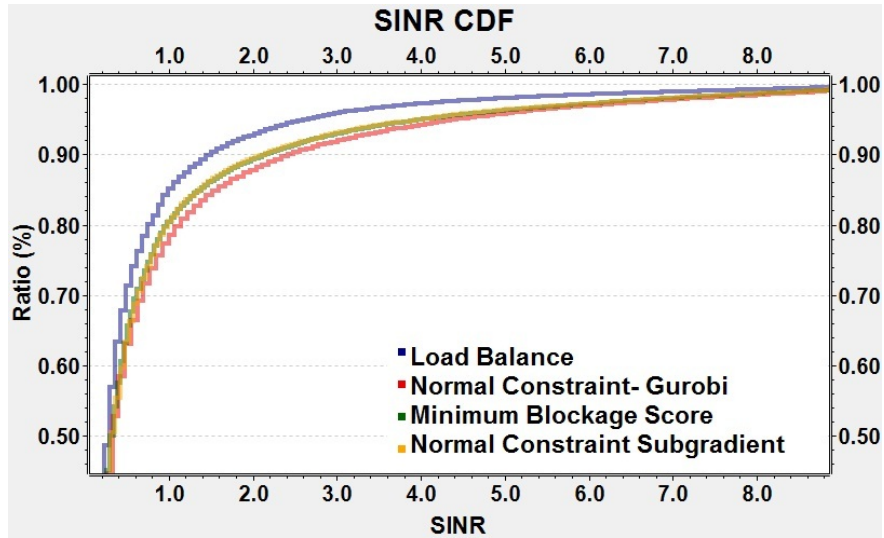


Figure 3.8 Cumulative Distribution Function for SINR

the solver is just a simple, primal solver, and we do not need to use either Gurobi or subgradient solvers. The normal constraint is the best algorithm according to the rate metric. The scalarization methods are performing better than blockage score and load balance methods except for the ASF subgradient. Load balance has the worst metric value.

As shown in Figure 3.10, the load balance algorithm has the best performance for minimizing the maximum load in the network. We have the same structure as Figure 3.9 here. Three scalarization methods have comparable performance, while all are outperforming blockage score algorithm. For subgradient

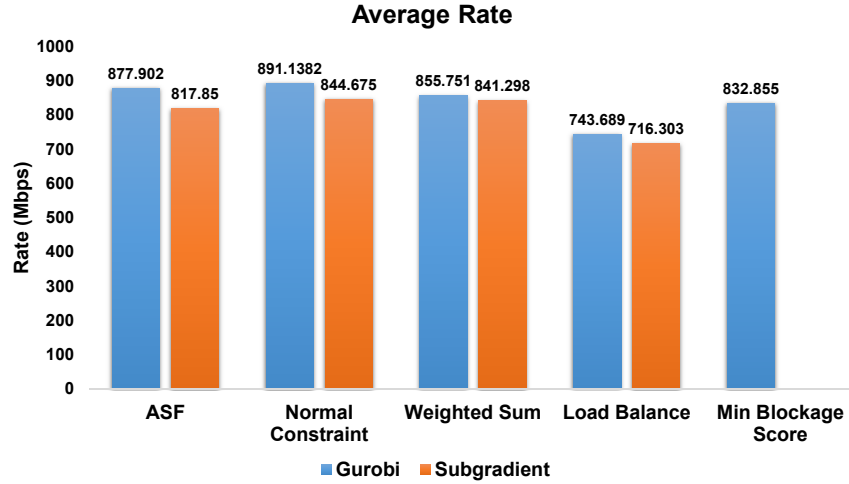


Figure 3.9 Average Rate for Different User Association Algorithms

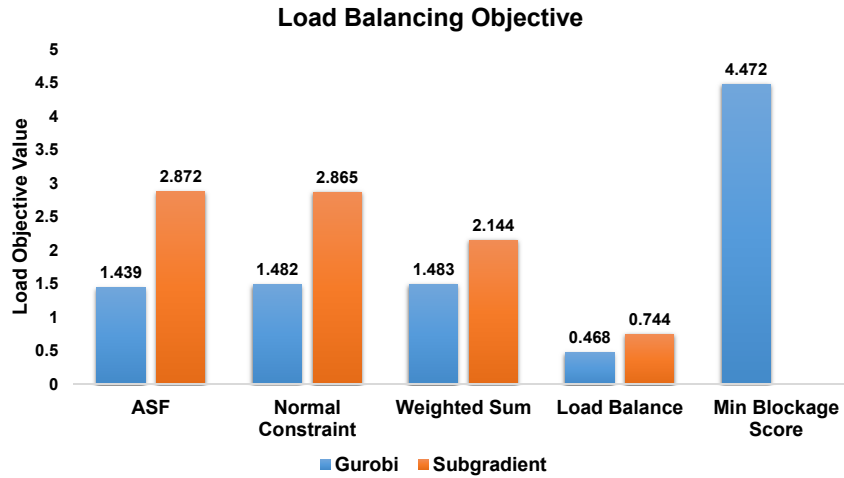


Figure 3.10 Average Maximum Load for Different User Association Algorithms

versions of scalarization methods, the weighted sum has the best performance compared to the other two methods. The optimal method, as expected, does not perform well in terms of the load distribution in the network. The optimal method only considers the blockage avoidance measures to define the behavior.

Figure 3.11 shows how the number of subproblems in a biobjective solver will affect the resulting load objective value. There are solutions with 10, 20, 30, and 40 subproblems for each scalarization and single objective methods that either subgradient or Gurobi solver solves. As shown in the figure, the general trend of load objective reduces when the number of subproblems (weight vectors) increases. It means that when there are more weight vectors, it is more probable that the solver finds better solutions. We can also see that the Gurobi solver results are better than the subgradient. For selecting the best solution from s solutions (s

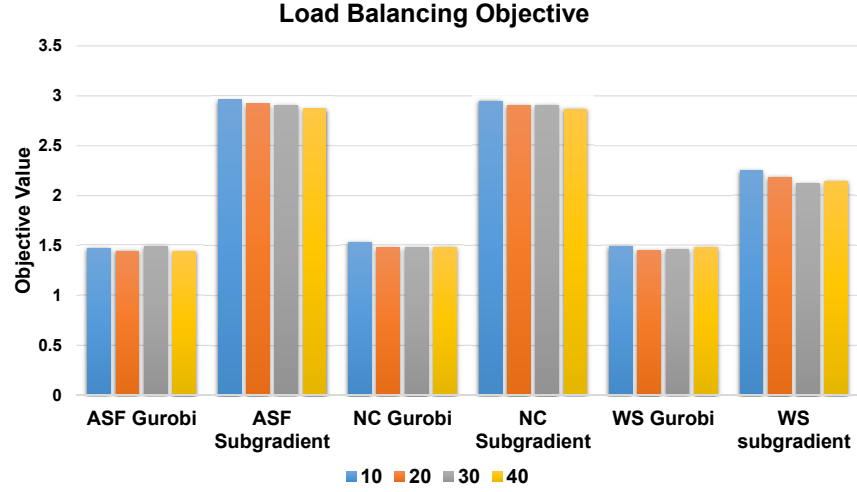


Figure 3.11 Average Maximum Load for Different Subproblem Sizes

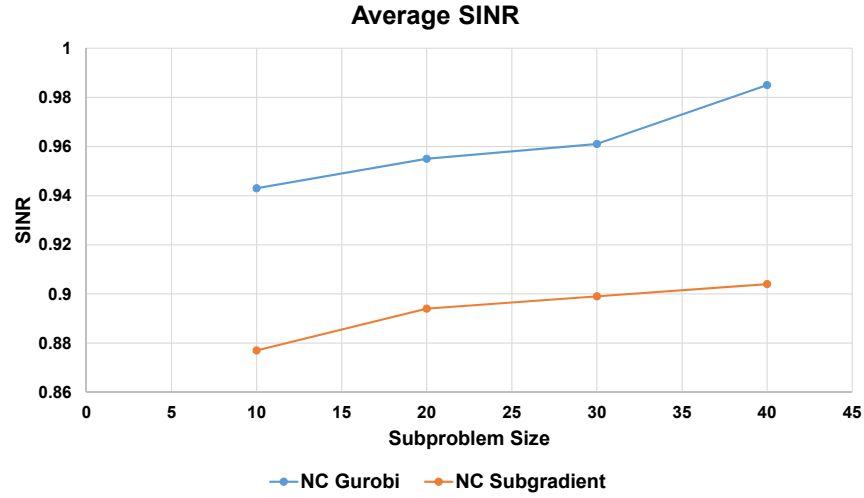


Figure 3.12 SINR Comparison for Normal Constraint Method

is the subproblem size), we choose a solution whose objective values are closest to the ideal point, $(0, 0)$ point.

We present the comparison between SINR values for two different solutions to normal constraint scalarization method in Figure 3.12. The figure shows that the primal solutions have better SINRs for all sub-problem sizes. However, the trend within each curve is almost increasing with increase in subproblem size.

In order to show the statistical behavior of the blockage model, we perform simulations for different numbers of the blocking objects (P). We show the results in Table 3.1.

In Table 3.1, Average Blockage is the total blockage in the network divided by the number of blocking

Table 3.1 Effect of blocking object count on blockage and blockage score

blocking object Count (P)	650	850	1050
Average Blockage	1235	1274	1281
Average Blockage Score	99033	115076	114429

objects and Average Blockage Score is the total blockage score in the network divided by the number of blocking objects. As we can see in this table, the number of blockages increases with the predictable number of blocking objects. The blockage score increases except for the blocking object count of 1050. We observe that the blockage score decreases while the blockage increases in a blocking object count change from 850 to 1050. This shows that the blockage score and the blockage count do not only depend on P but also on the distribution of the blocking objects and with what mobility model the blocking objects and the UEs move. Therefore, P is not the only contributing factor to the blockage score.

As shown in this section's results, we use three different scalarization methods to compare them against each other. For different parameters, different scalarization methods can have the best results with a small variation. The difference stems from the inherent differences in the scalarization methods.

We use the scalarization methods to convert the biobjective optimization problem to a single objective problem and we show that it is possible to provide a solution to user association problem while we optimize more than one objective function to benefit the network from more than one aspect. We provide the solution to each of the problems that have linear time complexity unlike the current solutions to combinatorial optimization problems that have exponential time complexity. We also prove that the duality gap is bounded for all of the solutions. In general, the WS scalarization method performs better for various metrics compared to other methods. It shows that the simple structure of WS captures the complexity of both of the objective functions. Since the lower number of subproblems yields results that are as good as the higher number of subproblems, the solver can use a small number of subproblems to solve the problems faster.

3.6 Summary

In this chapter, we further study the user association problem in mmW networks by considering two objective functions, one for minimizing blockage in the network and the other for minimizing maximum load across all base stations in the network. The results show that the biobjective approach achieves better performance on improving the load balance objective and reducing the blockage score objective, increasing SINR, and

keeping base stations in the networks less loaded. For the weighted sum scalarization method, we show that the solution to the dual problem via the subgradient method has better time complexity than solving the primal problem, but with the cost of losing some SINR value. A possible future direction of this research can be an investigation of other factors affecting link blockage chance and the importance of their role in this area.

CHAPTER 4

LEARNING-BASED LINK BLOCKAGE PREDICTION

We study the user association problems for mmW networks in previous two chapters. In this chapter, we focus on maintaining a robust link between a pair of nodes. The methods for maintaining a robust link do not try to associate a UE to a BS, but try to exploit the signal paths around either node for uninterrupted communication.

Link quality can be deteriorated when a link is blocked or when its associated node moves. It is important to discriminate between link blockage and link movement since they require different remedies. If link quality drop is due to link movement, the most probable cause is the misalignment of transmitter and receiver beams. A new beam adjustment can restore link quality. If link quality drop is due to link blockage, the nodes require a different remedy. One possible solution is switching to an unblocked path. An inappropriate solution, such as applying beam alignment for link blockage or path switching for link movement, may deteriorate the link quality drop. The nodes can use beam scanning in case of link blockage to find NLOS paths to switch and they can perform beam adjustment to fix beam misalignment. Therefore, a beam selection of neighboring beams to the current one can solve the problem. We should apply these two methods to appropriate cases since incorrect use of them results in more unnecessary overhead. Therefore, it is crucial to detect the cause of link quality.

In both behaviors mentioned as link blockage and link movement, the existing work to detect and discriminate them is to have a window of RSSI measurements over time [7]. In their approach, the algorithm calculates the variance of a series of RSSI measurements. The receiver measures RSSI values in fixed-length time windows. If the variance falls below a threshold, the receiver considers the time window as a link blockage. If the variance is more than another threshold, the receiver considers the window as a link movement.

There are raytracing methods in the literature that try to resolve the paths of Line-of-Sight (LOS) and NLOS for path switching purposes. We can apply these methods to simulation and analytical approaches. The accuracy of such methods drops over time and as the number of reflections increases over the geographical area [55]. Therefore, the raytracing algorithms cannot benefit from the low overhead methods for link quality

drop.

According to the experiments conducted in [7], to achieve an accuracy of close to 100%, the receiver will have to experience a delay of 30 milliseconds (ms). During this period, the link is in a halt state and cannot transmit data. For an IEEE802.11ad link, the maximum data rate is 7 Gbps [56]. In the mentioned method of classifying the RSSI variation window as a link blockage or link movement, the 30 ms means that the link losses 210 Mb in this time window. This amount of data that is unavailable for a service running on the receiver can affect the quality of experience (QoE) negatively. The interruption in data communication can have a negative effect on data-hungry applications like video streaming.

We develop a learning-based prediction framework that predicts link blockage and discriminates between link blockage and link movement in real-time without the need for having time-consuming RSSI variance calculation. In this framework, the overhead is generating the data point and predicting the label of that data point. The predictor predicts by measuring the diffraction values of different multipath components (MPC) formed around the receiver. Then trainer converts the information into a data point to classify it as blocked or not blocked. The accuracy of the classifier is close to 90%, and it incurs much less overhead on the communication system compared to the time window method. For simplicity, we refer to multipath components (MPCs) as paths throughout the chapter.

We summarize our contributions in this chapter as follows:

1. We develop a learning-based prediction method to classify link blockage and link movement efficiently and quickly.
2. We devise a data point structure to accommodate paths with different Angles of Arrival and develop a method to measure diffraction values of multipath components.
3. We investigate the learning features such as blockage burst feature to enhance prediction accuracy.
4. We evaluate different learning approaches to find the best training approach for link blockage/link movement prediction in terms of accuracy and speed.

We organize the rest of this chapter as follows. In Section 4.1, we discuss the related work on link blockage and link movement detection. In Section 4.2, we present an elaboration on the system and the required components. We discuss our proposed prediction framework in Section 4.3. We present the simulation results in Section 4.4.

4.1 Related Work

In this section, we discuss the state-of-the-art methods of link blockage and link movement detection and discrimination.

The authors in [57] proposed to detect link blockage by a sudden drop of SINR in static HDTV signal links. In more recent work, Sur et al. [58] proposed a method to predict the quality of all the paths around a receiver by probing just one path.

The closest work to our proposed method is from Nitsche et al. [59]. They propose a method to make link establishment in 802.11ad standard more efficient and faster by detecting whether link blockage occurs on a LOS path. However, their approach does not distinguish between link movement and link blockage. The authors evaluated the performance of the method in the form of LOS detection accuracy. Furthermore, the paper focuses on establishing an mmW link and does not consider maintaining the link in case of link blockage or movement. Our method predicts link blockage or movement to decide on the appropriate action. They also used the legacy 2.4/5 GHz band to do most of the sensing procedure, while our method is independent of the legacy frequency band.

Zhang et al. [60] presented a model to characterize the abrupt channel changes due to the receiver's blockage and orientation change. This model captures the abrupt changes of Angle of Arrival (AoA) and Angle of Departure (AoD) of the paths but does not accurately detect the movement of the link. In another paper [61], the authors propose a method to track signal-to-noise-ratio (SNR) in mmW networks, without addressing the problem of detecting link blockage and movement with the minimum amount of time consumed.

Oguma et al. [62] devised a reactive method for switching the base station when a link blockage occurs. In their work, they measure the throughput to detect only the link blockage. The authors did not consider the throughput degradation due to link movement. They also did not consider the time it takes to perform the detection procedure. Zarifneshat et al. [51] used blockage history to predict the possibility of future blockage incidents. Their method of link blockage detection did not consider the link quality degradation due to link movement.

Some research has been done in the field of 3D performance evaluation in wireless networks. Mordachev et al. [63] have proposed a statistical model for interference in spaces with various dimensions including

3D space. They use probabilistic geometry to obtain outage probability in wireless networks. However, interference is not a major issue compared to link blockage in the mmW because of directional signal propagation in mmW networks. For the mmW networks, Camisso et al. [64] provide a model for coverage probability in 2D and 3D propagation environments for the mmW networks. They provide the model for the reliability of the networks based on the interference as well. Samimi et al. [65] have gathered information on delay, angle of departure, and the angle of arrival for the mmW signal to design a model and simulation tool for the 3D channel in mmW networks that experts can use to design those networks.

Park et al. [66] measured the performance of different link blockage detection methods. The experiments show various time delays for different methods. The best time they measured is 15 ms for 100% accuracy. This time is only for detecting the link blockage. The methods cannot discriminate between signal strength variations due to link blockage and link movement. Yang et al. [67] have proposed a neural network approach to extrapolate the downlink channel information from uplink channel state information. They also propose a graph-based neural network to track the channel in a dynamic environment [68]. In our method, we use the signal pattern to detect the link blockage and link movement that happens in mmW networks. We define the problem in the mmW spectrum. This spectrum introduces new challenges that the researchers cannot address by methods designed for lower frequency bands [67, 68]. Therefore, the signal degradation due to link blockage and link movement is not as big of a problem as in lower frequency bands.

Alrabeiah et al. [69] have proposed an out-of-band method to detect the paths in mmW band like [59]. They use deep neural networks to achieve this goal. However, the authors used preexisting datasets to train their neural networks. There is no discussion of the overhead of training the neural networks in realistic settings. Ali et al. [70] have proposed an early link blockage warning method using the diffraction of sub-6 GHz signal for mmW signal. Since the authors have used the simulation platform to validate their method, they did not see the difference in the magnitude of diffraction between sub-6 GHz and mmW signals. In reality, the difference is not negligible, and this leads to inaccuracies of the method. They also did not measure the mmW band's diffraction in a real platform to see a meaningful difference. Gonzalez-Prelcic et al. [71] have discussed using the out-of-band signal for beam alignment in the mmW band. The authors do not mention the cost of beam tracking and try to reduce the beam tracking overhead by the out-of-band signal. Sur et al. [72] have used Wi-Fi (2.4 GHz) to find the strongest path in the mmW band and offload the traffic from the mmW band to Wi-Fi when there is no line-of-sight path in the mmW band. However,

they do not consider the beam tracking procedure. Relying only on the out-of-band signal to track beams in the mmW band can lead to inaccuracies as the out-of-band signal has its differences compared to the mmW band. Kumar et al. [73] have provided ray-tracing simulations to show how sparse the channel is in the mmW spectrum. They also use a sub-6 GHz channel to predict the behavior of the mmW channel. They use ray tracing to show a correlation between the two spectra.

Wang et al. [74] have proposed a regression model based on the signal strength over all possible beams to make a data point to be used to train the model. This gives them a model to predict the signal spectrum for beamforming in future instances. However, their method requires scanning all beams each time in model training that can be time-consuming. In another paper, Wang et al. [75] have proposed a classification and ranking method for beam selection in vehicular networks. Their method depends on the situational awareness provided to the system by the vehicular network. We cannot assume such information is available in all environments, like a sidewalk with pedestrians. Wu et al. [76] and Booth et al. [77] use probabilistic and signal processing methods to align beams for the transmitter and the receiver. The primary process is to select the optimum beams for both the transmitter and the receiver to maximize the link budget. While both methods try to maximize performance, they do not consider a proactive policy to avoid link blockage. Our method uses the channel conditions to sense that there is an impending link blockage.

Lin et al. [78] have proposed a dynamic base station formation algorithm to maximize users' rate. They apply the change in the base stations through the selection of them and the beams to use. Palacios et al. [79] propose a method of using localization techniques for better beam alignment in mmW. They use the direction of arrival of the signal for localization and ultimately perform fast handovers or beam alignment in mmW access points. Hashemi et al. [80] propose a multi-armed bandit model to reduce the exploration overhead via a stochastic optimization model. The goal of the optimization problem is to maximize the directivity gain. The model exploits the information of previous beam matchings as contextual information for future ones. Aykin et al. [81] have proposed a search scheme for multiple paths in mmW using multi-armed beams. They designed the beams to search multiple directions at once.

The above work considers either a link blockage or link movement. However, none of them take into account both jointly. In this chapter, we address the practical scenarios where both mobility and link blockage can occur. We also address the time delay that occurs during the detection of link blockage or link movement, which is very important as the proposed method targets to minimize the duration of the link outage.

4.2 System Primer and Design

This section discusses the system overview and different components required to realize our proposed system of link blockage and link movement detection. First, we present a primer on diffraction in mmW frequency. Second, we cover the method to compute the paths formed around a receiver. Finally, we introduce the learning methods.

To depict a high-level idea of the proposed system, we discuss the general aspects of it. First, the link blockage prediction system calculates the paths around each receiver. Then, it measures the diffraction value on each path. The data obtained from this procedure acts as one data point in the classifier's data space. The system feeds this data point to the classifier to determine its label, i.e., blocked or not blocked.

4.2.1 Diffraction in Millimeter-Wave Frequency

The link blockage prediction system performs the prediction task by measuring diffraction values on each path formed around the receiver. Since the link blockage affects the diffraction value on a link [10], we propose to use this value on different paths around a receiver to measure the chance of a LOS link between a sender and a receiver being blocked.

The experts define diffraction as the phenomenon of changing in the direction of a mechanical or electromagnetic wave when it reaches the edge of an obstacle. Physicists describe this phenomenon as bending the wave around corners of blocking object or an aperture into the shadowing region.

In mmW links, there is diffraction when an obstacle is close enough to the link. Since the transmitter and receiver antennas are antenna arrays for beamforming, Angle of Arrival (AoA) and Angle of Departure (AoD) are small. The LOS is the signal from the main lobe of the beam pattern in the transmitter that travels to get to the main lobe of the beam pattern in the receiver in a straight line. The authors of [10] have done experiments on α which is the angle between transmitter and receiver main lobe. For negative α values the receiver is in the shadow region, and for positive values, it is in the lit region.

The experiments show that for higher frequencies the diffraction is higher. For the metal blocking object, there is no signal penetration while there is some signal penetration in a wooden obstacle. They also compared their measurements with two well-known diffraction models. The first is Uniform Geometrical Theory of Diffraction (UTD) and the second is Knife Edge Diffraction (KED). The measurements of angle

experiments comply with the UTD model well.

In the second set of experiments in [10], the authors tested the effect of a metallic cuboid on a link between transmitter and receiver. The object acts as an obstacle, thus creating a shadow region behind it on the receiver side. They move the object on the x-axis while the transceivers sit on the y-axis. With x_0 being the center of the link, they move the object in small increments from negative x towards positive x. There is diffraction originating from both edges of the cuboid.

The measurements for the moving obstacle experiment comply with both UTC and KED diffraction models. In the shadow region, there is interference between the diffracted signal from both sides of the obstacle. The obstacle should be very close to the LOS, so the diffraction is perceivable (in order of 4-5cm). The less perception of diffraction is because of the lower diffraction values as the frequency increases. The decrease of diffraction causes the Fresnel zones around the link to be thinner. Therefore, an object should be close enough to LOS or any path so that the receiver can experience the diffraction.

For the human body as an obstacle, the authors in [10] also have some experiments. In this experiment, the authors model the human body as a double-edged knife. One edge is on the front of the body, and the other one is on the back. The proximity issue for diffraction is true in this experiment too. The human body should be close enough to the LOS path so that the receiver can detect diffraction. The authors conducted experiments in 60 GHz and 300 GHz.

The Knife Edge Diffraction (KED) model is a popular model for human originated diffraction. As stated in [10], the authors can define the model defined in (4.1).

$$\frac{E_{D,KED}}{E_{RX}} = \left(\frac{1+j}{2} \right) F(\nu) \quad (4.1a)$$

$$F(\nu) = \int_0^x \cos(t^2) dt + j \cdot \int_0^y \sin(t^2) dt \quad (4.1b)$$

$$\nu = x + jy \quad (4.1c)$$

In Equation (4.1), ν is a geographic parameter that is dependent on location of the obstacle with respect to the path that has diffraction. $F(\nu)$ is the Fresnel integral on complex number ν .

In another work, the authors presented a double-edged knife model for diffraction by a human body [37]. The authors consider the double edge property to both the front-back and top-down sides of the human body. This involves four sides of the human body to diffraction value. Their experiment complies with the KED

model they provided. The same property of getting close to the path for diffraction effect is true in their experiments.

4.2.2 Measuring Diffraction Value on a Multipath Component

This section elaborates on the method we use to measure diffraction on each path formed around a receiver. We use the method for diffraction value measurement. Our proposed method uses the diffraction values for predicting whether an mmW link quality drop is due to beam misalignment or link blockage.

The link blockage prediction system performs the prediction task by measuring diffraction values on each path formed around the receiver. Since the link blockage affects the diffraction value on a link [10], we propose to use this value on different paths around a receiver to measure the chance of a LOS link between a sender and a receiver being blocked.

We adapt the method we use to measure diffraction from a paper on indoor localization [82]. The authors implemented the method on commodity WiFi networks that work in the 2.4/5 GHz frequency band.

The authors define a power fading model (PFM) to include diffraction, path loss, and absorption fading. Based on the equations used in [82], the sources of power fading are propagation fading, diffraction fading, and target absorption fading. Depending on how the blocking object locates concerning the link, different summations of the mentioned fading sources represent the total power fading of a link. We use this model for diffraction fading analysis on the mmW link. The advantage of this power fading model is that it discriminates between different elements of the power fading process. The channel models used in the literature consider parameters that are not the focus of our chapter. They do not discriminate between different power fading factors as separate functions of distance and frequency. Therefore, we use the mentioned power fading model, which better serves the purpose of our defined solution.

The equation of diffraction fading used in [82] is as follows.

$$D_{ijt} = 20 \log \left(\frac{\sqrt{2}}{2} \left| \int_v^\infty \exp\left(\frac{-j \cdot \pi \cdot z^2}{2}\right) dz \right| \right) \quad (4.2)$$

$$v = h_t \sqrt{2(d_{it} + d_{jt}) / (\lambda \cdot d_{it} \cdot d_{jt})}$$

In Equation (4.2), D_{ijt} is diffraction value on link between transmitter i and receiver j caused by blocking object t . For parameter v , h_t is distance from highest point of blocking object to the link, d_{it} is distance between transmitter i and blocking object t and d_{jt} is the distance between receiver j and blocking object t .

The equation for diffraction fading is applicable to paths that create a single link between the transmitter and the receiver. For this purpose, Equation (4.2) holds for each path. Each path can be LOS or NLOS. If it is LOS, the transmitter i and receiver j are the transmitter and the receiver. If it is NLOS there are two cases. In the first case, we assume the blocking object is between a reflecting point and the receiver. In this case, we replace the transmitter i in Equation (4.2) by the reflecting point from which the signal path is reflected. In the second case, the blocking object is between the transmitter and the reflecting point. We treat this case like the LOS case in which a blocking object is between the transmitter and the receiver.

Before we continue with the total power fading model, we describe First Fresnel Zone (FFZ). The FFZ is an imaginary ellipsoid in 3D space. If the transmitter's signal is reflected off a blocking object in it, the received signal will have minimal phase shift (less than half of the signal wavelength) [83]. Other Fresnel Zones contain the previous one recursively. We consider the FFZ because as the frequency increases, the effect of blocking objects on the higher Fresnel Zones decreases [83]. This also applies to the 60 GHz millimeter wave band that we use in our simulations and experiments.

We denote the total power fading between transmitter i and receiver j by R_{ij} .

$$R_{ij} = \begin{cases} L_{ij} + D_{ijt} + A_t + \eta & LOS \\ L_{ij} + D_{ijt} + \eta & NLOS \text{ but in FFZ} \\ L_{ij} + \eta & Out \text{ of FFZ} \end{cases} \quad (4.3)$$

In Equation (4.3), L_{ij} is propagation fading of the link between transmitter i and receiver j , η is the measurement noise. The absorption fading of blocking object t , which the authors denote by A_t , and define as the signal absorbed by the blocking object when the blocking object is on the link's LOS path and blocked LOS path. There are three zones based on which the authors in [82] model the power fading. The absorption fading in the LOS zone of Equation (4.3) is due to the presence of a dynamic blocking object on a LOS path. A static blocking object did not block the LOS path. The algorithm does the detection of whether a blocking object is in any of those zones via a comparison of the difference of averaged Channel State Information (CSI) values. In order to set up this comparison, the authors in [82] defined different parameters. The first one is δ_{eff} which the authors define as the averaged standard deviation of CSI values over all subcarriers.

$$\delta_{eff} = \frac{1}{K} \sum_{k=1}^K \frac{f_k}{f_o} \times \delta_k \quad (4.4)$$

In Equation (4.4), K is number of subcarriers, f_o is central frequency, f_k is frequency for subcarrier number k and δ_k is the standard deviation of CSIs read in subcarrier k when there is no blocking object in the network. Raytracing algorithms can measure the CSI values of blocking object-empty networks for different network locations after measuring the CSI values at some anchor points. The network locations are different places within the deployed network that the algorithm measures the CSI or calculates through raytracing. The process of measuring CSI values of the empty network can also be done with several CSI measurements in the network. This process has its accuracy, but the overhead can be high for large environments (scalability issue). We can also use a hybrid method of both raytracing and in-field CSI measurement. We need to perform this process only once.

The second parameter defined in [82] for the purpose of power fading model selection is the effective CSI change, denoted by ΔCSI_{eff} .

$$\Delta CSI_{eff} = \frac{1}{K} \sum_{k=1}^K \frac{f_k}{f_o} \times (F_k - O_k) \quad (4.5)$$

In Equation (4.5), F_k and O_k are CSI values read from subcarrier k when a blocking object is in FFZ of a link and when there is no blocking object in the network, respectively. The authors also call O_k baseline CSI reading for subcarrier k .

Based on parameters defined so far, the authors in [82] formulated the conditions for selecting different power fading models as follows.

$$\begin{cases} |\Delta CSI_{eff}| > |A_t| & LOS \\ \delta_{eff} < |\Delta CSI_{eff}| \leq |A_t| & NLOS \text{ but in FFZ} \\ |\Delta CSI_{eff}| \leq \delta_{eff} & Out of FFZ \end{cases} \quad (4.6)$$

In Equation (4.6), if any of the conditions are met, then the blocking object is in that specific zone. In this way, the receiver can select the appropriate power fading model in Equation (4.3). We should note that the inequality $\delta_{eff} < |A_t|$ holds because the value δ_{eff} is a standard deviation of CSI values when there is no blocking object. Therefore, its magnitude is not high. On the other hand, $|A_t|$ is the absorption value of the target that is high especially in the mmW band that the human body absorbs most of the signal.

With N transmitters and M receivers, there are MN possible links in the network. For each link with a transmitter and receiver, there are two to three paths: LOS and the rest are NLOS [21]. There are two or

three paths that the wireless hardware resolves. In theory, the number of paths can be more than that, but the number of paths that the hardware can resolve is limited to two or three paths. The two to three paths resolved by the hardware are of different clusters [2]. The hardware cannot resolve the intra-cluster paths because their characteristics are very close to one another. When we discuss the paths, we mean the path clusters and not the clusters' multipath components. The resolution of the paths also means the resolution of the path clusters.

According to Equation (4.3), for LOS, there are three unknown parameters, including x and y coordinates of the blocking object and the effective height of blocking object h_t . We assume that there is one blocking object for a receiver that impacts the paths around it. For the LOS path, the system knows the location of both transmitter and receiver. For NLOS paths, the system knows the receiver's location, but it does not readily know the location of the reflective point from which the signal is bounced to the receiver.

The assumption of having one link blockage on each link may seem strong. In practice, in populated areas, the bodies of humans move in reasonable proximity. This closeness makes a group of people a bigger blocking object. Therefore, the assumption of having one blocking object in a specific direction is not strong enough to invalidate the method. Although there are cases of having more than one blocking object on a link, one can estimate the strongest paths reaching the receiver by performing the path discovery method for each blocking object separately. The height at which we deploy the base stations can be high and the assumption still holds. The proximity of the blocking object to the mobile device, that happens in the populated areas, can make the link blockage happen despite the height at which we deploy a base station.

We can obtain the location of the transmitter and the receiver quite easily. Since we design the method for the outdoor cellular network, the system can obtain the mobile nodes' location from the GPS of the device. If the GPS information is not available, the base station can measure the location of the mobile device through signal-based localization methods [84]. Such methods are especially favorable in mmW communication due to its directional links. The location of base stations (like picocells) is fixed and determined at the time of installation.

To discover the location of reflective points, Wei et al. [85] proposed a method that we use to locate the ambient reflectors in the environment. In this method, the authors introduce two primary mechanisms. The first mechanism finds different paths from the transmitter to the receiver. The second mechanism uses the parameters for multiple paths that the algorithm finds to reconstruct dominant reflectors. For finding

different paths' directions, the receiver receives the signal on each antenna element. The MUSIC algorithm estimates the directions of the paths for both transmitter and the receiver. Using phase shift in signals, the method can know each path's distance from the transmitter to the receiver.

In the second mechanism, we estimate each reflecting point's location using the parameters estimated in the first mechanism. The algorithm estimates the location of the receiver by using two signal paths from the transmitter to the receiver. In combination with the signal paths, the receiver's location can give us the location of the reflector. The mechanism continues with the estimation of the orientation and reflectivity of the reflectors. Finding the reflectors by one set of transmitter and receiver (BS and UE) can be used for other sets of BS and UE because the reflectors are of the propagation environment's static features. This algorithm's time complexity is not negligible and we need to consider it in the system overhead assessment. However, we should note that the system does this process of finding reflection points in an offline manner and it does not affect the algorithm's online performance, which is critical. The system performs the process once, and it can use the information for the future. There can be some regular updates as the static objects on the map may change over time.

Some regular updates of the reflector map are inevitable due to new furniture or buildings. For very dynamic objects like cars, trucks, etc., we have some new paths created due to the reflections of such objects' metallic surfaces. Moltchanov et al. [86] have provided a model for the length of the time interval of blockage and non-blockage as a function of blocking object and mobile node speed. Based on their model, the time interval of the blockage state of the link increases with the mobile node's speed. Let us consider the reflector a mobile node, according to their results. The time interval of the blocked state for a link that comes from a reflector on a moving vehicle is long enough to neglect its effect on the paths formed around a mobile device that a human carries. Moreover, the proposed algorithm's online part can adapt itself to major and general changes in dynamic reflectors in the long term.

As we will see further in Section 4.5, the third dimension's effect in our experiments is limited. In our hardware setup, the codebook used for scanning the space scans the elevation angle of zero, that is, the plane in which we place the devices. We conduct the majority of our experiments in indoor environments. The indoor environments tend to create more signal paths due to the presence of the ceiling. The several paths in an indoor environment may create the need for having high-resolution 3D raytracing. However, our experiments (Section 4.5.4, Figures 4.20 and 4.21) show that not considering the other elevations would not

decrease the prediction accuracy.

After getting the location of the reflectors of NLOS paths, we have three equations with three unknown parameters. Each path has one PFM equation with actual CSI reading. The system computes the unknown, which are x and y coordinates of the blocking object location and effective height of the blocking object, by solving each of the equation systems.

We must mention that the algorithm performs all these procedures if the conditions for the blocking object being in FFZ are true. If this is not the case, the diffraction value of the blocking object on the path is zero. Once the system knows and optimizes the unknown parameters, it computes the diffraction value of the blocking object on the MPC by using Equation (4.2).

For path detection, we use a method to discover the paths formed around a receiver due to beamformed signal transmission and reception. Sur et al. [58] proposed a method to discover the paths formed around a receiver. Their approach uses two main observations that the authors experimented with the validity of those observations. First, there is a limited number of paths formed around the receiver, and second, for a path that is covered by two or more beams, the path's qualities read by the beams often correlate. We use the method to compute the paths formed around the receiver.

In the following, we discuss a brief analysis of the link blockage probability. According to [14], the authors define the probability of a link being LOS as $p(R)$ in which R is the length of the link. The probability of a link being blocked (NLOS) is $1 - p(R)$. When the blocking objects are modeled as a rectangle like in our simulations, the LOS probability is $p(R) = e^{-\beta R}$ where β is a parameter that the system determines by the density and size of the rectangular blocking objects. In the equation for the LOS probability $\beta = 2\lambda(\mathbb{E}[W] + \mathbb{E}[L])/\pi$ [87]. W and L are the random variables for the width and the length of the blocking objects and λ is the density of the blocking objects. As we can see, the probability of LOS decreases as the length of the link increases since the blocking objects are more probable to block the longer links. The mean of both width and length of the blocking objects have an inverse relation with LOS probability. Higher blocking object density also decreases $p(R)$.

4.2.3 Discovering the Multipath Components Around Receiver

In this section, we discuss the method used to discover the paths formed around a receiver as a result of beamformed signal transmission and reception.

Sur et al. [58] have proposed a method to discover the paths formed around a receiver. Their method uses two main observations that the authors did experiments on the validity of those observations. First, there is a limited number of paths formed around the receiver, and second, for a path that is covered by two or more beams, the path's qualities read by the beams often correlate. They have also shown that the structure of the paths does not change in the presence of a human blockage. The only thing that changes is the quality of affected paths. They show that human blockage does not create new paths as the signal reflection of the human body is not strong.

To set up the prediction framework, the authors have proposed a model called path skeleton. Path skeleton is the parameters obtained from signal readings by changing beams around the receiver. Whenever the current path quality drops, the system can predict the quality of other paths by probing the current path and selecting the best one. The receiver can perform the complete scanning to obtain a path skeleton whenever the transceivers are going to establish a link. The path skeleton is valid until either transmitter or receiver moves. If that is the case, there is the need to renew the path skeleton by performing the scanning process again.

A receiver has K number of beams that V number of paths around it. When the transceivers need to establish a link, the receiver probes all the beams in the receiver's antenna to measure channel impulse response (CIR). In fact, CIR is aggregated effect of all paths reaching the receiver. There are two sides to the CIR value. First, the authors have modeled the CIR value in each beam as a function of the number of paths, the angle of the paths, their amplitude, and phase. On the other hand, there are actual measurements of the beam CIRs. The goal is to optimize the paths parameters by minimizing the difference between modeled and measured CIRs. The model of CIR is as follows.

$$h_k^{rep} = \sum_{i=1}^V A_k(\theta_i) \cdot a_i e^{j\phi_i} \quad (4.7)$$

In Equation (4.7), h_k^{rep} is modeled or reproduced value of CIR for receiver's beam number k , V is total number of paths reaching to the receiver, A_k is array factor or gain at a direction for beam number k , θ_i is the angle of arrival of path i , a_i and ϕ_i is the amplitude and phase of path i . For each beam, the method uses the same model. $H^{rep} = \{h_1^{rep}, \dots, h_k^{rep}\}$ is the vector of CIR models for all beams.

The measured CIRs for each beam are gathered in a vector $H^{ms} = \{h_1^{ms}, \dots, h_k^{ms}\}$. The optimization

problem we need to solve is,

$$\{\hat{p}_1, \dots, \hat{p}_V\} = \arg \min_{\{\hat{p}_1, \dots, \hat{p}_V\}} \|H^{ms} - H^{rep}\|^2 \quad (4.8)$$

In Equation (4.8), an optimizer computes V paths by minimizing the difference between modeled beam CIRs and their measured counterparts.

The authors of [58] define a binary variable. The task of this variable is to determine which path is blocked by an obstacle. Since h_k^{ms} is CIR for beam k and contains the effect of all paths on this beam, the authors add the binary variable to this model to determine obstacle blocks which path according to the CIR measured. In general, this part of their system is not directly related to our proposed system.

4.2.4 Learning and Classification Methods

The proposed prediction system for predicting link blockage has a learning system in its heart that learns the specifics of the diffraction values on different paths reaching the receiver. In this section, we discuss different learning modalities that we consider in the proposed framework. We also elaborate on different classification methods used in the proposed system.

4.2.4.1 Learning Modalities

The machine learning community calls two main training modes supervised and unsupervised [88]. In supervised training, each data point that is a point in data space consists of a feature vector and one or more labels. In this way, when a specific feature vector is ready, the trainer would know what the label of that data point is. On the other hand, in unsupervised training, the data point includes the feature vector and does not have the label.

In the problem that we try to solve, the data points consist of a feature vector that contains diffraction values and a label showing whether that feature vector indicates a link blockage situation or not. This structure of data point in the problem makes the system do a link blockage detection procedure once for each data point. It means the system does a link blockage detection procedure even though the purpose of our proposed system is to minimize the number of link blockage detection procedure in its traditional way (i.e., having a time window of RSSI recordings) and waste time. That is the exact procedure we want to

avoid by proposing the prediction framework. Therefore, the proposed prediction system needs to learn in an unsupervised way.

The other type of training method categorization is online or offline training [88]. In offline (batch) training, all data points required for training the classifier are available. When the data points train the classifier, it does the prediction task. While predicting the labels, the classifier does not change as it sees the new data points. Applications use this type of training when a big enough data set is ready at the time of training, and the data points in the future do not change much. Therefore, there is no need to update the classifier as new unlabeled data points are available.

In online (sequential) training, the data set required to train the classifier is not available at first to train the classifier. Instead, the data points become available one at a time. Such systems may have huge data sets that make it impossible to process the whole data set at once. Therefore, we require to process them incrementally [88]. The other characteristic of such systems is that they are dynamic. The parameters of data points tend to change over time. Therefore, it is reasonable to make changes to the classifier as it sees the new data points over time. As a new data point becomes available for label prediction, the classifier assigns a label to it, and then based on that label, it updates the classifier weights in order to include the changes the new data point introduces to the classifier.

4.2.4.2 Training Methods

In our proposed prediction framework, we use different training methods to compare their accuracy of link blockage prediction. In this section, we briefly explain the training methods.

The most popular unsupervised learning algorithm is clustering [88]. In this method, the clustering algorithm groups the data points that belong to the data space in such a way that the distance between points in the group is less than the distance between points of different groups.

In our prediction framework, we use the K-means clustering algorithm [88]. In this algorithm, K is the number of groups or clusters. In the problem we define, the number of clusters is two, with no link blockage and link blockage. We use this algorithm in two modes, online and offline. In offline mode, all points are ready to form the clusters. In online mode, the points become available one by one. The clustering algorithm updates the cluster centers after seeing each data point.

The other method used is the classification method. In the particular case that we use in our experiments [89], there are two phases. In the first phase, we train a cascade classifier with labeled data. In the second phase, unlabeled data points become available one by one, and the classifier predicts each data point's label. Then the trainer updates the classifier by the new data point.

The classifier that we train in the offline stage is a cascade classifier. In cascade classifier, there are several numbers of smaller classifiers called base classifiers [89]. The cascade classifier predicts label of the data point based on partial sums of the base classifier. Each base classifier t has a value of itself h_t and a threshold b_t . The classifier labels a data point \mathbf{x} positive if $\forall t \in \{1, \dots, T\}, H_t(\mathbf{x}) > b_t$ in which $H_t = \sum_{i=1}^{i=t} h_i$. T is the total number of base classifiers.

Each data point for training the cascade classifier has a weight. The training algorithm initializes the weights to a value that is a function of the number of data points. The training algorithm can use each feature of the data point for the current base classifier. The method divides the feature values into intervals to form a histogram. Then based on which interval a feature value falls, the trainer computes the base classifier for that feature. The trainer chooses best base classifier based on how well a base classifier can discriminate between positive and negative data points. The trainer chooses the best base classifier via computing a function of the weight of positive and negative samples. A small portion of data is enough to train the cascade classifier.

In the second stage, we do the online boosting. In this stage, as each data point is available, the trainer feeds it to the cascade classifier to determine its label. Then the trainer uses this label to update the base classifiers and their threshold values. The online boosting algorithm updates the weight distribution according to the predicted label of the new data point. Based on the updated weight distribution, the trainer also updates the value of the current base classifier. The online boosting algorithm updates weight of the sample by using an updated base classifier value. There is a mechanism to stop sample weight from growing excessively while the algorithm uses the base classifiers. The online updater also can update the length of the cascade classifier.

The classifier may have a prediction error. If the trainer only relies on the predicted label by the classifier, the model can lose its accuracy over time because the prediction error can accumulate. To decrease the effect of prediction errors on the model, the online trainer can use the real label obtained to train the model. The trainer obtains the real label using the legacy method of windowed recording of SNR values for detection of the link blockage. In this way, the trainer uses the real label periodically. This can calibrate the model and

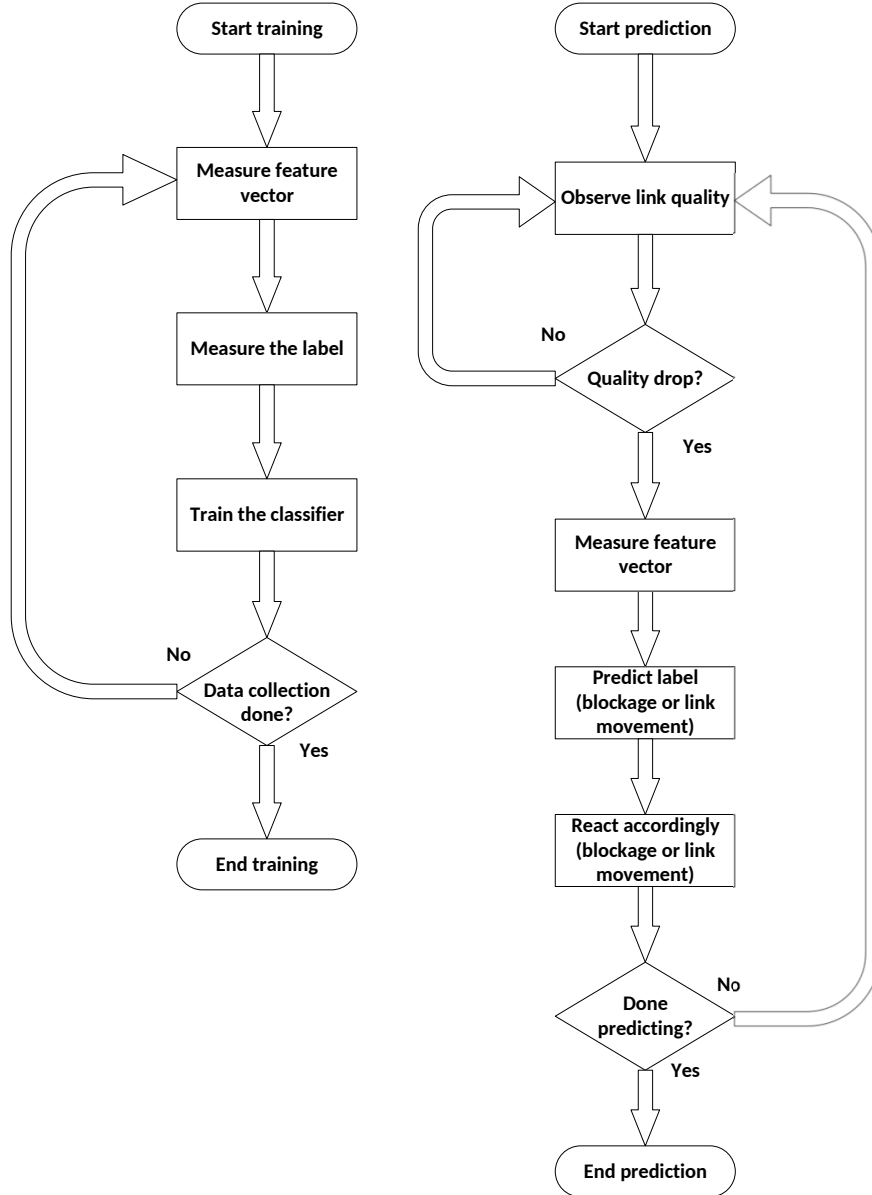


Figure 4.1 Flowchart of training and prediction phases

slow down the effect of the prediction error accumulation on the classifier.

The last set of classifiers used in our experiments is cross-validation with a group of classifiers. This experiment's difference from the previous classifier is that the trainer does not update the classifier when it predicts the label of a new data point. This set includes ensemble, k-nearest neighbors, and support vector machine [88].

We show a flowchart of both training and prediction phases of our proposed system in Figure 4.1.

We train a classifier to classify the feature vectors. The feature vectors contain the diffraction values

on the paths formed around the mobile receiver. blocking objects around the mobile receiver cause the diffractions.

The data set size dictates how much space we require for the training and maintaining the model. Although some of the methods we use for training rely on a limited data set, we use online training schemes that theoretically can run forever. We show in the evaluation section (Figure 4.10) that the size of the data set required to reach an acceptable accuracy is not large.

4.3 Feature Vector Generation

This section introduces a key element in the learning model, which is the feature vector generation. This section has two parts. In the first part, we consider how we make different features from diffraction values on the receiver's paths. In the second part, we discuss another feature that enhances the performance of the predictor. This feature considers the burst of link blockage incidents.

4.3.1 Feature Generation Based on Diffraction

The main point of using diffraction values for generating features in our proposed prediction framework is that whenever there is a link blockage, we have diffraction happening in the shadow region caused by the blocking object. We argue that by measuring the value of diffraction, we can predict the link blockage incident.

Because of the reflections of different static environmental objects like buildings, walls, and boards, the signal reaches the receiver in more than one path. Each path is like a link, and transceivers can use it for communication. A blocking object can have a diffraction effect on any of these paths. The receiver perceives the diffraction as a component of power fading, and it can calculate diffraction by using a power fading model.

To generate features from diffractions on links, we need to have a template for diffraction values. The paths reaching a receiver can change as either transmitter or receiver moves. The movement of transceivers makes the paths change their AoA. It is also possible that the number of paths changes while transceivers change location. The location change makes some paths disappear, and some others appear on a different AoA. To have a constant number of features for each data point and have each feature the same meaning, we propose to divide the angle space around a receiver to C sectors with equal angular width. Depending on

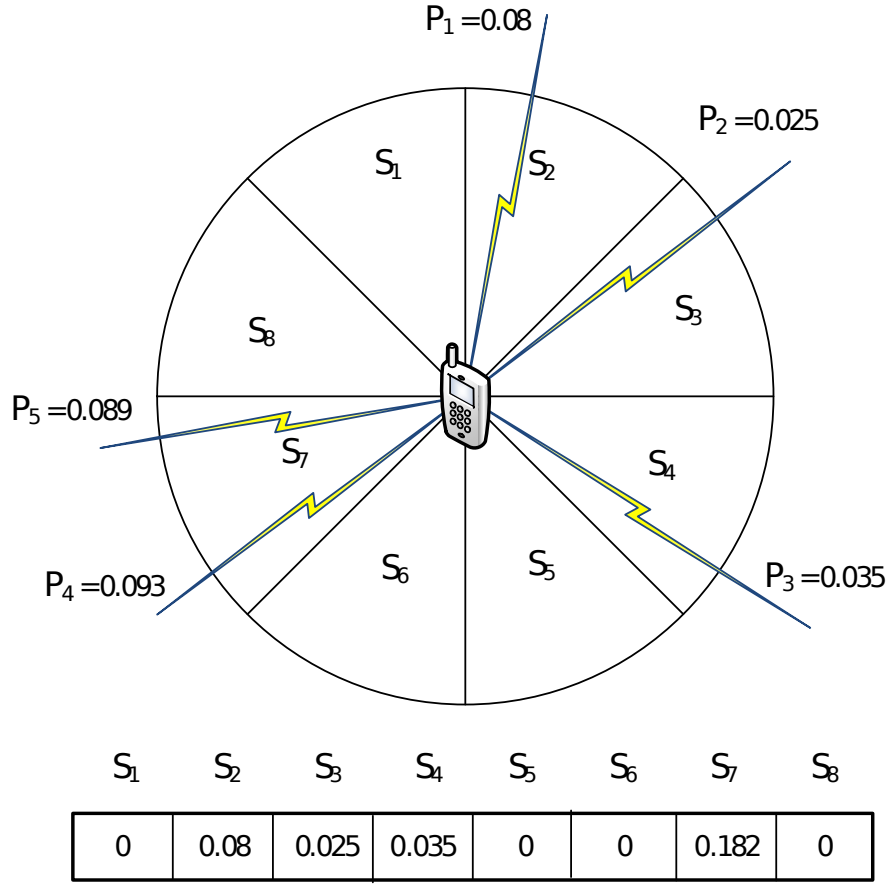


Figure 4.2 Sectors are drawn around receiver and features associated with them

the AoA of each path, each of them falls into one of these sectors. Some sectors may have more than one path, and some of them may have no paths.

One should note that both the AoA and AoD parameters can change with the movement of transceivers. For our application, we only need the AoA. However, due to channel reciprocity, the AoD in one link direction is the AoA when the link is reversed.

The trainer generates the data point features based on paths that fall into each sector around the receiver. If a sector has no paths within its angular limits, the respective feature gets zero. If a sector has one path with AoA within its limits, the respective feature value would be the diffraction value due to the presence of a blocking object in the proximity of the path. If there is more than one path falling in a sector, the respective feature is the sum of the diffraction of the sector's paths. The diffraction values are the absolute values of complex numbers that define diffraction. Figure 4.2 shows an example of the feature generation procedure from diffraction values on paths around a receiver.

Figure 4.2 illustrates an example of how the algorithm uses the diffraction values to make a feature vector. There are five paths formed around the receiver in the center. We name the paths $P = \{P_1, \dots, P_5\}$. In this example, we divide the space around the receiver into eight equally sized sectors called $S = \{S_1, \dots, S_8\}$. Each path has a diffraction value due to the presence of a blocking object in its proximity. Each path belongs to a sector depending on its AoA. Therefore, there are some sectors without any paths in them, some sectors have one, and some have more than one path in them. The algorithm adds the diffraction values of the paths P_4 and P_5 together to represent the seventh feature's value. The sectors $\{S_1, S_5, S_6, S_8\}$ do not contain any paths, so their respective features get zero. We show the feature vector at the bottom of Figure 4.2.

4.3.2 Burst Feature

A data point with just diffraction values as features cannot reflect the whole situation that the data point represents. A critical condition in the environment is the burst of link blockages. Suppose there is no consideration of burst of link blockage. A data point sampled from a very bursty time interval has the same effect on training the classifier as the same data point sampled from a time interval that link blockages are happening sparsely. This is due to the fact that the feature vector without the burst feature only has information of the diffraction values at the time of sampling and no memory incorporated in the data point. Therefore, it does not make any difference the trainer trains the classifier by either data points. The feature vector without the burst feature lacks the memory element of the burst feature. It only contains the diffraction values for that instance of time that the trainer collected the sample. For example, the trainer cannot see an impending link blockage (or a link blockage in the time that has just passed) just by looking at the diffraction features and it needs the memory element in the burst feature to include the impending link blockage in its picture of the environment that it sampled by the data point. Therefore, relying only on the diffraction values cannot reflect the complete information that we are sampling. Because of that, we use an extra feature to capture the link blockage burst aspect of the environment. Since the feature vector with the diffraction values only looks at the environment at the time of sampling, a memory of the previous link blockages can help the trainer see the environment in a more inclusive manner. This is how inclusion introduced by the burst feature helps the classifier enhance its accuracy.

There are two approaches to address the link blockage burst issue. The first one is to have a counter that counts the number of link blockages in a time window. This approach does not fully capture the link

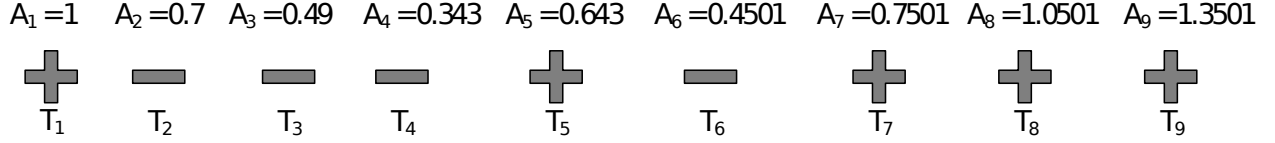


Figure 4.3 Example of EMA value change in a series of samples

blockage's burst as it cannot discriminate between different moments of link blockage bursts within the time window.

The second method is to use the Exponential Moving Average (EMA) [90]. The authors devised the method to dampen the effect of a data instance as time passes. The weighting factor decreases over time but does not reach zero. The value of 1 accounts for the link blockage incident, and 0 accounts for no link blockage (as in labels of data points). The equation which we use to define the burst feature is as follows [90].

$$F_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1 - \alpha) \cdot F_{t-1}, & t > 1 \end{cases} \quad (4.9)$$

In Equation (4.9), F_t is the value of exponential moving average of link blockage incidents at time t , Y_t is the link blockage incident at time t which can be 0 or 1 and α is the weighting decrease parameter which is between 0 and 1. If α is higher, it means the earlier observations lose effect in the average faster. The formula shows that when there is no link blockage, the EMA value decreases with the rate of $1 - \alpha$. When there is a link blockage, the formula adds the value of α to the EMA value. The classifier predicts the values of link blockage incidents as the last prediction.

To better describe how the burst feature can capture a link blockage burst, we provide the following example. Figure 4.3 shows an example of EMA value change with the change of link blockage incidence. Figure 4.3 shows a series of link blockage samples, and for each of them, we calculate the EMA value. In this example, we assume $\alpha = 0.3$. There are nine samples in this example. The plus sign (+) is an indicator of link blockage happened, and the negative sign (-) indicates that link blockage did not happen. As we can see for the first sample in T_1 , there is a link blockage, and according to Equation (4.9), EMA equals 1. For the samples in T_2 to T_4 , we see a decreasing amount of EMA. That is because time passes, but there is no link blockage. Thus the effect of link blockage in T_1 diminishes over time. At T_5 , there is a positive link blockage increasing the EMA with $\alpha = 0.3$. The increasing trend from T_7 to T_9 is obvious due to positive link blockage. The burst feature is one scalar, and we add it to the feature vector. As we can see in this

Algorithm 4.1 Training algorithm of the classifier

- 1: **if** *link quality drops* **then**
- 2: Measure the diffraction values and use burst feature (feature vector)
- 3: Decide on the reason of link quality drop by method [7] (the label)
- 4: Train the classifier
- 5: Update the burst feature by the measured label
- 6: **end if**

Algorithm 4.2 Prediction algorithm by the classifier

- 1: **if** *link quality drops* **then**
- 2: Measure the diffraction values and use burst feature (feature vector)
- 3: Predict label
- 4: Update the burst feature with the predicted label
- 5: Update the classifier's weight vector (online boosting)
- 6: **end if**

example, by having more recent link blockages the EMA increases and this increase indicates a more bursty time interval. When there is no link blockage, the EMA value decreases and this indicates a less bursty time interval. We then use the EMA value as the burst feature in the feature vector. Since this is a moving average, the average indicates the effect of link blockages for a limited time and the average forgets older link blockages. We use the parameter α to adjust how forgetful the EMA should be.

The general procedure of the prediction is as follows. When there is a degradation of the link quality, to detect whether the cause of this degradation is link blockage or link movement, the receiver measures the diffraction values of the paths around itself. Then, the algorithm generates the data point for the diffraction values according to the AoA of the paths. After that, the predictor receives the data point after being trained. The predictor (classifier) predicts the label of the data point. If it predicts link blockage, then the link quality degradation is due to link blockage. Therefore, the receiver should take appropriate actions to improve the link quality. Otherwise, the degradation is due to link movement, and the transceivers need to take appropriate action (like transceiver beam adjustment). Our proposed method's benefit is that it prevents excessive windowed RSSI measurements and leaves the link in the blocked state. We elaborate on the training and prediction procedures as follows. According to Algorithm 4.1, the system performs the training process when there is a link quality drop. The system updates the burst feature with the measured label and it uses it to construct the next feature vector.

In Algorithm 4.2, the trainer performs the online boosting procedure if the time has come based on

the frequency that it selects for the online boosting procedure. Since for online boosting, the trainer needs to measure the label traditionally. There is a trade-off between the frequency of online boosting and the accuracy of the prediction.

One can argue that the path discovery method introduced in [58] is time-consuming itself. We should state that the method described in that section is a precise method for the path discovery process. If that is not the primary concern and with some compromise on the accuracy of paths, we can use the method in [85] to achieve the same goal. With lower path accuracy, the method in [85] discovers reflecting points and then feeds them to a raytracing software to come up with the AoA of the paths and their qualities in different locations. The nodes in the network can perform this procedure once.

According to the training methods, the unsupervised training methods do not need a label for training. Therefore, they do not need any usage of the traditional method used in [7] to know the ground truth. We experiment with unsupervised training methods to see if we can solve the problem without the labels. Since acquiring the labels incurs overhead to the system, the unsupervised classification saves extra effort. We apply unsupervised learning (clustering) to this problem by assigning labels to clusters. We assign the labels to clusters by measuring the accuracy. If the accuracy is maximum for an assignment, the system chooses that assignment for all of the simulations.

For the supervised methods, whenever there is the need for a label, we use the method in [7] to get the label and assign it to the feature vector. As we see in the results section, the needed offline training is not much to achieve satisfactory accuracy.

The system's overhead to generate the feature vectors is not low for a naïve setup. The process of measuring CSI has the overhead of scanning on all antenna elements. However, we can further enhance the system by reducing the number of CSI measurements (measuring AoA or AoD) by choosing 10% of total points in the area for anchor points for CSI measurement. Then, using raytracing tools, it is possible to get the measurements of other locations without CSI measurement. It is also possible to use the reflective surfaces and their orientations to enhance the raytracing procedure's performance.

The proposed link blockage prediction algorithm can have various applications in link blockage avoidance. By detecting the link blockage in time, the transmitter can adapt the rate proactively, unlike the traditional methods that wait for a link quality degradation and lower the rate. The rate adaptation is one method of dealing with an impending link blockage. Another method is switching the beam to a direction

with a better-quality signal path without experiencing data loss. We can apply the same approach, but instead of switching to another beam, the receiver switches to another base station.

4.3.3 Time Complexity Analysis

This section elaborates on the time complexity of the method proposed for predicting link blockage and link movement. We provide a thorough time complexity discussion of all steps involved in the proposed method.

The first stage of the proposed prediction framework is training. In the training phase, the method does two main tasks. The first task is to find the signal paths around a receiver and compute the diffraction values due to blocking objects in the receiver's proximity. The path-finding procedure, as stated before, can have a high overhead if the method uses traditional SLS on every point in the area. However, we propose using the raytracing method to extend the effect of SLS in one location to other locations. This would decrease the number of SLSs and the overhead. For the feature vector generation in both training and testing phases, according to [82], the time for solving the optimization problem to measure the location of the targets is about 65 ms. This time is for the case of 21 clients. In our proposed solution, the number of clients (nodes) is one, but there are two or three signal paths. As mentioned in Section 4.2.2, there are three equations and three unknowns. Therefore, the time for solving the optimization problem is reduced significantly since we can solve the equation system one by one. For the training phase, we also need to know the label. We perform the traditional method introduced in [7] to know the label to train the classifier.

For the second phase, which is the prediction, we only need to generate the feature vector and then feed it to the classifier to classify. The classification does not take variable time, and it can be of $O(1)$ order. The algorithm performs the other parts before (reflector positioning and raytracing) and are not online. On the other hand, having a continuous SLS costs a high overhead. Besides that, the system can benefit from the added information of signal paths with lower overhead.

4.4 Simulation Results

In this section, we present the simulation results of our proposed prediction framework. We also include the implementation of the method on an mmW hardware platform. The platform consists of two tri-band routers that can communicate in the 60 GHz frequency band. In the following sections, we introduce the simulation framework and the hardware testbed.

4.4.1 Simulation Framework

The simulation framework has two parts. The first part of the simulation framework is the package responsible for creating the environment and moving the transceivers and blocking objects. The simulation framework also uses raytracing software [91] to generate the paths formed around the receivers due to objects and reflecting surfaces. Although the raytracing software is designed for indoor environments, it is possible to adjust different parameters to make the software adapt to the outdoor environments. For example, we adjust the reflectivity of the surface of blocking objects in the simulation area as the objects that we can find in an outdoor environment like concrete, glass, or metal. We also set the power of the transmitters according to outdoor BS settings. The raytracing software maintains spatial consistency. The environment is the same in all our simulations, and the mobility model for all mobile objects is the same throughout the simulations.

We collect the data for both LOS and NLOS areas. First, our mobility generator generates the locations on the mobile objects in the simulation area. Then, we feed those locations to the raytracing software to know the mmW channel for them. If a mobile user has a clear path to AP, then it will have a LOS path.

The first part of the simulator is also responsible for extracting the paths' diffraction values based on the Knife Edge Diffraction (KED) model. Then, it generates datasets for training and testing the predictors. The second part of the simulation framework receives the data sets generated by the first part. It uses them to train different learners and test their accuracy against the ground truth.

There are three parameters that we compute to compare the performance of different methods with each other. The first parameter is accuracy. Accuracy is the ratio of the number of data points predicted correctly according to the ground truth to the total number of data points. False Positive Ratio (FPR) is the ratio between the number of data points falsely predicted as positive (link blockage incident) and the total number of data points. False Negative Ratio (FNR) is the ratio between the number of data points falsely predicted negative (no link blockage) and the total number of data points.

The simulation environment comprises two fixed location transmitters on the north and south sides of the simulation area. Five receivers are moving east to west or in the opposite direction. Twelve blocking objects move from east to west or opposing directions. The number of data points in each data set is 500 for each receiver. Figure 4.4 shows different components in the simulation area. The figure shows the components' layout, and the number of each component is not necessarily accurate. Throughout the simulations, the simulator generates areas with both LOS and NLOS paths. Therefore, the trainer performs the samplings

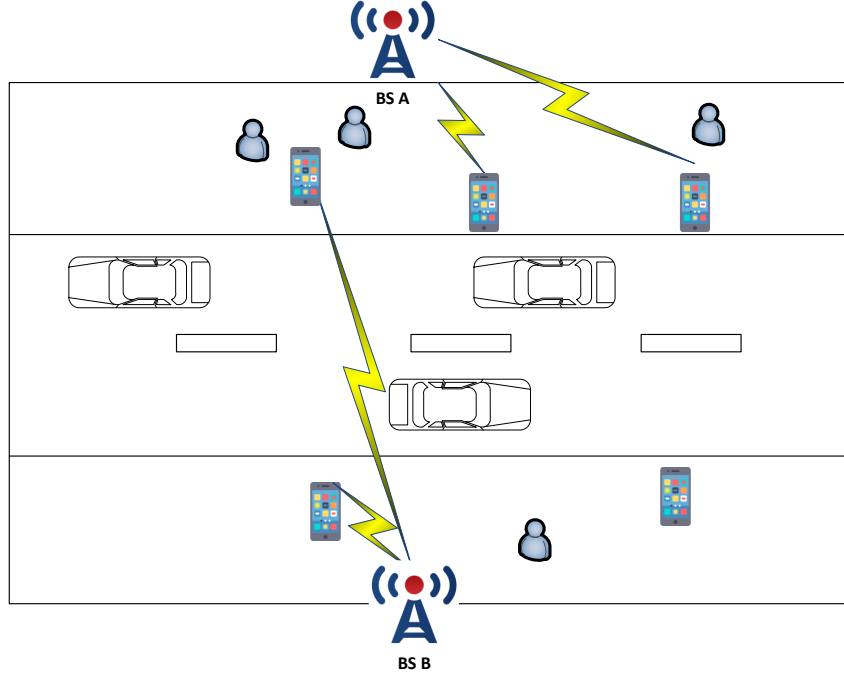


Figure 4.4 The layout of the simulation area

for both types of paths. Since the NLOS paths are weaker than LOS, the diffractions and consequently the feature vectors from diffractions are weaker and have limited contribution to training the model.

4.4.2 Results

The first experiment is about the cascade classifier's prediction accuracy we elaborate in Section 4.2.4. There are two types of experiment results presented here for this classifier. The first figure is the accuracy of the batch training and online boosting for each receiver when we use 30% of the data set for batch training and the rest for online boosting. According to Figure 4.5, the accuracy of prediction for all receivers is high. The highest is for receiver number 3, which is close to 90%. The accuracy is a significant improvement since it prevents the excessive RSSI recording for a significant number of the cases and shortens the detection to a label prediction. To further improve the accuracy, we should increase the variability of the feature vector across the dataset. It means that we should add other features to the feature vector like time and information of the movement of the mobile devices. We also included the effect of online boosting in the figure. For most of the receivers, the FNR is more than the FPR. It indicates that the cascade classifier tends to predict no link blockage rather than link blockage when predicts incorrectly.

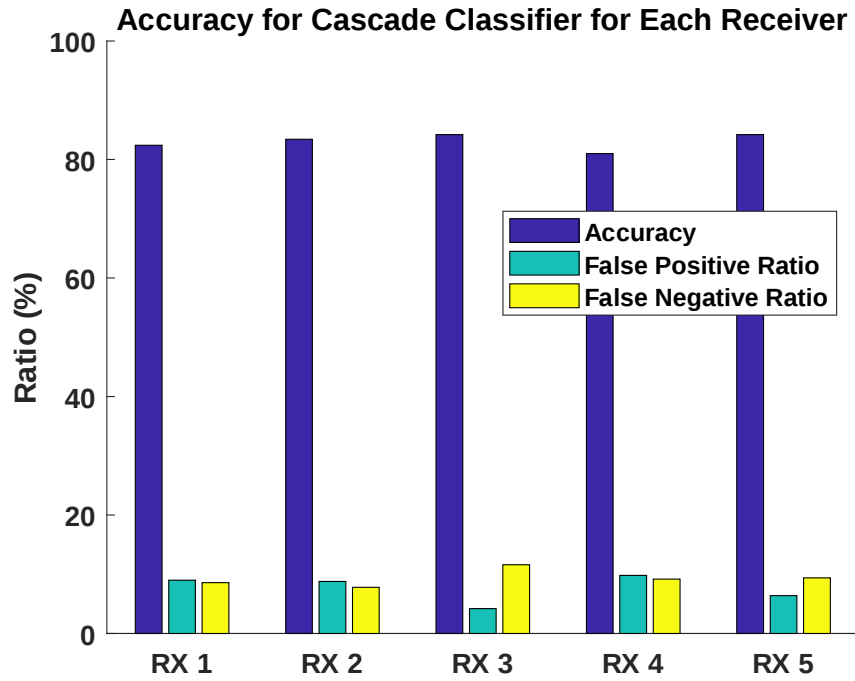


Figure 4.5 Accuracy of cascade classifier trained for each receiver

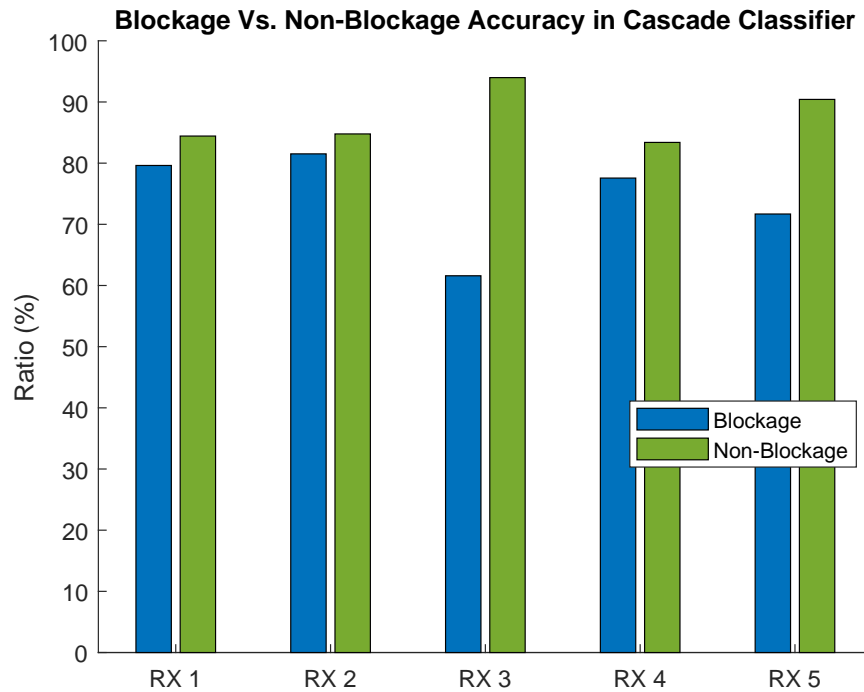


Figure 4.6 Link blockage vs. non-link blockage accuracy of cascade classifier trained for each receiver

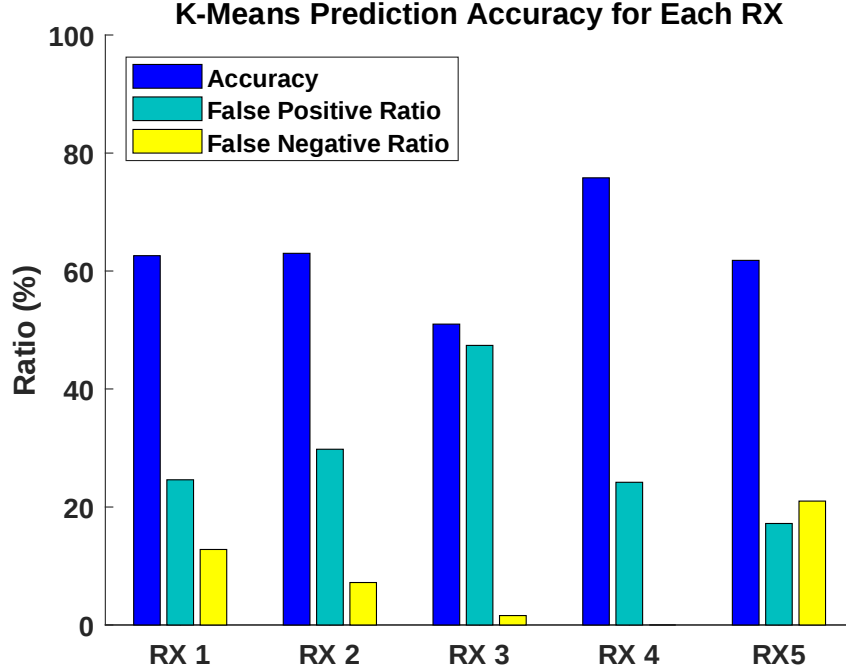


Figure 4.7 Accuracy of the K-Means clustering trained for each receiver

Figure 4.6 shows the accuracy of predicting link blockage versus non-link blockage labels. We can see that the accuracy of the majority of the receivers are about 80% and there is a balance between the accuracy of link blockage and non-link blockage labels.

Figure 4.7 depicts the accuracy of K-Means clustering after training with 30% of the data set and online updating the cluster centers with the rest of the dataset. It shows the results for all five receivers. We adopt the online updating of K-Means from [92]. We choose 30% after performing some experiments with other percentages for batch training. We observe that increasing the batch portion of the dataset does not increase the accuracy significantly. As shown in Figure 4.7, the results are worse than the cascade classifier. The accuracy of receiver number four is better than the rest. For receiver number three, accuracy and FPR are almost the same. It generally shows that clustering cannot be a good solution for our problem, although it is an unsupervised learning algorithm.

Since the clustering algorithm does not show the required accuracy for our defined prediction task, we continue with other learning algorithms. Figure 4.8 shows the accuracy of the cascade classifier with online boosting. We train the classifier by a data set that does not have the burst feature introduced in Section 4.3.2.

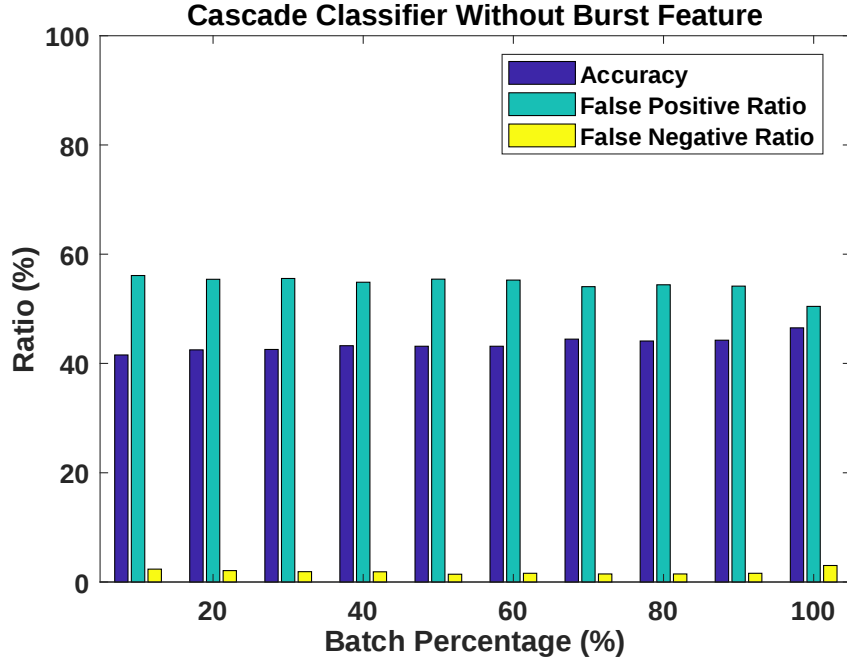


Figure 4.8 Accuracy of cascade classifier trained with no burst feature

As shown in Figure 4.8, the x-axis (Batch Percentage) is the percentage of the data set used for training cascade classifiers in batch mode. We use the rest of the data set for online boosting. We present the average of the metric values for different receivers. The figure shows that the data set without the burst feature is not very successful for predicting link blockage accurately. However, there is a slight increase in accuracy from lower batch percentages to higher ones. This increase indicates that the batch percentage is also an affecting factor in prediction accuracy.

We define two metrics. The first one is blockage accuracy and the second is the non-blockage accuracy. To compute blockage accuracy, we count the number of the positive (link blockage) labels that are predicted correctly by total number of blockage labels. The non-blockage accuracy is measured the same way for non-blockage label. We show the comparison between the accuracy of predicting blockage versus non-blockage labels in Figure 4.9. As we can see in the figure, the accuracy of blockage is much higher than the non-blockage label. It means that the classifier is strongly biased towards predicting blockage. We already show the low performance of the classifier in Figure 4.8.

Figure 4.10, depicts the same situation as Figure 4.8, but trained with data set that has burst feature.

The accuracy of the classifier's prediction trained by a dataset with the burst feature shows a significant improvement compared to Figure 4.8. The increased accuracy with the increase of the batch percentage

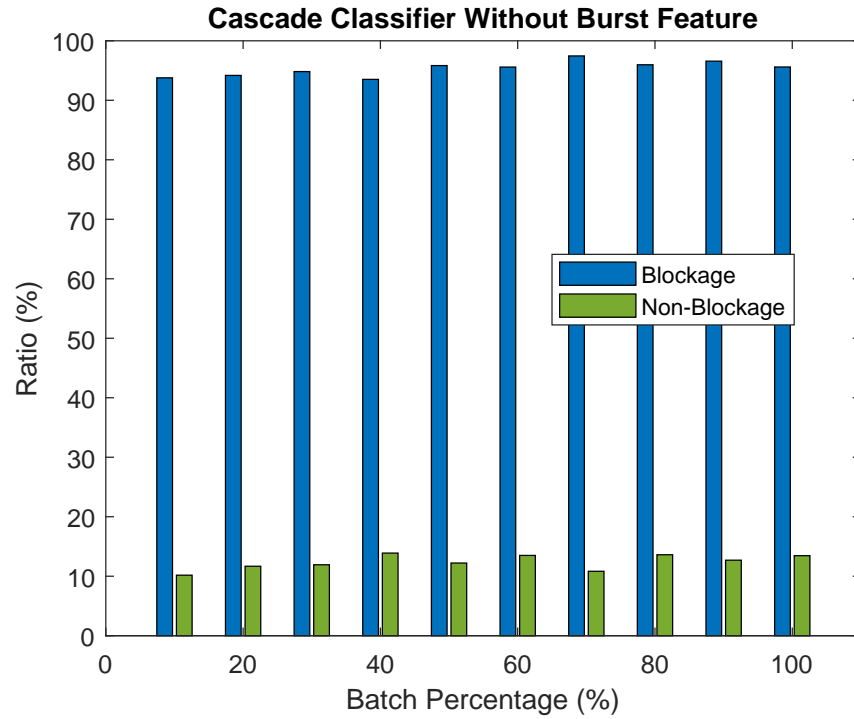


Figure 4.9 blockage vs. non-blockage accuracy of cascade classifier trained with no burst feature

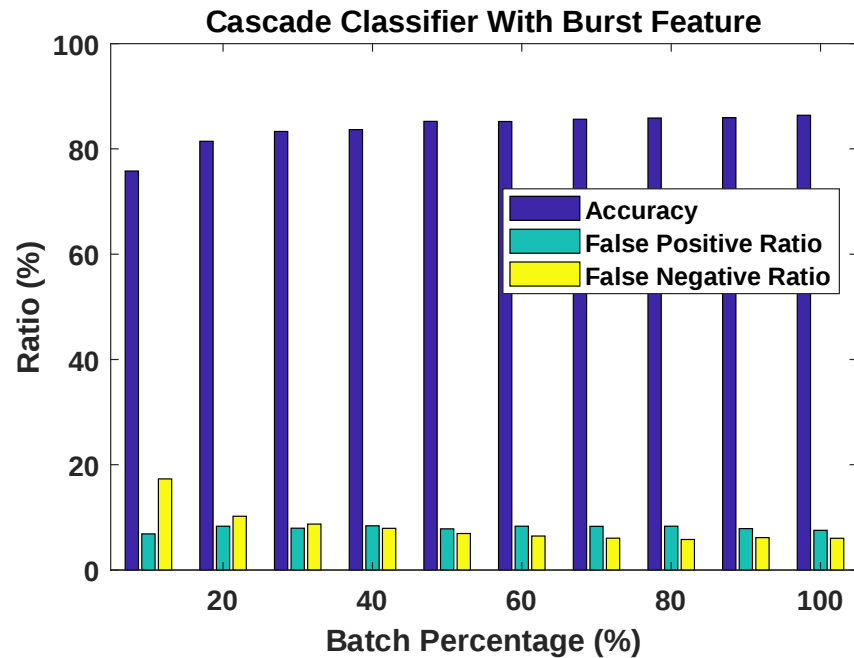


Figure 4.10 Accuracy of cascade classifier trained with burst feature

is still visible in this figure. As shown in this figure, a small training set batch percentage can result in satisfactory accuracy. We can consider the accuracy of the batch percentage of 30% and more satisfactory.

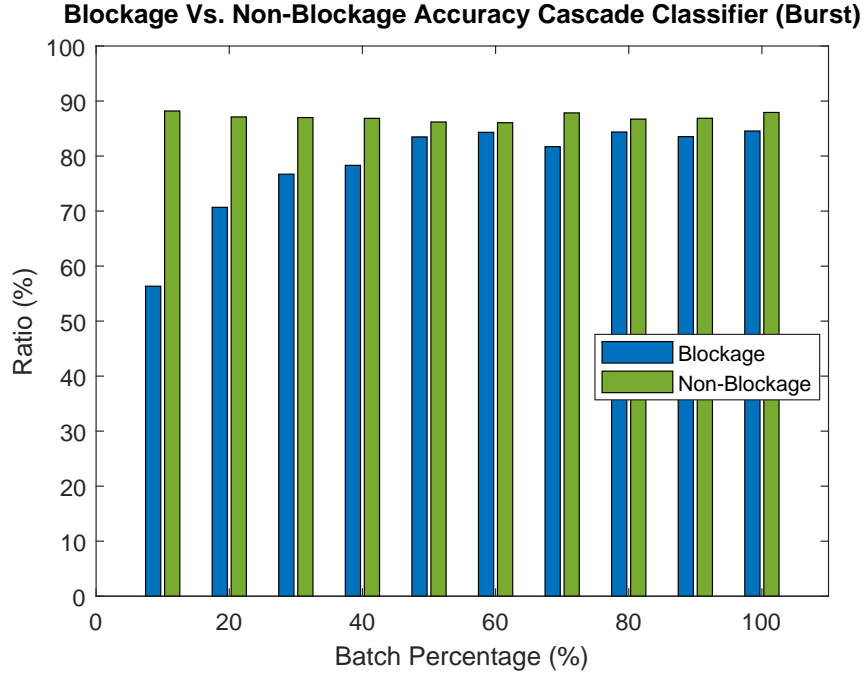


Figure 4.11 blockage vs. non-blockage accuracy of cascade classifier trained with burst feature

This shows the high convergence of the classifier to the satisfactory accuracy with the feature vectors created. The size of the dataset is 500, and 30% of that is 150 data points. By seeing the remaining 350 data points, the model can achieve its current accuracy of more than 80%. If there are other new data points, they further enhance the accuracy. Since we can achieve high accuracy by using 30% of the training data, the algorithm's convergence happens fast.

In Figure 4.10, we show that with a limited percentage of the data set used for training, we see an acceptable accuracy. If the size of the model grows beyond a threshold, we suggest limiting the size of the model. In this approach, the model drops the oldest data points and uses the newly introduced data points. In this way, the model size is limited to a specific value and does not grow beyond it. The recent data points control the accuracy more. Controlling the model size can prevent the growth of time complexity.

We compare the accuracy of prediction of blockage and non-blockage labels in Figure 4.11. It is evident that the burst feature improves the accuracy over no usage of the burst feature. We can also see that with increasing the percentage of the dataset used for batch training of the classifier, we can decrease the bias of the classifier.

To see the effect of just batch training, we experiment with three classification algorithms with and without the burst feature. The experiment includes measuring the metrics of training a classifier with

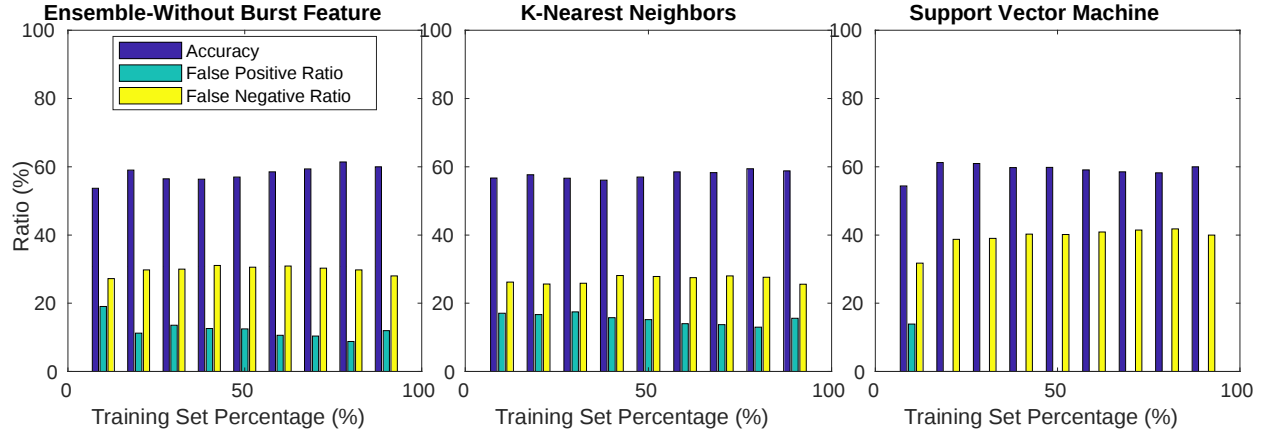


Figure 4.12 Effect of training set percentage on accuracy without online boosting and the burst feature

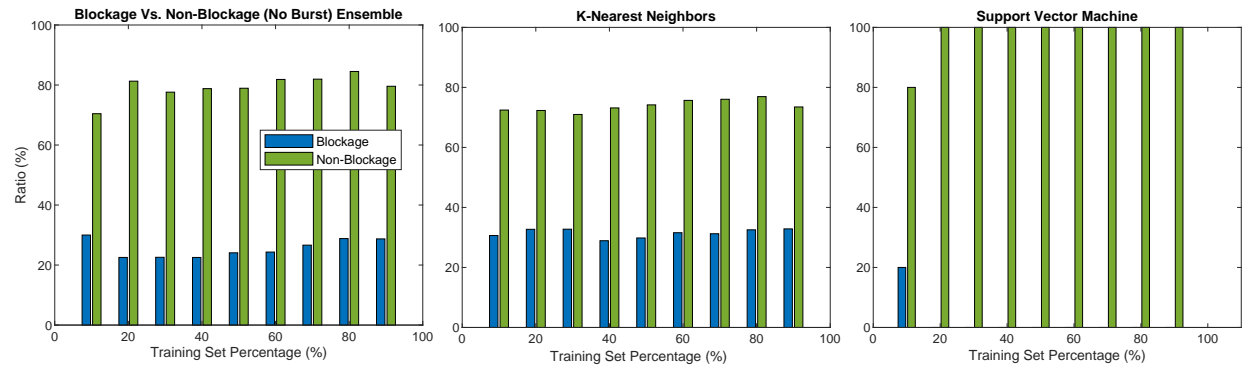


Figure 4.13 Effect of training set percentage on accuracy of blockage and non-blockage without online boosting and the burst feature

different percentages of the data set. Then, we use the whole data set to test the trained classifier. The classification algorithms include ensemble, K-nearest neighbors, and support vector machine. As depicted in Figure 4.12, the accuracy of support vector machine is better than the other two classifiers. The accuracy of all classifiers are not high enough and changing the training set percentage does not help much. Figure 4.14 shows the same experiment with the burst feature. In Figure 4.13, we show how the training set percentage to training different classifiers affects the prediction accuracy of blockage and non-blockage labels. The accuracy of predicting the non-blockage label is higher than the blockage one. This shows that without the burst feature the training has bias. As shown in Figure 4.14, the accuracy for the support vector machine is again slightly better than the other two classifiers. Figure 4.15 shows the blockage versus non-blockage label prediction accuracy with the burst feature. Again, the increased percentage of the dataset assigned to training causes the bias to decrease. All of the trained classifiers have worse performance than the cascade

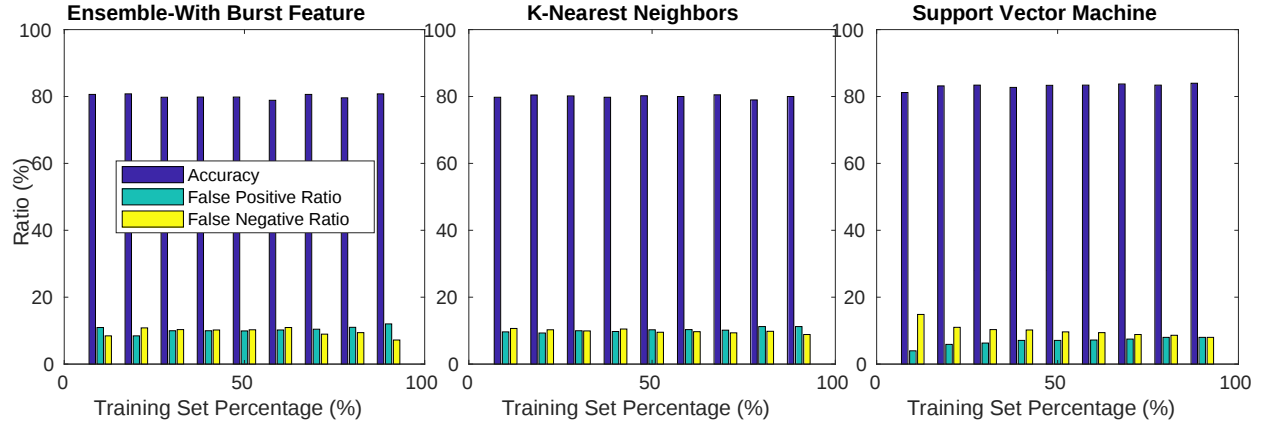


Figure 4.14 Effect of the training set percentage on accuracy without online boosting and with the burst feature

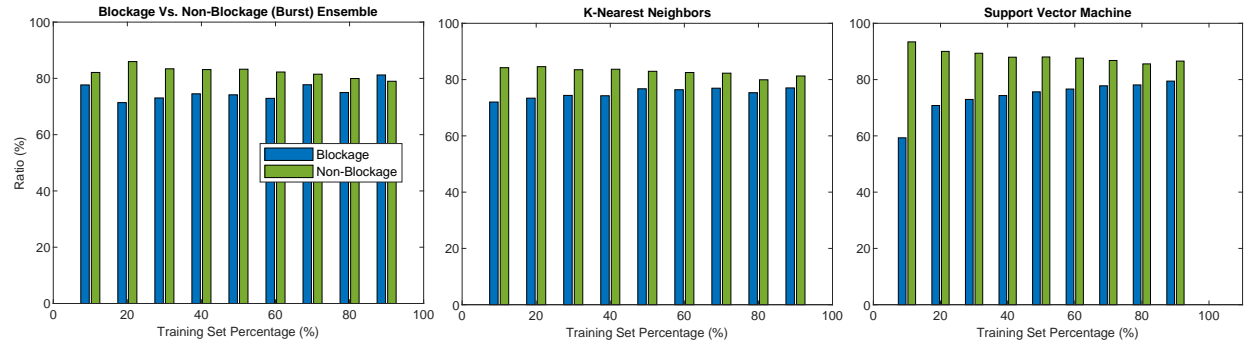


Figure 4.15 Effect of training set percentage on accuracy of blockage and non-blockage without online boosting and with the burst feature

classifier.

4.4.2.1 Results for different settings and parameters

To understand how effective online boosting can be in improving the accuracy, we experiment with different locations in the environment. Figure 4.16 shows the experiment setting. As depicted in Figure 4.16, there are three receivers, one transmitter and four static blocking objects called O1, O2, O3, and O4, and several dynamic blocking objects shown by solid squares without names. For generating data points, the only objects that we move are dynamic blocking objects. The purpose of the experiment is to observe the effect of the receiver's location on the classifier's accuracy trained for each receiver. Figure 4.17 shows the results of the location experiment.

As shown in Figure 4.17, when we train a classifier for the receiver in the center of the environment and

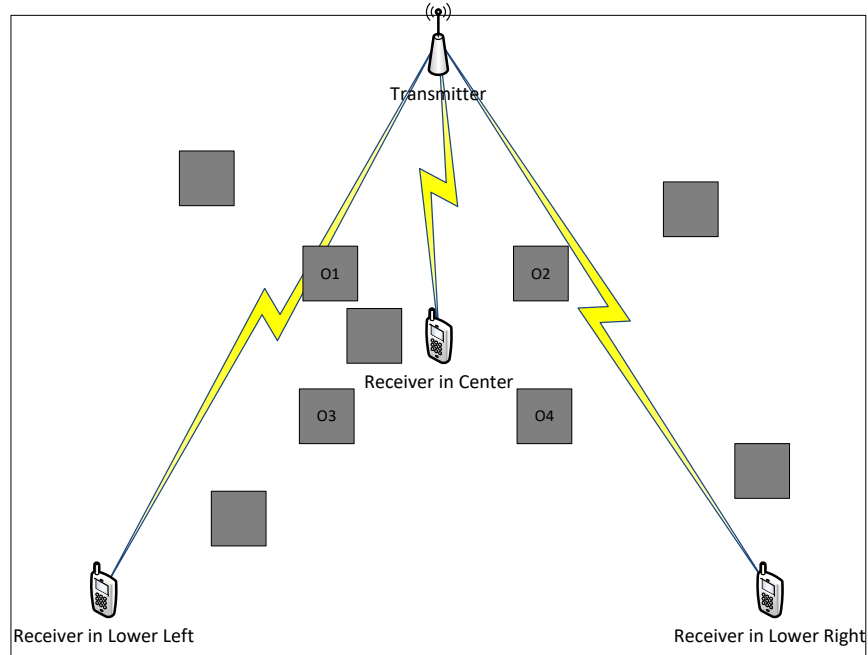


Figure 4.16 The location experiment setting

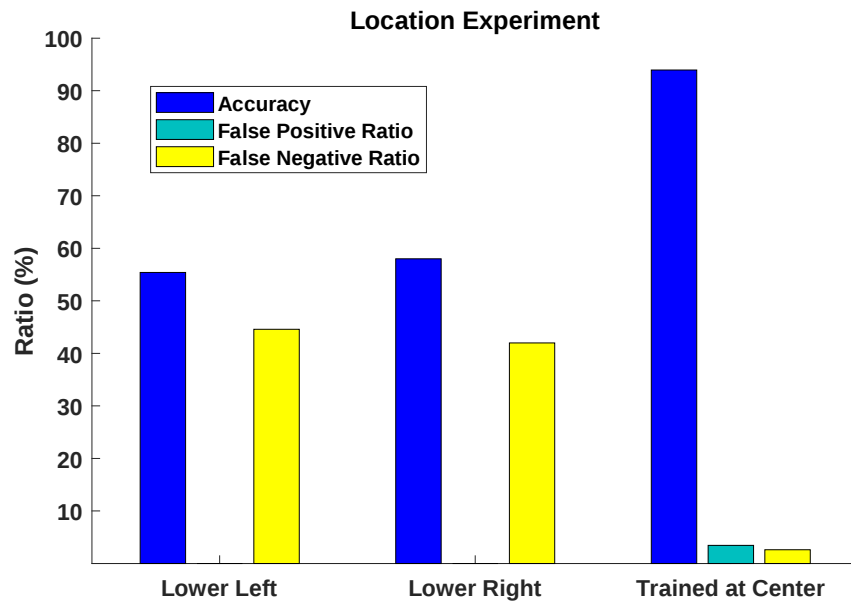


Figure 4.17 Measured metrics for location experiments

tested by the receivers in the other two locations (lower right and lower left), the accuracy of the classifier in the trained location is much more than the other locations. This observation shows that the problem we define has a dynamic nature. We can benefit from implementing an online booster to improve the classifier's accuracy over time, which we show in the previous results.

Table 4.1 Effect of relative permittivity value on prediction accuracy (%)

Relative Permittivity	2.24	4.24	5.24
Ensemble	85.8	86.4	86.4
KNN	84.8	85.6	84.5
SVM	85.0	85.2	86.2

We measure the accuracy in these simulations by considering the link blockage and not the link movement. We assume that when there is a link quality drop, either link blockage or link movement causes this. Based on the experiments in [7] this assumption is valid. Therefore, when we predict the link blockage, the positive label shows the link blockage, and the negative label indicates the link movement. The system performs training for the link blockage. If we involve the link movement in training, we do not expect to see the accuracy decrease as we generate the feature vectors solely based on diffraction. The diffraction happens when there are blocking objects around the link. Since, in the case of the link movement, we do not have blocking objects around the links, the feature vector for a link movement is far enough from the link blockage feature vector, and the classifier cannot predict these two interchangeably.

We investigate the effect of permittivity on the accuracy of prediction. The permittivity is the ability of the surfaces to polarize and store energy [93]. The higher relative permittivity results in a higher reflectivity coefficient. Hence more reflection happens on the surfaces. We change the permittivity of all surfaces of the room and the blocking objects. They all have the same permittivity. Then using the datasets in different environments, we measure the accuracy of Ensemble, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) classifiers. We show the results in Table 4.1. As shown in Table 4.1, the increase in permittivity causes the increase of classification accuracy in general. This is mainly because of an increase in the number of paths due to an increase in permittivity and the signal's reflection. The effect results in the generation of datasets that has more non-zero features in their data points. More non-zero features result in better training and prediction. However, the overhead of resolving the paths and measuring the diffractions increases.

We perform the simulations with different mobility models in the following. We use the random walk mobility model opposed to our line mobility model that we use for the previous simulations. Table 4.2 shows the accuracy of different classifiers for line and random walk mobility models [94, 95].

As shown in Table 4.2, when the mobile objects move by random walk model, the accuracy of the

Table 4.2 Prediction accuracy (%) vs. different mobility models

Mobility Model	Line	Random Walk
Ensemble	97.0	98.0
KNN	97.4	97.6
SVM	90.6	96.4

prediction is enhanced. Although the improvement is not much, it shows that our feature vector generation scheme works even better with the randomness introduced by the random walk model. The difference between the previous accuracy rates of the line mobility model and in Table 4.2 is because of the reduction of the blocking objects in the simulations. This reduction is necessary for generating the mobility model traces for the random walk model.

4.5 Detecting the Link Blockage Using Signal Patterns

In this section, we elaborate on the method we propose to assess the possibility of implementing the link blockage detection method on real hardware. We propose to use the physical layer parameters to compute the signal radiation pattern, and based on the pattern, the method detects whether a pattern belongs to the class of blocked or the class of unblocked. As stated at the beginning of this chapter, the reason for using the prediction framework is to avoid time-consuming sequential scanning of the space. In the case of detecting link blockage using signal patterns, the same argument holds. After obtaining the feature vector, the classifier predicts whether the signal pattern belongs to blocked or unblocked classes. The time-saving part comes from the fact that the nodes do not have to have time windows for detecting the link blockage when they use the classifier.

The proposed method in Section 4.3 shows the feature vector generation using the diffraction values. The system measures the diffraction value through optimizing it in the power fading model in Equation (4.3). In this section, we want to use the angular profile of signal strength around the receiver to generate feature vectors to train a classifier. Both feature vector generation methods share using the power strength parameter to obtain the feature values. We perform the experiments with the total signal strength and not the exact diffraction value due to some hardware limitations. According to Equation (4.3), the total power fading value does include the diffraction value. Therefore, the reason for including the experiments is to show that the proposed method of using diffraction to generate feature vectors works on actual hardware. In other words,

the experiments act as a feasibility study of the simulations.

In the proposed method, the transmitter and the receiver use the legacy IEEE 802.11ad procedures to exchange probe packets for getting the physical layer parameters. Using those parameters, both of the nodes can form the signal patterns formed around them. Using signal processing methods, we obtain the angular profile of the signal. We use the profile as the feature vector. We then use the feature vector for labeling based on the condition that we collect the feature vector. We use the resulting data point for training a classifier for classifying the same shape of the feature vector in the future.

4.5.1 Channel State Information Reconstruction

To get the signal pattern formed around the transmitter and the receiver, we use the method introduced in [96] to get the signal pattern using the antenna gains for different beams. We set up the codebook of each of the devices to have the maximum number of beams possible. It covers the half-plane of the space. The beams sweep half the space. Using the measured antenna complex gains and the beam shapes designed for the half-space, we can draw the signal patterns for the two sides of the link.

We obtain the Channel State Information (CSI) by reconstructing the sent signal on the receiver side. We use the CSI or the channel condition matrix for obtaining the signal paths. The signal paths are for both the receiver and the transmitter. The reconstruction process is composed of sending probe signals on each of the antenna elements of the transmitter's phased array antenna. The phased array of both the transmitter and the receiver of the devices used for the experiments has two-bit phase shifters. Each antenna element has one phase shifter. The number of bits for the phase shifter determines the beamforming resolution. For commodity devices, the cost efficiency dictates the phase shifters to have a limited number of bits for controlling them. The phase shifters' two bits can shift the phase of the signal transmitted from each antenna element with four different values.

In order to get the CSI value, the transmitter sends four-probe signals with different phases. The transmitter sends the signals with a reference signal, so the receiver measures the phase shift compared to that reference signal. Using this concept of transmitting probe and reference signal, the receiver can compute the phased antenna array's complex gains. The receiver then uses the antenna's complex gain for the reconstruction of the propagation channel or the CSI matrix. The CSI matrix describes how the transmitter transmits signal and the receiver receives it. It holds the required information necessary for computing the

strength, AoA, and the AoD of the transmitted signal. The channel matrix can also determine the number of signal paths formed around the transmitter and the receiver. The signal paths are the peaks of the signal strengths in different directions. The training algorithm uses these peaks to make better discrimination between the feature vectors of two different classes. We use the CSI to generate the feature vectors, which we call the signal patterns.

4.5.2 Link Blockage Detection Based on Signal Patterns

The proposed method for detecting the link blockage begins with the training process. In the training process, we obtain the signal pattern of the sent and the received signals. Then, depending on the condition under which we obtain the signal patterns, the system determines the signal pattern label. We use label one as no link blockage and two as a link blockage.

After collecting the required dataset for training, the training phase begins. In the training phase, the dataset trains a classifier. In our experiments, we use cross-validation to train and test the resulting classifier with the same dataset. The training is a supervised learning process as the feature vectors belong to different classes. For the label of the dataset, we use the legacy method of having a time window to determine the label of the feature vector.

The classifier predicts the label of each newly collected feature vector. The trainer then uses the predicted label as the condition (blocked or unblocked) in which it collected the feature vector. The feature vector in our experiments is composed of 75 values of the signal strength that covers one half-plane of space in one specific elevation. For the sake of simplicity, we use one elevation of 3D space and align the phased antenna arrays of the routers aligned with that elevation. We can extend the method to a range of elevations easily.

Before elaborating on the algorithm and the results of experiments, we present some experiments that we conduct to validate the usage of signal patterns in predicting the link blockage in different conditions. There are two nodes in each link, one is the access point (AP), and the other is the station (STA). Figure 4.18 shows the signal pattern of both of the devices when there is no blocking object on the way. For all of the signal pattern figures, the red curve is for the STA, and the blue curve depicts the AP's signal pattern. In Figure 4.18, the lobes of both AP and STA signals are pretty regular. There is one big main lobe in both devices' signal patterns. This shows that if there is no blocking object to the link, the signal patterns are regular and have minimal side lobes. The misalignment of the main lobes is due to the reflective surfaces in the indoor

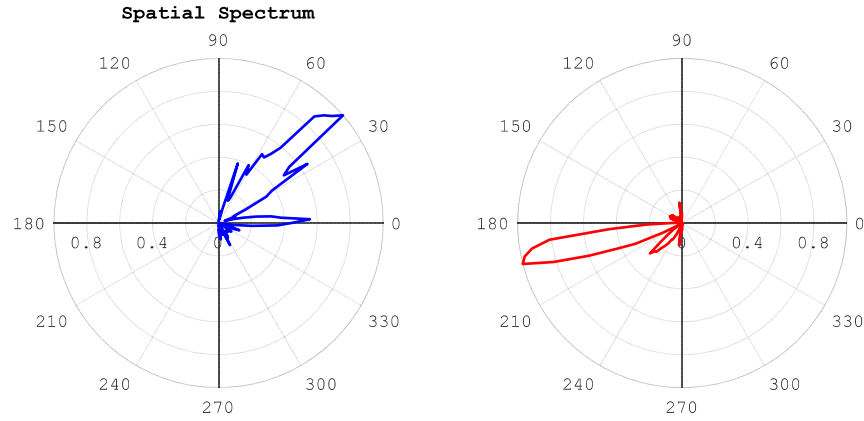


Figure 4.18 Signal pattern for an unblocked link

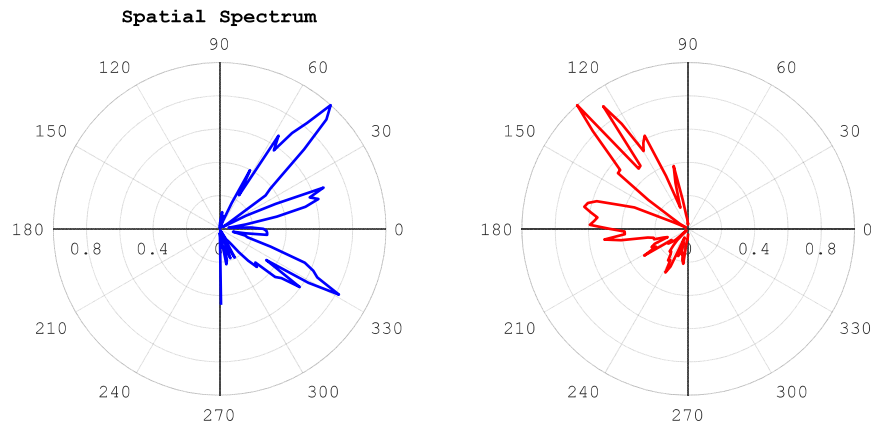


Figure 4.19 Signal pattern for a blocked link

environment. We conduct sampling in an indoor environment.

We consider the following condition as blocked. There is a blocking object that can block the signal and degrade the link quality significantly. In Figure 4.19, the signal patterns are not as regular as the unblocked condition shown in Figure 4.18. There are various side lobes and spikes in the signal patterns of both the AP and STA. The reason for such irregularities is the creation of the new signal paths when the original signal hits the blocking object. The blocking object surface absorbs some of the signal power and reflects the rest. The reflected signals are superposed with each other to create the spikes and the side lobes. As we can see, the signal patterns for blocked and unblocked signals create two different feature vectors that have a safe distance in data space. This ensures good training of the classifier and high precision of the prediction of the classes.

The data collection procedure takes milliseconds for each sampling. Therefore, the time consumption is

very low. The trainer performs the batch training offline; therefore, the time consumption is not a problem for the integrity of the operation. BS can seamlessly perform the process of sampling. There are mobile nodes that send periodic frames to the BS and receive from the BS. The system can use these frame exchanges for the data collection procedure. The algorithm needs no additional data transfers. The frequency of collecting data depends on the frequency of frames exchanged. At each location of a mobile node, there are regular frames exchanged with BS.

We do not use the CSI measurements for the simulations because we use a raytracing tool to get the paths to a mobile receiver. Therefore, we do not need to use CSI values for this purpose in the simulations. The CSI is data that we use for path detection in a possible hardware implementation of our method.

4.5.3 Data Collection

In this section, we discuss the data collection and preprocessing that we perform before the training and testing of the classifiers.

We perform the experiments in an indoor environment. We use the indoor environment to have more multipath to test the method in an uncontrolled situation. The environment is a standard room with different reflectors like the walls, furniture objects, and a glass window. We use two routers that we set one meter apart. The nodes are in a static state. They do not move as the methods that the operating system implements on the routers are not real-time.

To collect the data points (feature vectors), we set the transmitter in a specific location. Then, we locate the receiver in different locations in the environment. After that, we perform the signal sampling. We repeat this procedure to collect all of the data points that we need. In some locations, we add a human blocking object to create the link blockage that we require for the blocked label.

After the collection of data, we have the spatial spectrum for each data point. This is the raw data from the sampling stage. It is the angular spectrum of the signal strength. We normalize the spatial spectrum to keep the value of the signal strength between 0 to 1. Normalization is a common preprocessing method to avoid biased prediction by a specific feature.

4.5.4 Experimental Results

The hardware platform that we use to perform the experiments consists of two TP-Link TALON AD7200 tri-band routers. The router works in three frequency bands of 2.4 GHz, 5 GHz, and 60 GHz. The 60 GHz frequency band falls well in the millimeter-wave domain. The standard that uses the 60 GHz band is IEEE 802.11ad.

The operating system that we flash is a customized LEDE image [97]. The customized image adds some capabilities to the routers, like reporting the SNR for each beam and having the possibility of working with the system through the Linux environment. We connect the STA to the AP, and each time for getting the signal strength, the AP and the STA perform a Sector-Level Sweep (SLS) of the IEEE 802.11ad standard. We transfer the data collected on both of the nodes to a computer. The computer then processes them and creates the data point.

For hardware platform experiments, we generate datasets with two classes. The first class is when there is no blocking object in the way of communication. The blocking object in our experiments is the human body. For the first class (label 1), there is no blocking object between the AP and the STA. Then, the nodes initiate the SLS process. In the SLS scan of IEEE 802.11ad, the AP sends probe packets on all of its beams. In the meanwhile, the STA is in its quasi-omnidirectional mode. When a node is in its quasi-omnidirectional mode, the beam for receiving packets is wide and close to the omnidirectional beam. Of course, the quality of the signal received by such a beam is low. That is why the standard keeps the rate of the probe packets to the minimum supported rate. It ensures the correct reception of each of the packets. The AP sends on all of its beams, and STA receives on its quasi-omnidirectional beam. When this stage is finished, STA sends an acknowledgment packet that contains the beam index with the maximum SNR and the value of the maximum SNR of all of the probe packets STA received. In this way, the AP knows by which beam it should communicate with the STA. The only thing that is remaining is the best beam of the STA. To get the best beam of STA, STA sends probe packets on all of its beams, and the AP receives them by its quasi-omnidirectional beam.

We train three different classifiers with a dataset of 200 data points. We have two sets of data of the same size. One is the dataset that we collect from the AP. These are the SNR values of the probe packets. The STA sends the probe packets over all its possible beams. The second one is the same dataset in the opposite direction. The training dataset is 80% of the whole dataset, and we use the rest for testing as

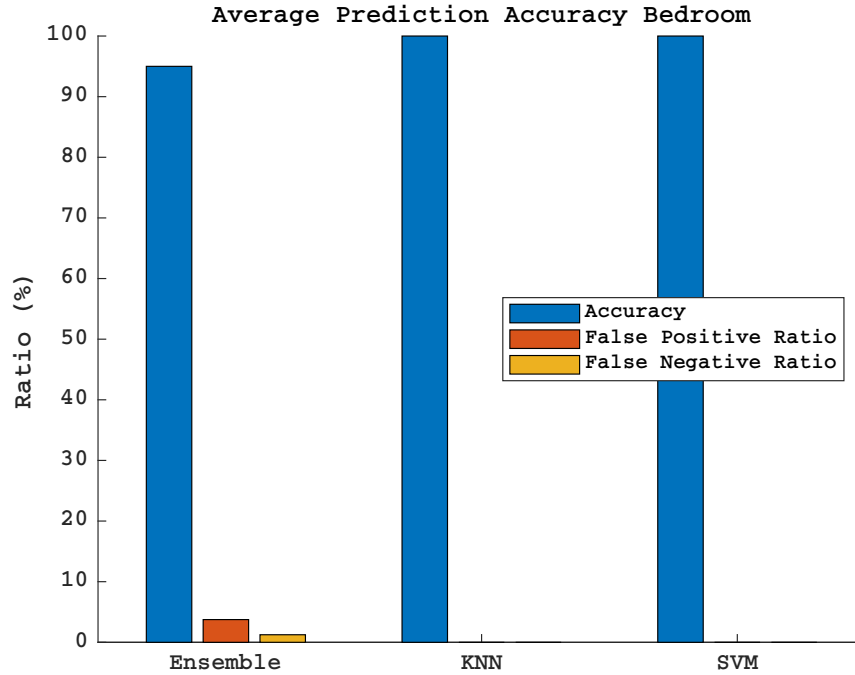


Figure 4.20 Average prediction accuracy in the bedroom

the cross-validation method. We sample the NLOS and LOS test cases independently and with the same quantity. For these experiments, we plot the average accuracy of AP and STA.

Figure 4.20 shows the average accuracy of the three classifiers on both AP and STA datasets. We train three classifiers of Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Ensemble. As shown in the figure, KNN and SVM have the highest accuracy with 100% of accuracy. There are two other parameters that we measure for the classifiers. The first one is False Positive Ratio (FPR), and the second one is False Negative Ratio (FNR). We define FPR as the ratio of test cases that the predicted class that is unblocked, but the ground truth is blocked to the total number of test cases. For the FNR, the definition is the opposite. The predicted class is blocked, but the ground truth is unblocked. For the Ensemble classifier, the accuracy is not perfect, and it is divided on FPR and FNR almost equally.

We do experiments in another indoor environment with the STA and AP nodes. We first make a 200-point dataset that includes both blocked and unblocked labels. Figure 4.21 shows the average accuracy of both AP and STA nodes in the living room environment. As shown in Figure 4.21, the accuracy for all of the classifiers is high like the bedroom results. This is due to the great discrimination of the blocked and unblocked data labels. It shows that we can rely on the classification of blocked and unblocked feature

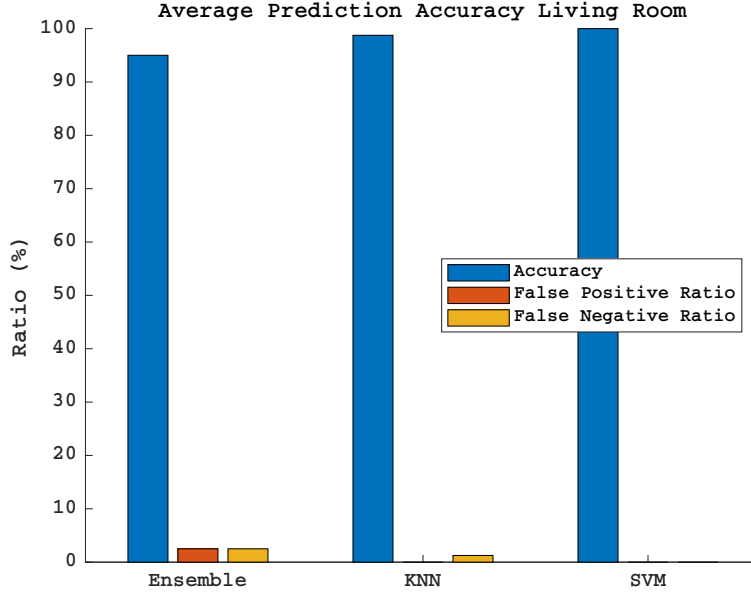


Figure 4.21 Average prediction accuracy in the living room

vectors.

We experiment with different human body orientations, and the results were not different significantly. The human in-move is another case that is worth investigating. Due to our hardware limitations, we could not conduct experiments with moving human. This is mainly because of the real-time nature of the experiments. The limitations in hardware do not let us perform real-time applications. With better hardware and firmware, it is possible to perform experiments with dynamic scenarios. We leave this for our future research efforts.

Comparing the results of the experiments and the simulations using CSI values, we see that the accuracy in the experiments is higher than in the simulations. In the experiments, we scan all the beams in the receiver. Therefore, it takes more time than the simulation method if we require it to run for a long time. In the simulations, the nodes are entirely dynamic, while in the hardware experiments, the nodes are static due to inherent hardware limitations for real-time implementation of the method. Therefore, for the comparison, we should note these points.

We measure the time delay for training and prediction of the methods on the hardware devices. We measure the time delays by millisecond. Table 4.3 shows these values. The time delay for performing the SLS procedure on the COTS devices we have is 490 ms. We measure this value using the *time* command in the bash script installed on the OpenWRT firmware of the routers. The time is for the association process to

Table 4.3 Time delays for training and prediction (ms)

Training Method	Ensemble	KNN	SVM
Training	5.38	0.14	0.66
Prediction	2.81	0.31	0.45

update the SNR values for beams and reading those values from the 60 GHz radio's firmware. The prediction time delay occurs in an online way but the trainer performs the training offline. Prediction delay in Table 4.3 shows only the prediction of the label and not the feature vector generation delay. We cannot measure the feature vector generation delay on the devices nor the simulation platform. Since in the simulation we use a raytracing tool for obtaining the paths, we are not able to measure time delay as it would happen in reality. For the experiments, we do not use the reflectors for detecting the paths and used scanning for this purpose. We provide a thorough discussion of the feature vector generation delay in Section 4.3.3.

We also perform another set of experiments to investigate the effect of orientation change on the classifiers' accuracy. We put the STA node to the northwest (NW) and west (W) of the AP node and generated the data points. We observe that the accuracy for all of the classifiers was 100%. The feature vectors' variation shows that it is robust to the orientation changes in a dynamic environment.

Another experiment that we conduct is the experiment for observing the effect of the distance on the classifiers' accuracy. We place the nodes (AP and STA) three meters apart and collected the data points. We observe that all of the classifiers' accuracy is equal to 98.75%, and the FPR is equal to 1.25%. Compared to the orientation experiment with the nodes one meter apart, we see a slight degradation of the accuracy due to multipath fading occurring in the environment. The excessive paths due to this type of fading can cause the classifier to make a mistake on how to predict the labels.

4.6 Summary

In this chapter, we propose a method of using diffraction values on the signal paths of a receiver to predict if a link blockage is going to happen or the link is going to move. We define a data space for training the machines that make such predictions. We use the data that were sampled to train a classifier and then tested its accuracy. We try various types of classifiers and showed that our proposed method of using diffraction values for link blockage prediction yields a high prediction accuracy and can be used to quickly predict link blockage. We also introduce the burst feature to add an extra dimension to the dataset. Adding this feature

resulted in a higher prediction accuracy.

CHAPTER 5

BEAM ADAPTATION USING OUT-OF-BAND SIGNAL FOR ROBUST MILLIMETER WAVE COMMUNICATION

In Chapter 4, we discuss a method to predict link blockage with high accuracy and low latency in mmW networks. However, the nodes that establish a link need to react efficiently to a link blockage to maintain the link quality.

Using the signal in the mmW band is a very common and reliable method to find the signal paths to align the beams in mmW forming a robust link. The drawback of this general category of solutions to the problem of beam adaptation in mmW communication is its high overhead and time consumption. The literature has applied this approach with different strategies to minimize the overhead it has on a mmW communication system.

Due to the high free space path loss of the mmW signal compared to the non-mmW signal [21], the phased array beamforms the mmW signal to concentrate the signal strength in a limited direction. The usage of beamformers or phased array antennas makes the communication in the mmW directional rather than omnidirectional communication in the non-mmW spectrum. Therefore, to have an omnidirectional image of the paths around the receiver, it needs to scan all the beams to cover 360° space. The act of scanning requires the transmitter to send a packet and the receiver to receive the same packet. Therefore, the scanning operation in mmW incurs a high overhead on the communication system. The nodes experience the overhead as a time delay in the link recovery process in case of the link movement or blockage [7].

The experiments show that the number of signal paths that are formed around a receiver are about two or three in most cases based on extensive outdoor measurements in [21]. In some cases, it is impossible to communicate through the LOS path; it is possible to exploit the NLOS paths created by blocking objects. Therefore, it is crucial to know the signal paths that reach a receiver. We call this procedure the *path discovery*.

There are two different types of path discovery methods. The first type is called *in-band* path discovery. In this category of methods, the scanning of all directions to find the paths that reach the receiver is done by using the signals in the mmW spectrum. The second type of the path discovery methods is called *out-of-band* path discovery, in which the receiver discovers the paths formed around itself in the mmW band by probing

the signals in another band (sensing band). The sensing band is a non-mmW band in which the reception of the signal is done omnidirectionally. In out-of-band path discovery methods, the node aligns the paths discovered in the sensing band to the paths in the mmW band.

The purpose of out-of-band method is to perform fast beam alignment and avoid the high-overhead operation of beam alignment in standard IEEE 802.11ad called Sector-Level Sweep (SLS). As stated before, one method to have low overhead beamforming and beam tracking is using the signal from the sensing band (out-of-band signal) to estimate the DoA of the signal in the mmW band. Since the sub-6 GHz band does not need beamforming for transmission and reception of signal, we can use it for single-transmit DoA estimation.

In this chapter, we develop an out-of-band path discovery method to find all the paths formed around a receiver. The method performs the probing for calculating the paths with no overhead on the mmW band. The receiver receives packets in the WiFi band (2.4 GHz or 5 GHz) and based on the Channel State Information (CSI) values, the MUSIC [98] and the ESPRIT algorithms [99] estimate all the paths formed around the receiver in the sensing band. We also experiment with other methods like Bartlett, Burg's maximum entropy, Capon, and linear prediction methods. Then, based on the Direction of Arrival (DoA) of the paths, algorithm selects the beam covering the strongest one. If a link quality degradation occurs on the current path, the receiver probes the rest of the paths to find the best candidate. Considering the number of the paths is two or three, the algorithm requires a minimal number of probing as it only probes the beams that cover the paths.

Figure 5.1 shows the different paths traveling from the access point to the receiver (client). By knowing the signal paths to a receiver, the receiver can change its current active beam to an unblocked path to avoid interruption of the ongoing communication.

We summarize the contributions of this chapter as follows:

1. We propose a new mmW band beam alignment method using out-of-band signals.
2. We use commercial hardware to perform measurements on both the out-of-band and mmW band.
3. The proposed method exploits all of the out-of-band paths in the mmW band beam alignment procedure.
4. We measure performance of our method with a TCP performance measurement tool.

We organize the rest of this chapter as follows. In Section 5.1, we overview the methods of the beam alignment. In Section 5.2, we summarize the direction of arrival estimation methods and the mmW channel.

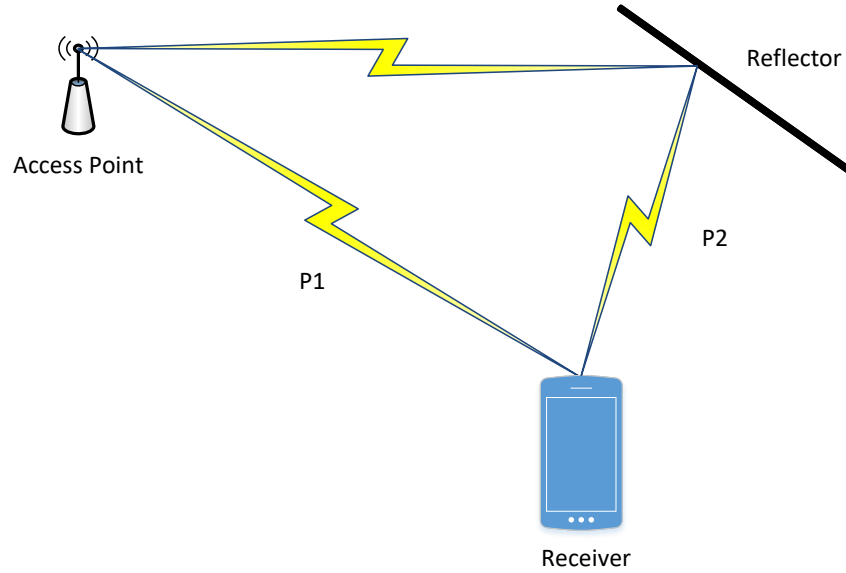


Figure 5.1 The paths formed around the receiver

We elaborate our method of out-of-band path discovery in Section 5.3. We discuss the hardware platform in Section 5.4.1. The results of the experiments come in Section 5.4.

5.1 Related Work

This section discusses different methods proposed in both path discovery categories, in-band, and out-of-band methods.

The first category is the in-band path discovery. Sur et al. [58] have proposed a method to discover the paths formed around a receiver. It requires a significant amount of mmW scanning and solving optimization problems. Teng et al. [85] propose a method that tries to draw the paths forming around a receiver by finding the reflectors based on some geometric equations.

The methods proposed in [100, 81, 76] probe more than one direction by sending special beams that have more than one lobe. They designed hardware that can generate a beam that has multiple lobes pointing towards the defined directions. The authors in [96] measured the CSI in 60 GHz. The expense of calculating CSI is very high in this method. There is a probabilistic method for beam alignment like [77]. The authors in [101] consider the problem of neighbor discovery in 60 GHz communication. The authors proposed to use active learning to find the best beam for decoding the beacons from an access point. They define a compressive sensing framework for the purpose of learning. The problem of consecutive beam probing still

persists. The out-of-band scheme can be applied in the problem of neighbor discovery as well. For the omnidirectional beam steering, Wang et al. [102] have proposed a method to steer the beam on different phased array antennas that they laid out to cover omnidirectionally. The scanning is present and they perform optimizations to decide which beams from which phased array antennas the algorithm should activate to reduce interference and maximize performance. With the out-of-band method, not only one can reach high performances given two antenna arrays for omnidirectional coverage, but they can also select the phased array antenna and its appropriate beam to activate.

Ghasempour et al. [103] have proposed a method for multi-stream mmW communications. The method finds both LOS and NLOS paths as it needs them for multi-stream communication. Jog et al. [104] have proposed a protocol for beam alignment between many access points (AP) and clients. It needs extensive communications. Sur et al. [105] have proposed a beam alignment method based on a limited number of probing in the mmW band. The number of probing, though limited, is still more than the number of probing in out-of-band methods. The model-driven methods like [106] try to limit the number of full-beam scanning by predicting the channel profile based on the correlation of the profiles of close locations. While it can reduce the number of beam scanning, it still needs to scan at certain reference locations and the prediction can be error-prone in cluttered indoor environments.

Zhou et al. [107] have proposed a method for enabling a robust link with the aid of a robot that acts as a relay. The relay installed on a robot needs to discover the reflecting points in the environment and plan the trajectory to deliver the maximum signal strength from the access point to the client. This method requires another mobile node that acts as a relay that needs to map the environment to find the reflecting points and plan the appropriate path. It needs excessive signal strength readings for mapping the environment. A more recent view is using the millimeter wave technology for a communication link from a vehicle. They have also done some measurements [108]. Wang et al. [109] have conducted a study of the performance of the mmW link under the mobility of a vehicle. The authors show the channel sparsity of the mmW spectrum in the outdoor environment with high mobility of the vehicle. The blockage can still occur, but the algorithm can steer the beam towards a non-blocked path once it happens. It shows that the out-of-band method can be useful in a dynamic outdoor environment given the limited number of paths in mmW.

To limit the number of mmW scanning, Nitsche et al. [59] have come up with the idea of detecting the LOS path in the mmW spectrum by scanning lower frequencies WiFi (2.4/5 GHz) as an out-of-band path

discovery method. The methods in [70, 110] use the out-of-band signal for early blockage warning as well. Haider et al. [111] have proposed to use visible light to steer the beam of a mmW client. Sur et al. [112] propose to find the best beam by out of band sensing and probing without overhead for mmW band. Similar to [59], the proposed method searches for one strong path. Woodford et al. [113] propose to use a LiDAR sensor to create a 3D map of the indoor environment to do raytracing of mmW signal. Their method requires a special LiDAR sensor and also fingerprinting (construction of the 3D model) of each new environment. Although they mention that mobile devices use LiDAR sensors, they use a hand-held standalone sensor in their experiences.

All of the mentioned methods do either probing in the mmW band or do not consider multiple mmW band paths. If they do, they do it with in-band overhead. In our proposed method, we use out-of-band frames to find the paths in the mmW band. Then, we use those paths to align the beams in the mmW band on them for enhanced mmW communication in case of blockage or movement. We also use a tool to measure the TCP performance of our method.

5.2 A Primer on the Direction of Arrival and the mmW Channel

In this section, we provide a background on two foundations of our proposed out-of-band path discovery method. First, we discuss the methods used in DoA methods for obtaining the DoA of the sensing band signals. We also give a quick review of how beam alignment in the IEEE 802.11ad standard works.

5.2.1 The Direction of Arrival in the Sensing Band

The researchers call one of the main categories of the methods for estimation of DoA Multiple Signal Classification (MUSIC) [114]. The problem of the DoA estimation is the estimation of the incident angle of the signal plane and the plane of passive sensors (antennas). We show the received signal and the angle it makes with the line of the antenna array in Figure 5.2. For demonstration purposes, we enlarge the antenna array of the sensing band of the receiver. The antenna patch of the mobile devices for both the sensing band and mmW band is fully embedded in the device's chassis. The angle θ is the azimuth of the DoA angle. We show the signal that the transmitter sends as sinusoids in the form of $x(t) = \alpha e^{j\beta} e^{j\omega t} = \rho e^{j\omega t}$. We show the time delay in the signal value as a phase shift: $x(t - t_0) = e^{-j\omega t_0} \rho e^{j\omega t} = e^{-j\omega t_0} s(t)$. We define a source signal as $s(t) = \rho e^{j\omega t}$. The source signal travels and reaches the receiver. The incident angle of the signal

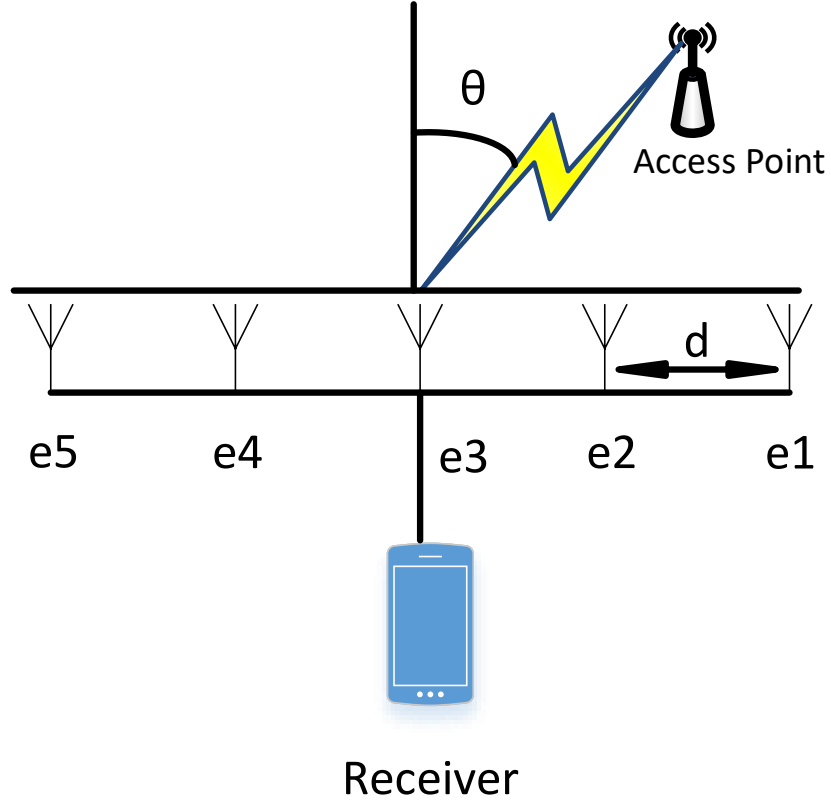


Figure 5.2 The incidence angle of signal and the antenna array

and the antenna array of the receiver is called θ . We show a schematic design of such a setting in Figure 5.2. The signal that is received by the first antenna element e_1 is $y_1(t) = x(t)$. The antenna element i receives the signal $x(t)$ with a delay defined by Equation (5.1).

$$\delta_i = \frac{(i-1)d \sin \theta}{c} \quad (5.1)$$

In Equation (5.1), d is the distance between antenna elements of element separation, θ is the DoA of the signal received by element e_i , and c is the propagation speed. The authors in [114] formulate the received signal by element e_i as,

$$y_i(t) = e^{-i\omega \delta_i} s(t) = e^{-j\omega \frac{(i-1)d \sin \theta}{c}} s(t) \quad (5.2)$$

If we put all of the signals received by N antenna elements we then have,

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\omega \frac{d \sin \theta}{c}} \\ e^{-j\omega \frac{2d \sin \theta}{c}} \\ \vdots \\ e^{-j\omega \frac{(N-1)d \sin \theta}{c}} \end{bmatrix} \times x(t) = \mathbf{a}(\theta) \times x(t) \quad (5.3)$$

In Equation (5.3), $\mathbf{a}(\theta)$ is called the *steering vector*. Due to multipath effect, there are more than one signal that reach to the receiver. Therefore, we modify the model in Equation (5.3) as in Equation (5.4).

$$\begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \begin{bmatrix} a(\theta_1) & \dots & a(\theta_M) \end{bmatrix} \times \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix} \quad (5.4)$$

In this equation, vector Y is the received signal by N antenna elements. Matrix A is the steering matrix of the antenna array, vector X is the transmitted signal, and vector v is the noise on each antenna element. Each column in the matrix A is one steering vector as defined in Equation (5.3). In Equation (5.4), $a(\theta_i)$ is the steering vector of the i^{th} signal path.

The solution to the problem tries to find the vector $\vec{\theta}$. Vector $\vec{\theta}$ contains the DoA of all of the M incident signals. The first step of the solution to the problem by MUSIC algorithm [98] is computing the correlation matrix of the received signal vector $y(t)$. The matrix is $R_y = AR_x A^H + \sigma_0^2 \mathbf{I}$. R_x is the correlation matrix of the transmitted signal. If $N > M$ then $AR_x A^H$ is singular meaning $\det[AR_x A^H] = \det[R_y - \sigma_0^2 \mathbf{I}] = 0$. Therefore, σ_0^2 is an eigenvalue of R_y . The N -dimensional received signal space is divided into signal subspace U_x and noise subspace U_v . The steering vector $a(\theta_i)$ is in signal subspace. Signal and noise subspace are orthogonal which implies that $a^H(\theta_i)U_v = 0$.

The algorithm searches in the space of θ and draws the spatial spectrum that is the signal power based on DoA. The peak on the spatial spectrum shows one DoA of the received signal. There are M peaks associated to M signal paths. The authors in [114] formulate the spatial spectrum function as,

$$P(\theta) = \frac{1}{a^H(\theta)U_v} \quad (5.5)$$

Although the MUSIC algorithm works with any array shapes, it should know the antenna element positions. The algorithm is susceptible to antenna element position. The MUSIC algorithm should also know the gain

and the phase of the antenna elements. The antenna elements need calibration for the correct performance of the algorithm. The other disadvantage of the MUSIC algorithm is its time complexity. Depending on the resolution of the search space (θ) for DoA, the time complexity can be high, and it needs much computational power.

The other algorithm that we use in estimating the DoA is called the ESPRIT algorithm [99]. In this algorithm, the pairs of antenna elements are called *doublet*. The doublets are pairs of antenna elements that are identical. A vector separates the doublets called $\vec{\Delta}$, which determines the displacement between the antenna elements. The phase, the gain, and the position of the doublet can be arbitrary. N sets of doublets have $2N$ antenna elements. The steering vector depends on the array geometry and it should know the array geometry. The signal equation is the same as defined in Equation (5.4). However, there are two sets of such equations, one for each of the identical subarrays. We use the vector $\vec{\Delta}$ to define the equations.

The algorithm computes the correlation matrix of $2N$ received signals. M eigenvectors of the correlation matrix corresponding to M largest eigenvalues form the signal subspace U_x . The remaining $2N - M$ eigenvectors form the noise subspace U_y . In the end, the algorithm computes the vector of the $\vec{\Phi}$ as a one-time computation. This vector is a function of DoA values for M signal paths. The ESPRIT algorithm does not search for any space for the solution. It is a closed-form solution. Due to the noisy measurements and errors in the calibration of the doublets, we use the total least squares ESPRIT.

As a comparison of the two DoA estimation methods discussed in this section, both of the methods are high resolution. It means that with the same number of antenna elements, the estimation has more accuracy as the DoA values can change with the smaller increments compared to the beamforming methods. Both MUSIC and ESPRIT methods can detect multiple signal paths as long as enough antenna elements are present.

Since MUSIC needs fewer antenna elements, in case of the limited number of antenna elements, the MUSIC algorithm has preference over the ESPRIT algorithm. On the other hand, ESPRIT needs less computational power compared to MUSIC and more antenna elements. Therefore, the trade-off between computational power (computational delay) and the number of antenna elements exists in choosing one of these two methods.

The distance between the antenna elements for these methods is 0.5λ in which λ is the wavelength of the signal. Authors in [115] say that the accuracy of MUSIC is higher than ESPRIT. We also compare these

two methods in our scenario with real-world data.

In Bartlett's method [116], the algorithm uses periodogram for the estimation of power spectral density. In this method, an average of uncorrelated measurements of the signal is used to estimate the power spectral density. The algorithm achieves this by averaging the periodograms which results in an asymptotically unbiased estimate of the power spectrum. To obtain uncorrelated data, the method divides the data record into subsequences. It performs the averaging over each subsequence.

In Burg's Maximum Entropy Method (MEM) [117], the method estimates the spectral density by maximizing the entropy of the sampled data of the signal. The signal is partial and noisy. There are two conditions for the entropy maximization problem. First, the power spectral density has to be non-negative. Second, the Fourier transform of the estimate has to match the sample autocorrelation. The closed-form of the solution to the estimation problem needs coefficients that it obtains through the iterative Burg's algorithm.

In Capon's method [118], the method estimates the power spectral density based on an estimation of the cross-power spectral density. The smallest minima of the function of the steering vector and the sample covariance matrix gives the direction of arrival of the signal.

The last DoA estimation method that we use is Linear Prediction Method (LPM) [119]. This method is based on minimizing the mean output signal power of the array elements. Which is subject to the constraint that the weight of the array elements is unity. The weight of the array elements is a function of the correlation matrix. The power spectrum is a function of the correlation matrix of the sensors (antenna elements) and the steering vector of the antenna elements.

5.2.2 A Primer on mmW Channel and Beam Pattern

This section skims through the fundamental concepts and their mathematical relations of mmW channel and beam pattern.

Due to the unique propagation characteristics of the mmW signal, devices use a particular type of antennas called phased arrays for transmission and reception. The phased arrays are composed of small patch antennas as antenna elements. The phased arrays mentioned make the *analog beamforming* possible. In analog beamforming, the phased array can activate only one beam pattern at a time. Based on the phase shifters' resolution, the beam pattern can cover a limited space around the phased array. The goal is to find

the sparse signal paths by scanning beam patterns (in-band path discovery) or getting DoA using signals of another frequency (out-of-band path discovery).

On the other hand, there is *digital beamforming* that can have more than one beam pattern active. In digital beamforming, each antenna element has its own RF chain. This can result in a more expensive phased array. Many commercial devices (including the devices we use) are not using digital beamforming due to the high cost of the beamformer.

The commercial devices can store several settings to the phase shifters as beam patterns. These vectors can set the phase shifters to specific values resulting in a beam pattern that points to a specific direction in the space. Therefore, each beam pattern can cover a fraction of the space around the receiver or the transmitter. On the other hand, according to measurements of mmW signal propagation, the receiver receives the mmW signal in 2 or 3 paths [21]. The goal is to find the sparse signal paths by either scanning beam patterns (in-band path discovery) or by getting DoA using signals of another frequency (out-of-band path discovery).

According to [96], the authors define the received signal \mathbf{y} in the receiver as,

$$\mathbf{y} = \mathbf{p}_r^H \mathbf{H} \mathbf{p}_t \mathbf{x} + \mathbf{p}_r^H \mathbf{N} \quad (5.6)$$

In Equation (5.6), \mathbf{p}_r is the complex gain of the receiver and \mathbf{p}_t is for the transmitter, \mathbf{x} is the transmitted signal, \mathbf{H} is the channel matrix that describes how the signal propagates from the transmitter to the receiver, and \mathbf{N} is the noise perceived by the receiver. The matrix \mathbf{H} describes the multipath effect of the transmitted signal as well. The channel matrix is a function of the number of paths from the transmitter to the receiver, the complex gain, and the transmitter and receiver's steering vectors. The steering vector is the same as the one in Equation (5.3). It is a function of DoA and determines the phase shift based on the distance.

The authors in [96] formulate the receiving channel as $\mathbf{h}_r = \mathbf{p}_r^H \mathbf{H}$. They also formulate the transmitting channel as $\mathbf{h}_t = \mathbf{H} \mathbf{p}_t$. Their algorithm uses these values for computing the received signal power. The magnitude of receiving and transmitting channels is the received signal power.

Due to cost efficiency considerations, the hardware components that we use have limited capabilities for commercial devices. In the device that we use, two bits in a word control the phase shifters, making them choose between four phases (from 0 to 2π radian). This limitation affects the accuracy of the beamformer in commercial devices.

5.2.3 Beam Alignment in IEEE802.11ad

In this section, we provide a quick review of the beam alignment in the IEEE 802.11ad standard. Since this is a widespread standard in mmW communication, it is useful to know the mechanism and compare it to other methods in terms of time complexity and overhead.

The IEEE 802.11ad standard calls the beam alignment Sector Level Sweep (SLS) [120]. SLS has two main phases. In the first phase, an *initiator* sends SLS frames in the beacon packets to the receiver or *responder*. The initiator starts by sending SLS frames on all of its defined beam patterns. The responder listens to all of these SLS frames with one quasi-Omni beam pattern. Since the bit rate for this transmission is at the lowest, the beam pattern on the responder can be wide because a small portion of the signal can deliver the SLS frames. When the initiator finishes sending all of the frames, the responder sends the initiator an SLS feedback frame that contains the initiator's beam index of the strongest frame received, alongside its SNR. In this way, the initiator knows which of its beam patterns it should use to communicate with the responder. The responder performs the same process so the responder knows through which of its beam patterns it should communicate with the initiator. We illustrate a general scheme of the first phase of the process in Figure 5.3. In Figure 5.3, Node A is the initiator, and node B is the responder. In the second phase, an optional beam fine-tuning takes place. The time complexity of the SLS beam alignment method is $O(B_i + B_r)$ in which B_i is the number of the beam patterns in the initiator, and B_r is the number of the beam patterns in the responder. Each new communication link requires a process of SLS. In the case of link breakage, after some delay, the nodes initiate the process. The SLS process is a high overhead beam alignment procedure that has a high occurrence rate. Therefore, it is important to look for methods that have lower overhead as the beam alignment procedures happen frequently.

5.2.4 Beam Adjustment Using Out-of-Band Measurements

In this section, we elaborate on the method that we use to adjust the beam of the mmW interface to the channel that the out-of-band interface estimates.

The out-of-band CSI measurements can lead to finding more than one signal direction in the sensing band. In our design, we use the weights of the antenna elements to create beams along with different directions in the 3D space with different azimuths and elevations. When the AP measures the DoA of the

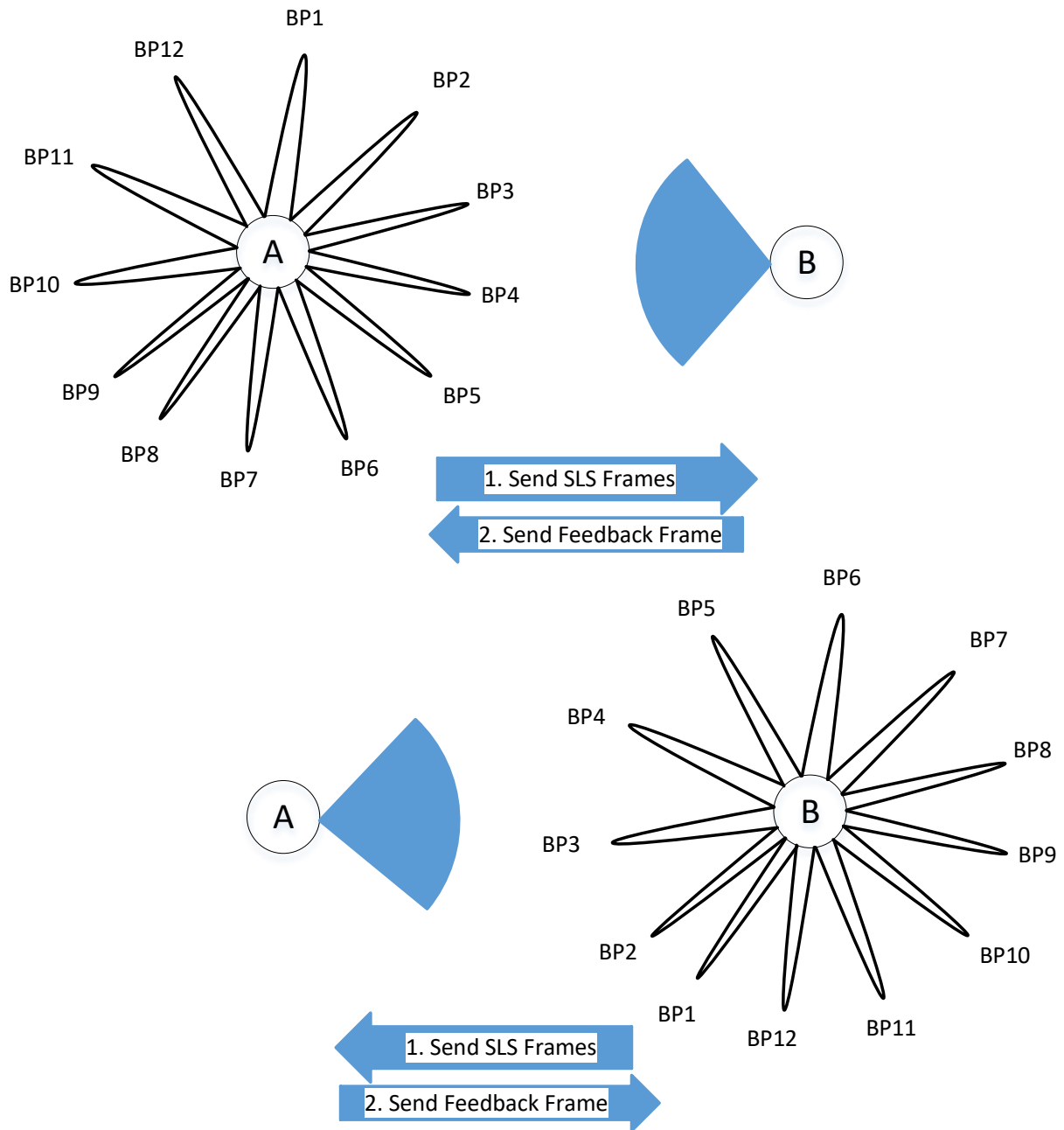


Figure 5.3 The general process of sector-level sweep

signal in the sensing band, the DoA estimating algorithm may find one or more than one signal arrival direction. In the case of finding one direction, the AP uses the appropriate antenna element weights for the phased array antenna to generate the beam that points to the direction of the received signal in the sensing band.

The other case is when the DoA estimating algorithm estimates more than one DoA from the sensing

band. In this case, the AP needs to use the weights of the antenna elements that create beams for different directions in space to create a beam that points to all the directions that the DoA estimating algorithm estimates. The combined beam may not have the directionality of a beam that points to one direction, but it can amplify the gain from several directions that is useful in an indoor multipath environment.

Devices use phased array antenna in mmW communication systems to form the beam toward a specific direction in space consists of several patch antenna elements. Each of such antenna elements is connected to a phase shifter to change the phase of the signal that it transmits. A phased array antenna has N antenna elements and K beams are precoded in its codebook. If we assume that the antenna elements' layout is a 1D linear formation and the space between each antenna element is the same, the authors in [121] compute the gain for the k^{th} beam for spatial direction θ as:

$$A_k(\theta) = \sum_{n=1}^N w(n, k) \cdot \exp(j2\pi nd \cos \theta / \lambda) \quad (5.7)$$

In Equation 5.7, $w(n, k)$ is the weight of the n^{th} antenna element for the k^{th} beam, d is the distance between each antenna element, and λ is the wavelength of the carrier signal. Activating different antenna elements satisfies the effect of the sum in Equation 5.2. The summation effect is based on the superposition of the signals that all of the antenna elements transmit. The exponential part of the equation is collectively called the steering vector of the phased array antenna and is dependent on the structure of the phased array.

We use the beam combining method [121] to achieve a beam that the device creates by combining two beams that each of them point to a different direction in space. The main use case of the beam combining procedure is in a mmW system that has two or more phased array antennas. However, in the system that we use for our experiments, this is not the case and the routers have one phased array antennas. The other method is to divide the antenna elements into two subarrays in which each subarray forms the beam independently. We choose the antenna elements for each subarray and use the weights for each antenna element in one subarray to form the beam in the direction that the DoA estimation method estimates.

For example, if the DoA estimation algorithm estimates that the signal arrives at two angles of θ_1 and θ_2 then the phase shifters of subarray A create a beam that points to direction θ_1 and the subarray B handles θ_2 . The periodic signal sent over the different antenna elements for each subarray generates the two beams on the same phased array antenna. Since the beam combining method reduces the number of antenna elements compared to the case of having all antenna elements for the same beam direction, it reduces the directivity

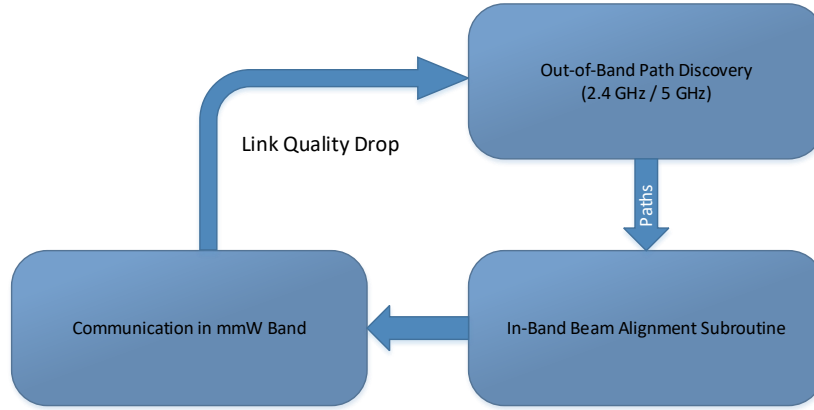


Figure 5.4 A flowchart of out-of-band path discovery

of the beams.

5.3 Out-of-Band Path Discovery

In this section, we describe the proposed method for the out-of-band path discovery.

We use the packets in another WiFi band like 2.4 GHz or 5 GHz to measure the proposed system's channel matrix. Using the algorithms for estimating DoA, the receiver can discover the paths around itself that are formed due to out-of-band signal propagation. The receiver then uses the discovered paths as guides for the in-band or the mmW band to align its beams based on them. The frames used in the non-mmW band can be periodic beacon frames sent regularly, or even the receiver can request ping frames from the transmitter as needed to update the discovered paths in the non-mmW band.

We depict a schematic of the system architecture of the out-of-band path discovery in Figure 5.4.

We show the process in Figure 5.4. For estimation, we assume that the DoA of the out-of-band signal is the same as the mmW band. When the algorithm discovers the new paths, it transfers them to another subroutine to align the beams. By knowing the phased array's antenna gains, it is possible to adjust the beams to the direction needed.

When the phased array aligns the beam in a specific direction, the communication begins. During the communication, the algorithm monitors a link quality measure like SNR for a drop in link quality. When the quality drops below a threshold, the control initiates another out-of-band path discovery process. Since the system wants to exploit multiple paths discovered in the non-mmW band, the initiation of the out-of-

band path discovery process does not occur unless there are no other paths that the node can use for the communication.

In case of a link quality drop in the current path in use, the receiver probes an alternative path to see if the path is strong enough for the communication. If it is, the current active beam switches over to the alternative path and continues the communication. If the alternative path is not suitable for the communication, the receiver initiates another out-of-band path discovery process to update the paths.

The proposed method can replace out-of-band path discovery's initial process with a legacy beam alignment of the IEEE 802.11ad standard. In this way, there is an in-band path discovery in the system, and the communication does not rely solely on out-of-band path discovery. This is due to some inaccuracies in the out-of-band path discovery. Because the signal of 2.4 GHz / 5 GHz frequencies can go through the walls and other blocking objects, but the mmW signal cannot. Consequently, the receiver can sense some non-mmW signal paths due to their diffraction around the blocking objects, but the equivalent paths do not exist in the mmW band. However, the problem of excess non-mmW paths has a simple solution. The receiver needs to check multiple paths discovered by the out-of-band path discovery. If a path does not have the required quality, the receiver initiates another out-of-band scan. The checking of the paths is as simple as switching to a beam that covers the path. Then, the receiver receives a beacon from the transmitter.

The other case of the mentioned scenario is when the out-of-band scan discovers two paths, but there are three paths in the mmW band. Since the non-mmW signal has both the reflection and the diffraction, this case is not possible. A portion of the non-mmW signal goes through blocking objects. The blocking object absorbs a portion of the signal, and the rest is reflected off the surface of the blocking object. On the other hand, the blocking object absorbs a portion of the mmW signal, and it reflects off the rest. If the mmW source is behind a blocking object, there is no path (or a very weak one). If the mmW source is in front of a blocking object, the reflections and the LOS path create the paths. The non-mmW signal also has reflections and LOS paths. Therefore, the mmW signal does not have any other means of propagation other than the ones the non-mmW signal has.

The Algorithm 5.1 shows how the beam alignment uses out-of-band path discovery to avoid high overhead in-band scanning.

In Algorithm 5.1, multiple paths that the out-of-band scanning discovers are separated into vectors of the spatial spectrum. Then, the algorithm uses each beam pattern's antenna gains to see how close a spatial

Algorithm 5.1 The general procedure of Out-of-Band Path Discovery

```

1: Align the beams using the IEEE 802.11ad legacy method
2: while true do
3:   if link quality drops below threshold then
4:     Ping the transmitter
5:     Perform DoA estimation {obtains  $\vec{\Theta}$ }
6:     for  $\forall \theta_i \in \vec{\Theta}$  do
7:        $\vec{\beta} = \arg \min_{\forall bp_j \in BP} \|\theta_i - bp_j\|$ 
8:     end for
9:      $\beta_{max} = \arg \max_{\forall \beta_i \in \vec{\beta}} SNR(\beta_i)$ 
10:    Use  $\beta_{max}$  for communication
11:   end if
12: end while

```

spectrum of each path is to the beam pattern. The algorithm chooses the closest beam pattern (the one with the minimum norm) for a beam pattern that best covers the path. Using a probing (shown by the function $SNR()$) for each of the selected beams, the node selects the path with maximum strength for the communication.

Since the number of paths discovered is limited (two or three), the number of probings to find β_{max} is limited to that number. Therefore, the proposed method incurs a constant number of probing on the mmW band communication system compared to the IEEE 802.11ad legacy beam alignment method.

5.4 Experimental Evaluations

This section discusses the experiment scenarios and how we use the hardware to generate the results. We also present the results of the experiments we conduct by the hardware platform. We first discuss the scenarios and how different components work together and then provide the results.

5.4.1 Experimental Hardware Platform

In this section, we discuss the hardware platform used to conduct experiments in this chapter.

In this chapter, we use commercial off-the-shelf (COTS) wireless hardware to validate our proposed method. We use two TP-LINK TALON AD7200 tri-band routers to act as the transmitter or the access point (AP) and the receiver or the station (STA). To access control on the routers and some MAC layer parameters,

we flash both routers with LEDE firmware images as a part of the TALON Tools project [97]. The second part of the hardware platform is the hardware and the tool for collecting the Channel State Information (CSI) matrices for estimating the DoA of the non-mmW signal. The tool [122] that we use works on Intel 5300 network interface card (NIC) to collect CSI data.

5.4.2 Evaluation Setup

There are four nodes in the network of the devices in our experimental setup. The two routers that act as the transmitter (AP) and receiver (STA) in the mmW band, the laptop that we use for collecting the CSI values, and the controller laptop that we use to coordinate all the other three devices and collect data from them.

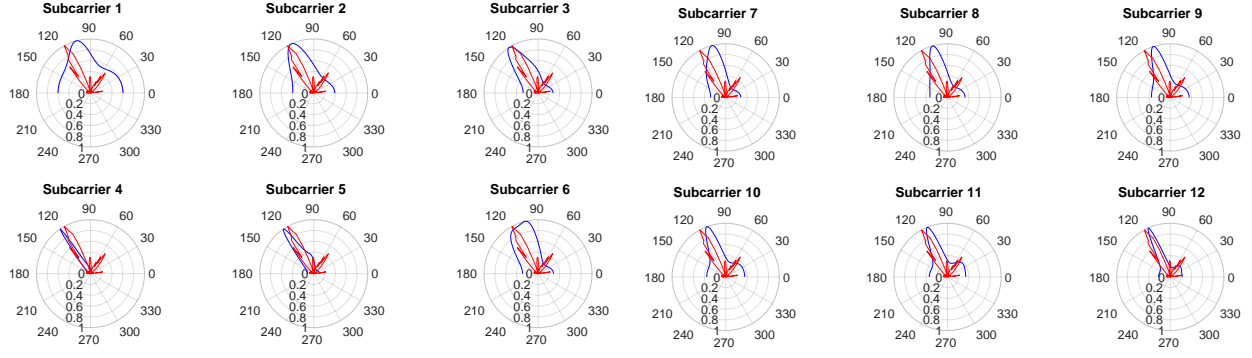
First, the controller laptop commands the routers to set up a mmW link between each other. Then, It instructs the AP router to ping the CSI laptop. Based on the CSI matrices obtained, the CSI laptop estimates the DoA of the non-mmW signal and the controller laptop uses it. The controller laptop also collects the SNR of the ground truth from the AP router. This gives the paths formed around the STA router based on the SLS procedure in the IEEE 802.11ad standard.

5.4.3 Methodology of the Communication Performance Assessment

To do the assessment, we measure the Transmission Control Protocol (TCP) performance of the link. We use a popular tool for the TCP performance measurement called iPerf3 [123].

The test starts by sending a ping packet from the transmitter to the receiver. The CSI laptop receives the ping packet, and the CSI laptop measures the CSI matrices for different subcarriers. There are 30 subcarriers, and each subcarrier has its own CSI matrix. Therefore, we can measure the DoA for each subcarrier. Due to hardware imperfections for the CSI measurement subsystem, we cannot get the best DoA measurement for all subcarriers. We mostly see the imperfect DoA measurement for higher subcarriers in the form of no maximum signal strength for any arrival direction. Using the DoA of each subcarrier, the receiver makes a beam for that point in that direction. The receiver uses the beam to communicate with the transmitter. The receiver performs the process we discuss and according to our measurements, it does not incur much computational overhead on the receiver.

For each bandwidth measurement, iPerf sends data from the transmitter to the receiver for ten seconds.



(a) The subcarriers 1 through 6

(b) Subcarriers 7 through 12

Figure 5.5 The spatial spectrum of the signals in the office cubicle

Every 0.1 seconds, the iPerf tool performs one bandwidth measurement. For each location of the transmitter, we perform one ten-second iPerf test for each subcarrier. Each subcarrier has its own DoA, which in turn, we translate to a beam for communication. We collect all of the bandwidth readings for evaluation.

5.4.4 Experiments Results

We conduct experiments on two different frequency bands for out-of-band path discovery. The WiFi network works with two frequency bands of 2.4 GHz and 5 GHz. We experiment with two different DoA estimation methods, MUSIC and ESPRIT, as described in Section 5.2.1. The first category is the signal spatial spectrum that shows the spatial behavior of in-band and out-band signals. The second category is the communication performance assessment.

5.4.4.1 Signal Spatial Spectrum

The first set of results is from an indoor location. In all of the polar plots in this section, the blue line indicates the spatial spectrum of the Out-of-Band signal, and the red line shows the spatial spectrum of the In-Band signal. For the following figures, the frequency of the Out-of-Band signal is 5 GHz, and for the In-Band signal, the frequency is 60 GHz. The subcarriers covered in the experiments are subcarrier numbers $S = \{-28, -26, -24, -22, -20, -18, -16, -14, -12, -10, -8, -6, -4, -2, -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 28\}$ in the channel 36 of 5 GHz band. We set the 2.4 GHz band to channel 11. Figure 5.5 shows the spatial spectrum samplings in an office cubicle. As shown in Figure 5.5, the spatial spectrum of the 5 GHz signal estimated by the subcarriers in the lower half of the whole subcarriers has good accordance

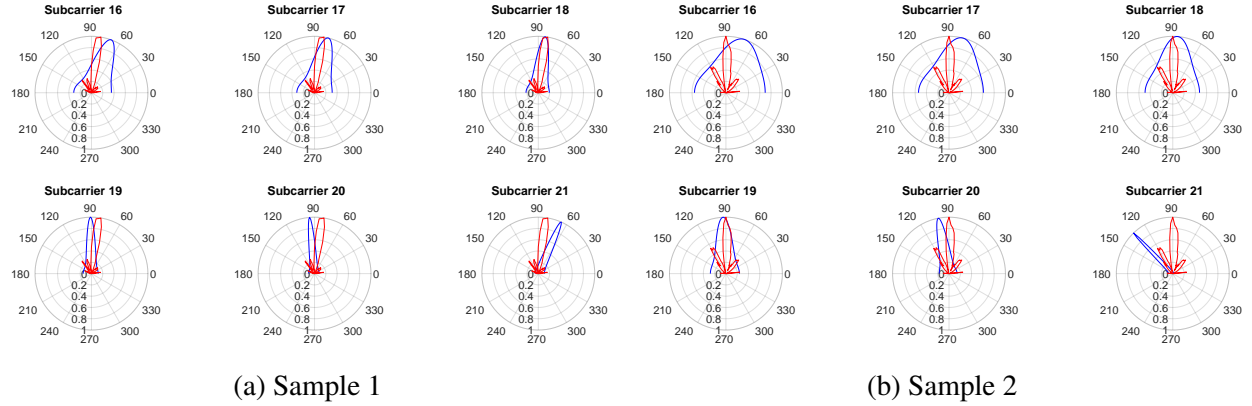


Figure 5.6 The spatial spectrum of the signals in the bedroom

with the spatial spectrum of the 60 GHz band. We show the accordance as the alignment of the in-band and out-of-band spatial spectrum. It is a good result since the setup of the devices is very tight in the office cubicle. Moreover, the office cubicle has minimal space with three walls surrounding the space. We expect to have the multipath fading in such limited space. On the other hand, the 60 GHz signal's spatial spectrum shows a pretty clean single lobe signal. One reason for such high accordance between the spatial spectrum of 5 GHz and 60 GHz signals is low multipath fading. The cubicle walls are made of fabric, which we expect it to reflect less signal compared to walls made from more solid materials.

Figure 5.6 shows the spatial spectrum of the signals in a room with hard walls. The walls are suitable surfaces to reflect the signal. The room is a bedroom with medium dimensions. The nodes are in the middle of the room within 1 to 2 meters apart, and there is furniture present. The out-of-band signal frequency is 5 GHz. In Figure 5.6, we show two samples of the same placement of the nodes. As we can see, the high subcarriers show better conformity to the ground truth of the 60 GHz signal spatial spectrum. There is some deviation from the ground truth in the subcarriers 20 and 21. The reason is that the signals in both bands are reflected off the hard walls of the bedroom. The walls of the bedroom are harder than the office cubicle. The nodes are pretty close to the solid wall in these samplings. In the sample 2 subcarriers 16, 17, and 18, the out-of-band shows that it can capture the multipath in 60 GHz well by spreading over both main and side lobes of the 60 GHz signal.

The next location we experiment is a living room with no furniture. We place the nodes in the center of the room. We use the 2.4 GHz band for out-of-band measurements. We place the antennas in a linear array with 0.5λ spacing. Figure 5.7 shows the results of experiments in the living room. The spatial spectrum of

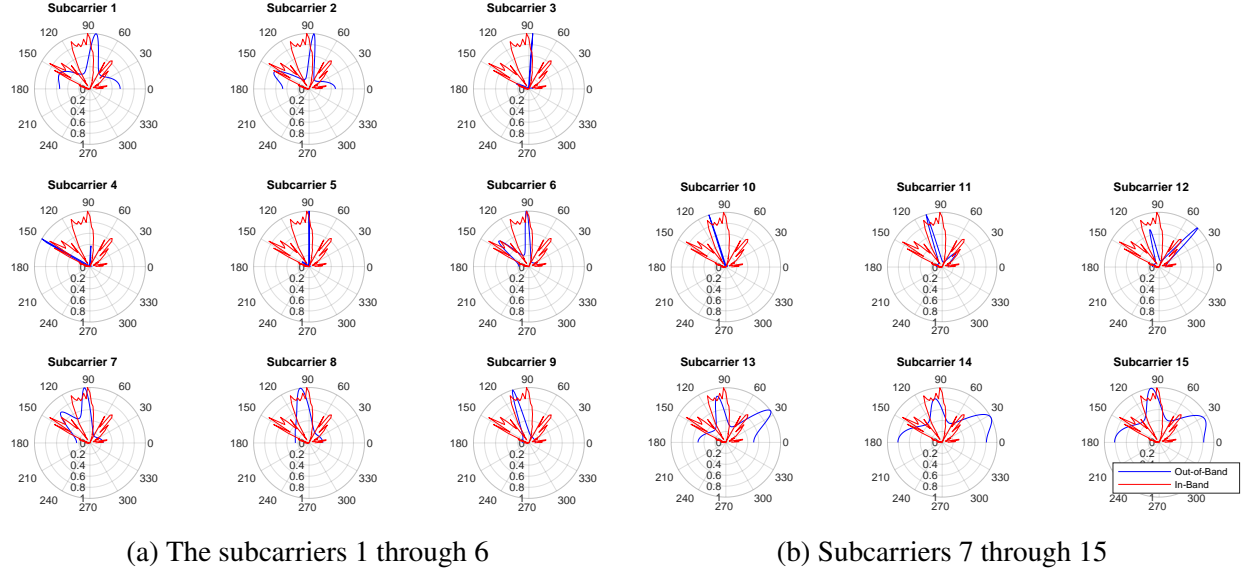


Figure 5.7 The spatial spectrum of the signals in the living room

the 60 GHz band in Figure 5.7 shows that there are multiple paths formed around the receiver in the living room. The main reason is the limited space around the nodes. Therefore, the reflected signal bounces off the walls several times until it gets to the receiver. Almost for all subcarriers, the out-of-band spatial spectrum estimates the 60 GHz band spatial spectrum well. Of course, out-of-band spatial spectrum does not properly cover some lobes. This is because of the DoA estimation methods' limitation and the hardware we use for collecting CSI matrices. The MUSIC algorithm can resolve two paths ($N - 1$) (or two peaks in the spatial spectrum) with three antenna elements Intel 5300 NIC has.

For the ESPRIT algorithm, the number of paths resolved is not improved compared to the MUSIC algorithm. The number of resolved paths by ESPRIT is $\lfloor \frac{N}{2} \rfloor$ with N being the number of the antenna elements. Figure 5.8 shows the spatial spectrum in the living room with both DoA estimation algorithms MUSIC and ESPRIT. The blue line in the graphs is the MUSIC algorithm, the green line is the ESPRIT algorithm, and the red line is in-band 60 GHz measurements. Since, unlike MUSIC, the ESPRIT is not a search algorithm, there is no spatial spectrum output generated by the ESPRIT algorithm. We only show the computed DoA by ESPRIT by the green radial line. As shown in Figure 5.8, both of the DoA estimation algorithms can estimate close to each other (with exception of subcarriers 8 and 9). The estimations are close to the spatial spectrum of 60 GHz band. The inaccuracy comes from both multipath fading in the 60 GHz band and the COTS hardware that we use in the experiments. The lower subcarriers show better accuracy

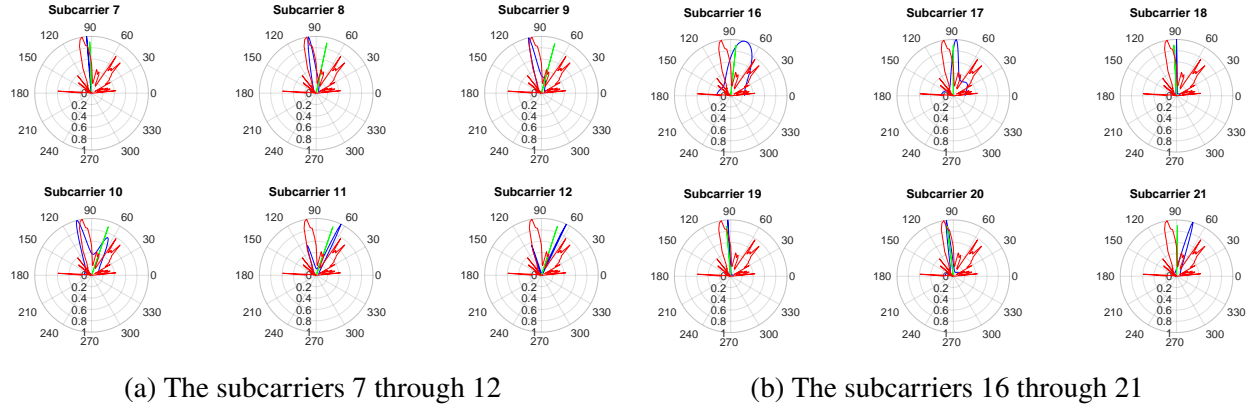


Figure 5.8 The spatial spectrum of the signals in the living room with ESPRIT

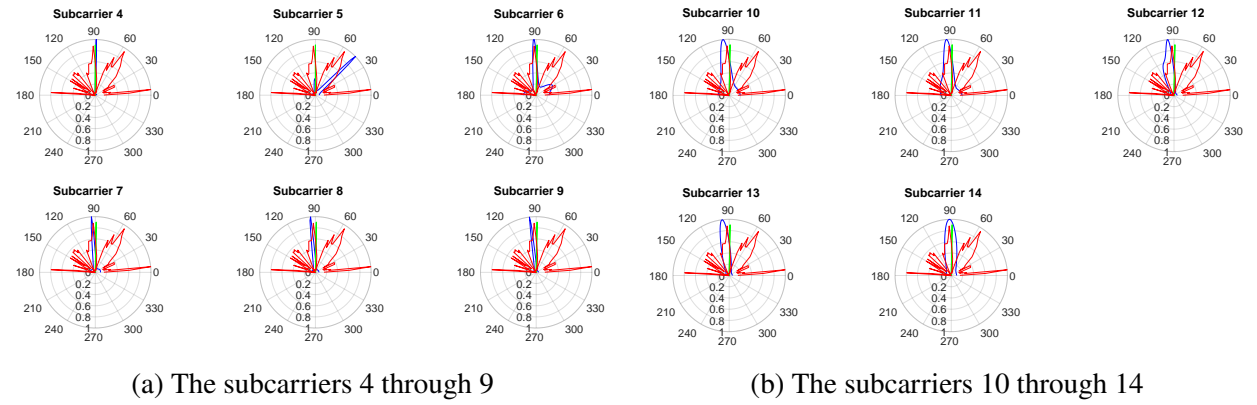


Figure 5.9 The spatial spectrum of the signals in the outdoor environment

due to the fact that the COTS hardware provides more precise CSI matrices for them.

We also perform experiments in the outdoor environment. The space is an outdoor tennis court without any blocking objects and walls around it. The purpose is to analyze the performance in outdoor and blocking object free environment. There are two courts side by side. Figure 5.9 shows the results. As Figure 5.9 shows, the 60 GHz signal's multipath effect is powerful in the outdoor environment due to the hard floor of the area. We position the nodes on the ground. However, the two DoA estimation methods can detect the peaks of signal in the 60 GHz band.

Due to limitations in hardware, we cannot perform the algorithms in a real-time fashion. This would limit the scenarios that we can create to assess our method, like having people walking. We leave this problem for our future research endeavors.

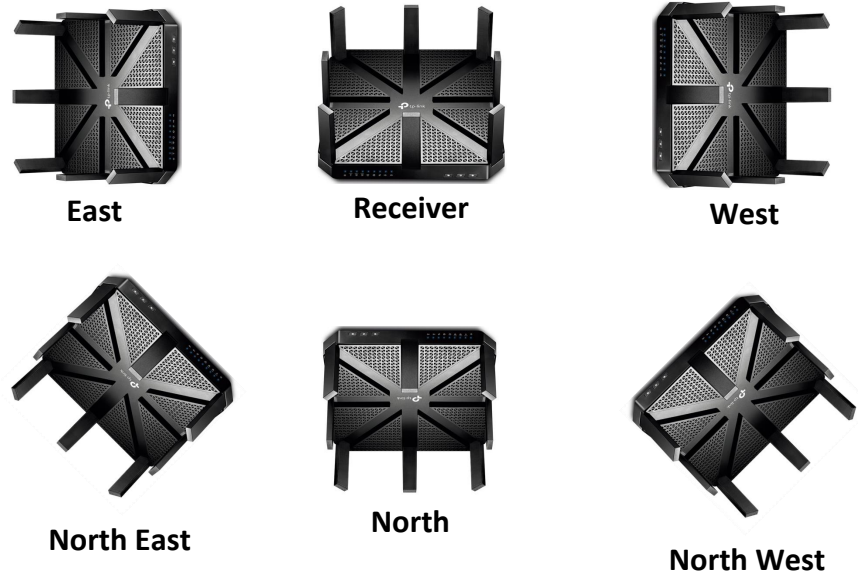


Figure 5.10 The topology of transmitter and receiver

5.4.4.2 Communication Performance Assessment

In this subsection, we discuss the performance evaluation of the communication link. The out-of-band paths make the communication link possible and there is no legacy beam sweep performed on the receiver side.

We perform the assessment for the links when we deploy the transmitter in different locations. We deploy the receiver in a fixed position and the transmitter changes its place for different evaluations. Figure 5.10 shows the topology of the transmitter in different locations with respect to the receiver.

The positions in Figure 5.10 are respective to the receiver. The locations of the transmitter are as follows: East (E), North East (NE), North (N), North West (NW), and West (W). Throughout this section, we refer to different transmitter locations when needed.

The first set of the results are from the evaluation that the transmitter is to the North of the receiver. In Figure 5.11, we show the cumulative distribution function (CDF) of bandwidth when the transmitter is to the North of the receiver (location N).

There are three types of beams. The first one is the beam that the MUSIC algorithm using the out-of-band signal generates. The second type is the beam that the ESPRIT method using the out-of-band signal generates. The third one is the original beam that is predefined in the router by the factory. As we can see in Figure 5.11, both MUSIC and ESPRIT methods generate the beams that yield better performance than the original beams of the router. The better result is due to the fact that the DoA estimation method works well

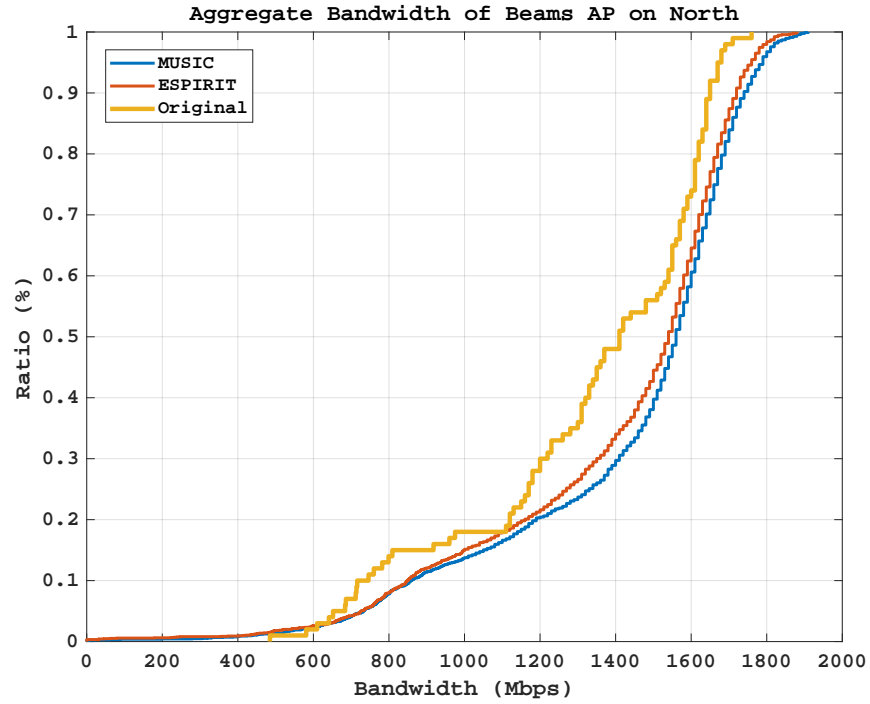


Figure 5.11 CDF over all subcarriers N

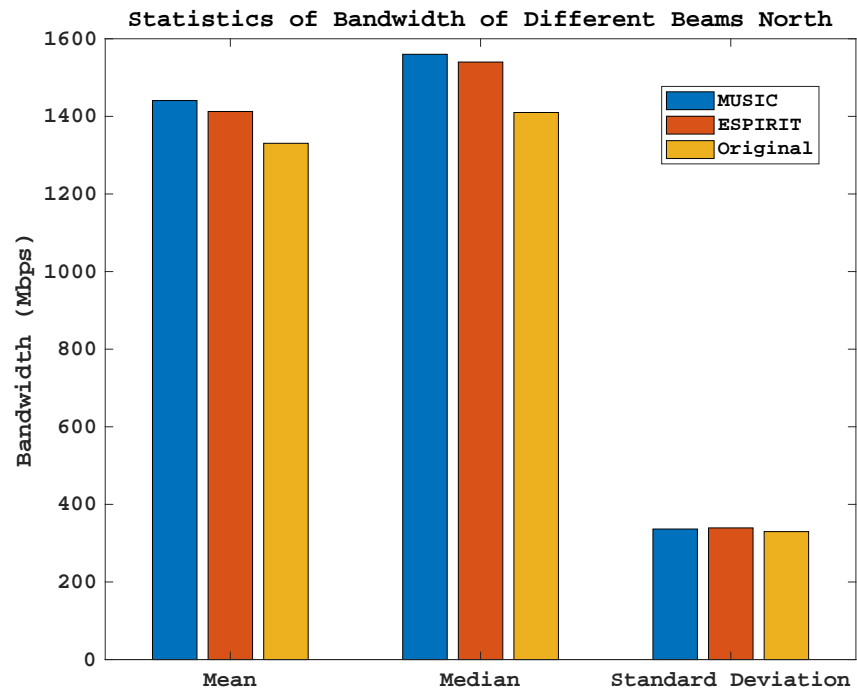


Figure 5.12 Statistics over all subcarriers N

in the direct link when both the phased array antennas of the routers are facing each other.

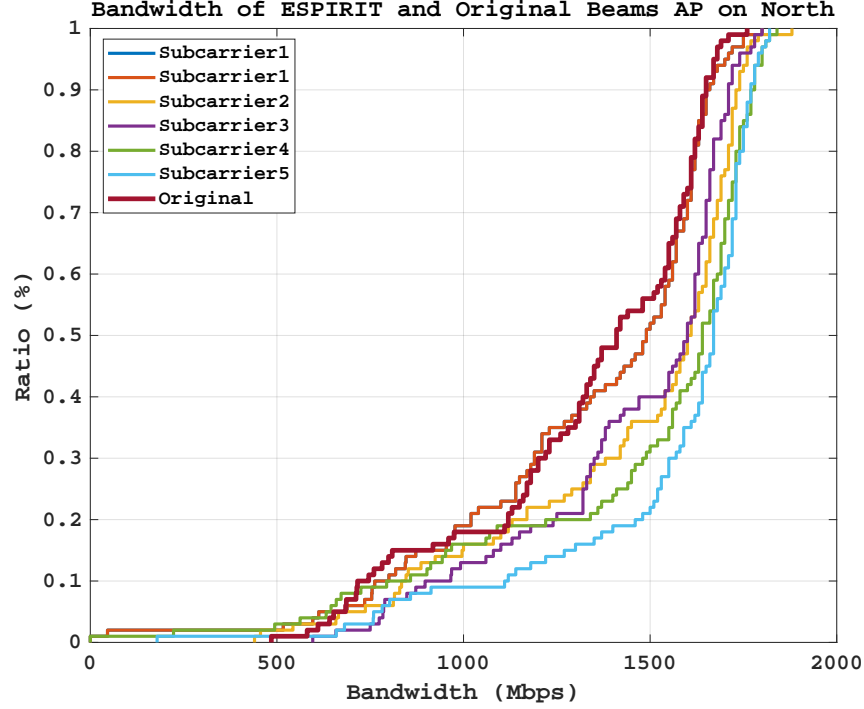


Figure 5.13 CDF over subcarriers 1 to 5 (ESPRIT) N

Figure 5.12 shows the statistics of the Figure 5.11. Although the results are from all of the subcarriers, the MUSIC and ESPRIT methods outperform the original beams.

Figures 5.13 and 5.14 show the CDF for the first five subcarriers for ESPRIT and MUSIC methods, respectively. We can see that ESPRIT performs slightly better than MUSIC.

Figure 5.15 shows that with deviating from the direction path, the performance gap between the out-of-band method and the legacy method shrinks. We still can see the better performance of the out-of-band method over the legacy method. We show the statistics for location NW in Figure 5.16.

Figure 5.16 shows that the mean and median for the original beam are less than MUSIC and ESPRIT algorithms for NW location.

For locations, E and W, MUSIC and ESPRIT methods' performance is not as good as the original beams. The reason for this is the shape of the original beams compared to the beams generated by out-of-band measurements. The original beams are more omnidirectional than the out-of-band generated beams. The phased array of the routers cannot generate beams that are highly stretched to the left or right.

In order to have a more inclusive performance comparison between DoA estimation methods introduced in Section 5.2.1, we plot the results of the performances of such DoA estimation methods in Figure 5.19.

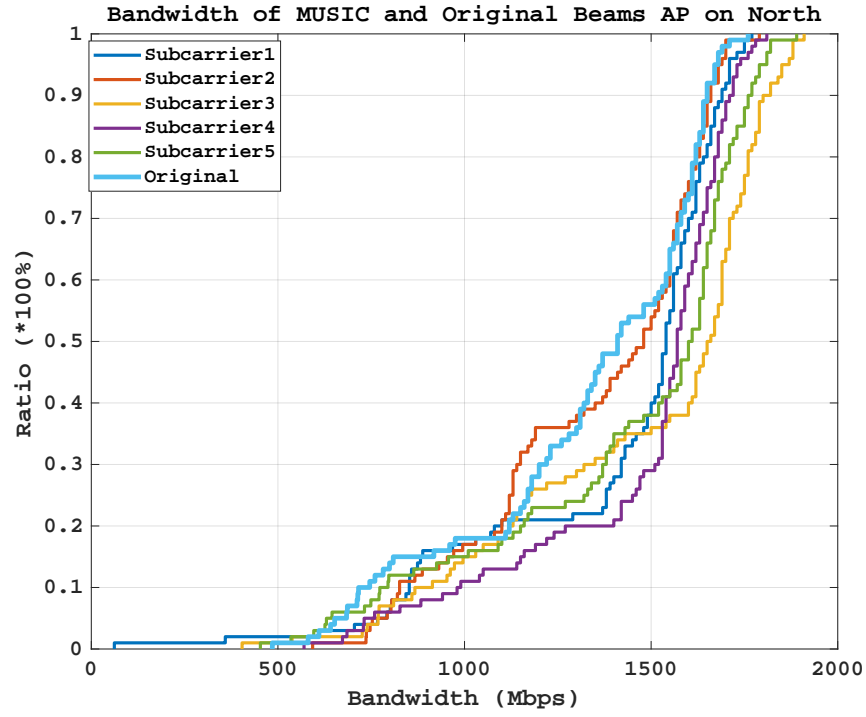


Figure 5.14 CDF over subcarriers 1 to 5 (MUSIC) N

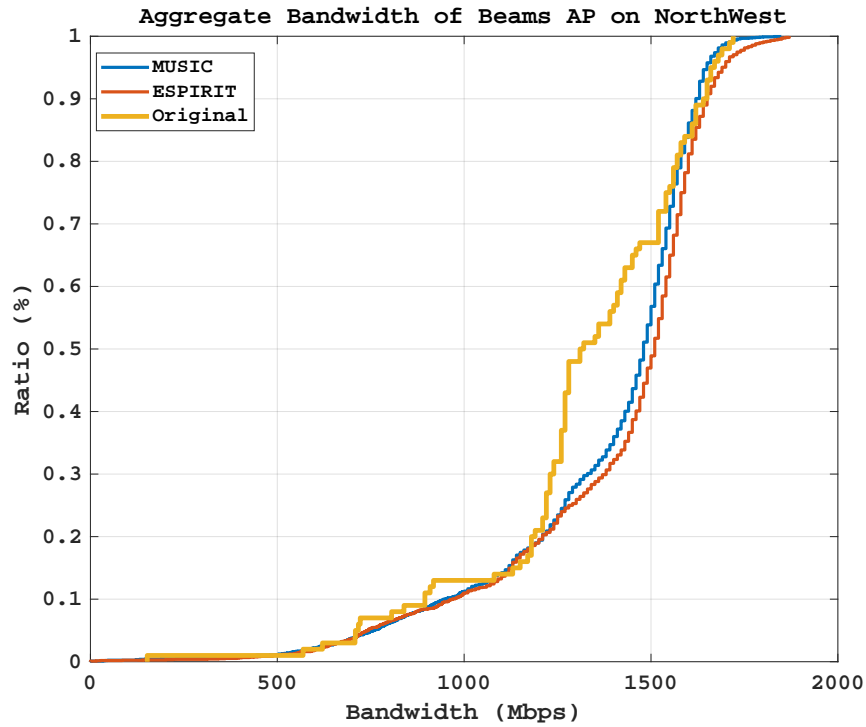


Figure 5.15 CDF over all subcarriers NW

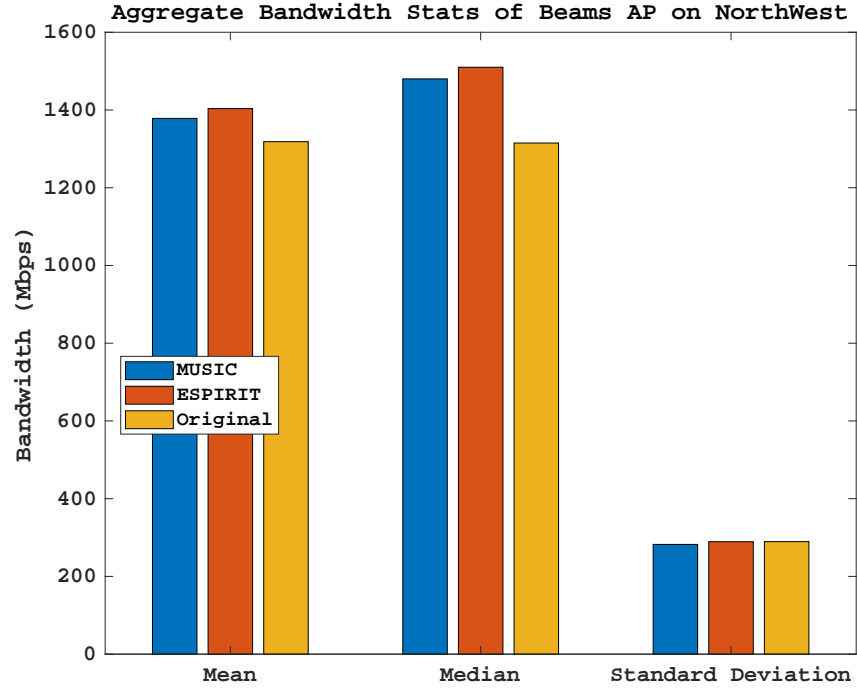


Figure 5.16 Statistics over all subcarriers NW

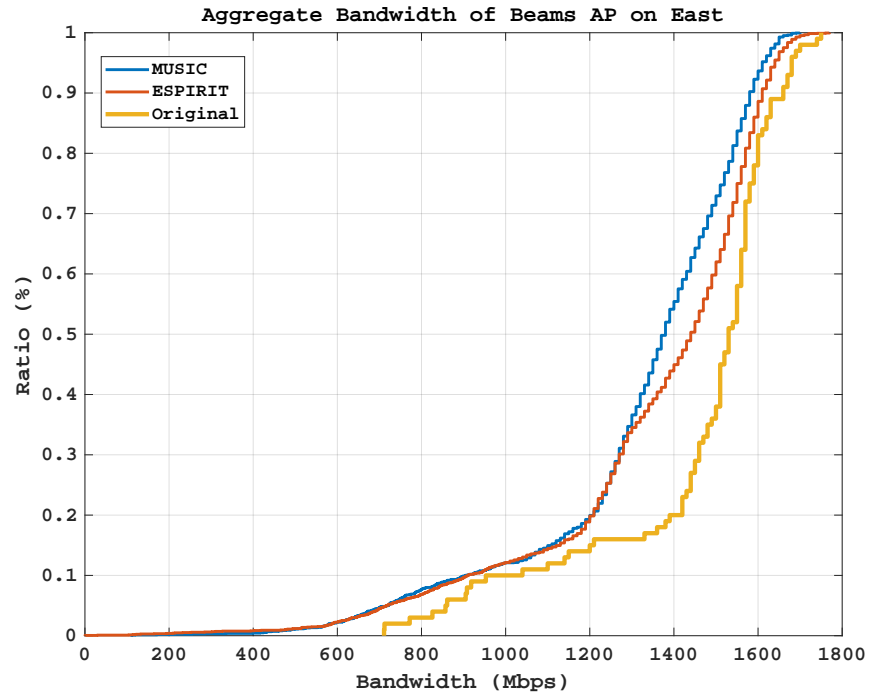


Figure 5.17 CDF over all subcarriers E

As we can see in Figure 5.19, the Capon method has the best performance in terms of bandwidth of the communication link. The beam that establishes the communication is created by the DoAs that this method

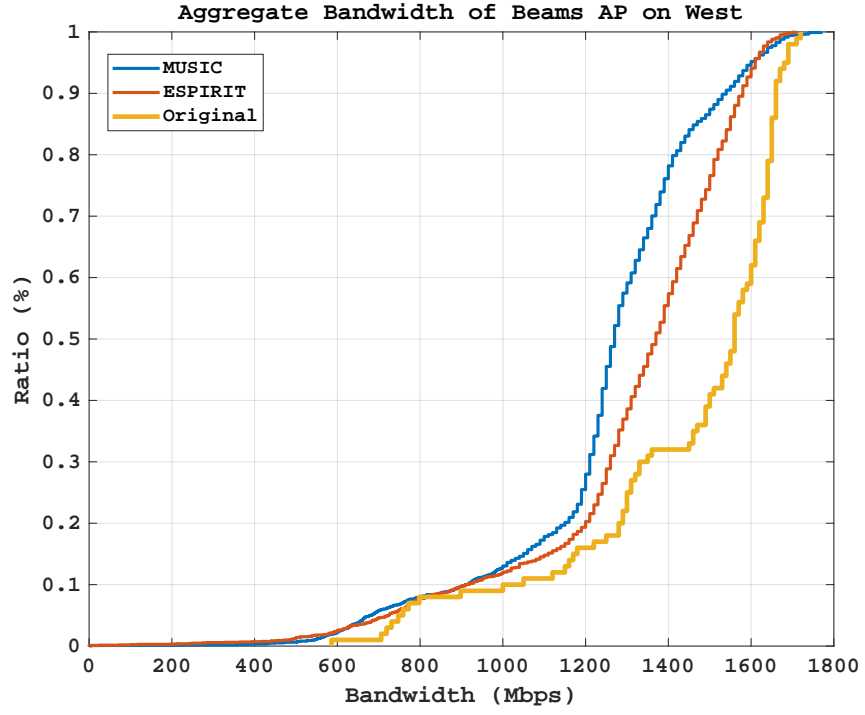


Figure 5.18 CDF over all subcarriers W

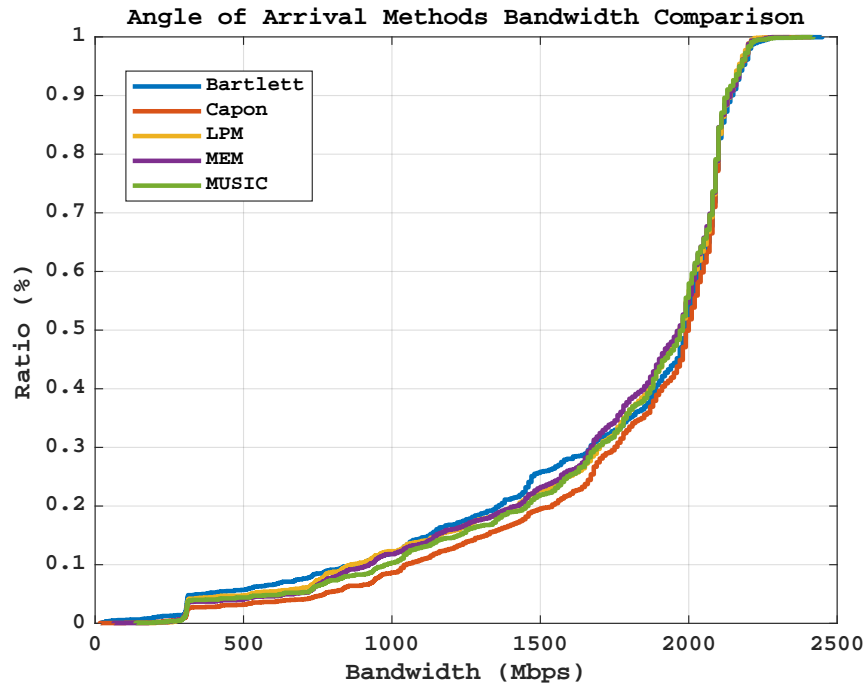


Figure 5.19 CDF of performance of the DoA estimation methods

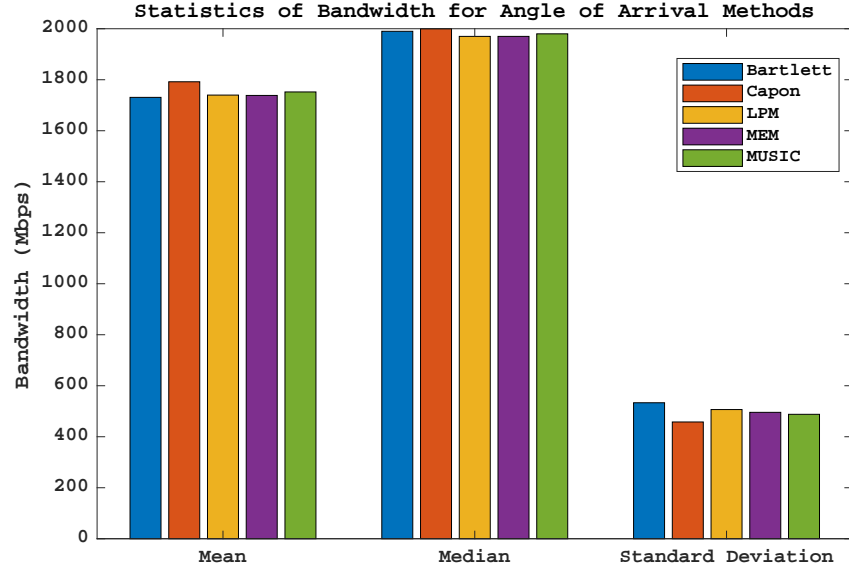


Figure 5.20 Statistics of performance of the DoA estimation methods

estimates. It also shows that this superiority is more meaningful for lower bandwidth values. The main reason is using the cross-power spectral density in the process of estimation. To better see the difference in the performances, we show some statistics in Figure 5.20. As shown in this figure, the Capon DoA method has the highest mean and median. It also has the lowest standard deviation that shows that it is more stable around the mean compared to other methods.

We compare the performance of our proposed out-of-band path discovery method with the Adaptive Codebook Optimization (ACO) method [96] and the IEEE 802.11ad Sector-level Sweep (SLS) method that is done in all of the WiFi devices that have mmW radio. We perform the orientation experiment for North, North West, and West AP orientation with respect to the STA. Figure 5.21 shows the CDF of the bandwidth when the AP is in the North of the STA. As shown in Figure 5.21, our proposed out-of-band method that uses the Capon method for DoA estimation has the best performance compared to ACO and IEEE 802.11ad. The out-of-band method's performance is better than the other two methods by a small amount. However, the low overhead of the out-of-band method makes this method more preferable over the ACO and SLS. We show the North West and West orientations in Figures 5.22 and 5.23.

For the North West orientation, the out-of-band method still performs better than the ACO and the IEEE 802.11ad. However, for the case of the AP positioned on the West (or exactly to the side) of the STA, the performance of the out-of-band method degrades. This degradation is mainly due to the limitation of the

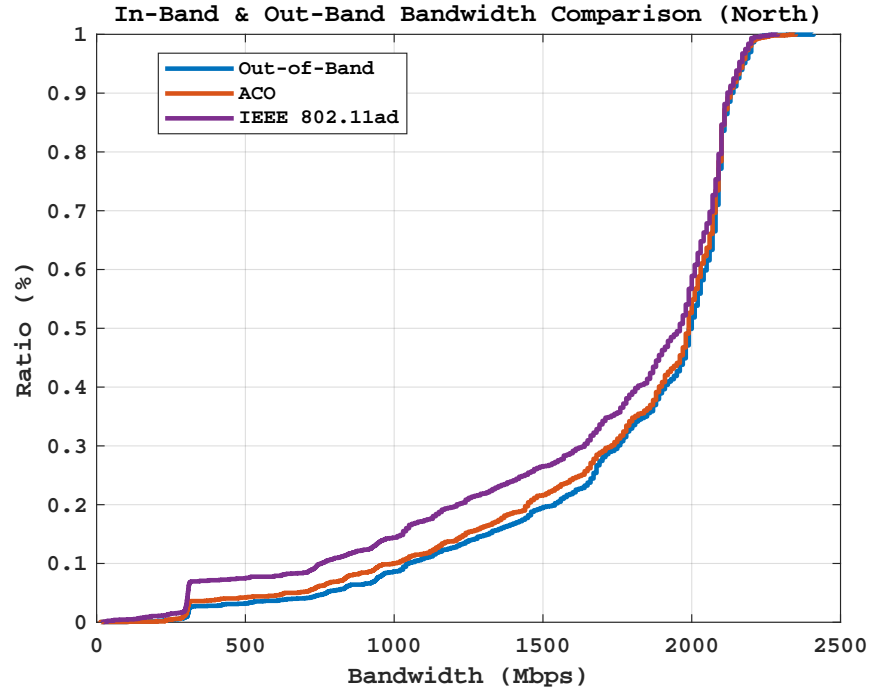


Figure 5.21 CDF of performance of the beam training methods (North)

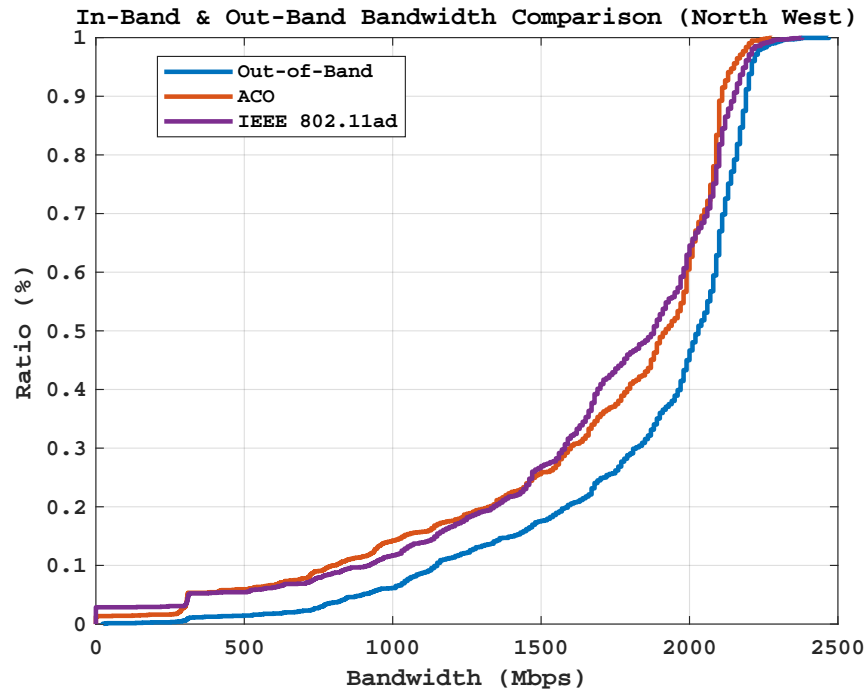


Figure 5.22 CDF of performance of the beam training methods (North West)

hardware and the DoA estimation methods when they try to estimate the DoA of a signal that comes close to the boundaries of the detectable angle spectrum which are 0° and 180° . To improve this limitation, we need

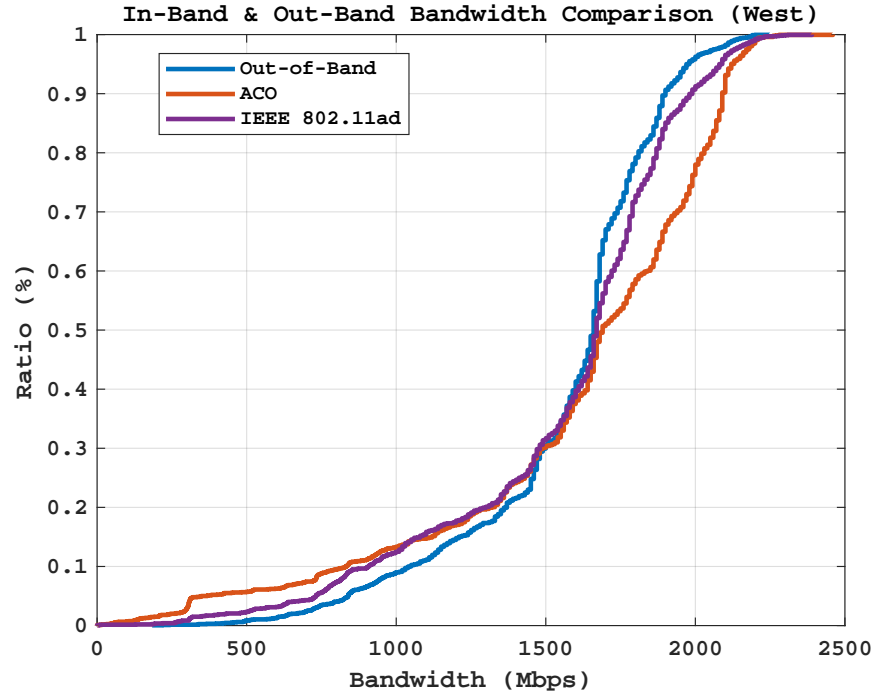


Figure 5.23 CDF of performance of the beam training methods (West)

Table 5.1 Statistics of the Orientation Experiment

Statistics	Mean (Mbps)			Median (Mbps)			Standard Deviation (Mbps)		
	ACO	OB	IEEE	ACO	OB	IEEE	ACO	OB	IEEE
<i>North</i>	1765	1792	1682	1990	2000	1960	484	458	560
<i>North West</i>	1683	1852	1679	1920	2030	1880	539	429	538
<i>West</i>	1603	1566	1563	1690	1660	1670	499	336	413

a new antenna array for the out-of-band path estimation to cover the side planes.

We show the statistics of the orientation experiments in Table 5.1. In this table, "OB" is the abbreviation of the out-of-band method.

As we can see in Table 5.1, as the AP moves from North to West, all of the statistics for the out-of-band decrease in general. This shows that there is a limitation of resolving paths for the out-of-band methods when the transmitter transmits from the sides. Although the same pattern is visible for the other two in-band methods, for West orientation the values of mean and mean and median for ACO are more than the out-of-band. This shows that ACO is more efficient in finding paths that are on the sides. We expect this result since ACO uses both front and back antenna elements for generating beams.

Figure 5.24 compares the performance of the beams generated from the paths that the CSI matrices of

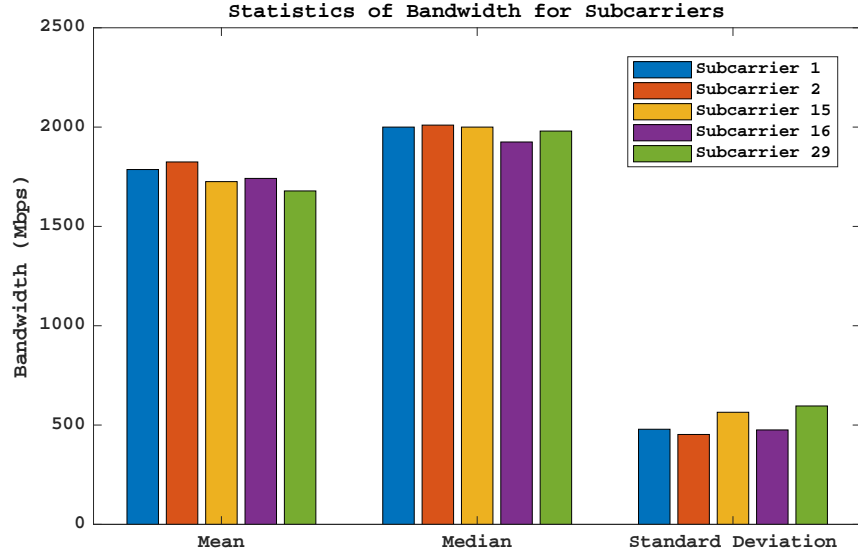


Figure 5.24 Statistics of performance of different subcarriers

different subcarriers in the sensing band discover.

As depicted in Figure 5.24, the performance of the out-of-band degrades as the subcarrier index increases. This means the CSI matrices measured on higher frequencies are noisier than the lower frequency matrices. This would cause mathematical problems for the DoA estimation methods. The fault originates from the CSI measurement device.

5.5 Summary

In this chapter, we study the use of another frequency spectrum to assist the beam alignment of the mmW band. Using COTS hardware, we do extensive experiments in different environments to show the possibility of such assistance. The indoor environment results show that the paths discovered using the out-of-band signal can be used for beam alignment in the mmW band. Due to the high level of multipath fading in the mmW band, the discovered paths have less compliance with the mmW spatial spectrum for the outdoor environment. Both of the DoA estimation methods show consistency in finding the paths and can capture the mmW band paths even with high multipath. There are some inaccuracies in the samplings as well. The inaccuracies come from the limited number of antenna elements and inherent errors of the COTS hardware. Using intelligent methods to mitigate the inaccuracies resulting from hardware limitations is left for future investigations.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this chapter, we first draw conclusions on the research described in this dissertation, and then, we discuss the future research directions from the current research.

6.1 Conclusion

We address several problems in mmW communication in this dissertation in different chapters. We discuss the conclusions of the chapters in the following sections.

6.1.1 Blockage Score and Load Balancing Joint Optimization

User association is considered as a method to maintain the link quality. We propose new methods of user association that are tailored to the propagation characteristics of the mmW signal.

To mitigate the effect of blockage on the quality of links in a mmW communication system, we propose a user association scheme that is aware of the blockage history in the neighborhood of a UE. We devise a new algorithm for the user association problem to mitigate the blockage count in mmW networks. We compare the performance of our proposed algorithm with other user association algorithms called Load Balancing, Closest BS, and Max SINR. According to different metrics, our proposed algorithm outperforms Load Balancing and Closest BS and Max SINR for blockage count by 21%. It also enhances the rate compared to experimented algorithms. We also propose a modification to our algorithm to enhance its time complexity. Experiments show that the modification can slightly improve the performance of the original algorithm with a better time complexity. We also embed the blockage scoring system we propose in the Load Balancing, but the experiments show a negative effect of using our blockage scores in the Load Balancing's performance.

We design the proposed algorithm to mitigate blockage for scenarios in which the blocking objects move in a semi-dynamic manner. That entails having human bodies moving in groups and predictable directions with limited deviation from the main path. For more dynamic settings, the overhead of the algorithm goes up.

When we use the proposed blockage mitigation approach, the load balance of the network may deteriorate.

This resulted in defining another problem and proposing the solution for it. To solve this problem, we define a biobjective optimization problem to optimize load balancing objective while minimizing the blockage score of the network. Since it is a biobjective optimization problem, we use a class of scalarization methods to convert the objective vector to an objective scalar. Then, for each of these scalarization methods, we define the single objective optimization problem. Through Lagrangian analysis, we provide the iterative and the linear solution for each of the scalarization methods. Since we use the dual Lagrangian approach for obtaining the solution, it is necessary to provide a sound analysis of the duality gap between the solutions of the primal and the Lagrangian dual. We provide such analysis for all of the scalarization methods that we use.

We show that the average blockage count for all of the scalarization methods is better than just the load balancing method. It is also evident from the results that the proposed dual Lagrangian solution is very close to the results of a solver that solves the primal optimization problems. It shows that there is a tight duality gap which we show in the duality gap analysis as well. The false positive and false negative metrics also show improvements over both optimizing the blockage score and the load balancing objectives individually. Since the blockage count is improved for the biobjective solutions, the rate for those methods is also improved compared to the load balancing method. The load objective is improved for the biobjective optimization methods compared to optimizing the blockage score. Not only the number of blocking objects is effective on the blockage score but also the distribution of the blocking objects can play an important role in changing the likelihood of blockage. Considering all metrics, we show that the weighted sum scalarization method can have better results compared to other scalarization methods.

6.1.2 Link-Aware mmW Communication Improvement

Link quality degradation in mmW networks is caused by different reasons. To have an effective solution, it is important to detect the reason of the link degradation such as link blockage or link movement. Therefore, we define the problem of predicting a blockage and discriminating between a link movement and a link blockage. To solve this problem, we propose a learning framework to train a classifier for predicting the blockage. We define a data space of the diffraction values on paths formed around a receiver. We also add a burst feature to the data space. This feature can indicate blockage burstiness at the moment that the algorithm samples a data point. The training algorithm trains the classifiers by data points generated in such a data space. The

learning modality is unsupervised online learning. We show that online boosting could have a positive effect on the accuracy of the classifiers. A cascade classifier with the online boosting has the best results. The pure batch training cannot have the accuracy required by the application we pursue. Therefore, an online enhancement can be useful in improving the accuracy. We provide the prediction accuracy with different surface permittivities. With the increase of this parameter, the accuracy of the prediction increases in general. Also, the mobility model of the mobile nodes has an effect on the prediction accuracy. With random walk, which is widely used for modeling human waking, the prediction accuracy is improved compared to a simple line movement. We also present the results from a hardware testbed. In these experiments, we measure the prediction accuracy of the blockage from the signal pattern that we obtain from the hardware devices. We show that all of the classifiers can reach high accuracy rates in predicting the blockage from the signal patterns.

After detecting that the reason for the link degradation is link blockage, nodes need to perform a beam alignment according to the updated signal paths with a low overhead. We propose an out-of-band method to find the signal paths that are formed around the receiver in the mmW spectrum from the CSI measured in the 2.4/5 GHz spectrum. The system works by measuring CSI from the sensing band and forming the beam in the mmW band on the DoAs that the CSI of the out-of-band estimates. The nodes then use the beam for communication in the mmW band. We provide two main types of results. The first type shows how the signal pattern from the out-of-band estimated CSI aligns with the mmW signal pattern. This shows that the out-of-band method that we propose can have high compliance with the mmW signal pattern. The second category of the evaluation results targets the performance of the mmW link that the estimated DoAs from the out-of-band signal created. Putting the transmitter in different orientations compared to the receiver shows that the performance of the out-of-band DoA estimation method is higher than the predefined beams of the devices. However, the performance decreases when the discovered paths are close to the sides of the out-of-band antenna array. This is due to the limitation of the DoA estimation methods. We also compare the out-of-band method with a well-known method of measuring the CSI in the mmW spectrum. The results show that the out-of-band method's performance is comparable to this in-band method. We compare various DoA estimation methods with each other to see how they differ in terms of the generated beam based on the output of the estimated DoA. Capon method of DoA estimation has better performance over other methods.

6.1.3 Contribution

In this dissertation, we discuss several challenges in mmW networks including user association to improve blockage and load balance, blockage prediction using diffraction, and mmW signal path discovery using out-of-band signal. We provide the first user association criterion specifically tailored to the mmW networks. The blockage score is a user association criterion that ranks the BSs based on the likelihood of the blockage if a UE connects to them. A UE is connected to a BS if it has the lowest blockage score. While the reduction of the blockage is of high interest in the mmW networks, optimizing the user association only based on the blockage score makes the network unbalanced. This is where the next contribution of this dissertation comes into play. The second contribution of this dissertation is defining a biobjective optimization problem to minimize the blockage score and balance the load in the network. Then, the solution to the Lagrangian dual problem of the biobjective problem is provided. The time complexity of this solution is linear while other solutions to the user association problems have exponential time complexity. In this dissertation, we prove that our solution does not differ much from the solution to the primal optimization problem. To perform beamforming efficiently in mmW networks, the nodes need to know the reason for a link degradation. The third contribution of this dissertation is a fast and accurate classification of the link degradation. We provide a successful application of machine learning methods to solve the problem of detecting link blockage and link movement in mmW networks. We provide a solution that is fast and accurate for detecting the link degradation cause. After detecting that the link degradation is due to a blockage, the node needs to find the alternative paths quickly to align its beams. The fourth contribution of this dissertation is providing a method for fast and low-overhead beam alignment. We provide a method to use the out-of-band signal for discovering the paths of the mmW signal. Our proposed method provides a fast and low-overhead method of the beam alignment in an environment in which the mmW channel changes fast. We also perform the experiments using the COTS devices to show the effectiveness of our solution.

6.2 Future Research Directions

In this section, we discuss several future research directions.

6.2.1 Path Discovery for User Association

In Chapters 2 and 3, we discuss our proposed methods of user association that uses blockage history to connect UEs to BSs, while it balances the load throughout the network. This metric is a heuristic and is prone to prediction error in high blockage situations. A more deterministic criterion is the signal paths from BSs to a UE that are discovered by out-of-band signal. Therefore, a BS selection metric is the mmW signal paths that reach the UE from different BSs in the communication range of the UE which are obtained by DoA estimation methods using out-of-band signals. The current state-of-the-art does not consider this metric for the user association problem. It is favorable to have the BS selection based on the paths with minimum overhead in both processing and mmW band scanning. The other metric is beamwidth. A wider beamwidth can make the mmW link less sensitive to the movement of UEs, but it can decrease the signal strength sensed by UE and BS's antenna array. On the other hand, a narrower beamwidth results in a better signal quality but it requires more precise beam tracking algorithms as the UE moves.

In this user association method, we build the pool of paths reaching to the UE from all of the BSs in the communication range (within the communication radius) of the UE. The BS with the strongest path is selected for handover. In this method, the UE goes to monitor mode to hear the packets that are not intended for the UE and captures them for calculating the CSI and the DoA estimation methods.

On the implementation side, the UE can only be associated to one BS. While a UE is not connected to a BS, it cannot change the beams with conventional methods of beam changing. We propose to have extensive research and development on methods of controlling the beam and getting signal strength while the UE is not connected to the BS. The methods can be difficult to develop since the programs in the procedures are working with the kernel of the operating system and even the firmware of the wireless hardware. Moreover, it is challenging to devise a method for processing the signal strength to determine the strongest path because the signal strength changes quickly especially in dynamic environments. It is essential to develop metrics that can capture stable signal paths for user association purposes.

6.2.2 Out-of-Band Method for Classic Problems in mmW

In Chapter 5, we discuss the problem of finding the signal paths in mmW spectrum using the signal from another frequency band. We only focus on the path finding problem in mmW. However, the problems in

mmW are not limited to finding the signal paths. In fact, classic problems in non-mmW networks can be valid problems in mmW networks as well. These problems can be more challenging in mmW spectrum due to the directional propagation of mmW signal.

Another research direction is using the out-of-band method for different problems that are defined for sub-6 GHz spectrum but with the mmW twist. The classic problems that have attracted a lot of research can be defined in the context of mmW with their unique challenges. One application is using the out-of-band method in neighbor discovery problem for the mmW spectrum. Since the beams with a limited view perform the mmW communications, it is even more challenging to coordinate two or more mmW nodes to discover each other. It is very useful to use the out-of-band method to replace the mmW band communication as a radar for detecting the neighbors.

Another application of the out-of-band method is in mmW Omni-directional communications. In some COTS devices, there is more than one phased array antenna. One can lay out these arrays in a way to cover 360° space and one phased array can cover one sector in space. The algorithm can use the out-of-band method to discover paths in all directions and it then uses those discovered paths to form the beams on different phased arrays for an Omni-directional coverage in mmW. Since the DoA estimation methods use the antenna arrays for sub-6 GHz frequencies to estimate DoAs for 180° space, it is essential to have two arrays for the out-of-band system to estimate the DoA of the signals for all 360° space. The challenge is aligning the DoAs of the paths discovered on different phased array antennas accurately.

Interference mitigation in sub-6 GHz is an important task that is essential to achieve the defined data rates. Although the communication is directional in mmW networks, the interference can still happen. The beams from different nodes can overlap and cause interference. Moreover, the interference can be severe due to high data rates in mmW communication because it can corrupt more data compared to sub-6 GHz networks. Therefore, it is essential to consider interference in mmW networks. The directionality of communication in mmW introduces the challenge of detecting interference and avoiding it. Using out-of-band signal for the purpose of detecting interference is a future direction of research.

In summary, the future research directions that the research in this dissertation inspires include the user association using the paths discovered via the out-of-band signal and employing the out-of-band method to solve classic wireless networks problems in the mmW spectrum. These problems include neighbor

discovery or interference control using out-of-band signal to limit the overhead on the mmW communication significantly. Since the communication in the mmW is directional, using multiple phased arrays to make omnidirectional communication possible in the mmW spectrum is of interest of future research. The approach of using the out-of-band signal in this problem is interesting to investigate since it introduces low-overhead solutions to the problems in the mmW networks. However, there are some challenges that need to be addressed. The most important is the alignment of signal of non-mmW on the mmW signal is critical in the out-of-band methods. The requirement of the precision for this alignment can be different in various applications.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] P. Jonsson, S. Carson, S. Davis, P. Linder, P. Lindberg, J. Ramiro, J. Outes, and A. Bhardwaj, "Ericsson mobility report," Ericsson, Tech. Rep., 2021.
- [2] 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 Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.
- [3] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo, Y. Hu, L. Kuai, Y. Yu, Z. Jiang, Z. Chen, J. Chen, Z. Yu, J. Zhai, N. Zhang, L. Tian, F. Wu, G. Yang, Z.-C. Hao, and J. Y. Zhou, "The role of millimeter-wave technologies in 5g/6g wireless communications," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 101–122, 2021.
- [4] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5g: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015. [Online]. Available: <http://dx.doi.org/10.1007/s11276-015-0942-z>
- [5] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.
- [6] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng, "Demystifying 60ghz outdoor picocells," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '14. New York, NY, USA: ACM, 2014, pp. 5–16. [Online]. Available: <http://doi.acm.org/10.1145/2639108.2639121>
- [7] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60 ghz indoor networking through flexible beams: A link-level profiling," in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '15. New York, NY, USA: ACM, 2015, pp. 71–84. [Online]. Available: <http://doi.acm.org/10.1145/2745844.2745858>
- [8] D. Liu, L. Wang, Y. Chen, M. El Kashlan, K. K. Wong, R. Schober, and L. Hanzo, "User association in 5g networks: A survey and an outlook," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1018–1044, Second quarter 2016.
- [9] K. Sui, S. Sun, Y. Azzabi, X. Zhang, Y. Zhao, J. Wang, Z. Li, and D. Pei, "Understanding the impact of ap density on wifi performance through real-world deployment," in *2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, June 2016, pp. 1–6.
- [10] M. Jacob, S. Priebe, R. Dickhoff, T. Kleine-Ostmann, T. Schrader, and T. Kurner, "Diffraction in mm and sub-mm wave indoor propagation channels," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 833–844, March 2012.
- [11] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, April 2012.
- [12] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3484–3495, October 2012.

- [13] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive mimo wireless networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1835–1850, March 2016.
- [14] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, Feb 2015.
- [15] H. Shokri-Ghadikolaie, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A mac layer perspective," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, Oct 2015.
- [16] G. Athanasiou, P. C. Weeraddana, and C. Fischione, "Auction-based resource allocation in millimeter-wave wireless access networks," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2108–2111, November 2013.
- [17] Y. Xu, G. Athanasiou, C. Fischione, and L. Tassiulas, "Distributed association control and relaying in millimeter wave wireless networks," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.
- [18] 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 Transactions on Networking*, vol. 23, no. 3, pp. 836–850, June 2015.
- [19] K. Sakaguchi, G. K. Tran, H. Shimodaira, S. Nanba, and T. Sakurai, "Millimeter-wave evolution for 5g cellular networks," *IEICE Transaction on Communication*, vol. E98-B, no. 3, pp. 388–402, 2015.
- [20] C. L. I, J. Huang, R. Duan, C. Cui, J. . Jiang, and L. Li, "Recent progress on c-ran centralization and cloudification," *IEEE Access*, vol. 2, pp. 1030–1039, 2014.
- [21] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [22] C. M. R. Institute, "C-ran: The road towards green ran, ver 3.0," *Beijing: China Mobile Res. Inst.*, 2013.
- [23] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 561–574, April 2012.
- [24] D. Fooladivanda, A. A. Daoud, and C. Rosenberg, "Joint channel allocation and user association for heterogeneous wireless cellular networks," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2011, pp. 384–390.
- [25] J. Ghimire and C. Rosenberg, "Resource allocation, transmission coordination and user association in heterogeneous networks: A flow-based unified approach," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1340–1351, March 2013.
- [26] L. Su, C. Yang, Z. Xu, and A. F. Molisch, "Energy-efficient downlink transmission with base station closing in small cell networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 4784–4788.

- [27] K. Sakaguchi, G. K. Tran, H. Shimodaira, S. Nanba, T. Sakurai, K. Takinami, I. Siaud, E. C. Strinati, A. Capone, I. Karls, R. Arefi, and T. Haustein, "Millimeter-wave evolution for 5g cellular networks," *IEICE TRANSACTIONS on Communications*, vol. E98-B, no. 3, pp. 388–402, March 2015.
- [28] B. Xu, Y. Chen, M. El Kashlan, T. Zhang, and K. K. Wong, "User association in massive mimo and mmwave enabled hetnets powered by renewable energy," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.
- [29] A. Alizadeh and M. Vu, "Load balancing user association in millimeter wave mimo networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 2932–2945, June 2019.
- [30] O. Semiari, W. Saad, and M. Bennis, "Downlink cell association and load balancing for joint millimeter wave-microwave cellular networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.
- [31] I. Guvenc, "Capacity and fairness analysis of heterogeneous networks with range expansion and interference coordination," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1084–1087, October 2011.
- [32] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Outage probability for heterogeneous cellular networks with biased cell association," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–5.
- [33] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eicic) in lte hetnets," *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 137–150, Feb 2014.
- [34] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. El Kashlan, "Distributed energy efficient fair user association in massive mimo enabled hetnets," *IEEE Communications Letters*, vol. 19, no. 10, pp. 1770–1773, Oct 2015.
- [35] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier hetnets with large-scale antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3251–3268, May 2016.
- [36] M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, E. Aryafar, S. Yeh, N. Himayat, S. Andreev, and Y. Koucheryavy, "Analysis of human-body blockage in urban millimeter-wave cellular communications," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–7.
- [37] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 ghz with a simple double knife-edge diffraction model and extension for directional antennas," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sept 2016, pp. 1–6.
- [38] J. Choi, "On the macro diversity with multiple bss to mitigate blockage in millimeter-wave communications," *IEEE Communications Letters*, vol. 18, no. 9, pp. 1653–1656, Sep. 2014.
- [39] S. Niknam, B. Natarajan, and R. Barazideh, "Interference analysis for finite-area 5g mmwave networks considering blockage effect," *IEEE Access*, vol. 6, pp. 23 470–23 479, 2018.
- [40] B. Han, L. Wang, and H. D. Schotten, "A 3d human body blockage model for outdoor millimeter-wave cellular communication," *Physical Communication*, vol. 25, pp. 502–510, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1874490717301416>

- [41] B. A. Bilgin, H. Ramezani, and O. B. Akan, "Human blockage model for indoor terahertz band communication," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.
- [42] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can terahertz provide high-rate reliable low latency communications for wireless vr?" 2021.
- [43] E. Triantaphyllou, *Multi-Criteria Decision Making Methods*. Boston, MA: Springer US, 2000, pp. 5–21.
- [44] A. Pascoletti and P. Serafini, "Scalarizing vector optimization problems," *Journal of Optimization Theory and Applications*, vol. 42, no. 4, pp. 499–524, 1984.
- [45] A. Messac, A. Ismail-Yahaya, and C. Mattson, "The normalized normal constraint method for generating the pareto frontier," *Structural and Multidisciplinary Optimization*, vol. 25, no. 2, pp. 86–98, 2003.
- [46] M. Ehrgott and X. Gandibleux, *Multiobjective Combinatorial Optimization — Theory, Methodology, and Applications*. Boston, MA: Springer US, 2002, pp. 369–444.
- [47] A. Wierzbicki, "The use of reference objectives in multiobjective optimisation," in *MCDM theory and Application, Proceedings*, ser. Lecture notes in economics and mathematical systems, F. G. and G. T., Eds., no. 177. Hagen: Springer Verlag, 1980, pp. 468–486.
- [48] Y. Nikulin, K. Miettinen, and M. M. Mäkelä, "A new achievement scalarizing function based on parameterization in multiobjective optimization," *OR Spectrum*, vol. 34, no. 1, pp. 69–87, Jan 2012. [Online]. Available: <https://doi.org/10.1007/s00291-010-0224-1>
- [49] C. Bazgan, S. Ruzika, C. Thielen, and D. Vanderpooten, "The power of the weighted sum scalarization for approximating multiobjective optimization problems," 2019.
- [50] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems," *SIAM J. on Optimization*, vol. 8, no. 3, 1998.
- [51] M. Zarifneshat, C. J. Liu, and L. Xiao, "A protocol for link blockage mitigation in mm-wave networks," in *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, Oct 2017, pp. 215–223.
- [52] "Gurobi optimization - the state-of-the-art mathematical programming solver," <http://www.gurobi.com/>, accessed: 21-5-2018.
- [53] "Omnet++ discrete event simulator - home," <https://www.omnetpp.org/>, accessed: 21-5-2018.
- [54] J. Ariyakhajorn, P. Wannawilai, and C. Sathitwiriawong, "A comparative study of random waypoint and gauss-markov mobility models in the performance evaluation of manet," in *2006 International Symposium on Communications and Information Technologies*, 2006, pp. 894–899.
- [55] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1445–1455, March 2014.

- [56] X. Zhu, A. Doufexi, and T. Kocak, "Throughput and coverage performance for ieee 802.11ad millimeter-wave wpans," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.
- [57] Z. Lan, C.-S. Sum, J. Wang, T. Baykas, F. Kojima, H. Nakase, and H. Harada, "Relay with deflection routing for effective throughput improvement in gbps millimeter-wave wpan systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1453–1465, October 2009.
- [58] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "Beamspy: Enabling robust 60 ghz links under blockage," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. Santa Clara, CA: USENIX Association, 2016, pp. 193–206. [Online]. Available: <https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/sur>
- [59] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with eyes closed: Mm-wave beam steering without in-band measurement," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, April 2015, pp. 2416–2424.
- [60] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.
- [61] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi, "Channel dynamics and snr tracking in millimeter wave cellular systems," in *European Wireless 2016; 22th European Wireless Conference*, May 2016, pp. 1–8.
- [62] Y. Oguma, T. Nishio, K. Yamamoto, and M. Morikura, "Implementation and evaluation of reactive base station selection for human blockage in mmwave communications," in *2015 21st Asia-Pacific Conference on Communications (APCC)*, Oct 2015, pp. 199–203.
- [63] V. Mordachev and S. Loyka, "On node density - outage probability tradeoff in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1120–1131, 2009.
- [64] M. Comisso and F. Babich, "Coverage analysis for 2d/3d millimeter wave peer-to-peer networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3613–3627, 2019.
- [65] M. K. Samimi and T. S. Rappaport, "3-d millimeter-wave statistical channel model for 5g wireless system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2225, 2016.
- [66] M. H. Park and Y. S. Choi, "Performance analysis of degradation detection method on millimeter wave channel," in *2015 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2015, pp. 971–973.
- [67] Y. Yang, S. Zhang, F. Gao, C. Xu, J. Ma, and O. A. Dobre, "Deep learning based antenna selection for channel extrapolation in fdd massive mimo," in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*, 2020, pp. 182–187.
- [68] Y. Yang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Graph neural network-based channel tracking for massive mimo networks," *IEEE Communications Letters*, vol. 24, no. 8, pp. 1747–1751, 2020.
- [69] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmwave beam and blockage prediction using sub-6ghz channels," 2019.

- [70] Z. Ali, A. Duel-Hallen, and H. Hallen, "Early warning of mmwave signal blockage and aoa transition using sub-6 ghz observations," *IEEE Communications Letters*, vol. 24, no. 1, p. 207–211, Jan 2020. [Online]. Available: <http://dx.doi.org/10.1109/LCOMM.2019.2952602>
- [71] N. Gonzalez-Prelcic, A. Ali, V. Va, and R. W. Heath, "Millimeter-wave communication with out-of-band information," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 140–146, 2017.
- [72] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, "Wifi-assisted 60 ghz wireless networks," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 28–41. [Online]. Available: <https://doi.org/10.1145/3117811.3117817>
- [73] C. K. Anjinappa and I. Guvenc, "Angular and temporal correlation of v2x channels across sub-6 ghz and mmwave bands," 2018.
- [74] Y. Wang, M. Narasimha, and R. W. Heath, "Mmwave beam prediction with situational awareness: A machine learning approach," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.
- [75] —, "Towards robustness: Machine learning for mmwave v2x with situational awareness," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 1577–1581.
- [76] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmwave beam alignment via correlated bandit learning," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5894–5908, 2019.
- [77] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1244–1248, 2019.
- [78] S. Lin and I. F. Akyildiz, "Dynamic base station formation for solving nlos problem in 5g millimeter-wave communication," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.
- [79] J. Palacios, G. Bielsa, P. Casari, and J. Widmer, "Communication-driven localization and mapping for millimeter wave networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 2402–2410.
- [80] M. Hashemi, A. Sabharwal, C. Emre Koksall, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 2393–2401.
- [81] I. Aykin, B. Akgun, and M. Krunz, "Smartlink: Exploiting channel clustering effects for reliable millimeter wave communications," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1117–1125.
- [82] J. Wang, H. Jiang, J. Xiong, K. Jamieson, X. Chen, D. Fang, and B. Xie, "Lifs: Low human-effort, device-free localization with fine-grained subcarrier information," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16. New York, NY, USA: ACM, 2016, pp. 243–256. [Online]. Available: <http://doi.acm.org/10.1145/2973750.2973776>

- [83] J. P. Lindsey, "The fresnel zone and its interpretive significance," *The Leading Edge*, vol. 8, no. 10, pp. 33–39, 1989.
- [84] J. Trogh, D. Plets, E. Surewaard, M. Spiessens, M. Versichele, L. Martens, and W. Joseph, "Outdoor location tracking of mobile devices in cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 115, May 2019. [Online]. Available: <https://doi.org/10.1186/s13638-019-1459-4>
- [85] T. Wei, A. Zhou, and X. Zhang, "Facilitating robust 60 ghz network deployment by sensing ambient reflectors," in *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*. Boston, MA: USENIX Association, 2017, pp. 213–226. [Online]. Available: <https://www.usenix.org/conference/nsdi17/technical-sessions/presentation/wei-teng>
- [86] D. Moltchanov, A. Ometov, P. Kustarev, O. Evsutin, J. Hosek, and Y. Koucheryavy, "Analytical tcp model for millimeter-wave 5g nr systems in dynamic human body blockage environment," *Sensors (Basel, Switzerland)*, vol. 20, no. 14, p. 3880, Jul 2020, 32664617[pmid]. [Online]. Available: <https://doi.org/10.3390/s20143880>
- [87] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5070–5083, 2014.
- [88] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [89] B. Wu and R. Nevatia, "Improving part based object detection by unsupervised, online boosting," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.
- [90] J. Hunter, "The exponentially weighted moving average," *JOURNAL OF QUALITY TECHNOLOGY*, vol. 18, no. 4, pp. 203–210, October 1986.
- [91] D. Steinmetzer, J. Classen, and M. Hollick, "mmtrace: Modeling millimeter-wave indoor propagation with image-based ray-tracing," in *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2016, pp. 429–434.
- [92] W. BARBAKH and C. FYFE, "Online clustering algorithms," *International Journal of Neural Systems*, vol. 18, no. 03, pp. 185–194, 2008, pMID: 18595148. [Online]. Available: <http://www.worldscientific.com/doi/abs/10.1142/S0129065708001518>
- [93] R. Zajicek and J. Vrba, *Broadband Complex Permittivity Determination for Biomedical Applications*. Books on Demand, 04 2010, pp. 365–386.
- [94] R. Roy, R., *Handbook of Mobile Ad Hoc Networks for Mobility Models*. Springer, 2011, ch. Random Walk Mobility.
- [95] H. Almogdady. (2021) Random mobility models (rwp,rd,rwk). [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/63194-random-mobility-models-rwp-rd-rwk>
- [96] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer, "Adaptive codebook optimization for beam training on off-the-shelf ieee 802.11ad devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 241–255. [Online]. Available: <http://doi.acm.org/10.1145/3241539.3241576>

- [97] D. Steinmetzer, D. Wegemer, and M. Hollick. (2018) Talon tools: The framework for practical ieee 802.11ad research. [Online]. Available: <https://seemoo.de/talon-tools/>
- [98] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.
- [99] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.
- [100] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18. New York, NY, USA: ACM, 2018, pp. 432–445. [Online]. Available: <http://doi.acm.org/10.1145/3230543.3230581>
- [101] A. Zhou, T. Wei, X. Zhang, and H. Ma, "Fastnd: Accelerating directional neighbor discovery for 60-ghz millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 5, pp. 2282–2295, 2018.
- [102] S. Wang, J. Huang, X. Zhang, H. Kim, and S. Dey, "X-array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: <https://doi.org/10.1145/3372224.3380882>
- [103] Y. Ghasempour, M. K. Haider, C. Cordeiro, D. Koutsonikolas, and E. Knightly, "Multi-stream beam-training for mmwave mimo networks," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 225–239. [Online]. Available: <http://doi.acm.org/10.1145/3241539.3241556>
- [104] S. Jog, J. Wang, J. Guan, T. Moon, H. Hassanieh, and R. R. Choudhury, "Many-to-many beam alignment in millimeter wave networks," in *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*. Boston, MA: USENIX Association, 2019, pp. 783–800. [Online]. Available: <https://www.usenix.org/conference/nsdi19/presentation/jog>
- [105] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, "Towards scalable and ubiquitous millimeter-wave wireless networks," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 257–271. [Online]. Available: <http://doi.acm.org/10.1145/3241539.3241579>
- [106] A. Zhou, X. Zhang, and H. Ma, "Beam-forecast: Facilitating mobile 60 ghz networks via model-driven beam steering," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.
- [107] A. Zhou, S. Xu, S. Wang, J. Huang, S. Yang, T. Wei, X. Zhang, and H. Ma, "Robot navigation in radio beam space: Leveraging robotic intelligence for seamless mmwave network coverage," in *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. Mobihoc '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 161–170. [Online]. Available: <https://doi.org/10.1145/3323679.3326514>
- [108] A. Zhou, S. Xu, S. Wang, J. Huang, S. Yang, T. Wei, and X. Zhang, "Robotic millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1534–1549, 2020.

- [109] S. Wang, J. Huang, and X. Zhang, “Demystifying millimeter-wave v2x: Towards robust and efficient directional connectivity under high mobility,” in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: <https://doi.org/10.1145/3372224.3419208>
- [110] M. Alrabeiah and A. Alkhateeb, “Deep learning for mmwave beam and blockage prediction using sub-6 ghz channels,” *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5504–5518, 2020.
- [111] M. K. Haider, Y. Ghasempour, D. Koutsonikolas, and E. W. Knightly, “Listeer: Mmwave beam acquisition and steering by tracking indicator leds on wireless aps,” in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’18. New York, NY, USA: ACM, 2018, pp. 273–288. [Online]. Available: <http://doi.acm.org/10.1145/3241539.3241542>
- [112] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, “Wifi-assisted 60 ghz wireless networks,” in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’17. New York, NY, USA: ACM, 2017, pp. 28–41. [Online]. Available: <http://doi.acm.org/10.1145/3117811.3117817>
- [113] T. Woodford, X. Zhang, E. Chai, K. Sundaresan, and A. Khojastepour, “Spacebeam: Lidar-driven one-shot mmwave beam management,” in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 389–401. [Online]. Available: <https://doi.org/10.1145/3458864.3466864>
- [114] P. Stoica and K. C. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, July 1990.
- [115] O. A. Oumar, M. F. Siyau, and T. P. Sattar, “Comparison between music and esprit direction of arrival estimation algorithms for wireless communication systems,” in *The First International Conference on Future Generation Communication Technologies*, Dec 2012, pp. 99–103.
- [116] M. S. Bartlett, “Smoothing periodograms from time-series with continuous spectra,” *Nature*, vol. 161, no. 4096, pp. 686–687, May 1948. [Online]. Available: <https://doi.org/10.1038/161686a0>
- [117] A. Martini, S. Schmidt, and W. D. Pozzo, “Maximum entropy spectral analysis: a case study,” 2021.
- [118] P. Stoica, P. Händel, and T. Söderström, “Study of capon method for array signal processing,” *Circuits, Systems and Signal Processing*, vol. 14, no. 6, pp. 749–770, Nov 1995. [Online]. Available: <https://doi.org/10.1007/BF01204683>
- [119] Y. Khmou, S. Safi, and M. Frikel, “Comparative study between several direction of arrival estimation methods,” *Journal of telecommunications and information technology*, vol. 1, pp. 41–48, 2014.
- [120] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, “Ieee 802.11ad: directional 60 ghz communication for multi-gigabit-per-second wi-fi [invited paper],” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, December 2014.
- [121] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson, 2014.

- [122] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, “Tool release: Gathering 802.11n traces with channel state information,” *ACM SIGCOMM CCR*, vol. 41, no. 1, p. 53, Jan. 2011.
- [123] (2020) iperf - the tcp, udp and sctp network bandwidth measurement tool. [Online]. Available: <https://iperf.fr/>