Machine-Type-Communication in 5G Cellular System

by

Yue Li

B. Eng., Beijing Institute of Technology, 2006

M. Eng., Beijing Institute of Technology, 2008

A Dissertation Submitted in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electric and Computer Engineering

Machine-Type-Communication in 5G Cellular System

by

Yue Li

B. Eng., Beijing Institute of Technology, 2006

M. Eng., Beijing Institute of Technology, 2008

Supervisory Committee

Dr. Lin Cai, Supervisor

(Department of Electrical and Computer Engineering)

Dr. Xiaodai Dong, Department Member

(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member

(Department of Mechanical Engineering)

ABSTRACT

The rapid development of Machine-Type-Communication (MTC) has brought big challenges to cellular networks such as super-dense devices and high-shadowing channels which may substantially decrease the spectrum efficiency and increase devices' power consumption. It is pressing to improve the transmission efficiency for MTC due to the limited wireless spectrum. Lower efficiency may also lead to longer transmission time and more energy consumption which conflict with MTC's requirement of lower power consumption.

In order to address the above issues, we propose to apply Network Coding (NC) and Device-to-Device (D2D) communications to MTC devices. Our approach introduces an additional delay for local packet exchange, which is acceptable given that MTC traffic typically has the feature of delay tolerance to certain degree. The benefit of the proposed approach is that the cellular transmissions are no longer user-specific, and thus an additional multi-user diversity gain is achieved. The cellular transmission efficiency will also be increased. How to apply the proposed approach for both downlink and uplink has been studied. For the downlink, in addition to the reduction of cellular resource consumption, the MTC devices' feedback load can also be significantly reduced because the cellular transmissions are not sensitive to user-specific errors. In the uplink, besides the enhanced transmission efficiency for full-buffer traffic, an additional small-data aggregation gain is achieved for MTC small-data traffic. Theoretical performance analyses for both downlink and uplink and the corresponding numerical evaluations are given.

Though the proposed NC and D2D approach can improve the transmission efficiency by exploring multi-user diversity gain, poor-quality MTC channels still exist which affect system performance. When the whole group MTC devices in an area experience high shadowing and penetration loss, we have to increase either the resource

consumption or the transmitting power to overcome the poor-quality channels. The existing small-cell solution can improve the MTC channel quality, but MTC's unique traffic characteristics and quality of service requirements, as well as other practical issues, make the small-cell deployment unprofitable. Therefore, we propose a solution using Floating Relay (FR) given the mature technologies of Unmanned Aerial Vehicle (UAV). We first target on the high-shadowing channels of the MTC devices and introduce the FR into the cellular system to improve the transmission efficiency and maximize the system capacity. An optimization problem, given the capacity limit of the FR's back-haul link and the maximum transmission power of each user, is formulated and then theoretically solved. An effective on-line flight path planning algorithm is also proposed.

Then, we extend the FR concept to a bigger picture and propose the UAV-assisted heterogeneous cellular solution. Detailed system design and comprehensive analyses on FR-cells deployment including frequency reuse, interference, backhaul resource allocation, and coverage are given. For UAV assisted networking systems, mobility and topology play important roles. How to dispatch a UAV to the optimal location in a mesh network to enhance the coverage and service of the existing network is a critical issue. Given the topology of existing service nodes, a new supplementary UAV can be sent to improve the quality of service especially for the users with poor-quality channels. The location of a newly added UAV is optimized to improve the service quality to the worst point.

In summary, we propose two means to improve the transmission efficiency for MTC in this thesis work. The NC and D2D approach can be used when some of the MTC devices have chances to experience better channels because of the fast fading and uneven shadowing. Otherwise, the FR can be applied to proactively improve the channel quality for MTC. The NC and D2D approach sticks to the latest

standard in the cellular system and thus provides a down-to-earth and backward-compatible MTC solution for 5G cellular system. The UAV-assisted heterogeneous cellular solution and UAV mesh networks can enable mobile Internet and ultra-reliable low latency communications, respectively. These solutions together effectively and efficiently support MTC, which is key to future proliferation of Internet of Things

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

my supervisor, Dr. Lin Cai, for her trust, supporting, mentoring, encouragement, and patience. She is like a lighthouse amidst the chaos, guiding me to roam in the sea of knowledge and make my vision of a better wireless network less remote. Sincerely, this is my most memorable 4-year in my life by now. I feel lucky and grateful to be her student.

Dr. Xiaodai Dong, Dr. Yang Shi, and Dr. Hai Jiang, for spending precious time to serve as my supervisory committee. Starting from the candidate exam, I have been following their valuable suggestions and finally reach the most important milestone of my life.

Dr. Hongchuan Yang, Dr. Wusheng Lu, and Dr. Jianping Pan, for their wonderful courses and valuable advices on my research.

my labmates, classmates, colleagues, and other friends, for getting along with me in these years. We grow up together, and make each other better people.

my parents, for giving my life.

my wife, for making the ride of my life worthwhile.

Yue Li, Victoria, BC, Canada

DEDICATION

To those who are working on the next generation of wireless technologies.

Chapter 1

Introduction

1.1 Background

Application demands, including their Quality of Service (QoS) requirements, and communication channel characteristics have been two most important factors that drive the evolution of cellular systems. In digital wireless communication era, to support ever-growing data rate requirements, the cellular system technologies have evolved from the Global System for Mobile communications (GSM) system mainly dealing with voice calls and short messages, to the Universal Mobile Telecommunications System (UMTS) supporting multimedia traffic up to 14 Mbps, and then to the current Long-Term Evolution (LTE) system which increases the data rate to 100 Mbps. The upcoming LTE-A and 5G aiming to further improve the spectrum efficiency to accommodate the latest data-thirsty applications such as high-resolution videos and virtual reality.

Wireless channels change dramatically in different scenarios and thus motivate different techniques. From the suburban area to cities, the cellular coverage changes from the homogeneous deployment consisting of only macro cells to the heterogeneous

case where small-cells are added to compensate the high path loss and shadowing in macro cells. Picocells are deployed indoors to compensate the high penetration loss through the wall of a building. Relay nodes are installed on the high-speed train to deal with the unstable channels caused by the high mobility and provide a better service to the users in the train.

Machine Type Communications (MTC), or Machine to Machine (M2M) communications, is a key to Internet of Thing (IoT) services and considered as one of the most important use cases in the 5G cellular system. They are quite different from traditional Human-to-Human (H2H) communications in terms of both the traffic and channel characteristics. The traffic generated by MTC devices typically has the features of small-data and delay tolerance, while the number of devices can be massive. Transmitting a large volume of low priority MTC small packets will decrease the system spectrum efficiency. In addition, due to their indoor or underground deployment, the channels of MTC devices have the characteristics of high penetrating loss and high shadowing, which significantly reduce the transmission efficiency and increase the energy consumption of MTC devices. Above challenges motivate us to study how to provide better support for MTC without sacrificing H2H users' experience in cellular systems.

Our solutions can be classified into two main categories as shown in Fig. 1.1. In the first category, we utilize the MTC's delay tolerance feature and group them together. In each group, they can exchange packets before and/or after the cellular transmission and then make the cellular transmission more efficient. Specifically, we apply Network Coding (NC) and Device-to-Device (D2D) communication to MTC without introducing additional networking nodes, where the transmissions are partially offloaded to the local D2D network and the cellular transmission efficiency is improved thanks to the additional multi-user diversity gain. The downlink and uplink cases are studied

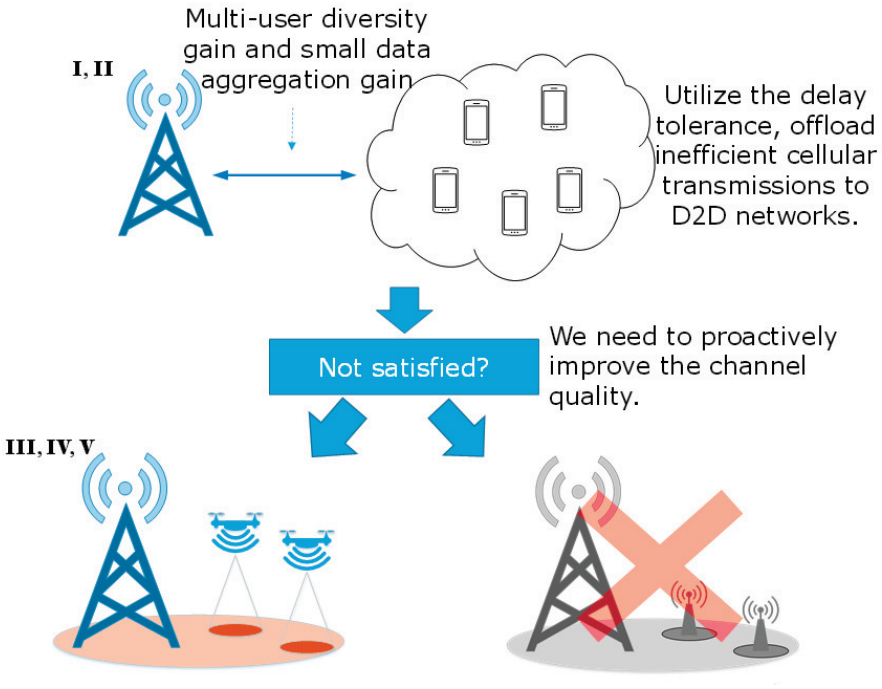


Figure 1.1: Arrangement of chapters.

in Chapter 2 and Chapter 3, respectively. The NC and D2D approach should be used when some of the MTC devices have chances to experience better channels because of the fast fading and uneven shadowing. When the whole group of MTC devices in an area experience high shadowing and penetration loss, the cellular transmissions have to increase either the resource consumption or the transmitting power to overcome the poor-quality channels. In these cases, the second category of solutions is needed. We have to pro-actively improve the channel condition by reducing the communication distance. Due the increasing difficulty of deploying new small-cells in the urban area, we propose to use the Unmanned Aerial Vehicle (UAV) based Floating Relay (FR). Thanks to the mobility of UAVs, when the FR is closer to MTC devices, it can mitigate the undesirable channels of the MTC devices and improve the spectrum efficiency. We focus on MTC devices and optimized the FR's location in Chapter 4. The usage of FRs is then generalized as the UAV-assisted base station serving both

MTC devices and Mobile Internet (MI) applications in Chapter 5. In Chapter 6, we further investigate UAV mesh network which not only provides the seamless coverage to low-priority MTC and MI traffic but also quickly responds to Ultra-Reliable Low Latency Communications.

1.2 Research Objectives and Contributions

1.2.1 Cooperative Device-to-Device Communication with Network Coding for Machine Type Communication Devices

In Chapter 2, we propose a downlink transmission approach leveraging cooperative D2D communications and network coding, which can largely reduce the cellular resource consumption and the total energy consumption. In the proposed approach, the base station generates and broadcasts linear combinations based on the packets requested by different user equipments (UEs) until at least one mature UE can recover all the original packets. Then, a selected mature UE broadcasts new linear combinations based on the recovered original packets to neighbors via D2D until all UEs can decode their packets. The contributions are as follows:

1. The detailed procedures, including the transmissions from both the base station and the mature UE, the selection of the mature UE, UEs' receiving from both the cellular and D2D, and the feedback, have been designed.

2. A feasible and backward-compatible system design including the necessary revisions on the protocol stack based on the current cellular system architecture has been provided.

3. The theoretical performance analyses, including the distribution of the transmission times in both the cellular and D2D networks, the MTC devices' power consumption, and the feedback load, have been derived.

4. Simulations have been conducted to validate the theoretical results and show the performance gains.

1.2.2 Cooperative Device-to-Device Communication for Uplink Transmission in Cellular System

In Chapter 3, a semi-centralized cooperative control method is proposed for the cellular uplink transmissions, where the UE relays are randomly selected according to a certain density decided by the base station. Two specific cooperative approaches based on D2D communications are proposed, which are the random UE relay approach and the one further applying network coding. The contributions are as follows:

1. To study D2D interference, we apply the stochastic geometry to derive the distribution of the D2D transmission distance given the Poisson Point Process (PPP), and then determine the D2D transmitting power and interference based on the distance.

2. Two distinct traffic models, i.e., the MTC traffic with small-data feature and the full-buffer traffic, have been applied to evaluate the system. The proposed approach has been modeled and analyzed based on them, respectively.

3. The theoretical performance analyses and simulations have been conducted to identify the performance gain. Some important guidelines for the base station, such as the optimal density of the UE relays and the applicabilities of two cooperative approaches in different scenarios, have been provided.

1.2.3 Power Allocation and Flight Path Planning for Floating Relay Supporting MTC Traffic in Cellular Systems

In Chapter 4, we target on addressing deep shadowing channels of MTC devices and introduce the FR into the cellular system to improve the transmission efficiency and maximize the system capacity. The contributions are as follows:

1. Considering the capacity limit of the FR's back-haul link and the maximum transmission power of each user, an optimization problem has been formulated to maximize the system capacity.

2. The optimization problem can be decoupled into two parts. We first obtain the optimal power allocation strategy given a fixed location of the FR. A theoretical optimal solution has been derived which minimizes the computation load and thus facilitate the on-line algorithm.

3. Next, two on-line FR placement algorithms have been designed for the unpredictable and predictable networks, respectively, where we calculate the direction of the FR's next movement based on the optimal capacity of the FR's current location.

4. The simulations focusing on different shadowing cases and the comparisons with other off-line approaches have been provided.

1.2.4 UAV-assisted Dynamic Coverage in Heterogeneous Cellular System

The growing popularity of mobile Internet and massive MTC with special traffic characteristics and locations have imposed huge challenges to current cellular networks. Deploying new base stations, however, becomes difficult and expensive, especially for

complicated urban scenarios and MTC traffic. The UAV-assisted heterogeneous cellular solution is proposed in Chapter 5. It utilizes UAV-based FR to deploy FR-cells inside the macro cell and thus achieves dynamic and adaptive coverage. Comprehensive analyses on FR-cells deployment have been provided and the contributions are as follows:

1. Given that the FR-cells reuse the uplink frequency bands used by macro cells, the mutual interference between the FR-cells and the macro cells, and that between the FR-cells, have been analyzed.

2. Given different traffic types generated in a FR-cell, four bandwidth allocation methods for the FR-backhauls have been proposed and analyzed.

3. Two FR-cell's coverage extension methods have been proposed to accommodate the changing topology and traffic distribution.

1.2.5 Placement of Supplementary Node in UAV Mesh Networks

For UAV mesh network, mobility and topology play important roles. How to dispatch a UAV to the optimal location to enhance the coverage and service of the existing network is a critical issue. Given the topology of existing service nodes, a new supplementary UAV can be sent to improve the quality of service especially for the users with poor-quality channels. In Chapter 6, the location of a newly added UAV is optimized to improve the service quality to the worst point. The contributions are as follows:

1. Considering the two-stage UAV mesh network, a min-max optimization problem has been formulated to minimize the longest service distance in an arbitrary acute triangle.

2. For the newly added service node, the closed-form expression of the optimal location within an isosceles acute triangle has been derived.

3. A real-time algorithm with low complexity has been proposed to find the optimal location of the new service node within an arbitrary acute triangle.

4. Comparisons with the exhaustive search and triangle's existing centers are conducted.

1.3 Abbreviations

Abbreviations used in the thesis are summarized in the following table:

Table 1.1: Abbreviations

Abbreviation	**Full Name**
3GPP	3rd Generation Partnership Project
GSM	Global System for Mobile communications
UMTS	Universal Mobile Telecommunications System
LTE	Long-Term Evolution
M2M	Machine to Machine
IoT	Internet of Thing
H2H	Human-to-Human
MI	Mobile Internet
MTC	Machine-Type-Communication
D2D	3rd Generation Partnership Project
NC	Network Coding
PUCCH	Physical Uplink Control Channel

PDCCH	Physical Downlink Control channel
PDSCH	Physical Downlink Shared channel
HARQ	Hybrid Automatic Repeat reQuest
CSI	Channel State Information
UE	User Equipment
eNB	evolved Node B
BLER	Block Error Rate
MCS	Modulation and Coding Scheme
AMC	Adaptive Modulation and Coding
DCI	Downlink Control Information
RAN	Radio Access Network
TTI	Transmission Time Interval
PRB	Physical Resource Block
RRC	Radio Resource Control
SFN	System Frame Number
RLC	Radio Link Control
PDU	Protocol Data Unit
RB	Radio Bearer
MAC	Media Access Control
PDCP	Packet Data Convergence Protocol
SDU	Service Data Units
QoS	Quality of Service
UAV	Unmanned Aerial Vehicle
FR	Floating Relay
PPP	Poisson Point Process

PDF	Probability Distribution Function
SNR	Signal-to-Noise Ratio
PMF	Probability Mass Function
NAS	Non-Access Stratum
PF	Proportional Fairness
SINR	Signal-to-Interference-plus-Noise Ratio
WSN	Wireless Sensor Network
FANET	Flying Ad-hoc NETwork
MANET	Mobile Ad-hoc NETwork
LOS	Line Of Sight
NLOS	Non Line Of Sight
KKT	Karush-Kuhn-Tucker
WCA	Weighted Coordinate Axes
A-WCA	Adaptive Weighted Coordinate Axes
DP	Dynamic Programming
UABS	UAV-Assisted Base Station
PRACH	Physical Random Access Channel
CP	Control Plane
SPS	Semi-Persistent Scheduling
URLLC	Ultra-Reliable Low Latency Communications
SR	Scheduling Request
BSR	Buffer Status Report
TDM	Time Division Multiplex

Chapter 2

Cooperative Device-to-Device Communication with Network Coding for Machine Type Communication Devices

2.1 Introduction

The Machine Type Communication is considered as a key technology enhancement for 5G systems [1]. In the future, the MTC devices will have the features of low power, a large volume of short-messages [2], delay tolerance, low mobility [3], etc. The requirements for low cost and coverage enhancements for MTC devices are also specified in [4]. In order to meet the link budget requirements for coverage enhancements, repetition by 50-100 times is needed for the Physical Uplink Control Channel (PUCCH) [4], which carries the feedbacks of Hybrid Automatic Repeat reQuest (HARQ-ACK) and the Channel State Information (CSI) for the downlink data

transmission. This inefficient uplink transmission will significantly increase a user equipment's (UE) power consumption and conflict with the requirement of low power MTC devices [2, 5]. The uplink spectrum efficiency will be also decreased [4].

Furthermore, the HARQ feedback for each individual repetition may be unnecessary and wasteful due to a very high Block Error Rate (BLER) and inefficient uplink feedback transmission. The fast feedback and retransmission mechanism in current HARQ cannot be fully supported. On the other hand, because of the long feedback delay, the CSI report may become expired, so it is difficult to apply the Adaptive Modulation and Coding (AMC). Also, due to the limited number of Downlink Control Information (DCI) that can be carried in one Physical Downlink Control channel (PDCCH) [6], the cell frequency band cannot be scheduled too fragmentarily, because it may lead to resource waste. Given that a MTC data packet may be too small to fully utilize the minimum resource block assigned, higher order modulation and coding rates are unnecessary.

Based on the practical constraints and difficulties above, in [7, 8], it was proposed to remove or disable the PUCCH for MTC UEs. Without PUCCH, the AMC and HARQ functions are disabled. This will bring some other benefits. For example, the intellectual property protections on these mature techniques can be bypassed, which facilitates the development of the MTC device market. From a technical perspective, the control messages for the AMC and the HARQ can be reduced. According to [9], more than 10 information bits can be saved in the DCI. On the other hand, the PDCCH, which carries the DCI, requires a much higher reliability than the data channel. According to the link budget analyses for MTC coverage enhancements, 100-200 repetitions are needed at a typical PDCCH configuration [4]. Thus, more than 10 bits reduction in the DCI is significant.

For data transmissions in the downlink, the Physical Downlink Shared Channel

(PDSCH) which carries user's data needs 100-200 times repetitions to meet the link budget of coverage enhancement requirement [4]. Given that the AMC and the HARQ cannot be fully utilized for MTC UEs and dynamical scheduling is also disabled due to the lack of CSI reporting, spectrum efficiency will be substantially degraded. With the rapid developments of the IoT, it is anticipated that MTC devices connected to the cellular network will rise tens to hundreds fold [10], the wireless spectrum will soon be in deficit for cellular systems. How to improve the transmission efficiency, especially to reduce the cellular resources consumption for the MTC UEs while reducing the UE power consumption is a pressing issue.

In the following, first, based on the MTC characteristics described above, we propose an efficient approach combined with Network Coding (NC) and Device-to-Device (D2D) communication for the multiple unicast scenario, which can substantially reduce the cellular resource consumption while the total UE energy consumption can also be reduced. Second, a feasible system design including the protocol stack is given, which is backward-compatible with the current LTE/LTE-A system and easy to implement. The evolved Node B (eNB) can fully control the transmission, and the security in Radio Access Network (RAN) will not be affected. Third, the closed-form probability mass functions of transmission times in both the cellular and D2D phases are derived. The design and analysis can be extended to multicast scenarios. The error rates in cellular and D2D transmissions and the feedback load are analyzed. Finally, numerical and simulation results corresponding to different channel settings are given, which can be used as references for the eNB to optimize the configuration of modulation and coding.

2.2 Related Work

Device-to-Device (D2D) communication has been accepted by the 3rd Generation Partnership Project (3GPP) as the feature of ProSe [11] in the latest release. In ProSe, one selected UE can broadcast messages to other UEs via D2D connections. The spectrum for D2D is allocated by the eNB to control the interference between the cellular and the D2D communications. Though D2D was initially proposed to disseminate critical messages to UEs when part of the network infrastructures are damaged in disasters, the same mechanism can also be used to address other scenarios such as local data exchange, UE relay and converged heterogeneous network [12, 13].

D2D is promising to support MTC. As MTC traffic is typically delay tolerant [3, 14], it can be offloaded to the D2D network to save cellular resources. For instance, [15] proposed to let a D2D transmitter act as a relay and apply superposition coding to piggyback the relay messages for downlink transmissions with its own D2D messages. In [16], D2D is used to balance the traffic load among different tier cells in LTE-A network. Gaming theory is applied in [17] to model the relay-assisted cooperative multicast. In this chapter, we consider more general and challenging downlink unicast/multicast scenarios and investigate how to use cooperative D2D to improve downlink transmission efficiency.

Considering the MTC UE's power consumption and the cost in [18, 19], it was found that D2D can reduce the UE power consumption. A self-sustainable D2D communication system powered by integrating ambient backscattering was studied in [20]. According to [4], the cost of the LTE modem is mainly affected by the maximum transceiver bandwidth, supported peak rate, antenna number and transmitting power. D2D transmission will use the same basic approach as the uplink transmission in LTE [21], and no significant additional cost is introduced.

On the other hand, applying Network Coding (NC) in higher layers to improve

efficiency has been heavily investigated [22–25]. The combination of NC and D2D was studied in [26] for local data exchange using D2D only. Cellular transmissions were considered in [27] where UEs need to share the information of missing packets in the D2D transmission phase, which results in additional overheads. In [28, 29], the Random Linear Network Coding (RLNC) was applied in the cellular broadcast channel. The RLNC code has a low complexity and good performance but needs the extra indicator overhead in the packet when we apply it to the LTE/LTE-A system due to the random linear coefficients. Although most of the research on combining NC and D2D focused on multicast scenarios, the multiple unicast scenario was studied in [30]. However, packet missing information should also be reported in the D2D transmissions, and it assumed error-free D2D, so the performance in realistic environments needs further investigation which motivates this work. In [31], the basic NC principles and grouping method for uplink transmissions were introduced. The combination of NC and D2D was applied for uplink in [32,33]. Different from [31–33], we focus on the downlink unicast scenario. NC is applied for the UE-relay node in [34] to combine the data come from the eNB, the edge UE and the relay UE, respectively, which is different from our scenario.

Comparing with the existing literature, the proposed solution has the following salient features. First, the imperfect channel condition leading to packet transmission errors has been considered in both the cellular and the D2D transmissions. Second, we have adopted the predefined linear combination coefficient to reduce the control overhead. Also, missing packets information is not required in the D2D transmission phase. Last but not the least, the features of MTC have been fully considered in the design.

2.3 System Model and Design

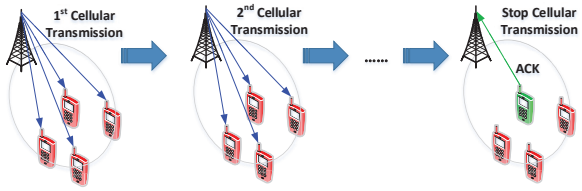
2.3.1 Preliminaries

The main idea of NC is that, instead of directly forwarding the messages, an intermediate node can code the information received before forwarding. With linear network coding [35], the intermediate nodes can generate new packets to transmit, using linear combinations of the original packets, where the coefficients selected and operations are in a Galois field. Using the network coded packets (linear combinations) and solving a system of linear equations, the receiver can recover the original packets once the number of linear combinations received constitute a full-rank matrix. For example, two linear combinations, $s_1 = \gamma_1^T[p_1\ p_2]^T, s_2 = \gamma_2^T[p_1\ p_2]^T$, are received, where p_1 and p_2 are two original packets, γ_1 and γ_2 are two linearly independent coefficient vectors. By applying Gauss-Jordan elimination, the two original packets can be recovered by $[p_1\ p_2]^T = ([\gamma_1\ \gamma_2]^T)^{-1}[s_1\ s_2]^T$.

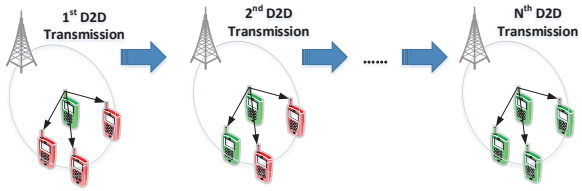
2.3.2 System Model and Transmission Procedure Design

To ensure MTC coverage, blind retransmissions, with or without different redundancy versions, have been proposed as Transmission Time Interval (TTI) bundling in 3GPP [4]. The HARQ soft combining can be applied for repetitions in one TTI bundled transmission. After one TTI bundled transmission, an ACK/NACK is reported. Considering the coverage enhancement requirement of MTC, the TTI bundled transmission may stick to a very low order modulation and channel coding rate [4]. The MTC UEs will perform channel estimation and report the CSI infrequently. The Modulation and Coding Scheme (MCS) and the Physical Resource Block (PRB) can also be assigned semi-statically. In this chapter, we consider the TTI bundling and semi-statical MCS as the baseline, and try to improve the downlink transmission

efficiency under this assumption.



(a) Cellular Phase



(b) D2D Phase

Figure 2.1: Transmission procedure.

In downlink transmissions, the eNB needs to deliver packets to multiple UEs. Several UEs are grouped such that the connectivity within the group is ensured, which is possible given the static locations of MTC UEs. We propose the transmissions in two phases, the cellular phase, and the D2D phase, as shown in Fig. 2.1 (a) and (b), respectively. In the cellular phase, the eNB generates linear combinations based on the packets needed by different UEs in a receiving group and broadcasts linear combinations to this group. After each transmission from the eNB, the UEs check the number of successfully received combinations and try to recover the original packets. If one UE has collected enough number of combinations (considered as a mature UE in this chapter), it sends an ACK to the eNB on a common feedback channel of this group. As long as the eNB is transmitting, each UE always performs the above receiving and feedback procedure regardless of other group members' behaviors. The eNB does not need to know which UE is mature, and thus the exact same content and MCS are applied on ACK, and receiving ACKs from several UEs simultaneously can be viewed as transmission diversity rather than collision. The eNB can roughly control the number of the mature UEs before stopping transmission based on the

number of received ACKs (in different slots) and the receiving power level of each ACK.

The coefficient vector of the linear combination of each transmission should be coordinated between UEs and the eNB. We propose to define several coefficient tables according to various dimensions in specifications, where the coefficient vectors are linearly independent to each other. The eNB notifies UEs the dimension of upcoming linear combinations via the Radio Resource Control (RRC) messages. For the MTC devices with fixed locations and traffic, the dimension keeps unchanged for a long duration once notified. A unique index is assigned to each coefficient vector in the table. There are many options in the current system to inform UEs the index without introducing additional overhead, such as implicitly inferring from the System Frame Number (SFN), or using the index in a deterministic order. The predefined coefficient tables may slightly increase the workload of standardization but reduce the communication overhead.

In the D2D phase, when the cellular transmission is terminated, one mature UE will be selected to broadcast new linear combinations to other members in this receiving group via the D2D network. The eNB will assign different waiting timers for different UEs semi-statically via the RRC procedures. When a UE is mature and the eNB has stopped transmission (no downlink transmission detected in semi-statically allocated resources), the timer will be started. The UE can broadcast in the D2D once the timer is expired. If any other UE starts the D2D transmission before its timer expires, the tagged UE shall abandon the D2D transmission to avoid collisions. The eNB may adjust the waiting timer according to the interference level. In most cases, the MTC devices have no mobility, so it is possible for the eNB to know the precise location of each device. Combining with the sounding reference signal from each device, for UEs closer to other group members and having lower uplink receiv-

ing powers (lower interference levels), a smaller value of the backoff timer will be assigned. The timers can also be adjusted by other concerns, such as the average energy consumption balance, which are beyond the scope.

We also introduce a maximum number of the D2D transmissions to prevent the deadlock in the D2D network. When the number of transmissions exceeds the maximum value and there are immature devices inside the group, the eNB will be aware of it because the Radio Link Control (RLC) entities of these immature devices in the eNB side cannot receive the status reports, and then the transmissions for these immature devices will fall back to the cellular mode.

The selected mature UE should avoid the coefficient vectors that were used by the eNB. The coordination of the coefficient vectors for one group in D2D phase can also be solved by time-based implicitly referring. The selected mature UE should ensure all the other UEs in this group can decode the linear combinations and receive the desired original packets successfully.

The above procedure can be further illustrated in the following example. M UEs constitute a receiving group, and each UE has to receive N different linear combinations to decode the original packets. M and N are two tunable parameters, and the value of N depends on M, the number of UEs in a group, and their traffic loads, so our work is general and applicable for many scenarios including homogeneous and heterogeneous UE traffic. For example, there are two UEs in a group ($M = 2$), one requests one packet and the other one requests two for each round. The eNB linearly combines the three packets by NC, each UE has to receive three different linear combinations to recover the packets they expect ($N = 3$). Also, our solution is applicable for both the unicast and multicast scenarios, as well as a mixture of them. For example, if two UEs request the same packets and the third UE requests another one per round ($M = 3$), linear combinations will be generated using the two original

packets, and each UE can recover all of them if two independent linear combinations are successfully received ($N = 2$).

We focus on the unicast MTC scenario which is a major motivation of the proposed solution. For the unicast MTC scenario where each UE requires only one packet from the eNB, linear combinations are generated based on M packets. If applying some traditional random linear coding such as the one in [36], N different linear combinations are needed to decode the M packets, where N is slightly larger than M. In this chapter, as explained in Sec. 2.3.2, we apply fixed coefficient tables, where the coefficient vectors are linearly independent to each other. Therefore, M received linear combinations are enough to build a full-rank invertible coefficient matrix, so the number of linear combinations needed equals M to recover M packets, i.e., $N = M$.

2.3.3 Protocol Design

In order to implement the proposed approach in the LTE/LTE-A system, a new layer is added in the eNB and UE sides to deal with the linear combination decoding and generating.

We introduce an NC layer to the current LTE protocol stack, as shown in Fig. 2.2. In the eNB side, fixed size RLC protocol data units (PDU) of different UEs are collected to generate linear combinations. The NC data stream is radio bearer (RB) specific, so after the multiplexing in the Media Access Control (MAC) PDU, the RLC PDUs from other RBs can coexist with the linear combination in one MAC PDU. Therefore, when the proposed approach is adopted, various types of data such as multicast and broadcast traffic can be sent to this group of UEs at the same time. When a UE receives a MAC PDU that contains one or more MAC service data units (SDU) that belong to the RBs with NC feature, the MAC SDUs are forwarded to the NC entities, where received combinations are collected in a receiving pool. Once there

are enough combinations, the original RLC PDUs can be recovered, and the desired RLC PDU is forwarded to the RLC layer and others will be discarded. Only the UE that is selected by the eNB can utilize all the recovered RLC PDUs to generate new linear combinations and broadcast them via D2D, which corresponds to the dashed line in Fig. 2.2.

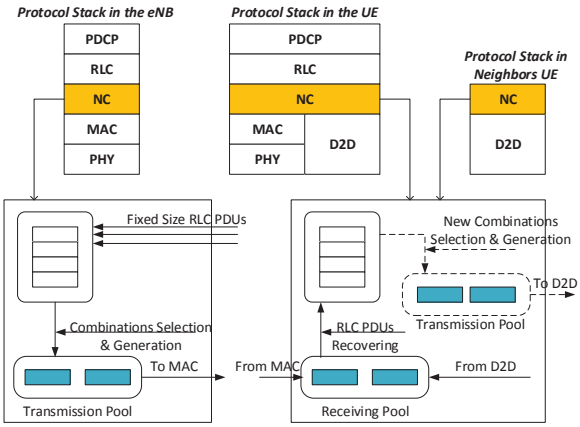


Figure 2.2: Design of the air interface protocol stack.

The linear combination generation function at the eNB side and the mature UE side is based on the predefined coefficient table, as explained in Sec. 2.3.2. After the eNB stops the cellular transmissions, the mature UE can first recover all the original packets by itself because a sufficient number of linear combinations have been collected in the cellular phase. Once the group of UEs is formed, each UE will be notified the dimension of the coefficient vectors and the coefficient table, so it knows the number of linear combinations needed to recover the original packets. After one packet is received, the checksum will be calculated in the channel decoding process to detect whether or not the packet is received error-free. Therefore, each UE can know how many packets (i.e., linear combinations) have been received successfully.

The recovering function at the mature UE side is done by solving a system of linear equations, which is similar to the example in Sec. 2.3.1. Then, the mature UE will generate new linear combinations (different from the eNB's as explained in

Sec. 2.3.2) based on the recovered original packets and send them to other group members via D2D. The immature UEs will recover the original packets based on the linear combinations received from both the eNB and the mature UE. The linear coefficients used by both the eNB and the mature UE are known so that a system of linear equations can be formed and then solved after a sufficient number of linear combinations have been received.

The complexities of the linear combination generation and recovering functions are $\mathcal{O}(N^2L)$ and $\mathcal{O}(N^3 + N^2L)$, respectively, where L is the length of the packet. For the recovering function, the term N^3 is the computational complexity to perform Gauss-Jordan elimination, and N^2L is the computational complexity of multiplying the inverted coefficient matrix with the received linear combinations [37]. The eNB has a higher computation capacity than MTC devices so that the bottle-neck is the recovering function of MTC devices. The Gauss-Jordan elimination/coefficient matrix inversion is the major part of the computation load. Considering the special features of the MTC traffic, L is small and $N = M$. In practice, due to the limited D2D communication distance, the group size, M, cannot be very large. Therefore, the rank of the coefficient matrix is small and the total computation load can be maintained at a low level. Besides limiting the group size, the prolonged processing time can also help to disperse the computation load. The packet recovery is not performed in each sub-frame. When a UE receives a sufficient number of linear combinations, it will send an ACK to the eNB. The ACK feedback is only determined by the number of linear combinations successfully received and independent to the packet recovering procedure, so the ACK can be sent in time. After that, the mature UE shall recover all the packets, regenerate new linear combinations, and start the D2D phase. However, this procedure has no strict time limitation. As described in Sec. 2.3.2, the eNB can set sufficiently large backoff timers for UEs to deal with the packet recovery.

To recover the original packets, a certain number of linear combinations need to be buffered in the UE side, which may increase the buffer size requirement for MTC UEs. For "Category 0" MTC UE in the latest LTE specification [38], 20000 bytes layer-2 buffer size is defined, which is large enough to store a sufficient number of linear combinations if the packet size is small. Therefore, we can fully utilize the existing layer 2 buffer for the MTC UE to support NC.

We cannot insert this NC layer to higher layers because the eNB should be aware of this proposed approach and control the transmissions based on the traffic load in the cellular network, availability of the D2D network, channel quality and Quality of Service (QoS) requirements. In addition, in the LTE/LTE-A system, the ciphering/deciphering is performed in the PDCP layer. In the proposed solution, although UEs will recover the original packets (RLC PDUs) of others, they cannot decipher the PDCP SDUs in the PDCP layer because they do not have others' keys. Also, if any malicious node transmits a corrupted version of linear combinations with the predefined coefficient vectors, the affected packets cannot go through the PDCP layer. Thus, our design can ensure that the security function in RAN is preserved.

Given the facts that network coding is a mature technology used in real systems [39], and other techniques used in the proposed solution are either compatible with the on-going industry standards, such as D2D in 3GPP, or belong to the existing techniques in the cellular systems. The proposed solution does not need significant changes in both the specifications and the hardware. In addition, given the new 5G air-interface project that launched by 3GPP [40], the architecture of the future cellular system will be more flexible to accept new technologies. Considering the urgency of the development of the MTC, the proposed solution is promising and beneficial.

2.4 Theoretical Analysis

Based on the above procedures, we analyze the performance of both transmission phases. We consider the total number of transmissions in the cellular and the D2D phases, respectively, as the performance metric. We also use the analytical results to calculate the total transmission energy consumption and compare it with the legacy system. We study the performance considering two cases in Sec. 2.4.1 and 2.4.2, respectively. The first one aims to minimize the cellular resources consumption, so the cellular phase transmissions to the group will be ended whenever there is one mature MTC device. In the second case, the cellular phase will be ended till there are at least K mature MTC devices ($K > 1$), which can offer the flexibility for the eNB to make the trade-off between cellular resources consumption and other concerns.

Symbols used in this section are summarized in follows,

Table 2.1: Symbol Notation List

Symbol	**Definition**
X_i	The minimum transmission times for the ith UE to become mature
P_{X_i}	The PMF of X_i
E_i	the average BLER of the ith UE
N	The number of linear combinations for a UE to become mature
M	The number of UEs
X	The cellular transmission times (Minimum cellular consumption)
P_X	The PMF of X
K	The required number of mature UEs
X_K	The cellular transmission times (Multiple mature UEs)
P_{X_K}	The PMF of X_K

X_{D_i}	The D2D transmission times for the ith UE to become mature
$P_{X_{D_i}}$	The PMF of X_{D_i}
X_D	The D2D transmission times to let all the UEs become mature
P_{X_D}	The PMF of X_D
$X_{legacy}^{(fb)}$	The uplink transmission times for feedback in the legacy system
$X^{(fb)}$	The uplink transmission times for feedback in the proposed approach
$X_D^{(fb)}$	The transmission times for feedback in D2D

2.4.1 Cellular Phase, Minimum Cellular Consumption

In this section, we assume that the eNB will stop transmission once there is at least one mature UE in the receiving group and thus the target of minimum cellular resource consumption can be achieved. This is the simplest case to start with. For a UE in the receiving group, when it receives N combinations it becomes mature. Because the TTI for MTC devices may be prolonged, as explained in Sec. 2.3.2, it is reasonable to assume that the MTC TTI (or bundled TTI) is longer than the channel coherence time, and thus the packet (linear combination) transmissions are independent to each other. If a UE becomes mature at the nth eNB transmission, the nth transmission must be successfully received. The total number of transmissions follows the negative binomial distribution. We denote the random variable X_i as the required minimum transmission times for the ith UE to become mature, $i = 1, 2, 3, ..., M$. Therefore, the Probability Mass Function (PMF) of X_i, which is denoted by $P_{X_i}(n)$, is given by

$$P_{X_i}(n) = \binom{n-1}{N-1}(1-E_i)^N E_i^{n-N}, \quad \text{for } n \in [N, \infty), \tag{2.1}$$

where E_i is the average BLER of the ith UE. It means that the transmissions may be failed with a probability of E_i. In the theoretical analyses and numerical evaluations, we use the average BLER of both the cellular and the D2D transmissions to reflect the imperfect channels.

We assume that each UE's channel is a stationary random process so that the Probability Distribution Function (PDF) of the ith UE's downlink receiving Signal-to-Noise Ratio (SNR) can be denoted by $f_{SNR_i}(r)$. As discussed in Sec. 2.3.2, the MCS will be configured semi-statically for the MTC devices. Given a certain MCS, the instantaneous BLER is a function of the instantaneous SNR, which can be denoted by $G_{MCS}(r)$. Therefore, E_i is given by

$$E_i = \int_0^\infty f_{SNR_i}(r) G_{MCS}(r) dr. \tag{2.2}$$

For MTC devices with fixed locations, we assume that the distribution of the SNR is not time varying and thus E_i is constant. One advantage of applying linear combinations is that only the number of received independent combinations affects X_i, which is determined by packet error rate or average BLER only. Consequently, we can simplify the channel model using a single parameter, average BLER, in the analysis.

For this case, if there is at least one UE can receive N combinations (become a mature UE), the eNB will stop transmission and offload the transmission to the D2D network. Random variable X is denoted as the required minimum transmission times in the cellular transmission phase, i.e., $X = \min(X_1, X_2, \ldots, X_M)$.

If the eNB stops transmission once there is at least one mature UE in the receiving

group, the PMF of X is given by

$$P_X(n) = \begin{cases} 1 - \prod_{i=1}^{M} S_i(n) - \sum_{v=N}^{n-1} P_X(v), & n \in [N, \infty), \\ 0, & n \in [0, N), \end{cases} \tag{2.3}$$

where $S_i(n) = 1 - \sum_{u=N}^{n} \binom{u-1}{N-1} (1 - E_i)^N E_i^{u-N}$.

If all the UEs' channels are independently and identically distributed (i.i.d), considered as the homogeneous BLER case, they will have the same average BLER, denoted by E, so that $P_X(n)$ can be simplified as in (2.4).

$$P_X(n) = \begin{cases} 1 - S(n)^M - \sum_{v=N}^{n-1} P_X(v), & n \in [N, \infty), \\ 0, & n \in [0, N), \end{cases} \tag{2.4}$$

where $S(n) = 1 - \sum_{u=N}^{n} \binom{u-1}{N-1} (1 - E)^N E^{u-N}$.

Proof. The PMF of X is derived in (2.5).

$$\begin{aligned} P_X(n) = 1 &- Pr\{\text{eNB stop after the } n\text{th transmission}\} \\ &- Pr\{\text{eNB stop before the } n\text{th transmission}\} \end{aligned} \tag{2.5}$$

From (2.5), we have

$$P_X(n) = \begin{cases} 1 - \prod_{i=1}^{M} S_i(n) - \sum_{v=N}^{n-1} P_X(v), & n \in [N, \infty), \\ 0, & n \in [0, N), \end{cases} \tag{2.6}$$

where $S_i(n) = 1 - \sum_{u=N}^{n} P_{X_i}(u) = 1 - \sum_{u=N}^{n} \binom{u-1}{N-1} (1 - E_i)^N E_i^{u-N}$. □

The assumption of homogeneous BLER can simplify the notations, and the above analytical framework is ready to be extended to remove this assumption. Furthermore, as shown in the numerical evaluation, comparing with the homogeneous BLER, the proposed approach can achieve a higher performance gain in the case of heterogeneous BLER. For these reasons, we skipped the derivations of the heterogeneous case. Based on the PMF of random variable X, the average or expectation of the total number of transmissions from the eNB is obtained below, which is an important performance metrics.

$$\mathbf{E}[X] = \sum_{n=N}^{\infty} nP_X(n). \tag{2.7}$$

2.4.2 Cellular Phase, Multiple Mature UEs

In this case, the eNB will not stop until there are at least K mature UEs in the receiving group. In practice, K is chosen based on many factors, such as the traffic, group size, QoS, interference level and so on, so that it is worthy to offer the flexibility in selecting K to make the trade-off between the cost in the D2D phase and that in the cellular phase. We assume that the cooperative D2D transmission will start only if the eNB has stopped transmitting to this group. We also assume the homogeneous BLER here to simplify the notations. We use X_K to denote the minimum transmission times for the eNB to let at least K UE to become mature.

If the eNB stops transmission once there is at least K mature UE in the receiving

group, the PMF of X_K is given by

$$P_{X_K}(n) = \begin{cases} 1 - \sum_{i=0}^{K-1} \binom{M}{i} \left[1 - \sum_{l=0}^{N-1} \binom{n}{l} (1-E)^l \ E^{n-l} \right]^i \\ \left[\sum_{l=0}^{N-1} \binom{n}{l} (1-E)^l \ E^{n-l} \right]^{M-i} - \sum_{j=N}^{n-1} P_{X_K}(j), \quad n \in [N, \infty), \\ 0, \qquad\qquad n \in [0, N). \end{cases} \tag{2.8}$$

Proof. Define event A_i as that i UEs are mature in the past n transmissions from the eNB, which means that these i UEs could become mature at any time before or at the nth transmission. We have

$$Pr\{A_i\} = \binom{M}{i} Pr\{B\}^i \left[1 - Pr\{B\}\right]^{M-i}, \tag{2.9}$$

where event B is defined as that one UE became mature in the past n transmissions from the eNB, and its probability is given by

$$Pr\{B\} = 1 - \sum_{l=0}^{N-1} \binom{n}{l} (1-E)^l \ E^{n-l}. \tag{2.10}$$

For the PMF of X_K, we have

$$P_{X_K}(n) = \begin{cases} 1 - \sum_{i=0}^{K-1} Pr\{A_i\} - \sum_{j=N}^{n-1} P_{X_K}(j), \quad n \in [N, \infty), \\ 0, \qquad\qquad\qquad\qquad\qquad\qquad n \in [0, N). \end{cases} \tag{2.11}$$

Plugging (2.9), (2.10) into (2.11), we can obtain (2.8). □

2.4.3 D2D Phase

In the D2D phase, one selected mature UE broadcasts new linear combinations via D2D, so the situation is similar to that in the cellular system, except that each UE will have a non-uniform initial status, i.e., each UE may have already successfully received several linear combinations before the D2D phase. It is possible to schedule concurrent D2D transmissions in a group for further enhancement, by which the transmission diversity gain is achieved.

After the cellular transmissions with at least one mature UE, the required minimum D2D transmission times for the ith UE to become mature is denoted by X_{D_i}, $i = 1, 2, 3, ..., M-1$. The PMF of X_{D_i} is given by

$$P_{X_{D_i}}(n) = \sum_{Q_i=1}^{\Omega} \left[\sum_{y=N}^{\infty} \binom{y}{N-Q_i} (1-E_i)^{N-Q_i} E_i^{y-N+Q_i} P_X(y) \right] \binom{n-1}{Q_i-1} (1-E_{D_i})^{Q_i} E_{D_i}^{n-Q_i}, n > 0, \tag{2.12}$$

where $\Omega = \max\{n, N\}$, and

$$P_{X_{D_i}}(0) = \sum_{y=N}^{\infty} \left[1 - \sum_{l=0}^{N-1} \binom{y}{l} (1-E_i)^l E_i^{y-l} \right] P_X(y). \tag{2.13}$$

Proof. As one mature UE has been selected to broadcast new linear combinations, there are upto $M-1$ UEs receiving the D2D transmission. Define event T_y as that the eNB stopped transmission at the yth cellular transmission. We have $Pr\{T_y\} = P_X(y)$, where $P_X(y)$ is the PMF of the minimum transmission times for the eNB to let at least one UE become mature derived in Sec. 2.4.1. Define event G as that the ith UE becomes mature at the nth D2D transmission, and event H_{Q_i} as that the ith UE needs Q_i linear combinations to become mature, i.e., this UE has successfully received

$N - Q_i$ different linear combinations from the eNB. We have

$$\begin{aligned} Pr\{H_{Q_i}\} &= \sum_{y=N}^{\infty} Pr\{H_{Q_i} \mid T_y\} Pr\{T_y\} \\ &= \sum_{y=N}^{\infty} \binom{y}{N-Q_i} (1-E_i)^{N-Q_i} E_i^{y-N+Q_i} P_X(y), \end{aligned} \tag{2.14}$$

$$Pr\{G \mid H_{Q_i}\} = \binom{n-1}{Q_i-1} (1-E_{D_i})^{Q_i} E_{D_i}^{n-Q_i}, \tag{2.15}$$

where E_{D_i} is the BLER of the D2D transmission for the ith UE.

The PMF of X_{D_i} is given by

$$P_{X_{D_i}}(n) = \sum_{Q_i=1}^{\Omega} Pr\{G \mid H_{Q_i}\} Pr\{H_{Q_i}\}, \tag{2.16}$$

where $\Omega = \max\{n, N\}$.

From (2.14), (2.15), (2.16), we can obtain (2.12).

The derivation above does not include the case of $n = 0$, i.e., a UE had already been mature before the D2D transmission. Define event H_0 as that the ith UE had already been mature via the cellular transmissions. Similarly to (2.14), we consider the $P_X(y)$ in the case of minimum cellular consumption to be equal to $Pr\{T_y\}$, and we have

$$Pr\{H_0 \mid T_y\} = 1 - \sum_{l=0}^{N-1} \binom{y}{l} (1-E_i)^l E_i^{y-l}. \tag{2.17}$$

The corresponding probability for $X_{D_i} = 0$ is given by

$$P_{X_{D_i}}(0) = \sum_{y=N}^{\infty} Pr\{H_0 \mid T_y\} Pr\{T_y\}. \tag{2.18}$$

Using (2.17) and (2.18), we can obtain (2.13). □

We use $X_D = \max(X_{D_1}, X_{D_2}, \ldots, X_{D_{M-1}})$ to denote the minimum D2D trans-

mission times to let all the UEs become mature. Homogeneous BLER is assumed in the D2D transmission, which is denoted as E_D. With the homogeneous BLER assumption, $P_{X_{D_i}}(n) = S_D(n), i = 1, 2, ..., M-1$, which is the same for all the UEs.

After the cellular transmissions with at least one mature UE, the PMF of X_D, the required minimum D2D transmission times for all the UEs to become mature, is given by

$$P_{X_D}(n) = \begin{cases} \sum_{i=1}^{M-1} \binom{M-1}{i} S_D(n)^i \left[\sum_{j=0}^{n-1} S_D(j)\right]^{M-1-i}, & n > 0, \\ S_D(0)^{M-1}, & n = 0. \end{cases} \tag{2.19}$$

Proof. It is easy to obtain the following equation,

$$P_{X_D}(n) = \begin{cases} \sum_{i=1}^{M-1} \binom{M-1}{i} Pr\{D_n\}^i Pr\{D_{bn}\}^{M-1-i}, & n > 0, \\ Pr\{D_0\}^{M-1}, & n = 0, \end{cases} \tag{2.20}$$

where D_n, D_{bn}, and D_0 are the event that one UE become mature at the nth D2D transmission, before the nth D2D transmission, and before any D2D transmission, respectively. By utilizing (2.12) and (2.13), we have $Pr\{D_n\} = S_D(n)$, $Pr\{D_{bn}\} = \sum_{j=0}^{n-1} S_D(j)$ and $Pr\{D_0\} = S_D(0)$. Plugging it in (2.20), we can obtain (2.19). □

Based on (2.19), the average D2D transmission times can be calculated by $\mathbf{E}[X_D] = \sum_{n=0}^{\infty} n P_{X_D}(n)$.

2.4.4 Legacy System

In this section, we derive the performance of the legacy system. Because of the lack of timely CSI reporting and thus the dynamical scheduling, the eNB has to send the packets to each UE one by one in the unicast scenario. Also, we do not

consider HARQ between the TTI bundled transmission (or we only adopt TYPE I HARQ which means we abandon packets when decoding is failed). The HARQ soft combining can still be applied for repetitions in one TTI bundled transmission as mentioned in Sec. 2.2. Therefore, the receiving behaviors at different bundled time slots are i.i.d., even for the same packet. The expected number of transmissions for the ith UE to successfully receive a packet is denoted by $\mathbf{E}[X_i]$. As X_i is a geometry random variable, its expectation can be obtained by

$$\mathbf{E}[X_i] = \frac{1}{1 - E_i}, \tag{2.21}$$

where E_i is the average BLER of the ith UE.

Considering there are M UEs in a receiving group, the expectation of the total number of transmissions in the legacy system $\mathbf{E}[X_{legacy}]$ can be given by $\frac{M}{1-E}$ and $\sum_{i=1}^{M} \frac{1}{1-E_i}$, for the homogeneous and heterogeneous BLER cases, respectively.

2.4.5 Legacy System combined with D2D

The approach of the legacy system combined with D2D is the basic UE cooperation method. In [41], the UE cooperation based on LTE/LTE-A system has been discussed. If there is a downlink packet for one of the UEs in a receiving group, the eNB will select the UE having the best cellular channel quality to send the packet. Then the packet will be relayed to the target UE via the D2D transmission. In this way, the multi-user diversity gain can be achieved. The procedures are shown in Fig. 2.3.

Because the eNB always selects the UE with the best channel, the expectation of the total number of transmissions in the cellular phase $\mathbf{E}[X]$ is given by

$$\mathbf{E}[X] = \frac{M}{1 - \min_i(E_i)}. \tag{2.22}$$

It is larger comparing with the result in (2.7), which means that this basic UE cooperation approach will cost more cellular resources than the proposed approach. The reason is that it is possible that one UE may experience fewer transmission times to mature than the UE with the lowest average BLER.

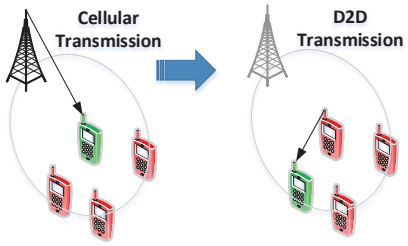


Figure 2.3: Procedures for basic UE cooperation.

Another advantage of the proposed solution is the feedback load. In order to let the eNB select the best UE, all the UEs in a receiving group have to report the channel quality frequently. Furthermore, as a corrupted packet cannot be recovered by the following packets, so that the ACK/NACK feedback is necessary. Based on the analysis of feedback channel PUCCH for MTC UEs in Sec. 2.1, the performance gain of this basic UE cooperation approach will be achieved at the expense of high UE power consumption and severe uplink spectrum efficiency degradation. On the contrary, using the proposed approach, the eNB does not need to know the exact channel quality of each UE in a group. It only needs to roughly know the average channel quality of this group and set a reasonable number of repetitions semi-statically. The ACK/NACK feedback is also substantially reduced. As mentioned in Sec. 2.3, only one ACK is reported in a common feedback channel when there is a mature UE. The detailed analysis of the feedback load is presented in Sec. 2.4.6.

For the D2D phase, in this basic UE cooperation method, the packets are relayed by the UE with the best channel. Therefore, there are $M-1$ packets to be transmitted in D2D. In the proposed approach, there are at most $\max_i(Q_i)$ packets to be

transmitted in D2D, where Q_i is the number of linear combinations needed by the ith UE to be mature in the D2D transmissions. Typically, $M-1 \geq \max_i(Q_i)$, so the proposed approach also has lower D2D cost. The event of $M-1 < \max_i(Q_i)$ only happens if one or more UEs have not successfully received any packets from the eNB at all.

The basic UE cooperation method may be improved by making other UEs also listen to the transmissions. A UE may not request D2D relaying if its own packet is successfully received during the cellular phase. However, it introduces more signaling exchanging and feedback overhead between UEs. In this chapter, we only compare the approach in [41], which has a mature design and is compatible with the LTE system, with the proposed approach.

Besides the legacy system combined with D2D only, the legacy system combined with NC only is another possible solution. The expectation of the total number of transmissions in cellular phase is $\mathbf{E}[X] = M/(1-\max_i(E_i))$, which is obviously even more than the transmission times in the legacy system. The advantage of this NC only approach is that the feedback load can be reduced comparing with the legacy system. The NC only approach is suitable for the uplink channel limited scenarios with delay tolerant traffic and the outmoded UEs which are manufactured according to the previous versions of the specification and do not support D2D transmissions. Because it cannot utilize the D2D feature, we do not consider this approach in this chapter.

2.4.6 Feedback Overhead

As discussed in Sec. 2.2, even though the blind retransmission can be applied, at least a feedback according to one bundled transmission is still necessary. Considering the large number of repetitions for MTC devices, the transmissions of ACK/NACK in

uplink will play an essential role in UE power consumption. In the legacy system, each point-to-point downlink transmission requires a feedback in the uplink, so that the number of feedback messages will be equal to that of the data messages. We use $X_{legacy}^{(fb)}$ to denote the uplink transmission times for feedback in the legacy system, and $\mathbf{E}[X_{legacy}^{(fb)}] = \mathbf{E}[X_{legacy}]$.

In the proposed approach, the feedback transmission times is equal to the number of mature UEs which cannot be controlled by the eNB very precisely. For example, if the eNB stops transmission when there is at least one mature UE, more than one UE may become mature simultaneously. Thus multiple feedback messages may be transmitted in the uplink simultaneously. The probability of that n UEs become mature simultaneously and the eNB stops after the rth transmission, which is denoted as $R(n, r)$, is shown below

$$\begin{aligned} R(n,r) &= \binom{M}{n} Pr\{\text{one UE mature at rth transmission}\}^{n} \\ &\qquad Pr\{\text{one UE mature after rth transmission}\}^{M\text{-}n} \\ &= \binom{M}{n} P_{X_i}(r)^n \left[1 - \sum_{u=N}^{r} P_{X_i}(u)\right]^{M-n}. \end{aligned} \tag{2.23}$$

Therefore, the probability of n UEs become mature simultaneously, $R(n)$, can be obtained by $R(n) = \sum_{r=N}^{\infty} R(n, r)$. The average number of transmissions for feedback is given by $\mathbf{E}[X^{(fb)}] = \sum_{n=1}^{M} nR(n)$.

For the D2D transmission, we apply the similar mechanism as that used in the cellular phase. In order to ensure that all the UEs in the D2D receiving group are mature, we assume that the ACK feedback in D2D is UE-specific. It can be realized by UE-specific feedback slot or CDMA-based solutions. If any ACK is lost, the selected mature UE has to keep transmitting which will result in deadlock. A maximum transmission number of linear combinations in D2D should be configured by the eNB

to avoid the above situation. A new control message could be also introduced for the selected mature UE to trigger all other immature group members to report NACK in a common resource. If there is no NACK detected in that common resource, the selected mature UE will stop broadcasting linear combinations. This mechanism can be applied when the selected mature UE has not received any ACK for a long time. Therefore, if the eNB will stop transmission when there is at least one mature UE, at most $M-1$ UEs will transmit feedback via D2D. We use $X_D^{(fb)}$ to denote the number of feedback transmissions in D2D. Due to the high reliability of the ACK feedback transmission, we do not consider the loss rate of ACK transmissions and assume $\mathbf{E}[X_D^{(fb)}] = M-1$, which is the worst case.

2.5 Performance Evaluation

2.5.1 Performance Metrics

In this section, we evaluate the performance of the proposed solution using numerical results obtained from analysis and use Monte Carlo simulation to validate the analysis. The unicast MTC scenario is applied, i.e., $N = M$. First, we evaluate the performance gain in term of the cellular resources consumption, where the transmission times from the eNB is a suitable metric. The performance gain is defined in (2.24), which measures the percentage of transmission saving in the cellular system by adopting the proposed approach.

$$GAIN_{cellular} = \frac{\mathbf{E}[X_{legacy}] - \mathbf{E}[X]}{\mathbf{E}[X_{legacy}]}. \tag{2.24}$$

The throughput can also be easily calculated by dividing the total data amount of M packets (multiplying packet size by the number of UEs) by the transmission

time (i.e., one LTE subframe time duration multiplied by the needed number of transmissions from the eNB). When we use less transmission times for a given amount of data, it can be converted to a higher throughput.

Based on (2.24), we first evaluate the performance gains for the case of minimum cellular consumption and general case (described in Sec. 2.4.1 and Sec. 2.4.2), respectively.

Since many existing solutions can be utilized to minimize the impact from the D2D interference[1], we do not consider the D2D interference and focus on the resources consumption in the cellular network in this chapter.

For fair comparison of the proposed solution with the legacy system without D2D, we use the total UE transmitting power consumption and total system transmitting power consumption to evaluate the performance gain, which are defined as

$$GAIN_{power}^{(UE)} = 1 - \frac{\rho\tau\beta\mathbf{E}[X^{(fb)}] + \mathbf{E}[X_D] + \beta\mathbf{E}[X_D^{(fb)}]}{\rho\tau\beta\mathbf{E}[X_{legacy}^{(fb)}]}, \tag{2.25}$$

$$GAIN_{power}^{(sys)} = 1 - \frac{\rho\tau(\alpha\mathbf{E}[X] + \beta\mathbf{E}[X^{(fb)}]) + \mathbf{E}[X_D] + \beta\mathbf{E}[X_D^{(fb)}]}{\rho\tau(\alpha\mathbf{E}[X_{legacy}] + \beta\mathbf{E}[X_{legacy}^{(fb)}])}, \tag{2.26}$$

where ρ is the ratio of UE transmitting power in the cellular and D2D networks. Generally speaking, $\rho \in [10, 100]$ [45]. We use $\rho = 20$ in the numerical evaluation. τ is the repetition times for MTC devices in both data transmission and ACK feedback. According to [4], we set $\tau = 100$. β is the ratio of transmitting power in data transmission and ACK feedback. Considering the short-message feature of the MTC

[1]Both the Cognitive Radio technologies using licensed bands and short-range radio technologies using unlicensed bands are possible ways to support the D2D transmissions, by which the interference to cellular users is negligible. Another option for operators is the network assisted D2D [42], where the D2D transmission will reuse the uplink cellular resources and the eNB will control the resource allocation. It can utilize the dynamical scheduling and D2D power control to minimize the negative effects to normal cellular transmissions. In [43, 44], a geometrical-based analysis showed that the system throughput can be further improved by utilizing the underlaying D2D communications with a guard distance. Furthermore, because of the relatively low D2D transmitting power and greater penetration losses on the radio interface by MTC UEs [42], their D2D interference is limited.

devices and the high reliability of the ACK feedback, we assume that the transmitting power consumed by one data packet will be similar to that consumed by one ACK feedback, and thus $\beta = 1$. α is the ratio of the eNB transmitting power and the UE transmitting power. We choose $\alpha = 20$ according to the maximum transmitting power of the eNB and the UE, which are 5W and 250mW [46], respectively. We consider the low transmitting power small eNB (37dBm, 5W) [47] in this chapter. For the regular eNB, the performance gain of the power consumption will be even higher.

(2.25) is the metric of UE transmitting power consumption gain, which is defined as the ratio of energy saving by using the proposed approach. From a UE's perspective, all of the transmitting power consumption in the legacy system is brought by the feedback, which is calculated based on the number of ACK/NACK transmissions and other scale factors, ρ, τ, β. In the proposed approach, the cellular feedback number is reduced but two other additional power consumption items are introduced, which are the D2D transmission and D2D feedback power consumptions. Therefore, there are two more items in the numerator of (2.25).

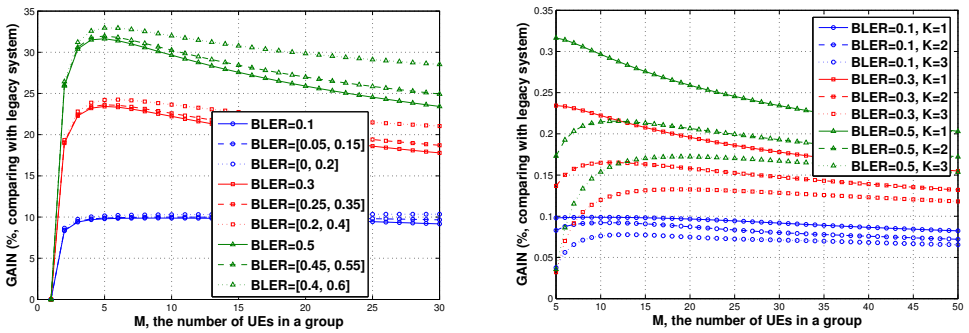
In (2.26), the transmitting power consumption of the whole system is evaluated, so that in addition to all the items listed in (2.25), the eNB transmitting power consumption is introduced for the legacy system and the proposed approach, respectively.

We do not consider the receiving energy consumption. By adopting the NC and D2D, each UE will receive more data packets than those in the legacy system. However, in the legacy LTE system, even when there is no data transmitted, UEs still have to receive and buffer the signal of the whole cell band in one subframe, demodulate and decode the first several OFDM symbols and blindly search whether there is a PDCCH (scheduling grant is included) masked with their Radio Network Temporary Identities (RNTI). Thus the most power-consuming part in the receiving process

cannot be avoided in both the proposed approach and the legacy system. Comparing with only receiving their own packets, additional baseband process is needed to receive multiple linear combinations. However, the power consumption of the receiving baseband is negligible comparing with the total UE power consumption [48].

2.5.2 Cellular Gain, Minimum Cellular Consumption

First, we consider the case in Sec. 2.4.1, where the eNB will stop transmission when there is at least one mature UE in the receiving group.



(a) Minimum cellular consumption (b) Multiple mature UEs

Figure 2.4: Performance gain in term of cellular resources consumption.

From Fig. 2.4 (a), comparing with the legacy system, the proposed approach can save 10% to 33% cellular resource consumptions according to different average BLER settings. The higher the BLER, the higher the gain, which means that our approach can achieve a much better performance in the situation of worse channel conditions. In the current system, such as LTE/LTE-A, the eNB may adjust the MCS to control the receiving BLER to improve the throughput. Given that our approach can save the transmission times significantly, the eNB can relax the target receiving BLER by using a higher order MCS or reducing the repetition times of a TTI bundled transmission for the MTC UEs. It means that more bits can be transmitted in each time-frequency unit, and then improve the overall spectrum efficiency.

The performance gain, as shown in Fig. 2.4 (a), is defined in (2.24), where $\mathbf{E}[X_{legacy}]$ has a linear relationship with M. Thanks to the high growth rate of the multi-user diversity gain w.r.t. M when M is small, the growth rate of the numerator in (2.24) is higher than that of the denominator. Thus, the performance gain increases w.r.t. M till reaching a peak. However, when M is relatively large, the multi-user diversity gain cannot be further enhanced, so the numerator increases slower than the denominator, and thus the performance gain decreases.

The comparisons between the homogeneous BLER and the heterogeneous BLER are also made in Fig. 2.4 (a). Obviously, the larger variance of the BLER, the higher performance gain we can achieve. The group of UEs may be close to each other so their path loss may be similar. However, considering the fast fading and shadowing, especially in the city scenarios, due to the complicated layout of buildings and vehicles, the overall signal attenuation to each UE may be quite different, and thus the proposed solution is very promising.

2.5.3 Cellular Gain, Multiple Mature UEs

Next, we consider the case corresponding to Sec. 2.4.2, where the eNB will stop transmission when there is at least K mature UEs in the receiving group.

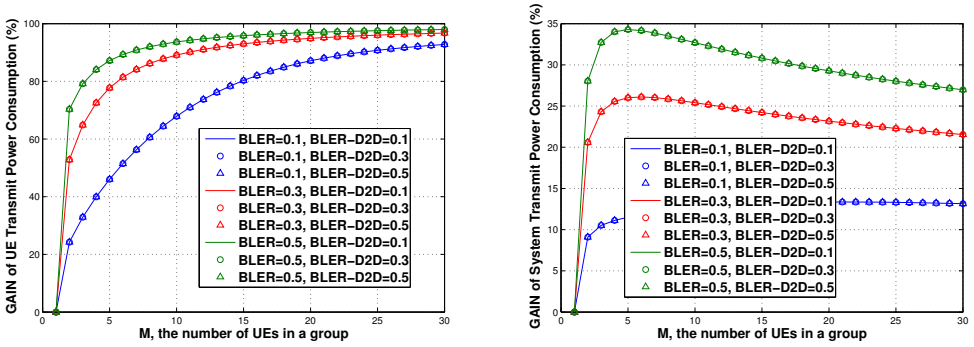
Fig. 2.4 (b) shows that to ensure more than one mature UE in the cellular phase, the performance gain will be reduced. Considering the interference control and guarantee of the QoS of certain traffic, the eNB may decide to spend more cellular resources to let more UEs become mature. According to the mature UE selection mechanism that described in Sec. 2.3.2, more mature UEs means more UEs to be selected, and thus with a higher probability, the UE having a lower interference to the cellular network or the UE having a better D2D channel condition (shorter waiting timer) will be selected.

In practice, the eNB will make the decision based on many factors, so that this part of simulations and analysis can offer the eNB a very good reference to make the trade-off between the cellular resources consumption and other goals.

2.5.4 Transmitting Power Gain

Based on the gain described in (2.25), (2.26) and the analysis in Sec. 2.4.3 and Sec. 2.4.6, we can obtain the following results as shown in Fig. 2.5.

From Figs. 2.5 (a) and (b), we can observe that given different BLER settings in the cellular and D2D transmissions, the UE transmitting power consumption can be significantly reduced. Also, about 10% to more than 30% system transmitting power can be saved by adopting the proposed approach. The proposed approach only needs some updates in software, which makes the power consumption gain more achievable.



(a) UE transmitting power gain (b) System transmitting power gain

Figure 2.5: Transmit Power Gain.

2.5.5 Comparison With Basic UE Cooperation

According to the analysis in Sec. 2.4.5, the comparison between the proposed approach and the basic UE cooperation approach is shown in Fig. 2.6.

We use (2.24) to calculate the performance gain in term of cellular resource con-

sumption for the basic UE cooperation method, and make the comparison with the proposed approach. We can observe from Fig. 2.6 that the proposed approach can achieve a higher performance gain given certain group size, and further increasing the group size results in a lower gain. From our analysis, the performance gain of the basic UE cooperation can never exceed the performance of the proposed approach, and we can also optimize the group size to maximize the performance gain.

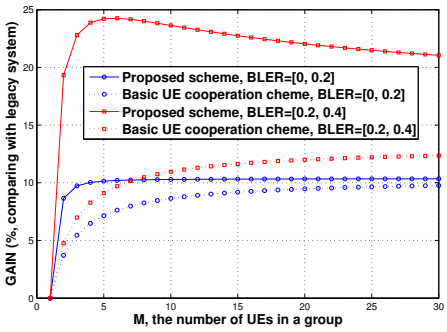


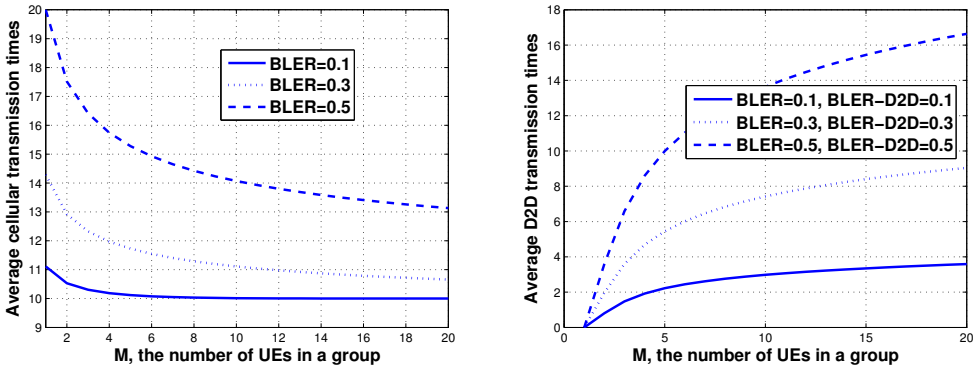
Figure 2.6: Performance comparison with basic UE cooperation.

2.5.6 Non-MTC Scenarios

In this subsection, we demonstrate the applicability of the theoretical analyses in the scenarios other than the unicast MTC. In Figs. 2.7, we fixed the number of needed linear combinations, $N = 10$, and then increase the number of UEs in a group, M, from 1 to 20. As discussed in Sec. 2.3.2, when $M < N$, it may be corresponding to the heterogeneous UE traffic scenario, because some UEs requires more than one packet. When $M > N$, it is a mixture of the unicast and multicast scenarios because some of the UEs ask for the same packets.

From Fig. 2.7 (a), given that the eNB will stop transmission when there is at least one mature UE in the receiving group, the needed number of cellular transmissions decreases with the increasing of M. It is because a higher multi-user diversity gain

can be achieved when the number the UEs is larger. The number of packets is constant regardless how many UEs are in this group. Once a mature UE successfully receives 10 different linear combinations, the cellular phase ends. Fig. 2.7 (b) shows the results of the D2D phase, where the needed number of transmissions increases with the increasing of M. A larger M means more UEs to be served by D2D, and thus more D2D transmissions are needed.



(a) The cellular phase (b) The D2D phase

Figure 2.7: The number of needed transmissions , $N = 10$.

2.5.7 Monte Carlo Simulation

Monte Carlo simulations are utilized to validate the theoretical results obtained from Sec. 2.4, and to justify the usage of average BLER rather than a specific channel fading model in the analysis. For each setting, we run the simulation 3000 times and obtain the average. We apply the following NLOS path-loss models for cellular and D2D links [46], respectively.

$$G_C(D_c)[\text{dB}] = -(44.9 - 6.55\log_{10}(h_{BS}))\log_{10}(D_c) - 34.46 - 5.83\log_{10}(h_{BS}), \quad (2.27)$$

$$G_D(D_d)[\text{dB}] = -36.8\log_{10}(D_d) - 43.8 - 5(n_{walls} - 1), \quad (2.28)$$

where h_{BS} is the height of the base station and n_{walls} is the number of walls between the D2D transmission. The channel gains, G_C and G_D, are functions of the transmission distances, D_c and D_d, for the cellular transmission and the D2D transmission, respectively. The Rayleigh fading channel is applied for both the cellular and the D2D transmissions as follows,

$$f_\gamma(x) = \frac{1}{\bar{\gamma}} e^{-\frac{x}{\bar{\gamma}}}, \tag{2.29}$$

where $\bar{\gamma}$ is the average received SNR.

We assume that one PRB is used for the MTC transmission so that the bandwidth, W, is fixed. Considering the noise spectrum density, n_0, the noise power equals $n_0 W$. Therefore, according to the transmission powers of the eNB and the D2D UE, denoted as P_{TC} and P_{TD}, and the path loss models shown above, the average received SNRs in both the cellular and the D2D transmission can be obtained. The parameter settings are summarized in Table 2.2.

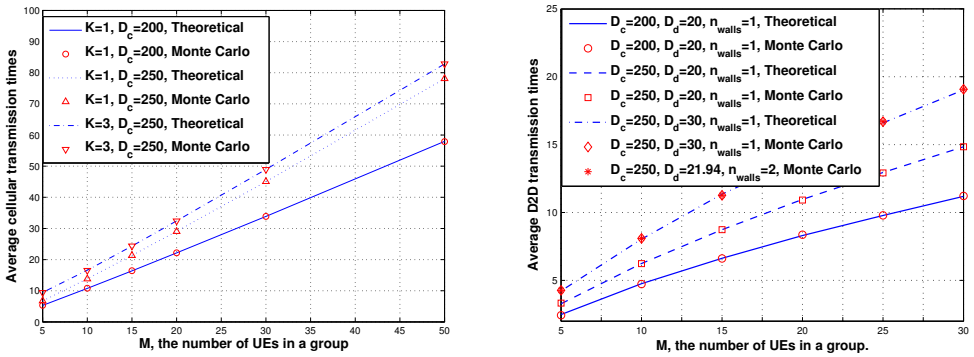
Table 2.2: Parameter Settings

Parameters	**Values**	**Description**
h_{BS}	25m	The height of the base station
n_{walls}	1, 2	The number of walls
D_c	200m, 250m	The cellular transmission distance
D_d	20m, 30m, 21.94m	The D2D transmission distance
W	180KHz	The bandwidth of 1 PRB
n_0	-17.4dBm/Hz	The noise spectrum density
P_{TC}	5W	The eNB's transmitting power
P_{TD}	12.5mW	The UE's transmitting power in D2D
MCS	QPSK, 1/3 coding rate	The fixed MCS for MTC devices
Block size	960bits	The number of information bits

Given the fixed MCS in Table 2.2, the SNR-BLER curve can be found in [49]. Therefore, the average BLERs can be calculated according to (2.2), equal 0.2606 and

0.4883 in the cellular transmissions for the distances of 200m and 250m, respectively. In the D2D transmissions, when $n_{walls} = 1$, they are 0.0565 and 0.2276 according to the distances of 20m and 30m, respectively. We set $n_{walls} = 2$ and $D_d = 21.94$ as another special channel setting which results in the same average BLER, 0.2276. These average BLERs are applied in our theoretical equations. We compare the theoretical results with that of Monte Carlo simulation to examine the consistency.

In Monte Carlo simulations, we drew an independent random variable following the distribution in (2.29) for each packet transmission to each UE, as the received SNR, and then obtained the corresponding instantaneous BLER based on the SNR-BLER relations in [49]. Given a certain instantaneous BLER, we used another uniformly distributed random variable $z, z \in [0, 1]$, to determine whether the current receiving is successful (succeed if $z >$ instantaneous BLER).



(a) The cellular phase (b) The D2D phase

Figure 2.8: Comparisons between the theoretical results and Monte Carlo simulations.

In Fig. 2.8, we compare the theoretical and simulation results based on the settings in Table 2.2. We only compare the D2D transmission times when $K = 1$, i.e., the simple case without the possible D2D transmission diversity gain. For the cellular transmission times, K are varied between 1 and 3. We can observe that the simulation results match well with the theoretical ones, which validates our analyses in Sec. 2.4.

The results also show that the performances of using average BLER and that of using a specific channel model consistent with each other, and given two different channel parameter settings, so long as they lead to the same average BLER, the final results in term of transmission times are the same. Therefore, our analysis based on average BLER is applicable, independent of the specific channel profile.

2.6 Conclusion

In this chapter, we have proposed a downlink transmission approach leveraging cooperative D2D communications and the network coding in order to reduce both the cellular resources consumption and the transmitting power consumption for MTC devices. A feasible protocol stack based on the modifications on current LTE system has been given and an analytical framework to quantify the system performance has been developed. As no hardware modification is required, the proposed solution is practical.

Chapter 3

Cooperative Device-to-Device Communication for Uplink Transmission in Cellular System

3.1 Introduction

In Chapter 2, we designed and analyzed the system applying NC and D2D for MTC in cellular downlink transmissions. Different from the downlink cases, the original packets to be transmitted in uplink are buffered in different devices and thus results in completely different transmission procedures. As mentioned in the previous chapter, the MTC has brought big challenges to cellular networks. Due to the complicated deployments and channel impairments for indoor/underground devices, one single transmission may need hundreds of repetitions to meet the link budget [4], which is especially undesirable for uplink because it conflicts with the requirement of low power consumption for most of the MTC devices [2]. In addition, serving a large number of the MTC devices will impose a huge pressure on the scheduler of the eNB.

Therefore, in this chapter, we further explore and investigate the feasibility to apply NC and D2D in cellular uplink transmissions.

The main objective of this chapter is to increase the transmission efficiency in cellular uplink while reducing the pressure on eNB's scheduler and devices' power consumption by exploring the delay tolerance feature of the MTC traffic. In this chapter, a semi-centralized cooperative control method is proposed to reduce the control/feedback overhead, where the UE relays are randomly selected according to the density decided by the eNB. We adopt both the random UE relay approach and the one further applying NC for cooperation, where the feature of delay tolerance is exploited by local packet exchange and the transmission efficiency is improved thanks to the multi-user diversity gain.

Two distinct traffic models, i.e., the MTC traffic with the small-data feature and the full-buffer traffic, are applied to examine the proposed approaches. In the current LTE/LTE-A system, the minimum resource allocation unit is one Physical Resource Block. If the packet size from the MTC device is too small, it cannot fully utilize one PRB especially when the channel condition is relatively good and a higher-order MCS can be used. This motivates us to investigate the effects of the proposed approaches for the MTC small-data traffic. Meanwhile, it is also worthy to examine the proposed approaches for other traffic models without the small-data feature when the small-data aggregation gain does not exist. Thus, the full-buffer traffic is considered. In the MTC scenarios, continuous data coming from delay insensitive videos and high resolution pictures can be modeled as full-buffer traffic. Therefore, two distinct traffic models, i.e., the MTC traffic with small-data feature [2] and the full-buffer traffic, are applied to examine the proposed approaches.

In the following, we propose an efficient semi-centralized cooperative control method for the uplink transmissions in cellular systems. Two specific cooperative approaches

based on D2D are proposed: one is the random UE relay approach and the other further applies NC. The proposed approaches can substantially increase the transmission efficiency while reducing the overhead, scheduling queue length, and devices' power consumption. A feasible system design including the protocol stack is given, which is backward-compatible with the current LTE/LTE-A system and easy to implement. Then, the system is modeled based on two distinct traffic models, i.e., the MTC small-data traffic and the full-buffer traffic. The D2D interference is considered and modeled applying stochastic geometry. Extensive simulations are conducted for the MTC small-data traffic in various scenarios to identify the performance gain and compared the two cooperative approaches. The performance with full-buffer traffic is analyzed theoretically. The closed-form results are obtained, according to which extensive numerical evaluations are then performed.

Extensive literature related to NC, D2D, and the application of NC and D2D in downlink transmissions has been discussed in Sec. 2.2, Chapter 2. For uplink, lots of researches focus on cooperative D2D communications but few of them consider to further introduce NC. In [50], the combination of UE relay and D2D was studied for uplink case, where the eNB selects the best UE relay but the global information is needed which increases the feedback load when the number of users is large. D2D interference cannot be ignored due to the fact that D2D transmissions will reuse the cellular uplink resource. A relay node selection method based on the D2D interference was proposed in [51], which also requires global information. In [31], the basic NC approach and a grouping method based on two users' complementary channels were introduced for uplink transmissions. [34] focused on a single link and applied NC for the UE relay node to combine the packets from the eNB, the edge UE, and the relay UE, which is different from our scenario.

Different from the previous work, in this chapter, we target on the combination of

NC and D2D in cellular uplink transmissions, and apply the semi-centralized grouping method based on the largest receiving power in the D2D links, which reduces the overhead as well as guarantees the quality of the D2D transmissions. Besides, both theoretical and numerical analyses are given from the system perspective, where the scheduling and interference are considered, rather than focusing on a single link.

3.2 System Design

3.2.1 Transmission Procedure Design

The eNB broadcasts a parameter a ($\in [0, 1]$) to all of the MTC UEs. Each UE can become a D2D agent according to a certain probability which is a function of a. A simple implementation is that each UE draws a random number between 0 and 1 and checks whether the random number is smaller than a to be a D2D agent, similar to the legacy Access Class Barring (ACB) approach in LTE/LTE-A. The UEs other than the D2D agents are named as capillary UEs.

We adopt the ACB-like approach because of the control overhead concern. If the eNB designates specific UEs as the D2D agents, first the eNB needs to collect all the requests from all active UEs and then the control information has to be sent back. It takes a long time and occupies a large amount of resources. This tedious procedure has to be repeated whenever the topology or the traffic changes and the groups need to be reformed. It is inefficient for the eNB to control the network with massive number of MTC UEs. On the contrary, there is only one parameter a being sent on the broadcast channel in the proposed approach.

Once the D2D agents are selected, each D2D agent broadcasts a preamble via D2D with a fixed transmitting power. All the capillary UEs find the nearest D2D agents according to the receiving power of the D2D preambles, and then establish the

D2D connections.

The D2D preambles are sent in the predetermined cellular uplink resources. All the D2D agents are synchronized by the eNB's downlink broadcasting channel, and the D2D preamble transmission procedure is triggered by either the updating of a or other broadcast notifications. Given that the eNB controls the timing of the preamble transmissions, it shall schedule the uplink cellular transmissions to allocate resources for preamble transmissions. During the time period that the topology is relatively stable, no D2D preamble or grouping is needed, until a number of new UEs arriving to the system and cannot find suitable D2D agent nearby. For the MTC scenario, given the fixed location of the devices, the overhead of preamble transmission and grouping is tolerable.

The D2D preamble transmissions may be spread using orthogonal code, time, and frequency to reduce the collision probability. The regular preamble structure in LTE system can be adopted, which includes 64 orthogonal Zadoff-Chu sequences. Similar to the Physical Random Access Channel (PRACH) structure in LTE, we can predetermine multiple frequency bands and sub-frames for the D2D preamble transmissions. Each D2D agent shall randomly select a Zadoff-Chu sequence, a frequency band, and a sub-frame. Given the limited transmission range of the preamble and the orthogonal resources used, the D2D preamble collisions can be well controlled.

If the location information is available, a capillary UE can select the nearest D2D agent by calculating the distance. Otherwise, according to the receiving power of the preambles, a capillary UE shall find the D2D agent corresponding to the highest average receiving power, which typically corresponds to the shortest transmission distance because the path loss may dominate the channel gain. Therefore, in the analysis, we assume that all the capillary UEs can find the nearest D2D agent for simplification.

Once a UE decides to connect the D2D agent, it shall exchange signaling messages with the D2D agent on specified resources. The resources can be implied by the selected Zadoff-Chu sequence and resource for the D2D preamble transmission. During the signaling exchange, the D2D preamble collisions can be further resolved, as well as the collisions among the capillary UEs in the same group. Besides, other parameters related to D2D data transmission can be negotiated, such as the maximum transmitting power and a predetermined schedule pattern inside the group. There are two types of collisions. One is the collision among different D2D agents, i.e., they select the same preamble and resource. Another one is the collision among capillary UEs in the same group, i.e, they are using the same resource to send signaling to the D2D agent. For the first one, random sequences generated by D2D agents (or their unique ID) are piggybacked in signaling and sent back to capillary UEs. If collision happens and a UE fails to decode the signaling, it shall notify that to the D2D agents. Thus, the colliding D2D agents can be aware of collisions and redo the preambles and resource selection. For the collision among capillary UEs, the random sequence can also be piggybacked, which is the same as the current contention-based random access mechanism in LTE system.

Each D2D agent and its connected capillary UEs can be considered as a cooperation group. We assume that the UEs' locations in a macro cell follow a Poisson Point Process (PPP), given the thinning procedure, the locations of the randomly selected D2D agents also follow a PPP. The macro area can be divided into many Voronoi cells, and each Voronoi cell contains one cooperation group. According to different traffic situations and UE densities, the probability of a UE becoming a D2D agent can be updated by the eNB.

In one Voronoi cell, there are two cooperative approaches for the uplink transmission. The first one is the NC approach, as shown in Fig. 3.1 (a), where the D2D agent

collects original packets from each capillary UE and then broadcasts all the packets back to the capillary UEs through the D2D network, so that each UE obtains all the packets of other UEs in this Voronoi cell. No matter which UE is scheduled, it will always transmit linear combinations of all the original packets. Therefore, the eNB can select the UE with the best channel condition in a Voronoi cell to transmit in the uplink, by which the multi-user diversity gain can be achieved.

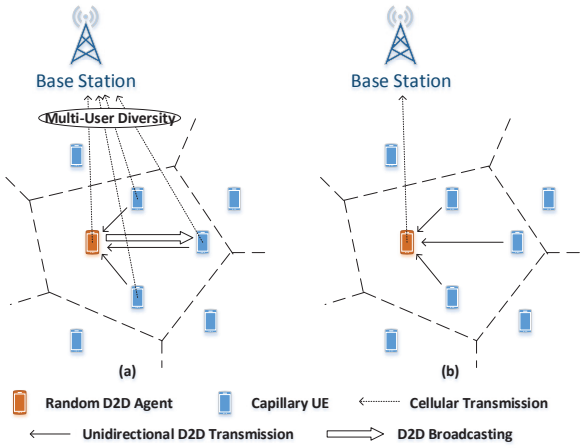


Figure 3.1: Two alternative semi-centralized control methods.

The second approach is shown in Fig. 3.1 (b) and is considered as the Random UE Relay (RUR) approach. The main difference is that the D2D agent will not broadcast the packets back to the capillary UEs and only the D2D agent can be scheduled for the cellular uplink transmission. The RUR approach can obtain the benefit of small-packet aggregation, but it cannot achieve the multi-user diversity gain. On the other hand, the signaling and feedback overhead is significantly reduced, because the eNB does not need to know the channel conditions and other status of the capillary UEs. Also, due to less D2D transmission times, the D2D interference level is reduced. Without a careful study, it is difficult to conclude which solution can achieve a better performance in a specific scenario.

It is possible to further explore the multi-user diversity gain in the RUR approach

by applying the dynamic scheduling within a group. However, coordinating all capillary UEs to send the packets to the scheduled UE dynamically can be costly and time-consuming, and thus becomes infeasible. Note that in FDD-LTE, the time for the scheduled UE to prepare the uplink transmission after receiving the scheduling message is 4ms only. Alternatively, each UE can store other's packets beforehand, which generates the same D2D traffic load and interference as the NC approach. From the received SINR's point of view, this method has the same performance as the NC approach (the same diversity gain and the same D2D interference), but the NC approach can reduce the control overhead thanks to its insensitiveness to the order of the (re)transmissions. The analytical results on the NC approach can be considered as the upper-bound of applying dynamic scheduling.

We propose the NC approach because it can effectively explore the multi-user diversity and small data aggregation gains, which improves the transmission efficiency, without introducing additional control overhead in regard of the re-transmission and reordering. We propose the RUR approach because it minimizes the control/feedback overhead and reduce the complexity of the eNB while maintaining the small data aggregation gain. The proposed approaches do not use multi-antenna techniques which may not be applicable for MTC devices. The transmission efficiency and the overhead/complexity due to massive MTC devices are two major issues to be addressed in the cellular system. We propose these two cooperative approaches to deal with them, respectively. Our analysis provides good guidelines to make the trade-off to select the better approach in different scenarios.

3.2.2 D2D Resource Allocation

In this chapter, we assume that the D2D transmissions are exactly underlying with cellular uplink transmissions, so no additional resource is occupied by D2D and thus

a fair comparison with the legacy system can be made. There is a mutual interference between D2D and cellular uplink transmissions and the resource allocation for the D2D has a great impact on the performance.

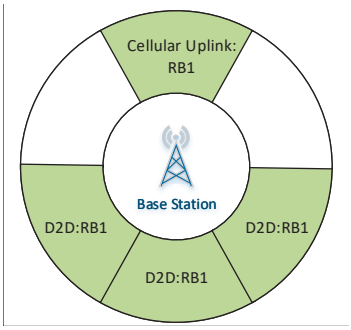


Figure 3.2: An example of the D2D resource allocation.

We assume that the eNB can mitigate the interference by a careful scheduling. A simple example of the scheduling is shown in Fig. 3.2, where we consider one PRB as the scheduling unit. The whole macro cell area can be divided into several sectors. When one PRB is occupied by a UE for cellular uplink transmissions, the sectors far away from it can reuse this PRB for D2D transmissions. Therefore, because of the relatively long distance, the interference from a cellular transmission to the D2D users is negligible. However, the distances to the eNB are similar for both the cellular UEs and the D2D UEs, so that the interference from D2D transmissions to cellular uplink transmissions cannot be neglected, which is modeled according to the worst case in the following sections.

3.2.3 Protocol Design

In the RUR approach, the D2D agent acts as a UE relay. Most of the functions used in this approach are either supported by the current LTE/LTE-A system or coming cellular standards such as the ProSe project in 3GPP [52]. For the capillary UE, it shall simply send the packets to the D2D agent. The D2D agent stores the packets

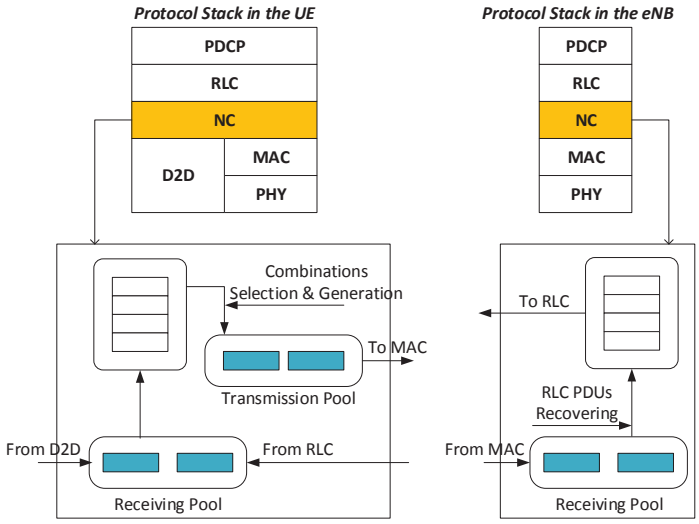


Figure 3.3: Design of the cellular air interface protocol stack.

from different capillary UEs in different dedicated Radio Link Control (RLC) entities. Once scheduled, it generates the Media Access Control (MAC) Protocol Data Unit (PDU) by multiplexing PDUs from different RLC entities according to the existing LTE priority handling algorithm.

For the NC approach, we insert a NC layer as shown in Fig. 3.3. Similarly to the first step of the RUR approach, all the capillary UEs send their packets (RLC PDUs) to the D2D agent. Then, the D2D agent broadcasts all the received packets, as well as its own packets, back to the capillary UEs. By now, all the packets need to be transmitted are stored by every UE in the group. When scheduled, the UE uses the packets stored in the receiving pool to generate the linear combinations and sends them to the eNB. We propose to define several coefficient tables according to various dimensions in specifications. In each table, linear independence is the guideline to design the detailed values of coefficient vectors. As long as N (the number of original packets) different coefficient vectors are linearly independent to each other, they can always build an invertible (full-rank) matrix, and thus Gauss-Jordan elimination can be applied to recover the original packets. A unique index

is assigned to each coefficient vector in the table. There are many options in the current system to inform UEs the index without introducing additional overhead, such as implicitly inferring from the System Frame Number (SFN), i.e., each SFN can be converted to a unique index, or using the index in a deterministic order. In order to facilitate the linear operation among packets, the fix-size RLC PDU is applied.

In the LTE/LTE-A system, the ciphering/deciphering is performed in the Packet Data Convergence Protocol (PDCP) layer. In the NC approach, though UEs received RLC PDUs from others, they cannot decipher the PDCP Service Data Units (SDUs) in the PDCP layer because they do not have others' keys. The keys used in the PDCP layer are negotiated by Non-Access Stratum (NAS) messages which are independent to the proposed approaches. Thus, our design can ensure that the existing security function in Radio Access Network (RAN) is preserved.

3.3 System Model For The MTC Small-Data Traffic

In this section, we consider the two-layer scheduling including the macro cell scheduling and that within one single Voronoi cell, and give the channel models for both cellular and D2D transmissions. After that, the D2D interference is modeled based on the stochastic geometry and the characteristics of the MTC small-data traffic.

Symbols used in this section are summarized in follows,

Table 3.1: Symbol Notation List (Small-Data Traffic)

Symbol	**Definition**

γ_k	The kth UE's instantaneous SNR
$\bar{\gamma}_k$	The kth UE's average SNR
$\gamma^{(j,i)}$	The received SNR in the eNB side from the jth, ith Voronoi cell
$F_{(\cdot)}$	The CDF of $(\cdot)$
$f_{(\cdot)}$	The PDF of $(\cdot)$
d_i	The distance from the ith Voronoi cell to the eNB
h_{BS}	The height of the base station
R_1	The distance from a capillary UE to its nearest D2D agent
R_2	The distance from a D2D agent to its furthest capillary UE
P_{D2D}	The D2D transmitting power
PL_{D2D}	The path loss between the D2D pair
I_{D2D}	The D2D interference to the cellular system
λ_D	The density of the D2D agent
λ	The density of UEs
M	The number of UEs in a Voronoi cell

3.3.1 Scheduling

In the macro cell scheduling, the users' fairness should be considered. The proportional fairness (PF) scheduling algorithm has been widely adopted, which can provide a reasonable trade-off between the throughput and fairness [53]. Considering the heterogeneous channel case, i.e., i.ni.d channels for different users, an alternative

normalized signal-to-noise ratio (SNR) scheduling algorithm can be used [54, 55], i.e.,

$$k^{(*)} = \arg\max_{k} \left\{ \frac{\gamma_k}{\bar{\gamma}_k} \right\} \tag{3.1}$$

where γ_k and $\bar{\gamma}_k$ are the kth UE's instantaneous and average SNR, respectively. $k^{(*)}$ denotes the scheduled UE.

We use the normalized SNR scheduling in the macro cell. Therefore, it can be assumed that there are several equivalent UEs in a macro cell, whose number equals the number of Voronoi cells in the macro cell. There is another layer of scheduling for the NC approach. The Max C/I scheduling algorithm is adopted in a Voronoi cell, as shown below, so that the scheduled SNR of each equivalent UE will be the same as the best UE in that Voronoi cell.

$$k'^{(*)} = \arg\max_{k'} \{\gamma_{k'}\} . \tag{3.2}$$

For the RUR approach, there is no scheduling in a Voronoi cell. The SNR of each equivalent UE is the SNR of the D2D agent in that Voronoi cell.

3.3.2 Channel Models

We assume independent log-normal shadowing and Rayleigh fading for each UE. Given that there are M capillary UEs in the ith Voronoi cell. The Cumulative Distribution Function (CDF) of the received SNR $\gamma^{(j,i)}$ in the eNB side from the jth UE, $j = 1, 2, ..., M+1$, in the ith Voronoi cell can be given by

$$F_{\gamma^{(j,i)}}(r \mid \bar{\gamma}_j) = 1 - e^{\frac{-r}{\bar{\gamma}_j}}, \tag{3.3}$$

where $\bar{\gamma}_j$ is the average SNR and is also a random variable follows log-normal distribution $\ln \mathcal{N}(\mu_i, \sigma)$ due to shadowing effect, which means $\bar{\gamma}_j$ is the average of small-scale fading. $F_{\gamma^{(j,i)}}(r \mid \bar{\gamma}_j)$ is the CDF of $\gamma^{(j,i)}$ given a fixed average SNR $\bar{\gamma}_j$. Its standard notation is $P(\gamma^{(j,i)} \leq r \mid \bar{\gamma}_j = x) = 1 - e^{\frac{-r}{x}}$, in order to simplify the notations in the following sections, we adopt the form shown in (3.3). We assume that UEs in the same Voronoi cell have approximately the same distance to the eNB. Therefore, $\mu_i = \bar{\gamma}(d_i)$, where $\bar{\gamma}(d_i)$ denotes the average received SNR based on the path loss and is a function of the distance from the ith Voronoi cell to the eNB, d_i. Given a fixed d_i, the CDF of $\bar{\gamma}_j$ is given by

$$F_{\bar{\gamma}_j}(x \mid d_i) = \Phi\left(\frac{10 \lg {}^{x}/_{\bar{\gamma}(d_i)}}{\sigma}\right), \tag{3.4}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

In this section, as we consider the MTC small-data traffic generated by indoor/undergroud devices, the Non-line-of-sight (NLOS) path loss channel model in [46] is applied and given by

$$G_D(x)[\text{dB}] = -[44.9 - 6.55 \log_{10}(h_{BS})] \log_{10}(x) - 34.46 - 5.83 \log_{10}(h_{BS}), \tag{3.5}$$

where $h_{BS} = 25m$ is the height of the base station.

3.3.3 Maximum D2D distance

Given that each capillary UE automatically connects to the nearest D2D agent, all the UEs will be clustered by the Voronoi cells. We denote the random variable R_1 as the distance from a capillary UE to its nearest D2D agent and R_2 as the distance from a D2D agent to the furthest capillary UE connecting to it. We first obtain the CDF of R_1 without macro cell boundary, i.e., infinite UEs following a PPP in an

infinite plane, as follows,

$$F_{R_1}(r) = 1 - \Pr[R_1 > r] = 1 - \left.\frac{(r^2\pi\lambda_D)^x e^{-r^2\pi\lambda_D}}{x!}\right|_{x=0} = 1 - e^{-r^2\pi\lambda_D}, \tag{3.6}$$

where λ_D is the density of the D2D agents.

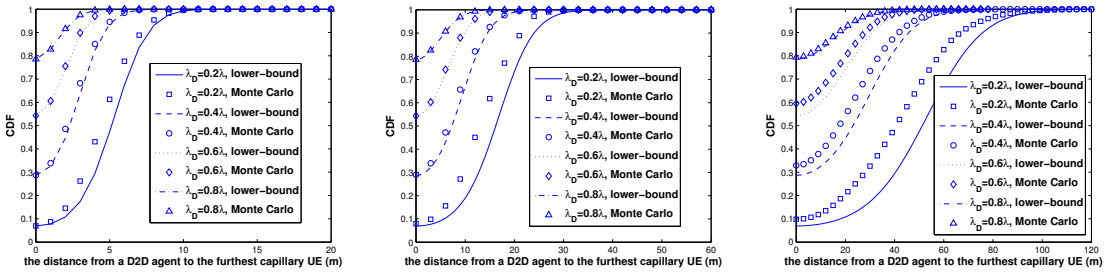
Given an infinite number of D2D agents and capillary UEs following the PPP with density λ_D and $\lambda - \lambda_D$, respectively, the CDF of the distance from a randomly chosen D2D agent to the furthest capillary UE connecting to it, R_2, is bounded by

$$\begin{aligned} F_{R_2}(r) &> \sum_{n=0}^{\infty} \Pr[M = n] F_{R_1}(r)^n \\ &= \sum_{n=0}^{\infty} \frac{3.5^{3.5}\Gamma(n+3.5)(\lambda-\lambda_D/\lambda_D)^n(1-e^{-r^2\pi\lambda_D})^n}{n!\Gamma(3.5)(\lambda-\lambda_D/\lambda_D+3.5)^{n+3.5}} = \hat{F}_{R_2}(r), \end{aligned} \tag{3.7}$$

where λ is the density of the UEs, including the D2D agents and capillary UEs, and M is the number of the capillary UE in a Voronoi cell. The proof of (3.7) can be found in Sec. 9.1.

When the size of each Voronoi cell is relatively small comparing with that of the macro cell, the proportion of the Voronoi cells cut by the macro cell boundary is small, so that (3.7) can well approximate the realistic situation that UEs are within a finite-size cell. We fix the radius of the macro cell to 500m and decrease λ from 10^{-1} to 10^{-3}. From Figs. 3.4 (a), (b), and (c), the gap between Monte Carlo results and the lower-bound in the infinite plane increase because a larger proportion of the Voronoi cells are affected by the macro cell boundary. Since we mainly focus on dense networks, the approximation is acceptable.

Thus, we apply the lower-bound $\hat{F}_{R_2}(r)$ for the following analyses, which overestimates the maximum D2D distance and can be used as the upper-bound of the D2D interference. The corresponding Probability Density Function (PDF) is obtained as



(a) $\lambda = 10^{-1}$ (b) $\lambda = 10^{-2}$ (c) $\lambda = 10^{-3}$

Figure 3.4: Monte Carlo verification on the lower-bound of the CDF.

follows,

$$\hat{f}_{R_2}(r) = \frac{\mathrm{d}\hat{F}_{R_2}(r)}{\mathrm{d}r} = \sum_{n=0}^{\infty} \Pr[M = n] n F_{R_1}(r)^{n-1} f_{R_1}, \ r > 0, \tag{3.8}$$

where the Probability Mass Function (PMF) of M is given by (33) in [56]. $\hat{f}_{R_2}(0)$ is an impulse function and can be obtained as follows,

$$\hat{f}_{R_2}(0) = \Pr[M = n] \lim_{r \to 0} \frac{\frac{\mathrm{d}}{\mathrm{d}r}\left[n 2\pi\lambda_D r e^{-r^2\pi\lambda_D}\right]}{\frac{\mathrm{d}}{\mathrm{d}r}\left[1 - e^{-r^2\pi\lambda_D}\right]}\bigg|_{n=0} = \Pr[M = 0]. \tag{3.9}$$

3.3.4 D2D transmitting power

In [46], the LTE uplink Open-loop Fraction Power Control (OFPC) approach is used to determine the transmitting power for D2D transmissions, which is given by

$$P_{D2D} = \min\{P_{\max}, P_0 + 10\log_{10} T + \alpha PL_{D2D}\}, \tag{3.10}$$

where $P_{\max}$ denotes the maximum UE transmitting power. P_0 and T are configured by the eNB and are given in Table 3.2. $PL_{D2D}(r)$ is the path loss between the D2D pair. α is selected from $[0, 1]$ to determine the compensation weight of the path loss [57].

Considering the fact that most of the MTC devices are placed in the indoor environment, we assume that all the D2D transmissions experience the NLOS channels and the PL_{D2D} can be obtained based on the distance between the D2D agent and the capillary UE, d_{d2d}, as follows [46],

$$PL_{D2D}(d_{d2d})(\text{dB}) = 36.8 \log_{10}(d_{d2d}) + 43.8. \tag{3.11}$$

For the MTC small-data traffic, we use the distance from a D2D agent to its furthest capillary UE, R_2, to determine the D2D transmitting power (mW) as follows,

$$P_{D2D} = 10^{0.1 \min\{24, -78 + \alpha(36.8 \log_{10}(R_2) + 43.8)\}}. \tag{3.12}$$

3.3.5 D2D Interference

In this chapter, we apply a simple rule to model the resource allocation for the D2D transmissions. In order to make a fair comparison with the current mechanism, the resource consumed for the MTC packets exchanging in the D2D network should be smaller than that for the same packets transmitting in the cellular uplink. In other words, the D2D transmission will be underlaid with the cellular transmissions for the same traffic without occupying additional cellular resource.

First, we use the received SNR (without the D2D interference) in the eNB side to perform the macro cell scheduling and then determine the MCS for the scheduled UE/Voronoi cell based on the SNR and estimated D2D interference level. We assume that the D2D transmission cannot use the last OFDM symbol of each subframe where UEs will transmit Sounding Reference Signal (SRS) for the uplink channel estimation, so that the eNB can still obtain the precise SNR of each UE. In order to ensure that all the D2D transmissions will be underlaid with the cellular transmissions, the number of

D2D transmissions could coexist per subframe, $\Omega^{(nc)}(M, MCS_i)$, in the NC approach is given by

$$\Omega^{(nc)}(M, MCS_i) = \frac{2M+1}{M+1} MCS_i, \tag{3.13}$$

where MCS_i is the MCS selected by the eNB for a specific Voronoi cell, which means i packets/linear combinations can be transmitted in one PRB in the cellular uplink. In the NC approach, as shown in Fig. 3.1 (a), there are two phases for the D2D transmissions, the D2D data collection and the D2D broadcasting. If there are totally M capillary UEs plus one D2D agent in the scheduled Voronoi cell, $2M+1$ D2D transmissions should be performed. We assume that each D2D transmission can only transmit one packet at one subframe. For example, when MCS_1 is selected, it needs $M+1$ cellular transmissions so that there should be $\frac{2M+1}{M+1}$ concurrent D2D transmissions on average in each subframe.

In practice, when MCS_0 is selected (SNR is too low and no transmission can succeed) for the cellular uplink, we can still allocate several concurrent D2D transmissions because they will not degrade the cellular transmission. Thus, (3.13) is the upper bound of the number of the concurrent D2D transmissions. Similarly, for the RUR approach, the upper bound of the number of the concurrent D2D transmissions can be given as follows,

$$\Omega^{(rur)}(M, MCS_i) = \frac{M}{M+1} MCS_i. \tag{3.14}$$

We only obtain the distribution of the maximum D2D distance for the Voronoi cells that have at least one capillary UE, so that we are using the conditional PDF

of R_2, which is

$$\hat{f}_{R_2}(r \mid r > 0) = \frac{\hat{f}_{R_2}(r)}{1 - \Pr[M = 0]}, \quad r > 0. \tag{3.15}$$

We assume that all the UEs in a Voronoi cell have the same distance to the eNB. For a scheduled Voronoi cell with M capillary UEs and given that MCS_i is selected, the estimated D2D interference can be approximated as,

$$\begin{aligned} I_{D2D}(M, MCS_i) \approx & \Omega^{(\cdot)}(M, MCS_i) \int_0^{\infty} \int_{D_0}^{D_R} P_{D2D}(r) \hat{f}_{R_2}(r \mid r > 0) G_D(x) f_d(x) \mathrm{d}x \mathrm{d}r \\ = & \Omega^{(\cdot)}(M, MCS_i) \int_{D_0}^{D_R} G_D(x) f_d(x) \mathrm{d}x \int_{0^+}^{\infty} P_{D2D}(r) \frac{\hat{f}_{R_2}(r)}{1 - \Pr[M = 0]} \mathrm{d}r, \end{aligned} \tag{3.16}$$

where $(\cdot)$ can be either (nc) or (rur). $f_d(x)$ is the PDF of the distance from a D2D agent to the eNB, $f_d(x) = 2x/(D_R^2 - D_0^2), x \in [D_R, D_0]$. D_R is the radius of the macro cell and D_0 is the guard distance between the D2D transmission and the eNB.

3.4 Performance Evaluation For The MTC Small-Data Traffic

3.4.1 Simulation Settings

Table 3.2: Parameter Settings

Parameters	**Values**	**Description**
λ	10^{-3}	The density of the UEs
λ_D	$a\lambda, a < 1$	The density of the D2D agents
D_0	300(m)	The minimum distance to the BS
D_R	500(m)	The maximum distance to the BS
A	$(D_R^2 - D_0^2)\pi/6$	The size of the simulation area
P_c	24(dBm)	The cellular transmitting power
B	1.8×10^5(Hz)	The bandwidth (1PRB)
N_0	$10^{-17.4}B$(mW)	Noise power
σ	$0, 1, 2$(dB)	Shadowing standard deviation
h_{BS}	25(m) [46]	The height of the BS
$P_{\max}$	24(dBm) [46]	Parameter in (3.10)
P_0	-78(dBm) [46]	Parameter in (3.10)
T	1 [46]	Parameter in (3.10)
α	$0.7, 0.8, 0.9$ [57]	Parameter in (3.10)

In the MTC small-data traffic scenario, when a UE send out all the packets, it will be removed from the scheduling list. Thus, the number and density of UEs are changing slot-to-slot, which brings difficulties to theoretical analyses and motivates

the following Monte Carlo simulations.

Table 3.3: MCS Table

Order	**MCS**	**SINR range (linear scale)**
0	transmission failed	$(0, 0.6]$
1	QPSK, Rate 1/3	$(0.6, 2.135]$
2	QPSK, Rate 2/3	$(2.135, 4.565]$
3	16QAM, Rate 1/2	$(4.565, 8.584]$
4	16QAM, Rate 2/3	$(8.584, 13.583]$
5	16QAM, Rate 4/5	$(13.583, 19.498]$
6	64QAM, Rate 2/3	$(19.498, \infty)$

In the following simulation, we let all the UEs follow a PPP in a certain area A with the density λ, and randomly select D2D agents according to the probability a. Thus, all the D2D agents follow a PPP with the density λ_D, $\lambda_D = a\lambda$. Each capillary UE connects to the nearest D2D agent. The D2D transmitting power is determined based on the maximum D2D distance within the Voronoi cell, R_2. The D2D interference is generated according to the approximation in (3.16). The received SNR from each UE is generated as a random number based on the Rayleigh fast fading and the log-normal shadowing in each subframe. After the macro scheduling, the scheduled UE's signal-to-interference-plus-noise ratio (SINR) is calculated based on its SNR and the D2D interference. The MCS in this subframe is then determined according to the SINR. Each UE has a certain number of packets to be sent. The simulation ends once all the UEs' buffer become empty. We use the transmission times that can be reduced comparing with the legacy LTE as the metric to evaluate the system. The fewer the transmission times, the higher the transmission efficiency

and the less the UE's power consumption. We define the gain as the ratio of the number of cellular transmissions that can be saved by the proposed approach to the number of transmissions needed in the legacy LTE (i.e., without D2D), as shown in (3.17), and use it as the performance index in the following simulation.

$$\text{Gain} = \frac{\#\text{ of transmissions in LTE} - \#\text{ of transmissions in the proposed approach}}{\#\text{ of transmissions in LTE}}. \tag{3.17}$$

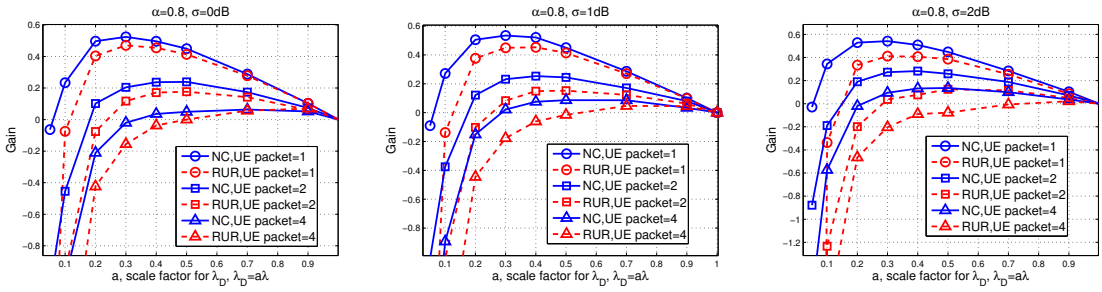
The parameters that are used in the simulation are summarized in Table 3.2. The Adaptive Modulation and Coding approach (AMC) is adopted. The MCSs that can be selected in the simulation are given by Table 3.3. The order of each MCS means the number of packets/linear combinations that can be transmitted by this MCS in one PRB. Based on the resource amount of one PRB and the resource occupied by the SRS and the Demodulation Reference Signal (DMRS) in the LTE uplink, the data amount of one MTC packet should be less than 88 bits. When the data amount of one UE is larger than 88 bits, it will be divided into several packets.

3.4.2 Simulation Results

In Fig. 3.5, $\alpha = 0.8$, which is the compensation weight in (3.10). Fig. 3.5 (a) is the homogeneous case with no shadowing. Figs. 3.5 (b) and (c) are the heterogeneous cases with the standard deviation of the shadowing equal to 1dB and 2dB, respectively. When each UE has only one packet to send, more than 50% and 40% transmission times can be saved by the NC approach and RUR approach, respectively. The performance gain decreases with the growth of the data volume per UE.

Besides that, the NC approach can achieve a higher performance gain when the standard deviation of the shadowing is larger. It is because the best channel in a

Voronoi cell is always selected and thus a higher multi-user diversity gain can be obtained in the heterogeneous case. For example, when the per-UE packet number is 1, the performance gain of the optimal point ($\lambda_D = 0.4$) is 46.87% for the homogeneous case. When the standard deviations of the shadowing are σ = 1dB and 2dB, the results are 48.78% ($\lambda_D = 0.4$) and 52.86% ($\lambda_D = 0.3$), respectively.



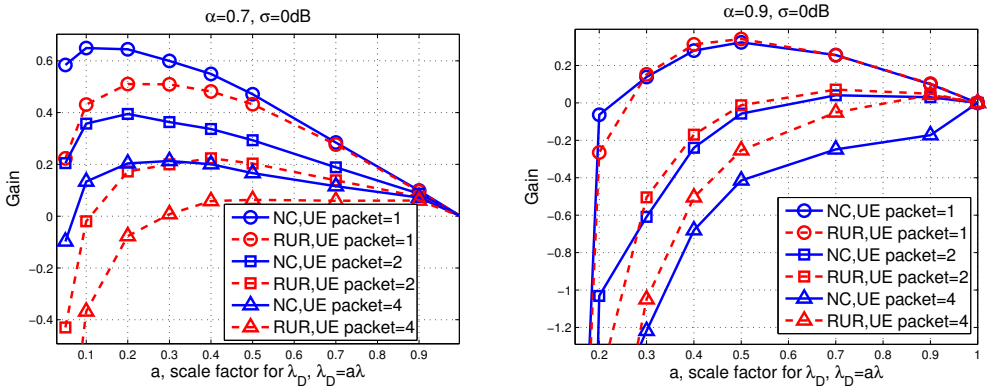
(a) (b) (c)

Figure 3.5: Performances of different shadowing setting.

On the contrary, with a higher shadowing the performance gain of the RUR approach is decreased. The reason is that a higher shadowing means a higher probability of a randomly selected D2D agent having a bad channel condition. Because this D2D agent is responsible to deliver packets from others, it will greatly degrade the system performance.

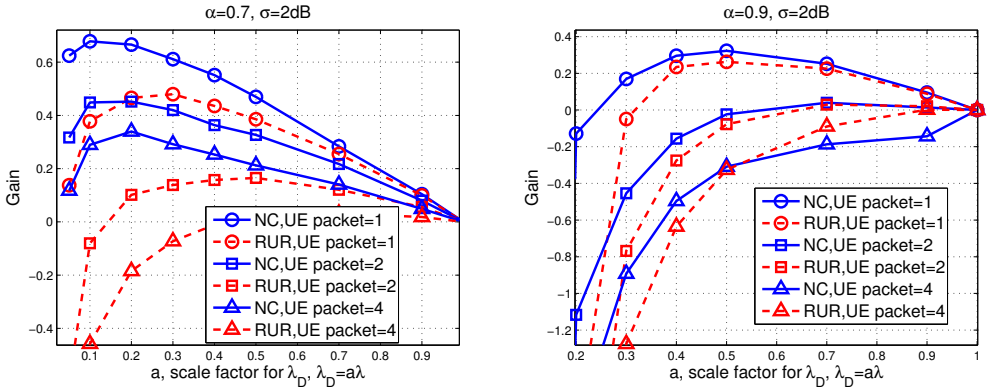
We also show the different results according to different D2D channel settings in Figs. 3.6. We use different α values in Table 3.2. By comparing these 3 figures, i.e., Fig. 3.6 (a), Fig. 3.5 (a), and Fig. 3.6 (b), we can observe that with the increasing of α the optimal λ_D becomes larger.

Given a certain transmission distance, a larger α means a larger D2D transmitting power, which will generate more interference. λ_D is negatively correlated with the D2D transmission distance. Therefore, when the D2D transmitting power is increased due to a larger α, the transmission distance has to decrease to reduce the D2D transmitting power according to the OFPC and control the interference level. Thus,



(a) (b)

Figure 3.6: Performances of different D2D channel setting (Homogeneous).



(a) (b)

Figure 3.7: Performances of different D2D channel setting (Heterogeneous).

the optimal point is moving toward a larger λ_D.

With the increase of α, the difference between the two approaches is getting smaller, and eventually the RUR approach becomes better than the NC approach (when $\alpha = 0.9$). It is because the NC approach needs additional D2D broadcasting stage, which generates a higher D2D interference. It brings a negative effect to the NC approach when the D2D transmitting power is relatively high comparing with the noise power and the cellular transmitting power. In this case, the D2D interference has offset all of the multi-user diversity gain achieved by the NC approach.

Figs. 3.7 repeat the simulations in Figs. 3.6 but with an on-average 2dB shadowing.

The trends are similar, where the optimal λ_D is getting larger with the increasing of the D2D transmitting power. Also, the conclusion we observed from Figs. 3.5 is confirmed again, i.e., the NC approach can obtain a higher gain in the heterogeneous case which is not achievable by the RUR approach.

In summary, the RUR approach is preferable if the standard deviation of the shadowing is small and the D2D transmitting power is high, and vice versa. For both of the two approaches, the optimal density of the D2D agents can be identified by the simulation results as well.

3.5 Theoretical Analyses For The Full-Buffer Traffic

We applied Monte Carlo simulations for the MTC small-data traffic scenario because the topology and the scheduling list are changing which make it difficult to analyze. In the full-buffer traffic scenario, UEs always have data to transmit and they will not be removed from the scheduling list. Thus, the topology of the network and the scheduling list are stable, which facilitates us to provide theoretical analyses and gain deeper insights in this section. Different from using needed transmission times as the metric in the Monte Carlo simulations for the MTC small-data traffic, the received SINR at the eNB side is considered as the metric for the full-buffer traffic. The performance evaluation is then decoupled with detailed MCS table and target Block Error Ratio (BLER) setting.

In this section, we apply the same channel models used for the MTC small-data traffic, which are given by (3.3). In order to further reduce the feedback load, it is assumed that the UE with the best large-scale channel gain in a Voronoi cell will be always scheduled, because in this way the eNB does not need to frequently monitor

the fast fading of each UE. In addition to Table 3.4, symbols used in this section are summarized in follows,

Table 3.4: Symbol Notation List (Full-Buffer Traffic)

Symbol	**Definition**
$\gamma^{(*,i)}$	The instantaneous SNR of the best UE in the ith Voronoi cell
$\bar{\gamma}^{(*,i)}$	The average SNR of the best UE in the ith Voronoi cell
$\gamma^{(**,i)}$	The received SNR at the eNB side from the ith Voronoi cell
L	The number of Voronoi cells covered by the base station

$\gamma^{(*,i)}$ and $\bar{\gamma}^{(*,i)}$ are the SNR and average SNR of the best UE in the ith Voronoi cell

We fix the number of the capillary UEs in the ith Voronoi cell to M, so that the CDF and PDF of the SNR with the best large scale channel gain is given by

$$F_{\bar{\gamma}^{(*,i)}}(x \mid d_i, M) = F_{\bar{\gamma}_j}(x \mid d_i)^{M+1} = \Phi\left(\frac{10\lg {}^{x}\!/_{\bar{\gamma}(d_i)}}{\sigma}\right)^{M+1}, \tag{3.18}$$

$$f_{\bar{\gamma}^{(*,i)}}(x \mid d_i, M) = \frac{\mathrm{d}F_{\bar{\gamma}^{(*,i)}}(x \mid d_i, M)}{\mathrm{d}x} = \frac{10(M+1)}{\sigma x \ln 10}\phi\left(\frac{10\lg {}^{x}\!/_{\bar{\gamma}(d_i)}}{\sigma}\right)\Phi\left(\frac{10\lg {}^{x}\!/_{\bar{\gamma}(d_i)}}{\sigma}\right)^{M}, \tag{3.19}$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution. We apply the same scheduling approach as that used for the MTC small-data traffic, i.e., the normalized SNR scheduling is applied in the macro cell and the Max C/I scheduling is used inside a Voronoi cell.

3.5.1 Network Coding

Given the full-buffer traffic and the NC approach, the CDF of the received SNR at the eNB side from the ith Voronoi cell is given by

$$\begin{aligned} & F^{nc}_{\gamma^{(**,i)}}(x \mid d_i) \\ =& \sum_{l=1}^{\infty} \left[\sum_{m=0}^{\infty} \Pr[M=m] \int_0^{\infty} \left(1-e^{-\frac{x}{u}}\right)^l f_{\bar{\gamma}^{(*,i)}}(u \mid d_i, m) \mathrm{d}u \right] \frac{(\lambda_D A)^l e^{-\lambda_D A}}{l!\left(1-e^{-\lambda_D A}\right)}, \end{aligned} \tag{3.20}$$

where the PMF of M, $\Pr[M=n]$, is given by (33) in [56], and A is the area of the macro cell. The proof of (3.20) is shown below.

Proof. We first assume that there are totally L equivalent UEs in the macro cell which also equals the number of Voronoi cells. The PDF of the SNR of each equivalent UE will be the same as that of the best UE in a Voronoi cell. Therefore, the conditional CDF of the received SNR at the eNB side from the ith Voronoi cell, $\gamma^{(**,i)}$, is given by

$$\begin{aligned} F^{nc}_{\gamma^{(**,i)}}(x \mid d_i, \bar{\gamma}^{(*,i)}, M, L) &= \Pr[\gamma^{(**,i)} < x \mid \text{the } i\text{th equivalent UE is selected}] \\ &= \frac{\Pr[\gamma^{(*,i)} < x, \frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}(M)} \text{is max}]}{\Pr[\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}(M)} \text{is max}]} \\ &= \frac{\int_0^{\frac{x}{\bar{\gamma}^{(*,i)}(M)}} f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(r \mid d_i) \prod_{\substack{j=1 \\ j\neq i}}^{L} F_{\frac{\gamma^{(*,j)}}{\bar{\gamma}^{(*,j)}}}(r \mid d_i)\, \mathrm{d}r}{\int_0^{\infty} f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(r \mid d_i) \prod_{\substack{j=1 \\ j\neq i}}^{L} F_{\frac{\gamma^{(*,j)}}{\bar{\gamma}^{(*,j)}}}(r \mid d_i)\, \mathrm{d}r}, \end{aligned} \tag{3.21}$$

where $\gamma^{(*,i)}$ and $\bar{\gamma}^{(*,i)}$ are the SNR and average SNR of the best UE in the ith Voronoi cell, respectively. $f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(\cdot)$ and $F_{\frac{\gamma^{(*,j)}}{\bar{\gamma}^{(*,j)}}}(\cdot)$ are the PDF and CDF of the normalized SNR of the best UE in the ith and jth Voronoi cell, respectively. $\bar{\gamma}^{(*,i)}(M)$ means the average SNR, a function of the number of capillary UEs in the ith Voronoi cell.

Thus, we first obtain the normalized SNR's PDF of the best UE in the ith Voronoi cell given M.

$$\begin{aligned} f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(r \mid d_i, M) =& f_{\gamma^{(*,i)}}(\bar{\gamma}^{(*,i)}(M)r \mid d_i)\left|\frac{\mathrm{d}}{\mathrm{d}r}\left(\bar{\gamma}^{(*,i)}(M)r\right)\right| \\ =& \frac{1}{\bar{\gamma}^{(*,i)}(M)} e^{-\frac{\bar{\gamma}^{(*,i)}(M)r}{\bar{\gamma}^{(*,i)}(M)}} \bar{\gamma}^{(*,i)}(M) = e^{-r}. \end{aligned} \tag{3.22}$$

The CDF can be obtained as follows,

$$F_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(r \mid d_i, M) = \int_0^r f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(x \mid d_i, M)\mathrm{d}x = 1 - e^{-r}. \tag{3.23}$$

It is clear that (3.22) and (3.23) are not related to d_i or M, which means that the distribution of the best UE's normalized SNR is i.i.d. for different Voronoi cells, so that we can change the notations of $f_{\frac{\gamma^{(*,i)}}{\bar{\gamma}^{(*,i)}}}(r \mid d_i)$ and $F_{\frac{\gamma^{(*,j)}}{\bar{\gamma}^{(*,j)}}}(r \mid d_i)$ in (3.21) to $f_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r)$ and $F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r)$, respectively. Therefore, (3.21) is rewritten as follows,

$$\begin{aligned} F^{nc}_{\gamma^{(**,i)}}(x \mid d_i, \bar{\gamma}^{(*,i)}, M, L) =& \frac{\int_0^{\frac{x}{\bar{\gamma}^{(*,i)}(M)}} f_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r) \prod_{\substack{j=1 \\ j\neq i}}^{L} F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r)\mathrm{d}r}{\int_0^{\infty} f_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r) \prod_{\substack{j=1 \\ j\neq i}}^{L} F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(r)\mathrm{d}r} \\ =& \frac{\frac{1}{L}\left[F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(\frac{x}{\bar{\gamma}^{(*,i)}(M)})\right]^L}{\frac{1}{L}\left[F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(\frac{y}{\bar{\gamma}^{(*,i)}(M)})\right]^L\Big|_0^{\infty}} = \left[F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(\frac{x}{\bar{\gamma}^{(*,i)}(M)})\right]^L. \end{aligned} \tag{3.24}$$

The conditional distribution of the average SNR of the best UE in this Voronoi cell, $\bar{\gamma}^{(*,i)}$, can be found in (3.19), we remove the conditions of $\bar{\gamma}^{(*,i)}$ and M successively,

$$F^{nc}_{\gamma^{(**,i)}}(x \mid d_i, M, L) = \int_0^{\infty} \left[F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(\frac{x}{u})\right]^L f_{\bar{\gamma}^{(*,i)}}(u \mid d_i, m)\mathrm{d}u. \tag{3.25}$$

$$F^{nc}_{\gamma^{(**,i)}}(x \mid d_i, L) = \sum_{m=0}^{\infty} \Pr[M=m] \int_0^{\infty} \left[F_{\frac{\gamma^{(*)}}{\bar{\gamma}^{(*)}}}(\frac{x}{u}) \right]^L f_{\bar{\gamma}^{(*,i)}}(u \mid d_i, m) \mathrm{d}u. \tag{3.26}$$

Assuming that there is at least one D2D agent existing, the condition of L in (3.26) can be further removed as shown below. According to the PPP model, $\Pr[L = l] = \frac{(\lambda_D A)^l e^{-\lambda_D A}}{l!}$. By substituting (3.26) into (3.27), (3.20) can be obtained.

$$F^{nc}_{\gamma^{(**,i)}}(x \mid d_i) = \sum_{l=1}^{\infty} F^{nc}_{\gamma^{(**,i)}}(x \mid d_i, l) \frac{\Pr[L=l]}{1 - \Pr[L=0]}. \tag{3.27}$$

□

3.5.2 Random UE Relay

Given the full-buffer traffic and the RUR approach, the CDF of the received SNR at the eNB side from the ith Voronoi cell is given by

$$F^{rur}_{\gamma^{(**,i)}}(x \mid d_i) = \sum_{n=1}^{\infty} \frac{(\lambda_D A)^n e^{-\lambda_D A}}{n! \left(1 - e^{-\lambda_D A}\right)} \int_0^{\infty} \left(1 - e^{-\frac{x}{u}}\right)^n f_{\bar{\gamma}_i}(u \mid d_i) \mathrm{d}u. \tag{3.28}$$

The proof of (3.28) is given below.

Proof. In the RUR approach, the eNB cannot select the best UE in a Voronoi cell, so that the CDF of the received SNR at the eNB side of the randomly selected UE in the ith Voronoi cell can be given by (3.3). The CDF of the normalized SNR of the randomly selected UE in a Voronoi cell can be easily obtained as follows, which is the same as that of the NC approach and is also i.i.d.. The condition of the distance from the ith Voronoi cell to the eNB, d_i, can be removed.

$$F_{\frac{\gamma,i}{\bar{\gamma},i}}(x \mid d_i) = 1 - e^{-x} = F_{\frac{\gamma}{\bar{\gamma}}}(x). \tag{3.29}$$

After the macro cell scheduling, the CDF of the received SNR in the eNB side from the ith Voronoi cell is given by

$$F^{rur}_{\gamma^{(**,i)}}(x \mid d_i) = \sum_{n=1}^{\infty} \frac{\Pr[L=n]}{1-\Pr[L=0]} \int_0^{\infty} \left[F_{\frac{\gamma}{\bar{\gamma}}}(\frac{x}{u})\right]^n f_{\bar{\gamma}_i}(u \mid d_i)\mathrm{d}u. \tag{3.30}$$

In the RUR case, we use the PDF of one single UE's average SNR $f_{\bar{\gamma}_i}(\cdot)$ instead of that of the best UE. By substituting (3.29) into (3.30), (3.28) can be obtained. □

3.5.3 Legacy System Without Cooperation

If there is no cooperation among UEs, only the macro scheduling will be performed and the eNB will always schedule the UE with the largest normalized SNR as shown in (3.1). The distribution of received SNR at the eNB side is similar to that of the RUR approach. The difference is that the user number in the macro scheduling list does not equal the number of the D2D agents in the RUR approach but equals the overall UE number in the legacy system.

Therefore, the CDF of the received SNR in the eNB side from the ith UE can be given by

$$F^{l}_{\gamma^{(**,i)}}(x \mid d_i) = \sum_{n=1}^{\infty} \frac{\Pr[U=n]}{1-\Pr[U=0]} \int_0^{\infty} \left[F_{\frac{\gamma}{\bar{\gamma}}}(\frac{x}{u})\right]^n f_{\bar{\gamma}_i}(u \mid d_i)\mathrm{d}u, \tag{3.31}$$

where U is the overall UE number and the PMF is given by $\Pr[U=n] = \frac{(\lambda A)^n e^{-\lambda A}}{n!}$.

3.5.4 D2D Interference

The maximum D2D distance is applied to calculate the D2D transmitting power for the MTC small-data traffic because the power control may not be performed in time and precisely for bursty traffic. However, the transmission is continuous in the full-

buffer scenario, and thus a more precise power control can be achieved. Therefore, we apply the OFPC approach combined with the distance from a capillary UE to the nearest D2D agent, which is denoted as R_1 in Sec. 3.3.3, and whose CDF is given by (3.6), to calculate the D2D transmitting power for the full-buffer traffic. The PDF of R_1 is given by $f_{R_1}(r) = \frac{\mathrm{d}F_{R_1}(r)}{\mathrm{d}r} = 2\pi r\lambda_D e^{-r^2\pi\lambda_D}$.

Given the arbitrary shape of cooperative area in practice, we use the PDF of R_1 in an infinite plane to approximate the D2D interference from a single source by

$$\begin{aligned} I_s &\approx \int_{D_0}^{D_R}\int_{0}^{\infty} P_{D2D}(r) f_{R_1}(r) G_D(x) f_d(x) \mathrm{d}r\mathrm{d}x \\ &= \int_{D_0}^{D_R} G_D(x) f_d(x)\mathrm{d}x \int_{0}^{\infty} P_{D2D}(r) f_{R_1}(r)\mathrm{d}r, \end{aligned} \tag{3.32}$$

where $G_D(x)$, $P_{D2D}(r)$ can be found in (3.5) and (3.10), respectively.

For D2D links, in addition to the NLOS channel model used in (3.11), the Line-Of-Sight (LOS) D2D channel is also considered in the performance evaluation of the full-buffer traffic, which is suitable for the devices in the outdoor scenarios and non-MTC scenarios. The path loss model [46] in dB scale is given by $PL_{D2D}(R_1) = 18.7\log_{10}(R_1) + 46.8$. The different models lead to different D2D transmitting power and interference level, which will make a big difference on the performance gains in the following numerical evaluations. Other than applying a specific MCS table for the MTC small-data traffic, we use another model to facilitate the evaluations of the theoretical results for the full-buffer traffic.

The number of concurrent D2D transmissions are highly related to the data rate in the macro cell. In a Voronoi cell, the D2D transmissions for certain packets should be finished before their cellular transmissions, so that the cooperation approaches can work. The D2D transmissions in one Voronoi cell can be underlaid with the cellular transmissions of other Voronoi cells. When the traffic arrival is continuous,

the procedure above can be wrapped around. Overall, the D2D transmissions cannot occupy more time-frequency resource than that used by the cellular transmissions for the same amount of data. Therefore, the higher data rate in cellular, the less resource can be utilized for the D2D transmission.

In the RUR approach, the D2D agent will collect the packets from other capillary UEs. If the data amount of each UE is s, the total data amount to be transmitted via the D2D in this Voronoi cell is sM. The data amount for the cellular uplink transmission is $s(M+1)$ because the D2D agent also has its own traffic. The ratio of the traffic amount in the D2D and the cellular is $M/(M+1)$, which is approximated to be 1 for simplicity. In other words, if the data rates in the D2D and the cellular are the same, $\Omega^{rur} = 1$. If the data rate in the cellular is c times of that in the D2D, $\Omega^{rur} = c$. In the NC approach, the D2D agent has to broadcast all of the packets back to each capillary UE, so that the ratio of the traffic amount in the D2D and the cellular is $(2M+1)/(M+1)$, which can be approximated as twice of that of the RUR approach.

The number of concurrent D2D transmissions is also related to the D2D resource allocation approach, i.e., whether the D2D transmissions are exactly underlaid with the cellular uplink transmissions. If all or some of the D2D transmissions can be offloaded to time-frequency resource other than that using by the cellular transmissions, such as D2D dedicated resource or using an unlicensed band, the number of concurrent D2D transmissions can be reduced.

Since the D2D resource allocation approach is not the main focus of this work, we use the extreme cases to evaluate the range of the performance gain of the proposed solutions. For example, we assume that the MCS of QPSK and 1/3 coding rate is adopted for the D2D transmission, which has the lowest modulation order and a relatively low coding rate in the LTE uplink. In the cellular, if the average MCS can

be as low as that of the D2D transmission or part of the D2D transmissions can be offloaded to another resource, the D2D concurrent transmission number will be much smaller, even zero.

We assume $\Omega^{rur} = 1$, $\Omega^{nc} = 2$ for the out-of-band D2D allocation case. If we select $\Omega^{rur} = 5$, $\Omega^{nc} = 10$, it means that all of the D2D transmissions can be underlaid with the cellular uplink and the average MCS of the cellular should be at least 64QAM and $1/2$ coding rate, which is the maximum modulation order in the LTE uplink and we believe that is large enough for the average MCS in the cellular for most of the cases. Therefore, the total D2D interference level can be approximated as $I_{D2D}^{(\cdot)} = \Omega^{(\cdot)} I_s$, where $(\cdot)$ can be either (nc) or (rur). We adjust $\Omega^{(\cdot)}$ to evaluate the performance gains in the following evaluation part.

3.5.5 SINR

Given the CDF of received SNR of each approach that shown in Sec.3.5.1, Sec.3.5.2 and Sec.3.5.3, and the approximation of the D2D interference, we can obtain the CDF of received SINR for the NC approach as follows,

$$F_{SINR}^{nc}(x \mid d_i) = F_{\gamma^{(**,i)}}^{nc}(\frac{N_0 + I_{D2D}^{nc}}{N_0} x \mid d_i). \tag{3.33}$$

The CDF of received SINR for the RUR approach can obtained by the same equation but replace $F_{\gamma^{(**,i)}}^{nc}(\cdot)$ and I_{D2D}^{nc} by $F_{\gamma^{(**,i)}}^{rur}(\cdot)$ and I_{D2D}^{rur}, respectively.

3.6 Performance Evaluation For The Full-Buffer Traffic

3.6.1 Parameter Settings

We use the received SINR at the eNB side from one Voronoi cell as the metric to evaluate the performance. In practice, given a fixed MCS table and a target BLER, a higher SINR results in a higher MCS for the uplink transmission in the tagged Voronoi cell. Thus a higher overall throughput of the macro cell can be obtained because of the equal scheduling opportunity for each Voronoi cell given the normalized SNR scheduling.

Besides the SINR gain, the control overhead can be also reduced by the proposed cooperative transmission approaches. Because of a higher transmission efficiency, the data amount can be transmitted by each scheduling grant is increased. Thus given the same data amount, the number of needed scheduling messages is reduced. Another kind of control overhead is the feedback load, which can also be reduced due to the facts that only the large scale channel gains and the D2D agents' channels are monitored in the NC approach and the RUR approach, respectively.

Different from the scenario of MTC small-data traffic, there is no small-data aggregation gain for the full-buffer traffic. The RUR approach cannot achieve a positive SINR gain but can reduce the queue length. The NC approach can achieve a positive SINR gain thanks to the multi-user diversity gain, especially in the heterogeneous case when σ is large. We set $\sigma = 4$dB [46] for the full-buffer traffic in this section. Other parameters used in the numerical evaluation are the same with those summarized in Table 3.2.

3.6.2 Monte Carlo Simulation

We have conducted Monte Carlo simulations to verify the key step of the conclusion in (3.20), i.e., eq. (3.25). The rest of proof procedures are just taking the expectation with respect to M and L. Three different cases are verified where $M = 5, 10, 20$ and $L = 20, 40, 100$, respectively. Because d_i only affects $\bar{\gamma}(d_i)$, we assume $\bar{\gamma}(d_i) = 5, 10, 15$ in the three cases. Other parameters are the same with those in Table 3.2. As shown in Fig. 3.8, the theoretical results match well with the results of Monte Carlo simulations. Because (3.28) and the result of legacy LTE adopt the same method and reuse the key part of (3.20), they can be verified by the same Monte Carlo simulations as well.

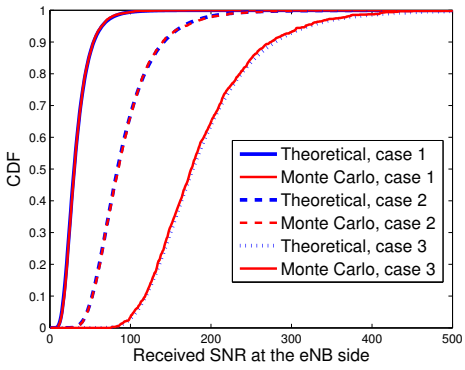
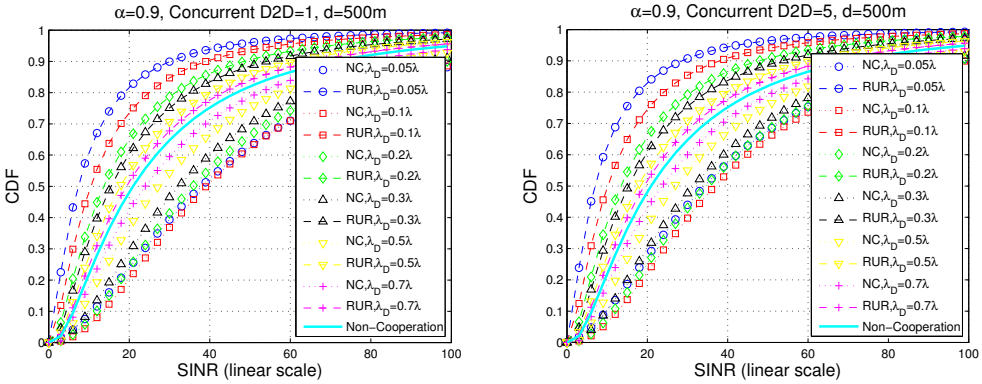


Figure 3.8: Monte Carlo verification for (3.20).

3.6.3 Numerical Results

The LOS D2D channel model is first applied. We fix the distance from one Voronoi cell to the eNB to 500m and examine the CDF of the received SINR at the eNB side from this Voronoi cell. The scenarios with the highest D2D interference level are shown in Fig. 3.9 (b), where $\alpha = 0.9$ and is the largest option in Table 3.2. According to the analyses in Sec. 3.5.4, $\Omega^{rur} = 5$ and $\Omega^{nc} = 2\Omega^{rur}$. Even in this less favourable case the NC approach can still achieve a significant gain. For instance, the 50 percentile and 20 percentile of SINR is increased by 71% and 100%, respectively,



(a) (b)

Figure 3.9: Evaluation results of the LOS D2D channel, $d_i = 500$m.

comparing with the Non-Cooperation system. It means that the NC approach is always a promising choice in the case of LOS D2D channel.

The NLOS D2D channel model is then applied. We choose the same settings for Ω^{nc} and Ω^{rur} but use different values of α. It can be observed from Figs. 3.10 that the SINR performances of the RUR approach are worse than the legacy system (Non-Cooperation). It is because the number of users to be scheduled is decreased thus a part of the multi-user diversity gain is lost. Also, an additional D2D interference is introduced by the cooperative transmissions. With the increasing of λ_D, more D2D agents can be scheduled and less D2D interference will be introduced due to shorter distances between capillary UEs and D2D agents, the CDF of the RUR approach is thus getting closer to that of the legacy system.

On the contrary, the NC approach can achieve a positive performance gain in the most cases, especially when the D2D interference can be maintained in a relatively low level. In Fig. 3.10 (a), a huge gain can be achieved. But with the rise of the D2D interference level (α and/or concurrent D2D number), the performance gain becomes smaller, as shown in Figs. 3.10 (b), (c) and (d). Eventually, in Figs. 3.10 (e) and (f), there is no positive gain.

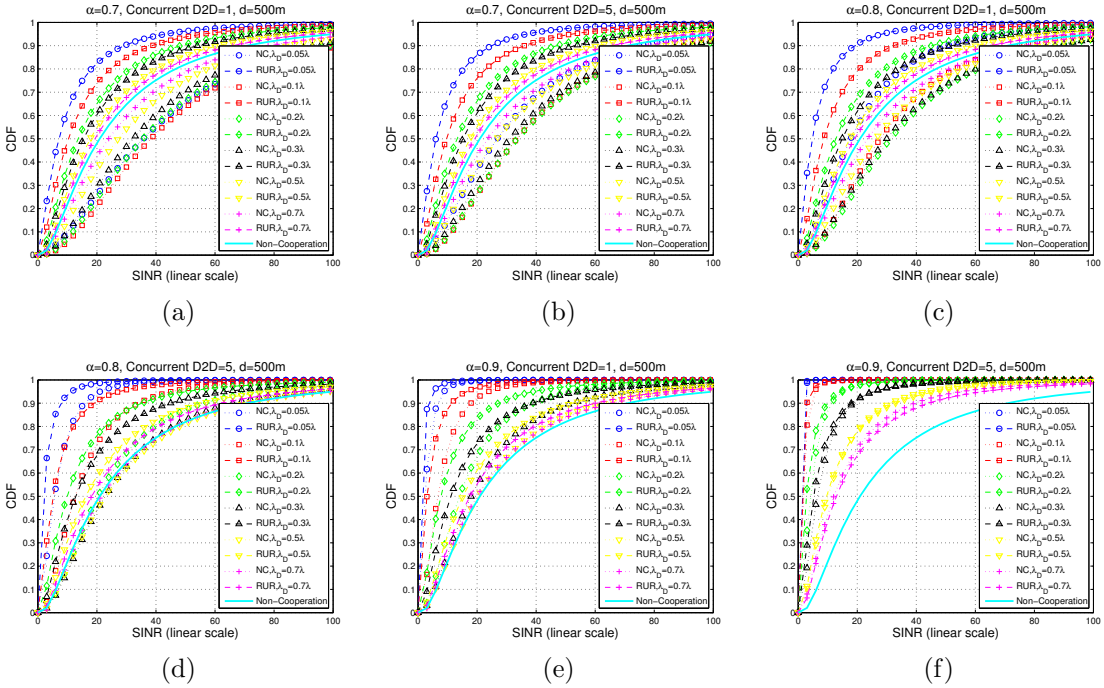


Figure 3.10: Evaluation results of the NLOS D2D channel, $d_i = 500$m.

The benefit of the NC approach comes from the best channel selection inside a Voronoi cell, which brings an additional multi-user diversity gain at the cost of the D2D interference. When the D2D transmitting power is low, the multi-user diversity gain will dominate and thus a SINR gain can be achieved. However, when the D2D interference is large enough, all of the multi-user diversity gain will be devoured by the interference.

In Fig. 3.10 (a), the NC approach can achieve the best performance when $\lambda_D = 0.1\lambda$. The difference between the curves of $\lambda_D = 0.1\lambda$ and $\lambda_D = 0.2\lambda$ becomes smaller when $\Omega^{(\cdot)}$ is increased in Fig. 3.10 (b). In Fig. 3.10 (c), $\alpha = 0.8$, the D2D interference level is further increased and 0.2λ is the optimal choice for λ_D. In the following figures, with the further increasing of the D2D interference level, the optimal λ_D is moving toward a larger value. It is because λ_D is negatively correlated with the distance between the capillary UEs and their D2D agents. Reducing the size of the

Voronoi cells can compensate the increasing of the D2D interference.

The NC approach cannot always achieve a better performance than the RUR approach. For example, in Fig. 3.10 (f), when $\lambda_D = 0.7\lambda$ the RUR approach is better than the NC approach. It is because that the D2D interference introduced by the NC approach is twice of that of the RUR approach. When the interference level is high, and there is not much multi-user diversity gain to be utilized by the NC approach due to the limited capillary UE number in each Voronoi cell, the doubled D2D interference will make the difference.

Although the performances of the cooperative transmission approaches are worse than the legacy system in the high D2D interference level cases (Figs. 3.10 (e) and (f)), these evaluations are still valuable due to other benefits brought by the cooperative approaches, such as the scheduling list reduction. Our analysis provides a clear relationship between the performance degradation and the scheduling list reduction, which is a good reference for the eNB to make the trade-off. For example, if the eNB decides to reduce the scheduling queue length by 30% in the scenario of Fig. 3.10 (f), how much performance gain needs to be sacrificed is shown and it is clear that the RUR approach is a better choice in this case.

In Figs. 3.11, we repeat the numerical evaluations for the scenarios of $d_i = 300$m. It can be observed that the performance gain and the optimal λ_D are the same with the results in Figs. 3.10, but the absolute values of the SINR are significantly increased. It is because the D2D interference comes from the whole cell which is not related to the position of the tagged Voronoi cell. Therefore, only the received power from the tagged Voronoi cell is increased due to a shorter transmission distance and thus the SINR is also increased.

In summary, for the full-buffer traffic scenario, the NC approach can always achieve a positive gain in the cases with LOS D2D channels or in the most cases with NLOS

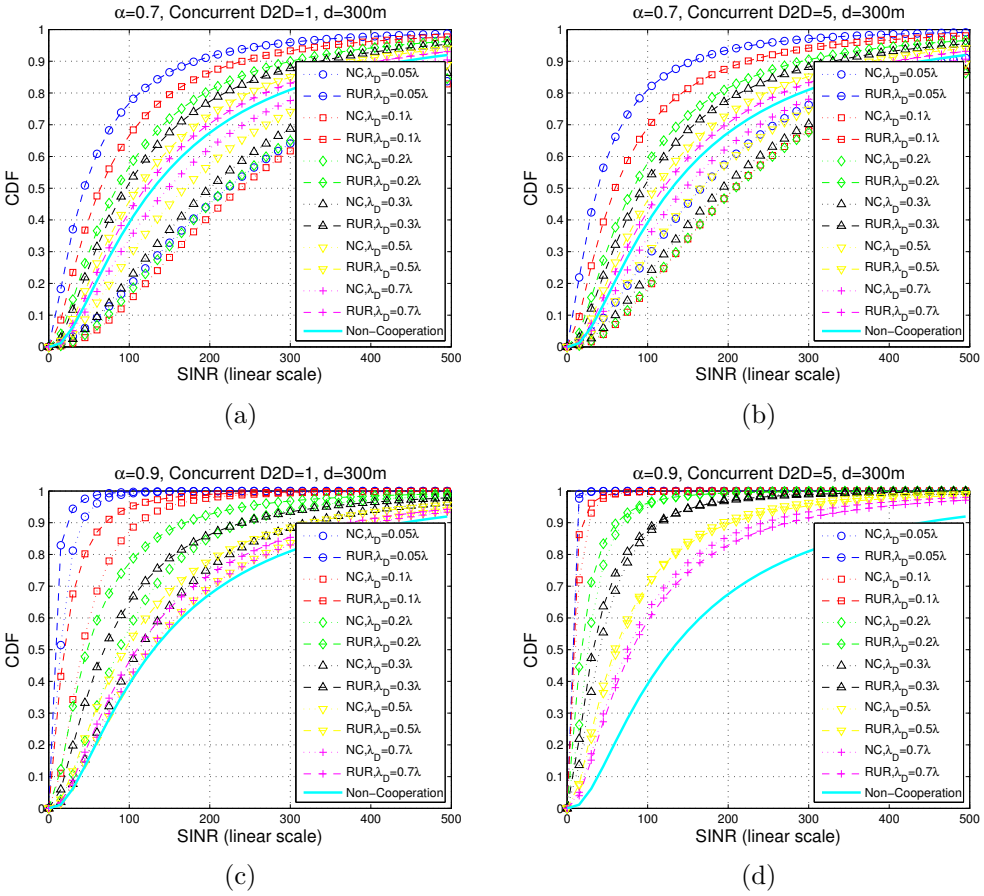


Figure 3.11: Evaluation results of the NLOS D2D channel, $d_i = 300$m.

D2D channels. The numerical results provide a good guideline for the eNB to control the density of the D2D agents. The RUR approach is a better choice if the eNB determines to reduce the scheduling queue length and the feedback load at the cost of the performance degradation. The quantitative performance degradations are identified by the numerical results, and thus help the eNB to make the trade-off.

3.7 Conclusion

In this chapter, a semi-centralized cooperative control method has been proposed for the cellular uplink transmission to address the new challenges brought by IoT, where the UE relays were randomly selected according to a certain density decided by the eNB. Two cooperative approaches based on the D2D have been proposed, which are the RUR approach and the NC approach. The proposed approaches have been analyzed given the D2D interference modeled by the stochastic geometry and two distinct traffic models, i.e., the MTC traffic with small-data feature and the full-buffer traffic. Intensive Monte Carlo simulations have been first conducted for the small-data traffic and then closed-form theoretical results have been derived for the full-buffer traffic. For both of the traffic models, the performance gains have been identified in various scenarios and the comparisons between two cooperative approaches have been made as well. Also, these results can provide an important guideline for the eNB to make the trade-off between the transmission efficiency and the scheduling queue length and determine the optimal density of the UE relays.

Chapter 4

Power Allocation and Flight Path Planning for Floating Relay Supporting MTC Traffic in Cellular Systems

4.1 Introduction

Though the approaches proposed in Chapter 2 and 3 can improve the transmission efficiency of MTC by exploring extra multi-user diversity gain, the undesirable MTC channels are not fundamentally changed. When all the MTC devices within an area experience high shadowing and penetration loss, the cellular transmissions still need to increase either the resource consumption or the transmitting power to overcome the poor-quality channels, even the device with the best channel is selected.

Deploying new small cells for MTC devices may be helpful. However, the unique traffic type of MTC makes the infrastructure of new small cells unprofitable. The

typical MTC traffic has the features of the small packet and low traffic volume [2], so the data amount contributed by MTC devices is limited, which makes the small-cell solution unprofitable. Besides that, deploying small cells becomes difficult and costly. The optimal location of a new cell to cover the users in the upper floors of skyscrapers may be in the mid-air and people are more sensitive to the radiation from a BS. All of these have made supporting MTC a major and tough issue for cellular systems.

The technologies of Unmanned Aerial Vehicle (UAV) had become mature in the last years. They are light, flexible and have a longer battery life than ever before [58], which motivates us to investigate the possibility of using UAV to carry a relay node and incorporating the Floating Relay (FR) in the cellular system.

The introduction of the FR is promising to support MTC in the cellular systems. It can approach the MTC devices to compensate the deep shadowing channels, increase the transmission/spectrum efficiency, and thus reduce the power consumption for MTC devices. Given that an FR can fly to almost everywhere, it can adjust its position according to the real-time traffic and the distribution of users' locations. Thanks to its 3-D mobility, the coverage offered by an FR may be much larger than that of a single small cell. Therefore, the cost of deploying new small cells can be greatly reduced.

However, how to find the optimal location of the FR and determine the flight path to achieve the best system performance in the cellular network remains open. Our work aims to fill the gap. In this chapter, targeting on the deep shadowing channels of the MTC devices, we introduce the FR to the cellular system and propose a mechanism for the FR to improve the transmission efficiency and maximize the system capacity. The main contributions are summarized as follows: Considering the capacity limit of the FR's back-haul link and the maximum transmission power of each user, an optimization problem is formulated to maximize the system capacity.

The optimal power allocation strategy, i.e., how to allocate the transmission power of each user for its existing serving BS and the FR, is derived. The maximum system throughput can be achieved given the location of the FR. Next, Based on the optimal power allocation, two effective on-line 3-D placement algorithms with low complexity are proposed for the FR to find the optimal location in unpredictable and predictable scenarios, respectively. Last but not the least, the comparisons with the other off-line approaches need to be provided.

4.2 Related Work

The idea of a moving network node itself goes back to 1998. A US patent by Charles D. Gavrilovich [59] required base stations to move along a roadway with traffic providing coverage for vehicles. The latest ones include the Project Loon [60] led by Google, where an LTE base station is carried by a balloon to provide the Internet access to rural and remote areas. Facebook has also launched their Aquila to expand the Internet's reach [61]. The Flying Ubiquitous Sensor Networks (FUSN) [62] has been one of the main applications of the combination of an aerial device and communication-capable node.

Many optimization problems using mobile sink in the Wireless Sensor Network (WSN) have been formulated and solved in the literature, among which balanced energy consumption is one of the main purposes of introducing mobile sinks to WSN. Routing methods and data collection strategies were proposed in [63–67] to increase the lifetime of network. The trade-off between energy saving and traffic delay was analyzed in [68]. [69] focused on the 3-D underwater scenario and proposed an efficient data collection approach to improve the lifetime as well as the throughput and the traffic delay. However, only linear route of the mobile sink was considered.

Besides lifetime, transmission efficiency and connectivity can also be improved by moving network nodes. [70] provided a comprehensive survey on Flying Ad-hoc NETwork (FANET). The MASP approach was proposed in [71] to increase the amount of collected data given the same energy consumption. In [72], authors implemented a Mobile Ad-hoc NETwork (MANET) to collect or disseminate content in complicated terrain. The system was then found useful in urban scenarios because of the existence of Line Of Sight (LOS) channel between Micro Air Vehicles (MAVs). This test was done by using eBee devices [73]. [74] proposed an algorithm placing a specified number of relay nodes to create a spanning network that minimizes the maximum link length. In [75], a multi-hop network was formed by the UAVs and thus provided connections between isolated sensor nodes and the base station. This work mainly focused on how to maintain the connectivity of the UAV-network.

Most of the cases studied in the literature focused on WSN and based on the IEEE 802.11 protocols. A few of them considered a better usage of the UAV in a cellular system. The project Loon was based on the LTE system, but the movement of a balloon cannot be dynamic so that the application scenario was restricted to the macro coverage extension for remote areas. The link level analyses based on the Nakagami fading model were given in [76], where the mobile relay was fixed on a train. [77] studied a wireless system where the UAV flies above distributed ground terminals, but the network topology was one-dimensional, and the flight path of the UAV was predetermined. In [78], three typical use cases of UAV-aided wireless communications, i.e., UAV-aided ubiquitous coverage, UAV-aided relaying, and UAV-aided information dissemination and data collection, were discussed. Besides that, UAV-enabled mobile relaying and D2D-enhanced UAV information dissemination were introduced as two key techniques.

However, in a complicated urban scenario where the basic coverage has already

been provided by existing stationary base stations, how to utilize floating network nodes to improve the system performance remains open, which motivates the work proposed in this chapter.

4.3 System Model

Fig. 4.1 gives an example of the system model. There are several existing BSs and UEs in a 3-D space. A UE shall connect to the serving BS providing the best average channel quality considering path loss and shadowing. Given the current location of the FR, each UE divides its transmission power into two parts: one is to communicate with the FR and the other is to the serving BS. The objective is to maximize the system throughput. Assuming that the UEs' power allocations are controlled by the network in a centralized way, we investigate the optimal power allocation strategy and the optimal location for the FR.

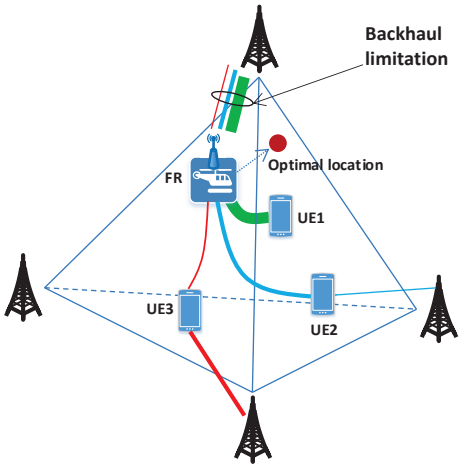


Figure 4.1: The floating relay with back-haul limit.

We denote the link between a UE and the FR as a front-haul link, and the one between a UE and its serving BS as a serving link. The maximum throughput of each single UE is not the objective of power allocation between front-hauls and serving

links, because the sum of front-hauls' data rate may exceed the back-haul's capacity. Due to this limit, UEs' power allocations are correlated to each other. The optimal solution is to maximize the system throughput, i.e., the sum of the data rates of all the UEs' front-haul links and serving links, rather than a single UE's throughput. Symbols used in this chapter are summarized in follows,

Table 4.1: Symbol Notation List

Symbol	**Definition**
C	The Shannon capacity of a wireless link connected to the UE
C'	The Shannon capacity of the back-haul link
$\boldsymbol{x}$	The location of the FR
$\boldsymbol{u}_i$	The location of the ith UE
$\boldsymbol{s}_i$	The location of the ith UE's serving BS
$\boldsymbol{s}_x$	The location of the FR's serving BS
α_i	The transmission power of the ith UE that is allocated to the FR
P_t	The total UE transmission power
P_t^F	The FR transmission power
$\mathbb{U}$	The UE set
d_i	The distances of the ith UE to the FR
D_i	The distances of the ith UE to its serving BS
D_x	The distance from the FR to its serving BS
$H_{(\cdot)}$	The equivalent normalized channel gain give the distance $(\cdot)$
α_i^*	The optimal power allocation of the ith UE
W	The Bandwidth
M	The number of UEs

$Z(\cdot)$	The maximized system throughput given the FR's location $(\cdot)$
R	The step length
$\boldsymbol{d}$	The direction of the FR's next move

We use link capacity, denoted by $C(d, p)$, to approximate the link data rate in this chapter. It is the Shannon capacity of a link given the transmission power p and distance d.

$$C(d, p) = W \log_2(1 + \frac{p \cdot g(d)}{N_0}), \tag{4.1}$$

where $g(d)$ is the channel and antenna gain, and N_0 is the power of noise. We do not consider the effect of fast fading because it is difficult to react according to fast fading. The channel gain is given by $g(d) = d^{-r}\Psi/K_0$, where r is the path loss exponent, Ψ represents the shadowing effect, and K_0 is a constant. Considering the nearfield effect, $g(d) = [\max(d, d_0)]^{-r}\Psi/K_0$, where d_0 is the reference distance.

In this chapter, the resources are statically allocated to users using orthogonal spectrum. Given the poor channels of MTC devices (high penetration loss and possible deep shadowing for devices inside buildings), the scheduling and feedback messages need to consume a large portion of the cellular resources to meet the link budget of the control channel, which is much higher than that of the data channel. Thus, the traditional dynamic scheduling may be too costly to apply. Therefore, the Physical Resource Block (PRB) is assigned statically (via high-layer messages) in the proposed solution to reduce the control overhead, so each user has a fixed bandwidth. The result with fixed bandwidth allocation is a valuable baseline, and the performance can be improved if coordinated scheduling between different BSs is possible. We do not

specify the multiple access approach in this work. As long as each user has orthogonal resources, such as frequency, code, and/or time, they will not interfere with each other.

The UAV broadcasts the reference signal as the Type 1a relay node in the LTE system [79]. The UE shall measure the channels to its serving BS and the FR, respectively, and then report the channel conditions and its locations to the serving BS. Neighboring BSs share the information via wired links (the X2 interface) without occupying any wireless resource. The FR's serving BS collects the information of the FR and UEs in neighboring cells, and then calculates the next move and notifies the FR. Given the procedures above, the overhead will not increase significantly because the UE nevertheless has to measure the channel and report it. Thus, we only increase the payload of the feedback message to include the information of the channel to the FR. Given that each UE's frequency band is statically allocated, the feedback message can be concise. The notification of the FR's next move may only contain a GPS coordinate (including altitude) and be sent once per several seconds.

The location of the FR greatly affects the channel gains of the back-haul and all the front-hauls, and thus the maximum system throughput varies. Our objective, finding the optimal location of the FR, can be formulated as a primal optimization problem (P1) with three constraints as follows,

$$\max_{\boldsymbol{\alpha},\boldsymbol{x}} \sum_{i\in\mathbb{U}} \left[C(||\boldsymbol{x}-\boldsymbol{u}_i||_2,\alpha_i)+C(||\boldsymbol{s}_i-\boldsymbol{u}_i||_2,P_t-\alpha_i)\right], \tag{P1}$$

$$\textbf{s.t.} \sum_{i\in\mathbb{U}} C(||\boldsymbol{x}-\boldsymbol{u}_i||_2,\alpha_i) \le C'(||\boldsymbol{x}-\boldsymbol{s}_x||_2,P_t^F,\beta), \tag{C1}$$

$$\alpha_i \le P_t, \quad \forall i \in \mathbb{U}, \tag{C2}$$

$$\alpha_i \ge 0, \quad \forall i \in \mathbb{U}, \tag{C3}$$

where $\boldsymbol{x}$ is the 3-D coordinate vector (location) of the FR, $\boldsymbol{u}_i$ is the location of the ith UE, $\boldsymbol{s}_i$ and $\boldsymbol{s}_x$ are the locations of the serving BSs selected by the ith UE and the FR, respectively. α_i denotes the transmission power of the ith UE that is allocated to the FR. P_t is the total UE transmission power and P_t^F is that of the FR. $\mathbb{U}$ is the UE set. In this chapter, we use the bold font denoting vectors, the double struck font denoting sets.

The constraint (C1) is the back-haul capacity limit, which means that the sum of front-hauls' capacities cannot exceed the back-haul's. β denotes that the bandwidth allocated to the FR's back-haul is β times of UE's bandwidth and $C'(d, p, \beta)$ is given by

$$C'(d, p, \beta) = \beta W \log_2(1 + \frac{p \cdot g(d)}{\beta N_0}). \tag{4.2}$$

(C2) and (C3) determine the valid range of the power allocated to the FR, which is non-negative and less than the maximum UE transmission power.

The global information is needed to make the decision for the FR's future moves. Since the traffic load and the topology of active nodes are changing (dynamical network), the global information is time-varying. Therefore, the FR may make the decision for either one step ahead given the unpredictable network or several steps ahead given the predictable network, i.e., the network is stable in a certain duration.

In the following sections, (P1) is decoupled and solved case by case. Given a fixed location of the FR, we first develop the optimal power allocation strategy, considering both the back-haul capacity constraint and maximum UE transmission power constraint, to maximize the system throughput. Next, if the network is unpredictable, given the optimal power allocation of each fixed location of the FR, we calculate the direction of the FR's next move. The near-optimal location for the FR can be found by a proposed on-line iterative algorithm. If the network is predictable, given

the stable duration, we find out the near-optimal location within the FR's maximum moving range by a proposed searching algorithm with adaptive step length. Then, the FR move to the converged near-optimal location directly.

4.4 Optimal Power Allocation

In this section, given the location of the FR, we obtain the optimal power allocation strategy. The problem (P1) is changed to (P2) as follows,

$$\max_{\boldsymbol{\alpha}} \ \sum_{i\in\mathbb{U}} \left[C(d_i,\alpha_i)+C(D_i,P_t-\alpha_i)\right], \tag{P2}$$

$$\text{s.t.} \ \sum_{i\in\mathbb{U}} C(d_i,\alpha_i) \le C'(D_x,P_t^F,\beta), \tag{C1}$$

$$\alpha_i \le P_t, \quad \forall i\in\mathbb{U}, \tag{C2}$$

$$\alpha_i \ge 0, \quad \forall i\in\mathbb{U}, \tag{C3}$$

where d_i and D_i are the distances of the ith UE to the FR and its serving BS, respectively. D_x is the distance from the FR to its serving BS. All the distances are known given the fixed location of the FR. Both the objective function and constraints in (P2) are convex [80].

To solve the problem, we consider different cases and solve them one by one. The constraint (C1) can be separated into two cases, which are back-haul unlimited and back-haul limited, respectively. In the first case, (C1) can be removed; in the second case, (C1) becomes (C1') as shown below.

$$\sum_{i\in\mathbb{U}} C(d_i,\alpha_i) = C'(D_x,P_t^F,\beta) \tag{C1'}$$

We first focus on the back-haul unlimited case. The constraints (C2) and (C3)

define the valid range of the power allocation. α_i with and without these two constraints may result in different front-haul capacities and thus affect the back-haul budget for others. However, given that the back-haul is unlimited, the data rate of each UE's front-haul is independent to others. Thus, the power allocations for UEs are not correlated, and we can find the optimal allocation for each UE individually using (P2'), where the summation in (P2) is removed, subject to (C2) and (C3).

$$\begin{aligned} &\max_{\alpha_i} \ C(d_i, \alpha_i) + C(D_i, P_t - \alpha_i), \quad \forall i \in \mathbb{U}. \\ &\textbf{s.t.} \ (\text{C2}), (\text{C3}) \end{aligned} \tag{P2'}$$

(P2') can be considered as the sum-rate maximization problem for two channel states with sum-power constraint. It is solved as follows. Substituting with (4.1), the equivalent expression of (P2') is given by

$$\begin{aligned} &\max_{\alpha_i} \ W \log_2(1 + H_{di}\alpha_i) + W \log_2(1 + H_{Di}(P_t - \alpha_i)) \\ \Longrightarrow &\max_{\alpha_i} \ \ln(1 + H_{di}\alpha_i) + \ln(1 + H_{Di}(P_t - \alpha_i)), \ \forall i \in \mathbb{U}. \end{aligned} \tag{4.3}$$

Taking a derivative of (4.3) w.r.t. α_i, and making it equal to zero, we have

$$\begin{aligned} &\frac{H_{di}}{1 + H_{di}\alpha_i} - \frac{H_{Di}}{1 + H_{Di}(P_t - \alpha_i)} = 0 \\ \Longrightarrow &\alpha_i = \frac{H_{di} - H_{Di} + H_{di}H_{Di}P_t}{2H_{di}H_{Di}}, \ \forall i \in \mathbb{U}. \end{aligned} \tag{4.4}$$

The power allocation α_i in (4.4) is optimal from a UE's perspective. We use α_i^* to denote the optimal power allocation in (P2), $\alpha_i^* = \alpha_i, \ \forall i \in \mathbb{U}$, if the constraints (C1), (C2) and (C3) are all satisfied.

If (C2) and (C3) are not satisfied, we should first draw each $\alpha_i, \forall i \in \mathbb{U}$, back into

the valid range, i.e., force the negative ones equal to zero and the ones larger than P_t equal to P_t, respectively. Due to the convexity of (P2'), if α_i calculated by (4.4) is negative, the maximum system throughput will be monotonically decreased when we increase α_i. Similarly, if we reduce α_i, which is originally larger than P_t, the maximum system throughput is also monotonically decreased. It can be concluded that if the ith UE's power allocation calculated by (4.4), α_i, is out of the valid range defined by (C2) and (C3), its value in the optimal solution of (P2'), denoted by α_i', is on the nearest boundary of the valid range, i.e., zero or P_t. Therefore, $\alpha_i' = \min[(|\alpha_i| + \alpha_i)/2, P_t]$.

Because UEs are not correlated with each other in this unconstrained problem (P2'), the above operation to α_i' will not affect other's optimal solution. After this process, (C2) and (C3) are satisfied. If (C1) is also satisfied, we conclude that this power allocation is the optimal solution to (P2) and $\alpha_i^* = \alpha_i',\ \forall i \in \mathbb{U}$.

If (C1) is not satisfied, it means that the back-haul capacity limit is violated. Due to the convexity of (P2), the optimal solution appears when the back-haul capacity is fully utilized by the front-hauls and thus (C1') is used to replace (C1). For this back-haul limited case, (P2) is converted to (P2") as follows,

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \ & \sum_{i \in \mathbb{U}} \left[C(d_i, \alpha_i) + C(D_i, P_t - \alpha_i) \right], \qquad \text{(P2")} \\ \text{s.t. } & (\text{C1}'), (\text{C2}), (\text{C3}) \end{aligned}$$

Substituting with (4.1), the Lagrange function is given by

$$\begin{aligned}\mathcal{L} = &\sum_{i\in\mathbb{U}}\left[W\log_2(1+H_{di}\alpha_i)+W\log_2(1+H_{Di}(P_t-\alpha_i))\right]\\ &+\lambda_0\left[\sum_{i\in\mathbb{U}}W\log_2(1+H_{di}\alpha_i)-C'(D_x,P_t^F,\beta)\right]\\ &+\sum_{i\in\mathbb{U}}\mu_i\alpha_i+\sum_{i\in\mathbb{U}}\mu_{i+M}(P-\alpha_i).\end{aligned} \tag{4.5}$$

Taking a derivative of (4.5) w.r.t. α_i, and making it equal to zero, we have

$$(\lambda_0+1)\frac{H_{di}}{H_{di}\alpha_i+1}-\frac{H_{Di}}{H_{Di}(P_t-\alpha_i)+1}+\mu_i-\mu_{i+M}=0. \tag{4.6}$$

According to the Karush-Kuhn-Tucker (KKT) conditions,

$$\begin{aligned}\mu_i\alpha_i &= 0,\ \ \forall i\in\mathbb{U}\\ \mu_{i+M}(P_t-\alpha_i) &= 0,\ \ \forall i\in\mathbb{U}\\ \mu_i &\geq 0,\ \ \forall i\in\mathbb{U}\\ \mu_{i+M} &\geq 0,\ \ \forall i\in\mathbb{U}.\end{aligned} \tag{4.7}$$

Therefore, when $\mu_{i+M}=0,\alpha_i=0\Longrightarrow\mu_i\geq 0$, the range of λ_0 can be found as follows,

$$\begin{aligned}&\left[(\lambda_0+1)\frac{H_{di}}{H_{di}\alpha_i+1}-\frac{H_{Di}}{H_{Di}(P_t-\alpha_i)+1}\right]_{\alpha_i=0}\leq 0\\ &\Longrightarrow\lambda_0\leq\frac{H_{Di}}{(H_{Di}P_t+1)H_{di}}-1.\end{aligned} \tag{4.8}$$

Similarly, when $\mu_i = 0, \alpha_i = P_t \Longrightarrow \mu_{i+M} \geq 0$, the range of λ_0 can be obtained by

$$\begin{aligned}\left[(\lambda_0+1)\frac{H_{di}}{H_{di}\alpha_i+1} - \frac{H_{Di}}{H_{Di}(P_t-\alpha_i)+1}\right]_{\alpha_i=P} &\geq 0 \\ \Longrightarrow \lambda_0 \geq \frac{(H_{di}P_t+1)H_{Di}}{H_{di}} - 1. &\end{aligned} \tag{4.9}$$

When $\mu_i = \mu_{i+M} = 0 \Longrightarrow \alpha_i \in [0, P_t]$, α_i can be written as a function of λ_0 as follows,

$$\begin{aligned}&(\lambda_0+1)\frac{H_{di}}{H_{di}\alpha_i+1} - \frac{H_{Di}}{H_{Di}(P_t-\alpha_i)+1} = 0 \\ =&\alpha_i H_{di}H_{Di}(\lambda_0+2) + H_{Di} - (\lambda_0+1)(H_{Di}P_t+1)H_{di} \\ \Longrightarrow &\alpha_i = \frac{(\lambda_0+1)(H_{Di}P_t+1)H_{di} - H_{Di}}{H_{di}H_{Di}(\lambda_0+2)}.\end{aligned} \tag{4.10}$$

The range of λ_0 are determined by (4.8) and (4.9) to ensure that α_i is on the boundary of the valid range, i.e., α_i equals either 0 or P_t. We assume that the total number of UEs in $\mathbb{U}$ is M and the value range of λ_0 can be divided into $2M+1$ intervals by M UEs' boundaries. For the jth interval of λ_0, $\mathbb{R}_j$, we define the UEs that satisfy (4.8) belong to $\mathbb{U}_j^0$, and those satisfying (4.9) belong to $\mathbb{U}_j^P$.

When $\lambda_0 \in \mathbb{R}_j$, $\alpha_i, \forall i \in \mathbb{U}_j^0 \cup \mathbb{U}_j^P$, can be fixed as either 0 or P_t. The following constraints are used to replace the constraints (C2) and (C3) in (P2").

$$\begin{aligned}\alpha_i &= 0, \forall i \in \mathbb{U}_j^0, \\ \alpha_i &= P_t, \forall i \in \mathbb{U}_j^P.\end{aligned} \tag{4.11}$$

The back-haul capacity should also be updated because it is deducted by the UEs

whose power are fixed to P_t. The updated constraint (C1') is given by

$$\sum_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}} W\log_2(1+H_{di}\alpha_i) = C'(D_x, P_t^F, \beta) - \sum_{i\in\mathbb{U}_j^P} W\log_2(1+H_{di}P_t). \quad (4.12)$$

We use (P2$_j$") to denote the new problem where the constraints in (P2") are updated as above.

Theorem 1. *The global maximizer of (P2$_j$") is given by*

$$\begin{aligned}
&\alpha_i = \frac{A(H_{Di}P_t+1)H_{di} - H_{Di}}{H_{di}H_{Di}}, \forall i \in \{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\},\\
&\alpha_i = 0, \forall i \in \mathbb{U}_j^0,\\
&\alpha_i = P_t, \forall i \in \mathbb{U}_j^P,
\end{aligned} \quad (4.13)$$

where

$$A = \left[\left(\frac{2^{\frac{C'(D_x,P_t^F,\beta)-\sum_{i\in\mathbb{U}_j^P} W\log_2(1+H_{di}P_t)}{W}}}{\prod_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}} \frac{(H_{di}+H_{Di}+H_{di}H_{Di}P_t)}{H_{Di}}}\right)^{\frac{1}{M_j'}}\right]^+,$$

and M_j' is the total number of UEs in $\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}$.

Proof. For $\alpha_i, \forall i \in \{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}$, it can be represented by (4.10). Substituting (4.10) into the updated constraint (4.12), we have

$$\begin{aligned}
\prod_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}} [1+H_{di}\alpha_i] &= \prod_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}} \frac{(\lambda_0+1)(H_{di}+H_{Di}+H_{di}H_{Di}P_t)}{(\lambda_0+2)H_{Di}}\\
&= 2^{\frac{C'(D_x,P_t^F,\beta)-\sum_{i\in\mathbb{U}_j^P} W\log_2(1+H_{di}P_t)}{W}}.
\end{aligned} \quad (4.14)$$

Given the total number of UEs in $\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}$, M_j', (4.14) can be rewritten as

$$\left(\frac{\lambda_0+1}{\lambda_0+2}\right)^{M_j'}\prod_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}}\frac{(H_{di}+H_{Di}+H_{di}H_{Di}P_t)}{H_{Di}}=2^{\frac{C'(D_x,P_t^F,\beta)-\sum_{i\in\mathbb{U}_j^P}W\log_2(1+H_{di}P_t)}{W}}. \tag{4.15}$$

$$\text{Define } A=\left[\left(\frac{2^{\frac{C'(D_x,P_t^F,\beta)-\sum_{i\in\mathbb{U}_j^P}W\log_2(1+H_{di}P_t)}{W}}}{\prod_{i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}}\frac{(H_{di}+H_{Di}+H_{di}H_{Di}P_t)}{H_{Di}}}\right)^{\frac{1}{M_j'}}\right]^+=\frac{\lambda_0+1}{\lambda_0+2}. \tag{4.16}$$

Therefore, λ_0 can be obtained as follows,

$$\lambda_0=\frac{2A-1}{1-A}. \tag{4.17}$$

By substituting the above result into (4.10) and combining with new constraints in (4.11), the optimal power allocation of (P2$_j$") can be obtained. □

Given the optimal power allocation in Theorem 1, the system throughput assuming that λ_0 belongs to the jth interval, denoted by $C^{(j)}$, can be calculated by substituting the power allocation into the objective function of (P2").

Given (4.17), we need to further check whether $\lambda_0\in\mathbb{R}_j$. If λ_0 is out of the range of $\mathbb{R}_j$, the result contradicts with the settings of $\mathbb{U}_j^0$ and $\mathbb{U}_j^P$, so that there is no KKT point exists and λ_0 corresponding to the global maximizer of (P2") cannot be in $\mathbb{R}_j$. We set $C^{(j)}=0$ for this case. It should be noticed that for some of the intervals, $\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P=\phi$ and $C'(D_x,P_t^F,\beta)-\sum_{i\in\mathbb{U}_j^P}W\log_2(1+H_{di}P_t)>0$, or the deducted back-haul capacity is negative, or $\alpha_i\in(-\infty,0]\cup[P_t,\infty), i\in\{\mathbb{U}-\mathbb{U}_j^0\cup\mathbb{U}_j^P\}$. These cases also mean that the KKT point does not exist in the current interval and we set $C^{(j)}=0$ for these cases.

At most one KKT point can be found in each interval. If we add a minus to the objective function of (P2"), the Hessian matrix is positive definite. Thus, all the KKT points of (P2") are either one of the local maximizers or the global maximizer. The KKT point corresponding to the maximum system throughput is the global maximizer. After we obtain $C^{(j)}$ for all the intervals, where $j = 1, 2, ..., 2M + 1$, the maximum $C^{(j)}$ among all the intervals is the maximum system throughput of (P2").

Remark 1. *The global maximizer of (P2") is obtained as follows,*

$$\boldsymbol{\alpha}'' = \arg\max_{\boldsymbol{\alpha}} \left[C^{(1)}, C^{(2)}, ..., C^{(2M+1)}\right]. \tag{4.18}$$

Remark 2. *The procedures of obtaining the optimal power allocation of (P2),* $\boldsymbol{\alpha}^*$*, given the location of the FR, can be summarized as:*

1. *Solve (P2') for each UE. Let* $\alpha_i' = \min[(|\alpha_i| + \alpha_i)/2, P_t], \forall i \in \mathbb{U}$.
2. *Check whether the constraint (C1) is satisfied by* $\boldsymbol{\alpha}'$.
   - *If so,* $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}'$;
   - *Otherwise, solve (P2$_j$"),* $j = 1, 2, ..., 2M + 1$*, according to Theorem 1 and then obtain* $\boldsymbol{\alpha}''$ *based on Remark 1.* $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}''$.

4.5 3-D Placement Algorithm

4.5.1 Weighted Coordinate Axes (WCA)

In this subsection, the unpredictable network is assumed, i.e., we have no knowledge about how drastically and frequently the network changes. The FR's optimal location within 10-meter and 100-meter moving ranges may not be on the same path. Only the current information is applied to seek for a temporarily better location. Based on

this instantaneous global information and the optimal power allocation given a fixed FR's location, we design the algorithms to determine the direction of the FR's next move.

We propose the Weighted Coordinate Axes (WCA) algorithm inspired by the gradient ascent method. When the topology is stable, the FR shall move to a local maximizer as soon as possible. The gradient ascent method is a simple but effective way to find the local optimal point. The new location can be obtained based on the previous location as follows,

$$\boldsymbol{x}' = \boldsymbol{x} + R\nabla Z(\boldsymbol{x}), \tag{4.19}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the locations of the FR in the current step and the next step, respectively. R is the length of the step. $Z(\boldsymbol{x})$ denotes the maximum system throughput given the FR's location $\boldsymbol{x}$, which is obtained in Sec. 4.4.

Given that the gradient of $Z(\boldsymbol{x})$ is difficult to derive, we use numerical method to approximate the gradient of each step. It is first assumed that the FR moves along one of the 3-D axes toward positive and negative infinity for a short distance, Δ, respectively. The system throughput increments by these movements can be obtained. The difference between the throughput increments on two directions can be considered as the weight of this axis. The weighted sum of all the axes is the approximation of the gradient at the current location and results in the direction for the next step.

In a 3-D coordinate system, the weight of the x axis, w_1, is given by

$$w_1 = Z(\boldsymbol{x} + \Delta[1\ 0\ 0]) - Z(\boldsymbol{x} - \Delta[1\ 0\ 0]). \tag{4.20}$$

Similarly, w_2 and w_3 can be obtained. The direction of the FR's next step is given

by $\boldsymbol{d} = [w_1\ w_2\ w_3]$.

A weight can be either positive or negative. A negative weight means that the system throughput is decreasing in that axis' positive direction. Based on the direction, $\boldsymbol{d}$, and the fixed step length, R, the coordinate of the FR in the next step, $\boldsymbol{x}'$, is obtained by

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{R\boldsymbol{d}}{||\boldsymbol{d}||}. \tag{4.21}$$

The WCA algorithm can be extended to other cases. For example, in practice, the FR may be restricted to a certain height due to safety concerns. In this case, the dimension is reduced to 2-D. By applying the same method shown above but making $w_3 = 0$ in $\boldsymbol{d}$, the next step, $\boldsymbol{x}'$, can be obtained. When No-fly zones applied in practice, (4.21) should be further checked for each step to ensure $\boldsymbol{x}'$ is in the feasible region. In order to avoid zigzagging flight path, additional conditions can be added to stop the WCA algorithm. One feasible way is to calculate the throughput increment for each direction. For example, $I_1^+ = Z(\boldsymbol{x}+\Delta[1\ 0\ 0]) - Z(\boldsymbol{x})$ and $I_1^- = Z(\boldsymbol{x}-\Delta[1\ 0\ 0]) - Z(\boldsymbol{x})$ are the increments of the positive and negative directions on the x axis, respectively. If all the six increments along three axes are negative, the current location is almost on a local maximizer and the WCA algorithm stops.

4.5.2 Adaptive Weighted Coordinate Axes (A-WCA)

Besides the study on unpredictable network, we propose the A-WCA algorithm to deal with the predictable network, i.e., the network keeps stable for a certain time and this duration is predictable. Given the stable duration and the moving speed of the FR, the maximum number of steps and moving distance can be obtained. The objective of the A-WCA algorithm is to find the optimal location within a certain

distance or a given number of steps.

Different from the one-step optimization in WCA, the FR moves to the target location directly after the A-WCA calculation. It means the FR does not move while the A-WCA's searching process. Thus, without the constraint of the FR's moving speed, an adaptive step length can be adopted to improve the efficiency, as shown in Fig. 4.2. Given the maximum moving distance, the minimum distance from the location of the ith step to the boundary, denoted as $S_i, i = 1, 2, ...$, can be obtained as follows,

$$S_i = \Phi - ||\boldsymbol{x}_{i-1} - \boldsymbol{x}_0||, \forall i > 0, \tag{4.22}$$

where Φ is the FR's maximum moving distance within the predicted network stable duration. $\boldsymbol{x}_i$ is the location at the ith step and $\boldsymbol{x}_0$ is the initial location.

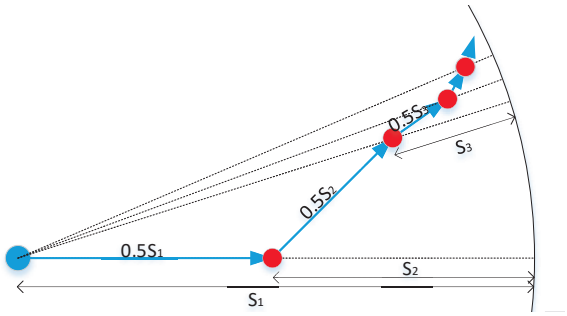


Figure 4.2: An example of the adaptive step length.

The step length of the ith step is given by $R_i = S_i/2$. Specifically, $R_1 = S_1/2 = \Phi/2$. The coordinate updating equation in searching process is given by

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \frac{(\Phi - ||\boldsymbol{x}_{i-1} - \boldsymbol{x}_0||)\boldsymbol{d}}{2||\boldsymbol{d}||}, \forall i > 0. \tag{4.23}$$

By applying the proposed A-WCA algorithm, the FR is guaranteed to stay within a sphere with radius Φ. The searching process will end when $R_i < \Gamma$, where Γ is a

predetermined threshold.

If the A-WCA algorithm targets on a point on the spherical boundary, the step length decreases exponentially and the searching process stops quickly. However, when the algorithm targets on a local maximizer within the sphere, the A-WCA may never stop due to its fluctuating step length. We thereby need a mechanism to reduce the step length in this case. Similar to the pocket algorithm, i.e., keep the best result in the pocket and replace it once a better one appears, we maintain the minimum distance from the searched location to the sphere boundary in the pocket during the searching process. Whenever there is a smaller distance generated by the new location, the distance in the pocket will be replaced. The step length is reduced by half if the pocket has not been updated for successive Λ steps. The detailed procedure of the A-WCA algorithm is summarized in Algorithm 1.

Algorithm 1 The A-WCA Algorithm

procedure A-WCA SEARCH
    Initialization:
    $i = 0$, $R_0 = \frac{\Phi}{2}$, Counter= 0, Pocket= Φ, F=0
    while $R_i \geq \Gamma$ **do**
        $i = i + 1$
        Update S_i and $\boldsymbol{x}_i$ according to (4.22) and (4.23)
        if F=0 **then**
            $R_i = \frac{S_i}{2}$
        end if
        if R_i <Pocket **then**
            Pocket=R_i, Counter=0
        else
            Counter=Counter+1
        end if
        if Counter$\geq \Lambda$ **then**
            $R_i = \frac{R_i}{2}$, F=1
        end if
    end while
end procedure

4.5.3 Direct Method

To examine the WCA algorithm in terms of the convergence speed, we compare it with an off-line method. In this off-line method, we first obtain the location corresponding to the highest system throughput using the WCA algorithm within a certain steps, and then let the FR move to it directly along a straight line.

In practice, the direct off-line method is difficult to apply because it significantly increases the computation load of the FR and the long-term prediction may not be accurate in dynamical networks. In the following performance analysis, we consider this method as an upper bound of the convergence speed for the WCA algorithm.

4.5.4 Dynamic Programming Method

In unpredictable networks, we want to examine how much accumulated throughput can be achieved before the next change. The accumulated throughput is given by $Acc(n) = \sum_{f=0}^{n} \sum_{i \in \mathbb{U}} \left[C(||\boldsymbol{x}^f - \boldsymbol{u}_i||_2, \alpha_i^f) + C(||\boldsymbol{s}_i - \boldsymbol{u}_i||_2, P_t - \alpha_i^f) \right]$, where $\boldsymbol{x}^f$ and α_i^f denote the location of the FR and the ith UE's power allocation in the fth step, respectively. The accumulated throughput includes every steps along the flight path starting from the 0th step (the initial location) to the nth step (the current location). For this metric, the Dynamic Programming (DP) is applied as a benchmark for comparison.

The DP is well known as a tool for the route optimization that provides the maximum utility [81]. However, it is suitable for the problems with discrete routing nodes. In this chapter, the movements of the FR are not restricted between discrete points, which means that the number of routing nodes is infinite. To simplify the problem, we first build up a cuboid 3-D grid where the straight line connecting the initial location of the FR and the target location is the body diagonal.

We use $Z_{i,j,k}$ to denote the system throughput of the routing nodes (i, j, k) on the

3-D grid, which can be calculated as explained in Sec. 4.4. The routing nodes start from the node $(0,0,0)$, the initial location of the FR, and end at the target (I,J,K). There are $(I+1)(J+1)(K+1)$ nodes in total on the 3-D grid. $U^*_{i,j,k}$ denotes the maximum accumulated system throughput from $(0,0,0)$ to (i,j,k). The recursive equation is given by

$$U^*_{i,j,k} = \mathbf{max}\left[U^*_{i-1,j,k}, U^*_{i,j-1,k}, U^*_{i,j,k-1}\right] + Z_{i,j,k}, \forall i \in [1,I], \forall j \in [1,J], \forall k \in [1,K]. \tag{4.24}$$

When any one or two variables in (i,j,k) equal zero, the recursive equation will be modified. For example, if $i=0$,

$$U^*_{0,j,k} = \mathbf{max}\left[U^*_{0,j-1,k}, U^*_{0,j,k-1}\right] + Z_{i,j,k}, \forall j \in [1,J], \forall k \in [1,K]. \tag{4.25}$$

If $i=j=0$,

$$U^*_{0,0,k} = U^*_{0,0,k-1} + Z_{i,j,k}, \forall k \in [1,K]. \tag{4.26}$$

Other cases are similar. Besides that, we have $U^*_{0,0,0} = Z_{0,0,0}$.

Given the recursive equations above, the maximum accumulated system throughput at each routing node on the 3-D grid can be obtained. By the backward search from the target location and pick up the previous node corresponding to the maximum accumulated system throughput one by one, the optimal route is obtained. This route is denoted by a set of routing nodes' coordinates that it passes by, $\mathbb{H}$: $\{\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_{I+J+K}\}$. $\boldsymbol{h}_1$ is the initial location of the FR and $\boldsymbol{h}_{I+J+K}$ is the destination. Every component in vector $\boldsymbol{h}_{I+J+K} - \boldsymbol{h}_s$ is equal to or greater than that in $\boldsymbol{h}_{I+J+K} - \boldsymbol{h}_{s+1}, \forall s \in [1, I+J+K-1]$.

The route $\mathbb{H}$ goes along the 3-D grid and the distance between two adjacent routing

nodes in $\mathbb{H}$ may not be the same as the step length of the FR. Therefore, we use the current location of the FR, $\boldsymbol{x}$, as the center to form a sphere, with a radius of step length R. Several intersections of this sphere and route $\mathbb{H}$ can be obtained and the one with the shortest distance to the destination will be the location of the FR's next step. This route is considered as the Smoothed DP (SDP) route. A simple example in the 2-D plane can be found in Fig. 4.3.

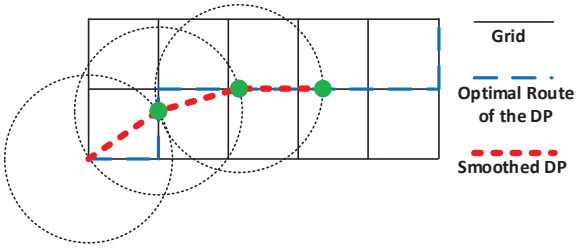


Figure 4.3: An example the smoothed DP route.

4.6 Performance Analysis

4.6.1 One-snapshot Simulation

A snapshot of random UE distribution (8 UEs) is shown in Fig. 4.4, the flight path of the FR obtained by the WCA algorithm is marked. It can be observed that after several steps of moving, the FR converges to one point. The flight path is close to that of the direct off-line method, as shown in the figure.

In this simulation, the bandwidth for each UE is $W = 1$MHz, about 6 Physical Resource Blocks (PRB) in the LTE system. The bandwidth for the FR is 5 times of UE's, i.e., $\beta = 5$. The maximum transmission powers are $P_t = 200$mW and $P_t^F = 1000$mW for the UE and the FR, respectively. The noise spectrum density is $n_0 = 10^{-17.4}$mW/Hz and $N_0 = Wn_0$. The NLOS channel in [46] is adopted, where $r = 3.68$, $K_0 = 10^{4.38}$. We do not consider the random shadowing in this case, thus $\Psi = 1$. The reference distance is $d_0 = 10$m. Four fixed BSs' coordinates are

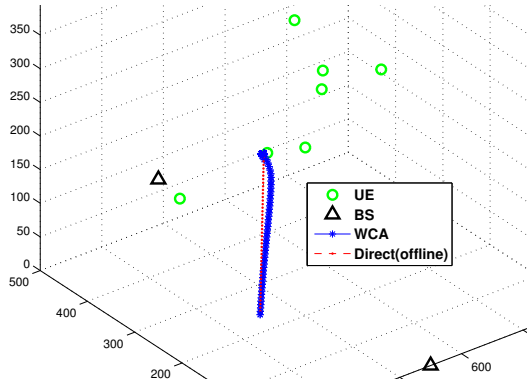
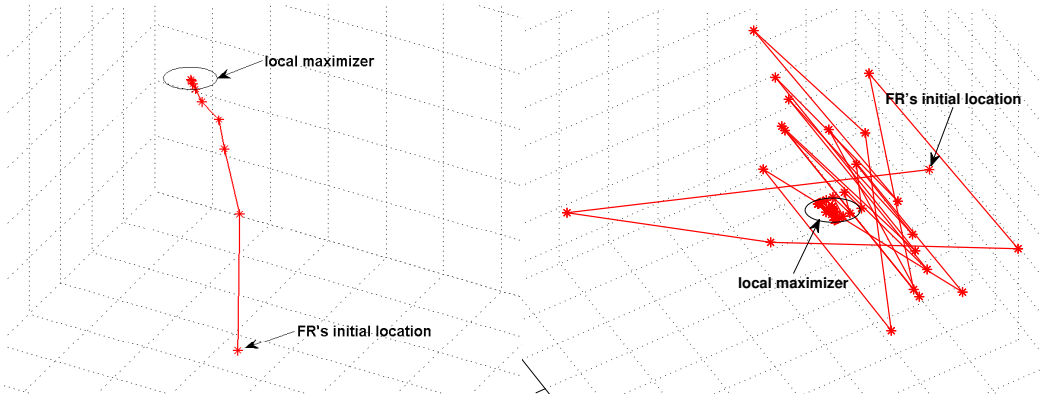


Figure 4.4: An example of the FR's movement based on the WCA algorithm.

[500 500 500], [0 250 250], [500 0 0] and [1000 0 250]. Each axis of a UE's coordinate is randomly selected within $(0, 500)$. The initial location of the FR follows the same distribution. Because of the complexity of the radio bearer reconfiguration for the relay node, the FR cannot switch a serving BS dynamically. Therefore, in this chapter, we assume a fixed serving BS for the FR, and the FR selects the nearest BS (based on FR's initial location) as its serving BS in the following simulations.

Snapshots of the A-WCA's searching process given a stable duration of 30 seconds are shown in Figs. 4.5. The FR needs to find the optimal location within its maximum moving distance which is 30 meters assuming 1m/s moving speed.



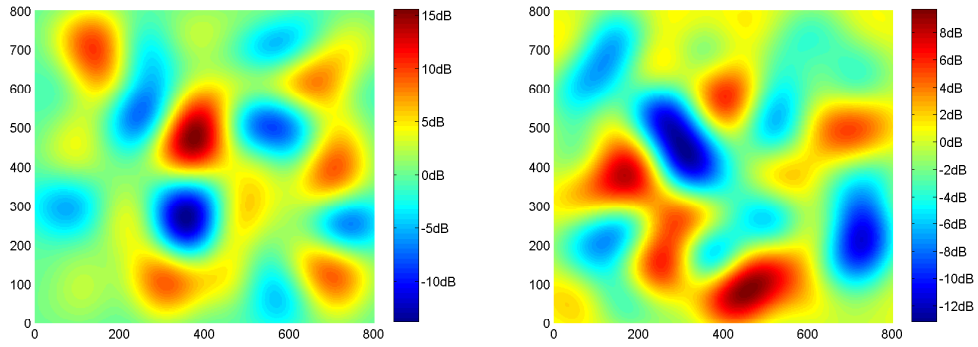
(a) Local optimal on the boundary (b) Local optimal inside the feasible region

Figure 4.5: Examples of the A-WCA's searching process.

Fig. 4.5 (a) is a typical case where the network's stable duration is relatively short

and the FR has a limited moving distance. In this case, the FR has a higher chance to target on a maximizer outside the moving range, and finally it will stop at a point on the boundary after the A-WCA search. The step length decreases exponentially for most of the time and thus the process stops quickly. We set the threshold, $\Gamma = 0.1$m, for the step length in the simulation. It takes 9 iterations in Fig. 4.5 (a) to finish the searching.

When the stable duration is longer and can support the FR to move further, the FR is more possible to target on a maximizer within the maximum moving range as shown in Fig. 4.5 (b). Though fluctuating at the beginning, the searching trace can still converge to a local maximizer after reducing the step length by half several times compulsively according to Algorithm 1. Given $\Lambda = 5$ in the simulation, it takes 55 iterations in Fig. 4.5 (b) to finish the searching. It should be noticed that the FR will fly to the converged local maximizer directly, rather than moving along the searching trace.

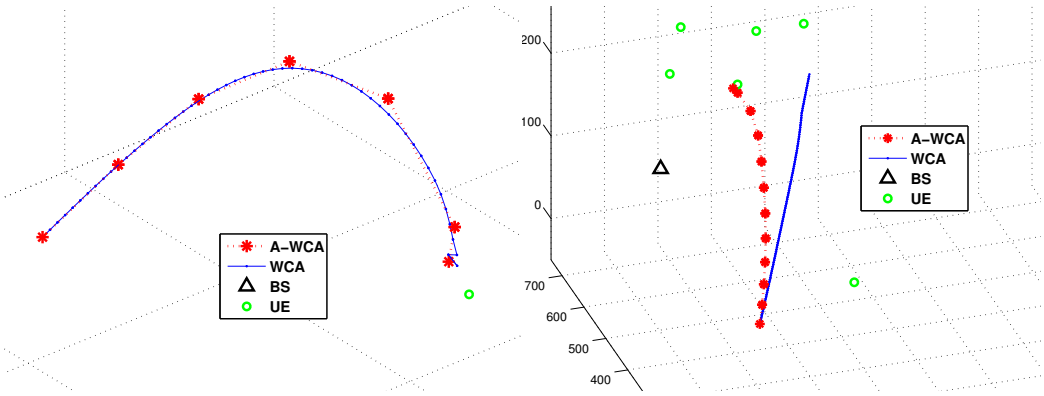


(a) (b)

Figure 4.6: Examples of the random shadowing map.

The NLOS path loss is applied in Fig. 4.4, which can be considered as the homogeneous shadowing case. In order to examine the proposed approaches in a harsher environment, a random shadowing is also adopted. We generate the shadowing map according to [82]. It follows Gaussian distribution in logarithmic scale with a standard

deviation of 4dB. We generate it for each UE because they have fixed locations while the FR is moving in the simulation. Considering the tower blocks in the city area, the shade mostly exists horizontally but not vertically so that the 2-D shadowing map is used in this chapter. UEs' shadowing maps are different, due to their locations, but highly correlated because they are blocked by the same building. We use two extreme cases, i.e., all the UEs share the identical shadowing map and each UE has an independent shadowing map, in the following simulations.



(a) Identical shadowing (b) Independent shadowing

Figure 4.7: Examples of the FR's movement based on the WCA and the A-WCA algorithms.

Given a predicted 30-second network stable duration, we apply the A-WCA algorithm successively for each 30-second period until reach a maximizer inside the feasible region (like the case in Fig. 4.5 (b)) rather than on the boundary. The flight paths of the A-WCA and WCA algorithms are marked in Figs. 4.7 (a) and (b) for the identical and the independent shadowing, respectively. Mostly, the two paths are consistent with each other in trend when using identical shadowing map, but they are likely to deviate and end up at different locations in the case of independent shadowing. $Z(\boldsymbol{x})$ becomes very rugged if users' shadowing are independent. Thus, the A-WCA algorithms is likely to climb toward a maximizer different from the WCA's target because of its long step length at the beginning of the searching process.

Convergence Speed

In Fig. 4.8, the system throughput on each step according to the flight path in Fig. 4.4 is shown and the comparison between the WCA algorithm and the direct off-line method are made for different step length settings. As shown in the figure, we choose $R = 3, 5, 10$, respectively. With the decreasing of the step length R, the gap between these two methods is slightly increased, but they are still very close to each other, which means that the convergence speed of the WCA algorithm is almost the same with that of the upper-bound (direct off-line method).

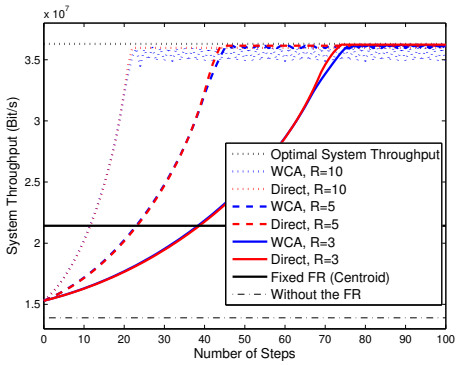
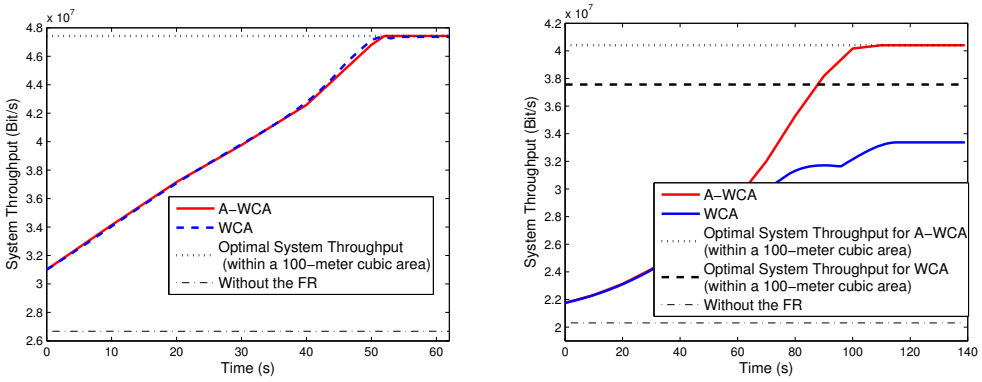


Figure 4.8: Comparisons between the WCA and the optimal results, in terms of the convergence speed and the maximum system throughput (homogeneous shadowing).

A smaller step length will result in more steps to reach the maximum system throughput in Fig. 4.8. However, it will not lead to a longer time consumption in practice. In this simulation, the throughput reaches the maximum at the 79th, 47th and 23th step for the cases of $R = 3$, $R = 5$ and $R = 10$, respectively. Considering the step length, the actual moving distances of the FR are 237, 235 and 230 for the three cases. If the moving speed of the FR is constant, the time consumptions of different step lengths are similar.

The A-WCA's convergence speed is the upper-bound within a certain range because its flight path is a straight line within the maximum moving distance of the stable duration. It can be compared with the WCA if they target on the same maxi-

mizer. When the A-WCA is applied successively, the overall flight path is very close to that of the WCA, as shown in Fig. 4.7 (a). The system capacities along the paths are shown in Fig. 4.9 (a), and they have similar convergence speed. In Fig. 4.9, because of different step lengths, the time elapsing is adopted as the x-axis. In the random shadowing cases we also applied the condition mentioned in Sec. 4.5.1 to avoid the zigzagging flight path. By comparing with Fig. 4.8, we can observe that the fluctuation is mitigated in In Fig. 4.9.



(a) Identical shadowing (b) Independent shadowing

Figure 4.9: Comparisons between the WCA and the A-WCA.

Maximum System Throughput

Besides the convergence speed, we also examine how close the WCA algorithm can get to the optimal location (the optimal system throughput). The optimal system throughput is calculated by the exhaustive search method based on discrete points. All the points are on the intersections of a uniform 3-D grid and the side length of each cell is small enough (0.5m).

It can be observed from Fig. 4.4 that the FR can converge to the vicinity of the optimal location and keep moving around it, which results in the zigzag in throughput curves in Fig. 4.8. We also use a fixed FR located at the centroid of the tetrahedron formed by four existing BSs, which intuitively is a good choice to add a new network

node, as another benchmark. According to Fig. 4.8, this solution is far from what can be achieved by the proposed WCA algorithm.

From Fig. 4.8, when R is small, the FR can get closer to the optimal location and thus the maximum system throughput achieved by the WCA algorithm further approaches the optimal throughput. Based on the discussion above, a smaller step length can make the FR get closer to the optimal location and will not have a big impact on the convergence speed and time consumption, but a smaller step length will lead to a heavier computation load and increase the signaling/feedback load.

The A-WCA algorithm can achieve a better performance than the WCA for most of the time. When they target on the same maximizer, the A-WCA can get closer to it thanks to its smaller searching step length ($\Gamma = 0.1$m in the simulation), comparing with the fixed step length of the WCA ($R = 3$m in the simulation), and thus achieve a slightly higher system throughput as shown in Fig. 4.9 (a). If $Z(\boldsymbol{x})$ is very rugged such as the case of independent shadowing, the WCA may either be trapped by a local maximizer (R is too small) or converge slowly (R is too big). On the contrary, the A-WCA has a higher chance to jump out of a small maximizer and converge quickly to a larger one thanks to its adaptive step length in searching process.

Fig. 4.9 (b) shows the system throughput along the flight paths in Fig. 4.7 (b), and the A-WCA converges to a maximizer with a higher system throughput. Given the difficulty of finding the global optimal location when random shadowing is applied, we investigate whether the location achieved by the proposed algorithm is the global optimal within a 100-meter cubic area. Assuming that the algorithm achieved location is at (a, b, c), the 100-meter cubic area is $([a-50, a+50], [b-50, b+50], [c-50, c+50])$. The granularity of the exhaustive search in a 100-meter cubic area is 0.1m. The figure shows that the A-WCA converges to a global maximizer within a 100-meter cubic area in this one-snapshot simulation but the WCA does not.

Accumulated Throughput

The same parameter settings for the one-snapshot simulation are applied to examine the accumulated throughput achieved by the proposed WCA algorithm. In this section, the direct off-line method is considered as the baseline. We introduce the gain as the metric which is defined as $\text{Gain}(n) = (Acc^{(*)}(n)/Acc_{direct}(n) - 1) \times 100\%$, where $Acc^{(*)}(n)$ denotes the accumulated throughput till the nth step of either the WCA algorithm or the SDP approach, and $Acc_{direct}(n)$ is that of the direct off-line method.

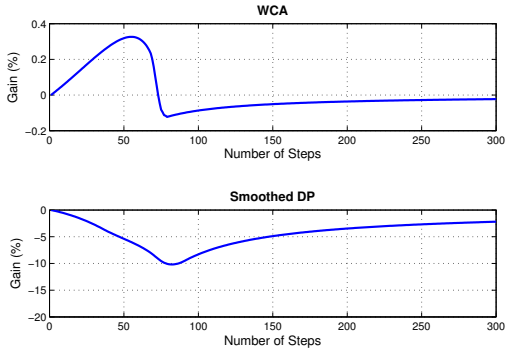


Figure 4.10: Comparisons between the WCA, the smoothed DP, and the direct off-line approaches, in term of accumulated throughput.

In Fig. 4.10, it can be observed that the WCA algorithm can obtain a positive gain at the earlier stage. It is because in each step the FR moves toward the direction with a higher system throughput increment while the direct off-line method ignores the throughput of each step and results in a lower growth rate of the accumulated throughput at the beginning. Later on, the direct off-line method reaches the optimal location sooner than the WCA algorithm, and thus its growth rate will be higher than that of the WCA algorithm. This period corresponds to the sharp decreasing of the gain in Fig. 4.10. After the WCA reaches the peak, the growth rates of the two methods are the same. The gain will slowly converge to zero with the increasing number of steps.

On the contrary, the SDP cannot obtain a positive gain in this particular one-snapshot simulation. The main disadvantage of the DP is that it is originally designed based on the discrete points on the 3-D grid, which makes the total length of the route (flight path) much longer than that of the other two methods. A longer moving distance reduces the growth rate of the accumulated throughput and thus leads to the performance degradation. The one-snapshot simulation is conducted given the homogeneous shadowing. More cases will be considered in Monte Carlo simulations.

We do not investigate the A-WCA's performance in term of the accumulated throughput, because it is only meaningful when the A-WCA and WCA are targeting on the same maximizer and the A-WCA is equivalent to the direct off-line method in this case. Otherwise, it will be more valuable to examine the maximum system throughput, i.e., which one can find a maximizer with a higher system throughput, rather than the accumulated throughput along the path.

4.6.2 Monte Carlo Simulation

Maximum System Throughput

Additional Monte Carlo simulations were conducted to further examine the gap between the maximum value obtained from the proposed algorithms and the optimal result. Four fixed BSs' coordinates are [500 500 500], [500 0 0], [0 500 0] and [0 0 500]. The value of each axis in UE's coordinate is a random variable and follows the uniform distribution within $(0, 500)$. The initial location of the FR follows the same distribution.

In Fig. 4.11, given homogeneous shadowing, the system throughput achieved by the A-WCA and WCA algorithms over the optimal one are shown. We set a 30-second stable duration for the A-WCA algorithm, and different step lengths are applied for the WCA algorithm. The results are the average over 100 rounds of random Monte

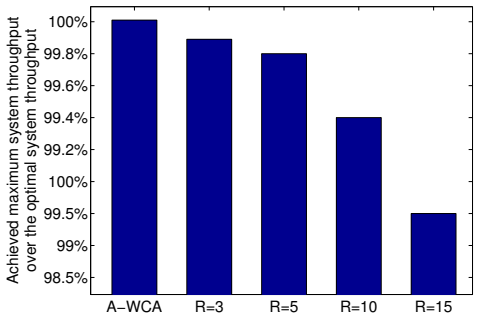
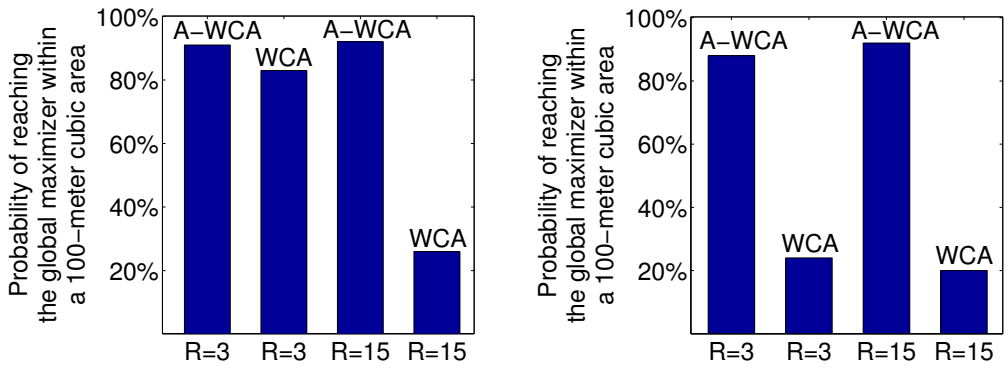


Figure 4.11: Comparisons between the WCA and the optimal system throughput given the homogeneous shadowing.

Carlo simulations. Due to the complexity of the exhaustive search for the optimal system throughput, the number of rounds cannot be too large, but they are sufficient to reveal the trends. Fig. 4.11 shows that by applying the WCA algorithm, the FR can be very close to the optimal location and achieve near-optimal system throughput. With the increasing of the step length, the gap between the WCA algorithm and the optimal throughput will increase. The A-WCA can achieve a better performance thanks to its smaller searching step length ($\Gamma = 0.1$m in the simulation). The A-WCA's result is even higher than the optimal result (the exhaustive search) because of the coarse granularity of the exhaustive search (0.5m).

Both the A-WCA and WCA algorithms can converge to the global optimal location in most of the cases when homogeneous shadowing is applied. However, it is difficult in the random shadowing case due to the ruggedness of $Z(\boldsymbol{x})$. We investigate how possible the proposed algorithms can find a global maximizer within a 100-meter cubic area. We define a successful case as the location obtained by the algorithm is within 5m from the optimal location. The results of 100-round Monte Carlo simulations are summarized in Fig. 4.12.

Fig. 4.12 (a) is the identical shadowing case. We apply different step length R for the WCA algorithm and assume a $10R$ maximum moving distance for the A-WCA. The A-WCA is applied successively until reach a maximizer inside the feasible region



(a) Identical shadowing (b) Independent shadowing

Figure 4.12: Comparisons between the WCA and the A-WCA.

rather than on the boundary. We can observe that the A-WCA can always achieve a better performance thanks to its adaptive step length of the searching process. It allows the A-WCA to jump out of a local maximizer and then converge quickly to a better one. This observation is consistent with that of the one-snapshot simulation. Besides, the A-WCA becomes slightly better when we increase the maximum moving distance or the network's stable duration. It is because the initial step length of the searching process is related to the maximum moving range. With a larger step, it will have a lower chance to be trapped in a local maximizer. The WCA algorithm is close to the A-WCA when $R = 3$, but deteriorates significantly when R is larger because it is difficult to converge given a fixed large step length.

In the independent shadowing case, the results of the A-WCA have the same trend but the WCA becomes worse, as shown in Fig. 4.12 (b). Reducing the step length is not very helpful for the WCA algorithm given $Z(\boldsymbol{x})$ is extremely rugged. The FR is easy to be misled by local maximizers and failed to converge to the global optimal location in certain area.

It can be concluded that when $Z(\boldsymbol{x})$ is smooth, such as the homogeneous and identical shadowing cases, the WCA algorithm can achieve a similar performance to the A-WCA's by wisely choosing the step length. The knowledge of the network's

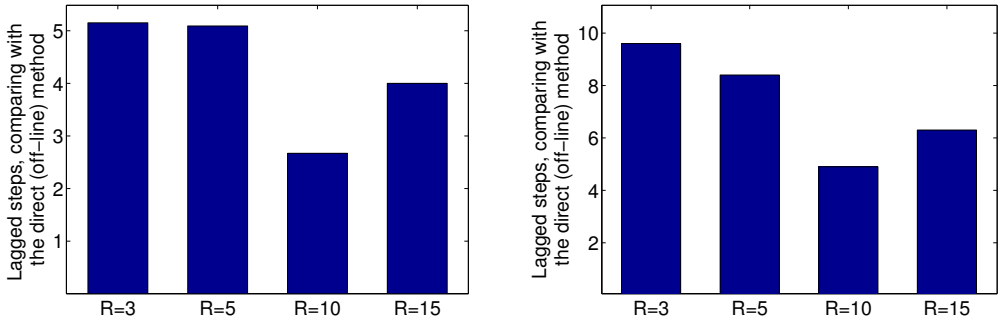
changing will not greatly improve the performance in these scenarios. In the independent shadowing case where $Z(\boldsymbol{x})$ is very rugged, the A-WCA shows a significant advantage. Thus, collecting the statistic information and predicting the network's changing pattern is necessary.

In Monte Carlo simulations, we only study the A-WCA in term of the maximum system throughput, i.e., whether it can find the optimal location within a 100-meter cubic area, especially in the random shadowing case. The convergence speed and the accumulated throughput are meaningful only if the A-WCA and the WCA target on the same maximizer. When the network's stable duration is short, the A-WCA's flight path is close to the WCA's. Otherwise, it gets closer to the direct off-line method. We can gain the insight on the A-WCA from the studies toward the WCA and the direct off-line method. Therefore, we mainly focus on the WCA algorithm in the following Monte Carlo simulations.

Convergence Speed

We use Monte Carlo simulations to verify the observation of convergence speed in the previous one-snapshot simulation. Fig. 4.13 shows how many extra steps are needed on average for the WCA algorithm to reach 99% of the maximum system throughput of the direct off-line method. It should be noticed that the WCA algorithm may reach the peak sooner than the direct off-line method. In these cases, the number of extra steps is set to zero.

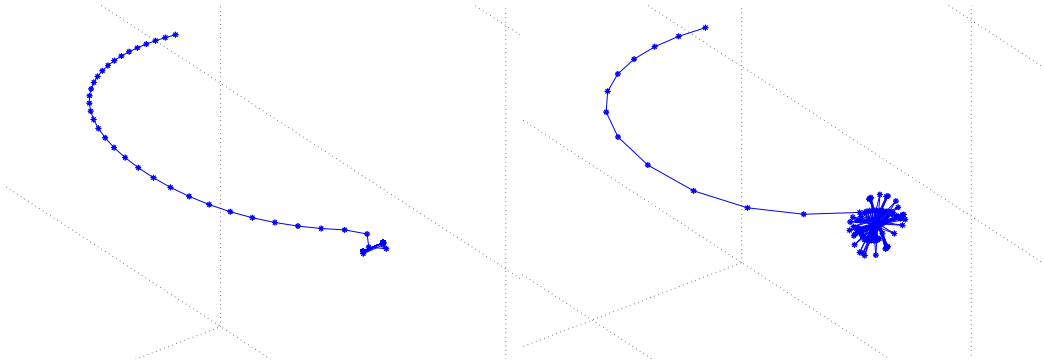
As shown in Fig. 4.13 (a), the WCA algorithm only lags behind the off-line method for several steps given the homogeneous shadowing, which is consistent with the observation in Fig. 4.8. With the increase of step length, the lags decrease in most of the cases. The same trend can be observed in Fig. 4.13 (b), where the independent shadowing is applied. The WCA's flight path becomes circuitous in this case, which



(a) Homogeneous shadowing (b) Independent shadowing

Figure 4.13: Comparisons between the WCA and the direct off-line method, in term of step number needed to reach the peak.

results in larger lags.



(a) $R = 3$ (b) $R = 15$

Figure 4.14: An example of the WCA's flight path when $R = 3, 15$.

In both of the cases, when the step length is too big ($R = 15$ in the figure), the lag will increase. It is because the location of the FR fluctuates sharply when the step length is too large, which makes it difficult to converge to the target location. An example is shown in Fig. 4.14 (b). By comparing with the case of a smaller step length, it is difficult for the FR to converge when $R = 15$.

Accumulated Throughput

In this section, we further verify the range of the gain of the accumulated throughput defined in Sec. 4.6.1. In the one-snapshot simulation shown in Fig. 4.10, the maximum and minimum gains can be identified. We record both of them for the WCA algorithm and the SDP method, respectively, in each simulation round, and then take the average. The results are summarized in Table 4.2.

Table 4.2: The range of the gain

Algorithm	WCA	SDP
Maximum Gain (Averaged over 100 rounds) Homogeneous shadowing	1.95%	0.80%
Minimum Gain (Averaged over 100 rounds) Homogeneous shadowing	-0.82%	-3.38%
Maximum Gain (Averaged over 100 rounds) Independent shadowing	1.03%	0.61%
Minimum Gain (Averaged over 100 rounds) Independent shadowing	-2.77%	-2.98%

As shown in Table 4.2, the WCA algorithm can achieve a better performance than the SDP method. In the homogeneous shadowing case, if the network changes frequently (limited steps with a valid global information), the WCA algorithm can obtain a positive gain, which is less than 1.95% on average. Otherwise, the WCA slightly lose the accumulated throughput, less than 0.82% on average, comparing with the direct off-line method. The performance is worse in the independent shadowing case because of more circuitous flight path and the longer delay shown in Fig. 4.13.

According to the simulation results, we can conclude that, no matter how long the global information is valid, the WCA algorithm is as good as or even better than the off-line approaches from the perspective of the accumulated throughput.

4.7 Conclusion

In this chapter, considering the high penetration loss/deep shadowing channels of the MTC devices, we introduced the FR into the cellular system to improve the transmission efficiency and maximize the system throughput. Given the capacity limit of the FR's back-haul link and the maximum transmission power of each user, an optimization problem has been formulated to maximize the system throughput. The optimal power allocation strategy has been derived, and two effective on-line FR placement algorithms, i.e., the WCA for the unpredictable network and the A-WCA targeting on the predictable network, have been proposed. Two comparative off-line approaches, i.e., the direct and the SDP methods, have also been described.

Extensive simulations have been conducted. The results show that the WCA algorithm can reach the near-optimal system throughput given the homogeneous shadowing and can find the global optimal location within a certain range for most of the identical shadowing cases. Collecting the statistic information and predicting the network's changing pattern is desirable in the independent shadowing case, where the WCA is easily misled by local maximizers and the A-WCA shows a significant advantage thanks to its adaptive step length in the searching process.

More simulations focusing on the WCA algorithm have been included. When the homogeneous shadowing is applied, the convergence speed of the WCA algorithm is close to the upper-bound and its accumulated throughput along the flight path is as good as or even better than the off-line approaches. However, the WCA's performance is worse in the independent shadowing case, which provides an insight on how the performance varies according to the ruggedness of the maximum system throughput over a 3-D area. It can be considered as the reference to determine the necessity of introducing the network prediction and the A-WCA algorithm.

Chapter 5

UAV-assisted Dynamic Coverage in Heterogeneous Cellular System

5.1 Introduction

In Chapter 4, based on the origin of MTC, we focused on devices' power allocation and the FR's flight path planning issue. In this chapter, we extend the availability of the FR and propose a system design applying FR-cells in the cellular system and targeting on a broader application. Comprehensive analyses on the deployment of FR-cells is given.

Due to the rapid development of Mobile Internet (MI) and Internet of Things, the volume and characteristics of traffic carried by the wireless links have been changed dramatically and led to congestion. In addition to problems caused by MTC devices mentioned in previous chapters, exchanging multimedia information by handheld devices also significantly increase the traffic volume in both the uplink and downlink, challenging the existing cellular system. New BSs or relays are required to mitigate the capacity shortage of existing cells and enhance the coverage. However, as ex-

plained in Chapter 4, deploying new BSs has become increasingly difficult and costly, especially in the complicated urban scenarios [83], and may not be profitable to solve the bursty MTC traffic.

Given the current high product maturity of the UAV, the urgency of the increasing volume of MI traffic and extreme channels of the MTC devices, as well as the problems of deploying traditional BS mentioned above, motivate us to investigate the possibility of introducing the UAV-based FR in the cellular system.

Some existing works have already employed the UAV in wireless communication systems, mainly the wireless sensor network [75,84,85]. Some specific outcomes could be applied in the cellular network, such as the positioning problem analyzed in [86] and [87], the optimal multi-hop path obtained in [88], and the optimized topology of UAVs in [89]. But many practical issues, such as frequency reuse, inter-cell interference, backhaul and traffic model in the cellular system, need to be further investigated. [90] showed the throughput improvement of a few simple cases of using FR in the cellular system. In [91], UAV was applied for Public Safety Communications (PSCs) in the cellular system. However, the full coordination between the macro BS and the FR remains unsolved. A neural-based cost function was formulated and then minimized in [92] to find out the optimal mapping approach from UAVs to demand areas. More comprehensive interference models other than the mutual interference between UAVs and the analyses on resource allocation for UAVs backhauls are still open issues.

In this chapter, we propose the UAV-assisted base station (UABS) to solve the problems brought by the increasing traffic volume of MI and by serving MTC devices with special traffic characteristics and locations. It enables heterogeneous deployment inside the macro cell and achieves dynamic and adaptive coverage. Key Issues related to the deployment of FR-cells including frequency reuse, interference, backhaul re-

source allocation, and coverage need to be analyzed comprehensively.

5.2 UAV-Assisted Base-Station

The UABS is a BS centralized controlled system, where the coverage can be dynamically adjusted leveraging the mobility of UAV-based FR. The BS monitors the traffic within the macro cell and sends out one or more FRs when needed, as shown in Fig. 5.1.

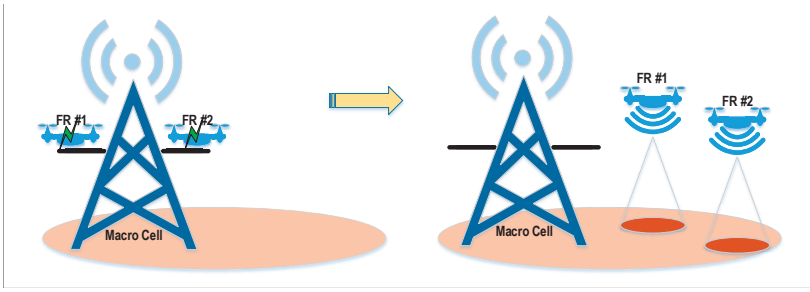


Figure 5.1: Comparisons between the WCA and the optimal capacity.

The macro BS monitors the buffer status of each user and the contentions on the Physical Random Access Channel (PRACH) to determine the timing of sending out the FRs. Suggested coverages of FR-cells can be made based on the real-time users locations and the statistic history record. In addition, most of the MTC devices have predefined transmission schedule and fixed locations, so the FRs can be placed beforehand. When an FR-cell is no longer needed, it will be recalled by the macro BS. There is a small garage on the BS tower for FRs to park. Their battery will be recharged when entering the garage.

The monitoring and information collecting functions are available in current LTE/LTE-A BSs. Three additional updates are needed to deploy UABS:

1. The algorithm for the macro BS to determine the optimal deployment of the FR-cells.

2. A separated Control Plane (CP) protocol between the macro BS and the FR, which enables the control of the FRs behavior and collects possible feedback from the FR.

3. The parking garage with a recharging function for the FRs.

Among the required updates above, update 3 is hardware-related. It can be implemented by non-contact recharging. It is out of the scope of this chapter. Updates 1 and 2 are software-related. For the new CP protocol mentioned in update 2, there are lots of existing mechanisms, e.g., adding a new layer on top of the radio access network protocol stack. The control information for the FR will be piggybacked in the RRC messages, similar to the transmission of the non-access stratum signaling in LTE/LTE-A. An alternative approach is to enhance the existing RAN protocol, such as adding new information elements in an RRC message or MAC headers. The detailed contents of the control information should be related to the UAV movement and will not be examined in this chapter.

We assume that update 2 have been implemented and the FR can fly to anywhere according to the macro BSs order. In this chapter, we focus on the issues related to update 1, including frequency reuse, interference, backhaul resource allocation, and coverage.

5.3 Frequency Reuse and Interferences

Within the coverage of a UABS, FR-cells reuse the uplink frequency bands used by other macro cells. We assume that the adjacent macro cells use different frequency bands and a certain reuse distance should be maintained by the topology. The transmissions inside an FR-cell can reuse the mechanism of D2D communication which has been used for many purposes in heterogeneous networks, such as load balancing [93]

and improving the connectivity to Internet [94].

An example of the frequency reuse and corresponding interferences is illustrated by Fig. 2, where FR-cell #1 and #2 are controlled by Macro-cell #1 and they reuse the uplink frequency of Macro-cell #2. Three types of interferences are generated in such situations as discussed below.

5.3.1 The Interference from the FR-cells to the Macro Cells

The signal transmitted inside an FR-cell can be received by other macro BSs, e.g., Macro-cell #2 in Fig. 5.2. At the macro BS side, the uplink signal from a macro user may be interfered by the transmissions in the FR-cells using the same time-frequency resource.

FR-cells are similar to or even smaller than small-cells. Because it can fly along a series of traffic dense areas and even proactively approach a specific user with very bad channel condition, rather than increase the transmission power. This strategy is especially suitable for the MTC scenario. Therefore, the transmission power used in FR-cell will be maintained at a very low level. If the topology has guaranteed a sufficient reuse distance, this type of interference can be negligible in practice.

5.3.2 The Mutual Interference between the FR-cells

If multiple FR-cells are close to each other and use the same frequency, mutual interference will be generated. One effective way to avoid this type of interference is to let different FR-cells use different frequency bands. An example is illustrated in Fig. 5.3 (a) where the reuse factor equals 1/7.

In this case, one single macro cell, the one using frequency band #1 in Fig. 5.3 (a), is divided into six frequency-reusing areas. The six reusing areas use six different frequency bands, #2-#7, which are used by adjacent macro cells. Given that any two of

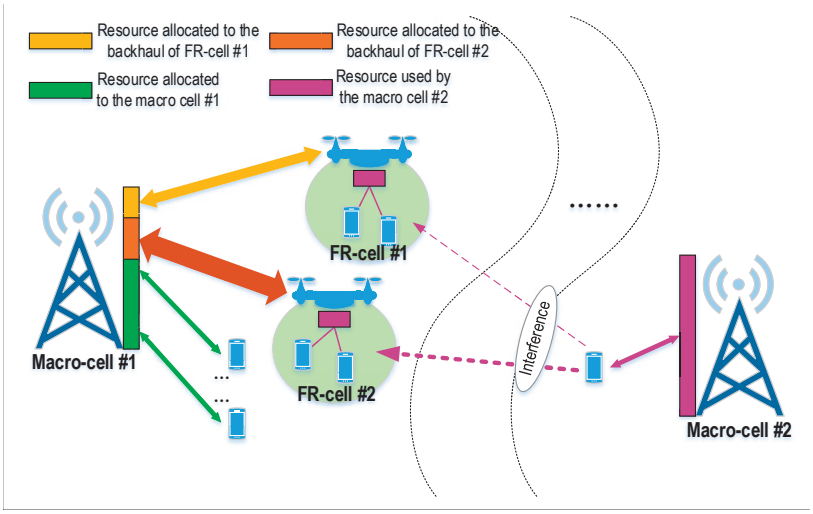


Figure 5.2: Frequency allocation for the FR-cells and corresponding interferences.

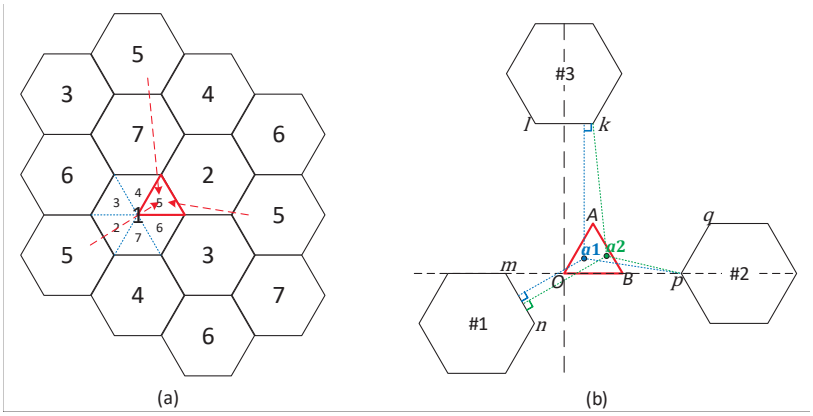


Figure 5.3: An example of the frequency allocation and the topology of interference.

the reusing areas will apply different frequencies, if at most one FR is deployed in each reusing area, the mutual interference can be largely eliminated. Alternatively, one single frequency band could be divided into several sub-bands for multiple FR-cells in the same frequency-reusing area, by which their mutual interference is eliminated.

Another approach is to allow the interference within a given threshold by adjusting the transmission power of each FR-cell to achieve an acceptable Signal-to-Interference-plus-Noise Ratio (SINR).

5.3.3 The Interference from the Macro Cells to the FR-cells

As shown in Fig. 5.3 (a), the FR-cells can also be interfered by uplink transmissions initiated by macro users in nearby macro cells. Most of the macro users on the cell edge apply a larger transmission power to overcome the path loss of a long transmission distance, which results in a higher interference to the FR-cells. Considering the relatively low transmission power inside the FR-cells, the SINR will be largely degraded.

Therefore, among the three types of interference, the one from the macro cells to the FR-cells is the most critical and cannot be ignored. The best time and location of FR deployment and the arrangement of the frequency-reusing areas highly depend on the modeling of this interference. We use the topology shown in Fig. 5.3 (a) to build the model.

We focus on the area reusing the frequency band #5, as shown in Fig. 5.3 (b). The result can be applied to other areas. We consider the worst case in this model, where there is always a macro user transmitting at the same time-frequency resource in each of the three closest macro cells. Furthermore, given any point inside the triangle $\triangle OAB$, the shortest distances from it to the three interfering macro cells are applied.

The point O is put on the origin of a rectangular coordinate system and the edge $\overline{OB}$ is on the x axis. For a point a inside $\triangle OAB$, the shortest distances to the macro cell #1 and #2 are always equal to the perpendicular distance to edge $\overline{mn}$ and the distance to point p, respectively. The shortest distance to the macro cell #3 can be either the distance to $\overline{lk}$ or that to k depending on the location of point a, as illustrated by $a1$ and $a2$ in Fig. 5.3 (b). These distances are easy to calculate based on geometry.

We assume that the received power is only determined by the path loss which

is a function of the distance. Given the transmission power of the macro users, the interference from one macro cell can be obtained. The interference power from three nearest macro cells accumulate on the air interface, and thus we can have the total interference I.

In practice, this model of the worst case can be applied when there is no information of the interference available at the macro BS side. After sending out the FRs for several times, the interference level at a specific time and place can be measured by the FRs and reported to the BS, according to which a statistical interference map can be built and maintained. Then, more precise deployment of the FR-cells is feasible. Furthermore, if the resource scheduling decisions can be shared among adjacent macro cells, the interference will be significantly reduced by utilizing the unoccupied time-frequency resources of the neighbor macro cells for FR-cells.

The unlicensed bands can be used for FR-cells. In that case, the mutual interference between the macro cells and the FR-cells is simplified. The QoS in unlicensed bands cannot be guaranteed, so the cellular system can utilize the unlicensed bands as a complementary solution but cannot fully rely on them for serving users.

5.4 Backhaul

A certain amount of time-frequency resources should be assigned for the FRs wireless backhauls (FR-backhauls) by the macro BS, as illustrated in Fig. 5.2. The macro BS needs to make a trade-off between the service quality of the FR-cells and that of other macro users. Several candidate methods of the resource allocation for FR-backhauls are analyzed as follows.

5.4.1 Minimum Fixed Bandwidth for the FR-backhaul

Allocating a fixed bandwidth for the FR-backhaul is the simplest way and easy for management. The bandwidth can be predefined or broadcasted to all the FR-cells, and the position of each FR-backhaul in the frequency domain can be inferred by FR-cells identity or other unique information. Once an FR-backhaul is established, it will be maintained and kept unchanged until the FR is called back. The Semi-Persistent Scheduling (SPS) in LTE can be further applied in the FR-backhauls where resource block assignments and MCS remain fixed for a certain period. The control signaling overhead is minimized in this method.

The disadvantage of this method is also obvious. To avoid the potential waste of the resource and guarantee the QoS of macro users, the FR-backhaul has to be allocated with a minimum bandwidth, which results in the capacity limitation of the FR-cell. The QoS of many types of traffic, especially Ultra-Reliable Low Latency Communications (URLLC), cannot be guaranteed. For MTC devices with low priority and delay-tolerant traffic, this method is the best thanks to its simplicity and low overhead.

5.4.2 Always-satisfied Bandwidth for the FR-backhaul

In this approach, the macro BS considers each FR as macro users but with the highest priority and the LTE dynamic scheduling is used. Thus the traditional scheduling grant, frequent Channel Status Indicator (CSI) feedbacks and other necessary control messages such as Scheduling Request (SR) and Buffer Status Report (BSR) will be transmitted. The control overhead is increased compared with the method of fixed bandwidth. However, the overall amount of control overhead is still reduced comparing with the case without FR-cells where the macro BS should consume a control overhead for each user.

Assigning the highest priority to the FR-backhauls is easy for the scheduler. But the QoS of macro users will be greatly affected if the traffic volume carried by the FR-backhauls is too large. The fairness between macro users and the users in FR-cells cannot be guaranteed as well. The always-satisfied bandwidth should be considered only if the traffic volume in the FR-cell is small and the highest scheduling priority is necessary, such as the URLLC scenario.

5.4.3 Traffic-aware Adaptive Bandwidth for the FR-backhaul

The traffic-aware adaptive bandwidth allocation method combines the above two approaches and is suitable for the mixture of the low priority MTC traffic and the URLLC traffic.

The macro BS allocates a minimum fixed bandwidth to each FR-backhaul to ensure the services to low priority MTC devices while reducing the impact on other macro users as much as possible. Once packets with a high priority are received, it will switch to the dynamic scheduling with always-satisfied bandwidth. The dynamic scheduling for the FR-backhaul is based on the BSR of high-priority traffic, and low-priority traffic stored in the FR still need to wait for the predefined SPS.

In this way, the macro users will be affected only if there is high-priority traffic in the FR-cells, which is typically infrequent and has a small volume. If without FR-cells, the service for the high-priority traffic may cause an even higher resource consumption due to a longer transmission distance than that in the FR-cell.

5.4.4 Optimized Fixed Bandwidth for the FR-backhaul

In practice, the minimum fixed bandwidth should be determined based on the QoS requirements of the MTC traffic, i.e., how much delay they can tolerate. However, to serve other types of traffic with more stringent QoS requirements but not as high

as that of the URLLC traffic, such as streaming media, the limited FR-backhauls will become the bottleneck and affect the QoS. The optimized bandwidth solution is proposed to enable these services in the FR-cells.

In this case, minimizing the usage of the resources in a macro cell is no longer the objective of the FR-backhauls design. Instead, we aim to improve the overall spectrum efficiency and achieve the maximum throughput. We take the uplink transmission as an example to further explain it. Assuming that there are totally N FR-cells, allocating the optimal bandwidth for FR-backhauls, $\boldsymbol{B}$, can be formulated as the following optimization problem,

$$\begin{aligned} \max_{\boldsymbol{B}} \quad & \left[W - \sum_{i=1}^{N} B_i\right] \eta_M + \sum_{i=1}^{N} B_i \log_2 \left[1 + \frac{P_{bh} g_{bh_i}}{n_0 B_i}\right], \\ \mathbf{s.t.} \quad & \sum_{i=1}^{N} B_i \leq \Gamma, \\ & 0 \leq B_i \leq C_i, \quad \forall i = 1, ..., N, \end{aligned} \tag{5.1}$$

where W is the bandwidth of a macro cell and B_i is the bandwidth allocated to the i-th FR-backhaul, P_{bh} is the transmission power of the FR on the backhaul, g_{bh_i} denotes channel gain of the i-th FR-backhaul, and n_0 denotes the noise spectrum density.

In the objective function, η_M is the average spectrum efficiency of the macro cell given a certain scheduling algorithm, and is assumed a constant for simplicity. When the number of users in the macro cell is relatively large compared with that is covered by the FR-cells, the multi-user diversity gain achieved by the macro cell will not be substantially reduced by offloading traffic to FR-cells. Also, η_M will not change much with the changing of the macro bandwidth. Though the gain that comes from scheduling over frequency selective fading channels may be shrunk if the bandwidth

is reduced, a threshold of the total bandwidth used by the FR-backhauls, Γ, can be set to ensure a large enough bandwidth left to the macro cell. Furthermore, when the users with bad channels are covered by the FR-cells and no longer scheduled by the macro scheduler, the overall spectrum efficiency of the macro cell will be increased.

In this problem, the first constraint is the threshold of the total bandwidth used by the FR-backhauls. The second ensures that the bandwidth allocated to an FR-backhaul will not exceed the average capacity of this FR-cell, so not to waste the resource. As described above, the D2D based transmission mechanism is used inside an FR-cell and thus the transmission for different users should be performed in a time division multiplex (TDM) manner. The average capacity of the i-th FR-cell can be simply modeled as $C_i = W' \int_0^\infty \log_2[1 + [P_{fr}x/(I_i + n_0 W')] f_{g_i}(x) \mathrm{d}x$, where W' is the bandwidth reused by one FR-cell. It may not equal the whole bandwidth of a macro cell, W, because it may be shared by multiple FR-cells in the same frequency reusing area. P_{fr} denotes the transmission power inside a FR-cell. The channel gain g_i is a random variable whose probability distribution function (PDF) is denoted by $f_{g_i}(x)$. In practice, when the PDF is not available, an average value can be used to obtain the approximation. I_i is the interference received by the i-th FR-cell. Either the interference we derived previously considering the extreme case or the statistical interference map can be applied, so I_i is assumed a constant for a specific location of the FR-cell. Therefore, given a certain location and the channel information of the i-th FR-cell, C_i can be calculated.

Taking the second derivative of the Lagrangian function w.r.t. B_i, we find that the second derivative is negative and the Hessian matrix is negative definite. Therefore, this problem can be proved to be a convex optimization problem, and lots of existing tools, such as CVX and fmincon, can be utilized to find out the optimal numerical results.

5.5 Coverage

The macro BS determines the suggested coverage based on the historical statistical data or the real-time distribution of the traffic volume. However, the estimation based on the historical data may not be precise. Updating the suggested coverage location to the FR real-timely will significantly increase the control overhead not only on the FR-backhaul but also for the macro users because the precise locations need to be collected frequently. The FRs are sent out to mitigate the overload or congestion for the macro cell, imposing extra burdens on the macro users may compromise the performance gain achieved by the FR-cells.

Therefore, we propose the BS semi-controlled coverage extension methods to improve the utilization of the FR-cells and reduce the control overhead. When the macro BS activates or updates an FR-cell service, the initial point, as well as the cruising mode, should be included in the control message. The options of the cruising mode may include but are not limited to:

1. no coverage extension
2. $(L, \boldsymbol{d})$ extension
3. adaptive $(L, \boldsymbol{d})$ extension

Using option 1, the FR stays at the initial point given by the macro BS. It is suitable for the scenario of stationary MTC devices whose locations are known and close to each other. When the traffic is dynamic or it is difficult to obtain the precise real-time locations of users, the following two options should be used.

In option 2, the FR will first cruise along a circle centered at the initial point with a radius of R, which is the radius of the original coverage of the FR-cell, and thus form a new serving area where the coverage is extended from R to $2R$, as shown

in Fig. 5.4 (a). This area is denoted as the first layer coverage. After that, the FR extends the radius of the cruising circle from R to d, which forms the second layer coverage. Assuming $R = 30$m and the flying speed of the FR is fixed to 1m/s, given any user inside either the first or second layer coverage, the time covered by the FR-cell according to different d is given in Fig. 5.5 (a), which is a function of the distance to the initial point. This time is counted within the period that the FR finishes both the first and second layer cruising once.

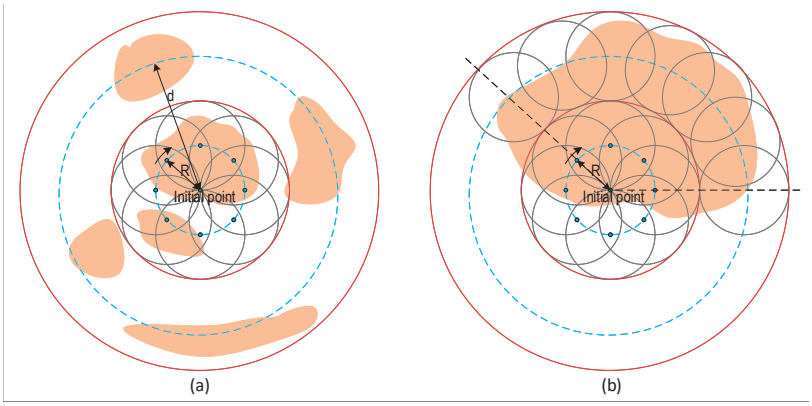
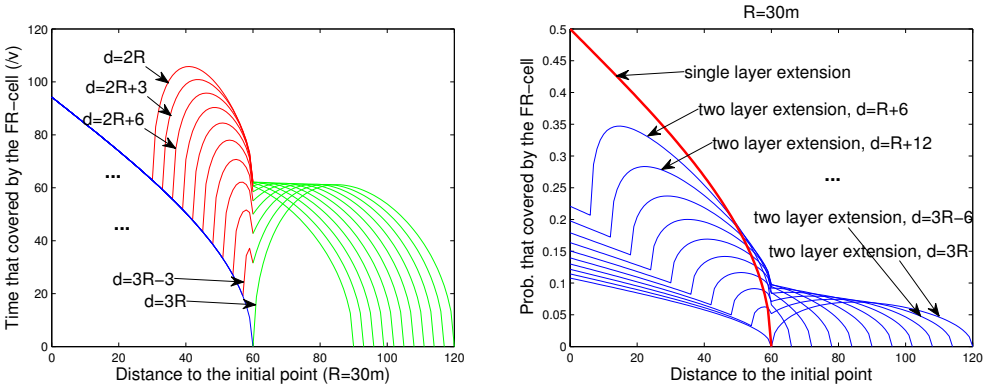


Figure 5.4: Examples of the extended coverage of the FR-cell.

In Fig. 5.5 (a), the blue and green curves denote the first and second layer coverages, respectively, and the red curves present the overlapped area. A seamless coverage can be guaranteed when $d < 3R$. With the increasing of d, the FR connection time (time covered by the FR-cell) in the overlapped area decreases while the total coverage is increased. The selection of d should be constrained by the acceptable minimum FR connection time, which is valuable for the MTC traffic where a certain amount of data are waiting for transmission.

For other types of traffic with continuous arrivals, such as streaming media, the FR connection probability (probability of being covered by the FR-cell) is more important than connection time, which is shown in Fig. 5.5 (b). Compared with the single layer extension, the average FR connection probability decreases in the two layer



(a) FR connection time (b) FR connection Prob.

Figure 5.5: FR connection time/Prob. vs. distance to the initial point.

extension, but the coverage is extended. The different settings of d result in different performance, as shown in the figure, which is the guideline for the macro BS to make the trade-off between the coverage and the single users enhancement.

In addition to the two-layer extension discussed here, more-layer extension can be configured by the parameters $(L, \boldsymbol{d})$, where L represents the number of layers, $\boldsymbol{d} = [d_1, d_2, , d_{L-1}]$, and d_i denotes the radius of the $(i+1)$-th layers cruising circle. The optimizations on $(L, \boldsymbol{d})$ given certain objectives are still open issues and beckon for future investigation.

Option 2 is suitable for the scenario where the users are scattered randomly, such as people in a shopping mall, and the macro BS is difficult to obtain the precise locations of them, as shown in Fig. 5.4 (a). However, when the users are clustered but their locations may change over time, blindly extending the coverage may lead to a very low utilization of the FR-cell. For example, in some activities such as weddings and conferences, all the guests are moving from one place to another simultaneously.

Therefore, we further propose option 3. In this option, the FR will finish the first layer cruising as option 2 and monitor the traffic density along the circle. For the second layer, the FR only extends the coverage within the sector where the first layer

traffic density is higher than a threshold, as shown in Fig. 5.4 (b). In the following layers, the FR applies the same method until either the L-th layer is reached or there is no place with a traffic density that exceeds the threshold found in the current layer.

5.6 Conclusion

In this chapter, we proposed the UABS to solve the problems brought by the increasing traffic volume of MI and by serving MTC devices with special traffic characteristics and locations. It utilizes UAV based FR to enable heterogeneous deployment of additional FR-cells inside the macro cell and achieves dynamic and adaptive coverage. We focused on how to deploy the FR-cells and the associated issues. Comprehensive analyses on FR-cells deployment including frequency reuse, interference, backhaul resource allocation, and coverage were given.

Chapter 6

Placement of Supplementary Node in UAV Mesh Networks

6.1 Introduction

We focused on MTC devices and optimized the FR's location in Chapter 4. The usage of FRs was then generalized as the UABS serving both MTC devices and MI applications in Chapter 5. In this chapter, we further investigate UAV mesh network which not only provides the seamless coverage to low-priority MTC and MI traffic but also quickly responds to Ultra-Reliable Low Latency Communications (URLLC) triggered by major incidents, such as uncontrolled emissions of liquid or gaseous contaminants in the cases of volcanic eruptions, wild fires, industrial incidents, crimes, and terrorist attacks. The use of UAV mesh network in possible incident areas can greatly facilitate the tasks in police departments, fire brigades, and other homeland security organizations [95].

There are two operating stages. The first is to form a sparse UAV mesh network which maximizes the coverage. Given the randomness and low frequency of inci-

dents, we limit the number of UAVs used in this stage to control the operation cost, and an appropriate topology of UAVs can be maintained to achieve a better performance [96]. Given a fixed number of UAVs, the isometric grid topology maximizes the coverage, which has been widely used in wireless communication systems. Once an incident happens, it can be tracked down to a triangle of three UAVs, which is ideally equilateral. In practice, the topology of UAV mesh network may be different due to geographical constraints or obstacles like buildings and trees, and thus the triangles in the mesh network may not be perfectly equilateral but likely to be acute, as shown in Fig. 6.1.

When a warning area is identified, the system shall respond quickly and enter the second stage, where it prepares to receive more incoming high-priority sensing data. The first-stage UAV mesh network is sparsely deployed to save the cost, so it may not be sufficient to guarantee the low latency and high reliability inside a warning area due to the long transmission distance. Also, because the coverage in other areas should be maintained, a new supplementary UAV will be sent to the warning area to support the URLLC traffic.

The QoS inside the warning area is critical, and thus the point with the worst transmission service quality, i.e., the worst point, will greatly affect the performance because it determines the existence of blind points. In this chapter, we optimize the location of the newly added UAV within the warning area to maximize the transmission service quality to the worst point.

The transmission service quality depends on channel status. Considering mechanical constraints and huge amount of necessary feedback, it is difficult for a UAV to adjust location dynamically according to fast fading. As a starting point, in this chapter, we apply a homogeneous shadowing and only focus on path loss. Therefore, we assume that the shorter the distance is, the better the performance is. Assuming

that the locations of sensor nodes or incidents are uniformly distributed, and each sensor is always served by the nearest UAV, the problem can be converted to minimizing the longest transmission distance inside an acute triangle determined by three nearest UAVs in the mesh. Because the location of an incident's occurrence varies time by time and the tracked target may be moving constantly, the first-stage UAV swarm may move around an area for the optimal coverage, which results in mobile mesh backbone and a dynamic shape of an acute triangular warning area that the supplementary UAV should be added to. Hence, a real-time algorithm with a low complexity is required to quickly find the optimal location for the newly added UAV given the changing topology.

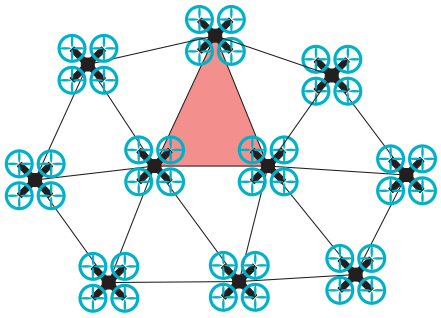


Figure 6.1: A triangle warning area.

Without the loss of generality, we denote each UAV as a service node and define the distance from a certain location to the nearest UAV as the service distance. In this chapter, considering the two-stage UAV mesh network, a min-max optimization problem is formulated to minimize the longest service distance in an arbitrary acute triangle. For the newly added service node, the closed-form expression of the optimal location within an isosceles acute triangle is derived. Then, a real-time algorithm with low complexity, i.e., the equal service distance (ESD) algorithm, is proposed to find the optimal location of the new service node within an arbitrary acute triangle. Finally, comparisons with the exhaustive search and triangle's existing centers, including incenter, circumcenter, and centroid, are conducted. By simulations, the

optimality of the proposed method is validated and the impacts of UAV's speed and processing time on the performance are investigated.

6.2 Related Work

In [97], multiple UAVs cruise in a target area following optimized trajectories and thus minimize the maximum uncovered time duration of each cell. A feasible distribution of the UAV swarm to properly cover the target area and the trajectory for each UAV to reach the target location were proposed in [98]. A novel metric to depict the quality of UAVs' deployment, i.e., the deployment entropy, was introduced in [96], and different deployments were evaluated accordingly. Most of the existing work focused on how to achieve a better coverage given a certain number of UAVs, which is the first stage described in Sec. 6.1.

The UAV mesh network shares some similarities with the mobile sink in the wireless sensor network (WSN). The main purpose of introducing mobile sinks is to balance the energy consumption. Routing methods and data collection strategies were proposed in [63–67, 99] to increase the lifetime of network. The trade-off between energy saving and traffic delay was analyzed in [68]. We also find the similarity between the UAV mesh network and the coverage problem in the cellular network [100]. [101], [102], and [103] focused on maximizing the number of covered users. The interference and backhaul constraints were considered in [104] and [105], respectively, and users' energy consumption was further optimized in [106] and [107]. In both of the UAV mesh network and other UAV-related systems, many existing work can be applied to the first stage of the UAV mesh network introduced in Sec. 6.1, but few of them can guide the procedure of the second stage.

In this chapter, we focus on the second stage of the UAV mesh network and op-

timize the location of the newly added UAV within a triangular warning area. By using the general concept, i.e., service nodes, instead of UAVs, the problem can be mathematically abstracted as finding the optimal location to add a point inside a triangle, which is similar to the facility location problem. In the past decades, the facility location problem had been extensively studied based on different objectives. Most of them are minimizing the sum of the transportation costs from one point to multiple destinations, which is known as the Fermat problem [108], the Weber problem [109], and the attraction-repulsion problem [110] according to different definitions of the transportation cost, respectively. Different from these classic problems, we aim to minimize the maximum service distance from any point inside the triangle to its nearest service node, rather than that between two service nodes.

The operation cost of a service node was then considered with the objective of minimizing the overall cost. Depending on whether a service node has limited capacity, the problem can be classified as capacitated and uncapacitated cases [111]. In [112] and [113], the two-level and multi-level uncapacitated facility location problem were studied, respectively. A linear programming based algorithm was proposed for capacitated case in [114], the financial costs and carbon emission were further considered in [115]. Unlike assuming fixed sets of locations (or candidates of locations) for facilities and clients in existing works, we target on the scenario where the location of the newly added service node is selected from a continuous plane. The worst point, which can be considered as the client in capacitated/uncapacitated facility location problems, may change according to the different locations of the new service node.

6.3 System Model

6.3.1 Preliminary

As shown in Fig. 6.1, once the warning area is determined, a new supplementary UAV, i.e., a new service node, will try to keep staying at the optimal location within the area regardless the change of topology. The new service node updates its location periodically. We define one period as a slot. At the beginning of each slot, the optimal location is updated according to the latest topology of mesh. The new service node moves to the updated location directly after the calculation and stays until the next slot. Due to the altitude regulation of UAV system [116] and complex terrain in low altitude area, the UAV mesh network is likely to operate at an altitude lower than 120m (the maximum altitude [116]) but still high enough to avoid most obstacles. For simplicity, we assume all the UAVs have the same altitude.

Given the same altitude, the projected distance on a horizontal plane monotonically increases with the service distance. Thus, the problem can be studied in a 2-D plane by projecting all the users and service nodes in the same 2-D plane. The objective is simplified to improve the transmission/service quality of the worst point inside a 2-D triangle area. The transmission/service quality is determined by the service distance, i.e., the shorter the distance, the better the quality. Users are uniformly distributed, which means every point inside the triangle area needs to be considered with equal priority. We only focus on the acute triangle in this chapter, and each user selects the service node based on the shortest service distance.

To find out the worst point and the longest service distance, we use Fig. 6.2 as an example. Given α, β, the angles of the altitude through vertex A and two sides $||AB||, ||AC||$, the length of $||AB||$, a triangle $\triangle ABC$ can be determined. Any point O inside $\triangle ABC$ can be denoted as a function of ϕ_1 and ϕ_2. Given the principle of service

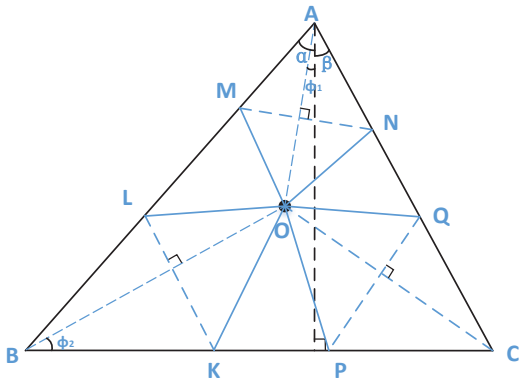


Figure 6.2: A triangle formed by three existing service nodes.

node selection, the triangle area is divided into several voronoi cells, and line segments MN, LK, and QP are the cell edges. It is obvious that $\max(||OM||, ||ON||) \geq ||OX||$, where X is any point on MN. Similar observations can be found on LK and QP as well. Therefore, we can conclude that all the vertexes of voronoi cells are candidates of the worst point because one of them lead to the longest service distance. As shown in Fig. 6.2, there are 6 candidate points, i.e., M, L, K, P, Q, and N. They are on sides of the triangle and are defined as on-boundary vertexes.

Besides that, in some cases, two adjacent vertexes on the triangle's boundary may converge to one vertex inside the triangle. In Figs. 6.3, we show examples where the vertexes K, P and N, Q converge to the inner vertexes (vertexes inside the triangle) K' and N', respectively. Similarly, M, L may also converge to M'. In total, there are 9 vertexes considered as the candidates of the worst point. We divide them into three vertex sets: $\mathbb{V}^{(1)} = \{M, L, M'\}$, $\mathbb{V}^{(2)} = \{K, P, K'\}$, $\mathbb{V}^{(3)} = \{N, Q, N'\}$. Each set consist of two on-boundary vertexes and one inner vertex. When the inner vertex exists, the on-boundary vertexes in the same set are no longer the valid voronoi vertexes because it conflicts with the property of voronoi cells, i.e., the distances from a voronoi vertex to adjacent service nodes should be the same. Therefore, we have the following Lemma 1.

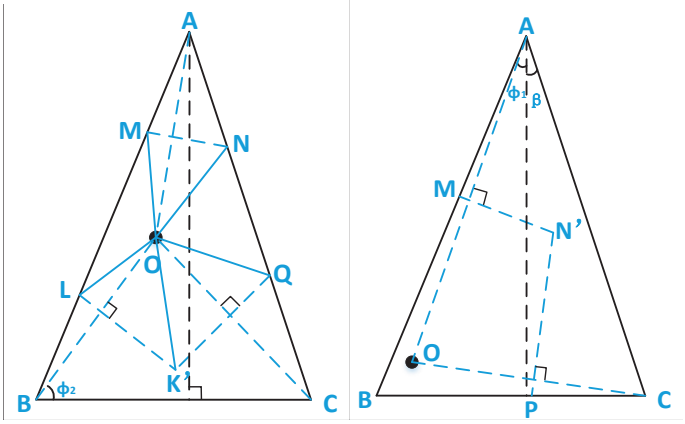


Figure 6.3: Examples of K' and N'.

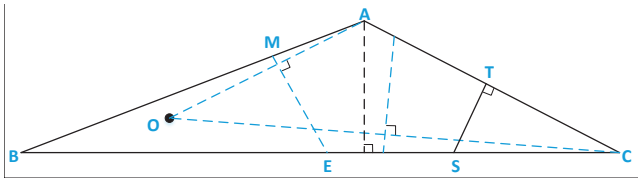
Lemma 1. *For a vertex set $\mathbb{V}^{(i)}, i = 1, 2, 3$, the voronoi vertexes can be either the one inner vertex if existed, or the two on-boundary vertexes otherwise.*

Lemma 2. *For $\mathbb{V}^{(1)}$, $\mathbb{V}^{(2)}$, and $\mathbb{V}^{(3)}$, the inner vertexes, i.e., M', K', N', exist only if $\angle AOB \leq \frac{\pi}{2}$, $\angle BOC \leq \frac{\pi}{2}$, and $\angle AOC \leq \frac{\pi}{2}$, respectively.*

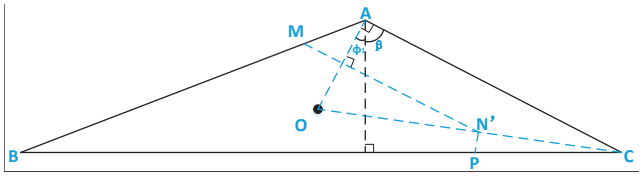
Proof. Taking Fig. 6.3 (a) as an example, given the property of voronoi cells, K' is the circumcenter of $\triangle BOC$. When $\angle BOC \leq \frac{\pi}{2}$, the circumcenter is inside $\triangle BOC$ and thus K' exists. Similar conclusions can be made for M' and N' as well. □

Lemma 3. *In an arbitrary acute triangle, at least one vertex from each vertex sets, i.e., $\mathbb{V}^{(i)}, i = 1, 2, 3$, exist, regardless the location of the new service node.*

Proof. In some cases, any vertexes from a certain set cannot be found in a triangle. An example is shown in Fig. 6.4 (a), where N, Q, N' are missing. Two new voronoi vertexes S and T have been introduced and determined only by the existing service nodes A and C. Furthermore, OA's mid-perpendicular line no longer has the intersection with AC (in the triangle), and a new intersection with BC, vertex E, is introduced.



(a)



(b)

Figure 6.4: Example of the disappearance of $\mathbb{V}_3$.

In Fig. 6.4 (b), It is observed that N' is the circumcenter of $\triangle AOC$. It will be outside $\triangle AOC$ and below OC only if $\phi_1 + \beta > \frac{\pi}{2}$. Given that $\triangle AOC$ is an acute triangle, $\phi_1 + \beta$ is always smaller than $\frac{\pi}{2}$. Therefore, N' can never move across OC and BC. The situation in Fig. 6.4 (a) will never happens in an acute triangle. Thus, given OC and OA are in the triangle, N' always exists if $\angle AOC$ is acute, N, Q always exist if $\angle AOC$ is obtuse. Similarly, the existence of vertexes from other sets can be proved by the same way. □

Lemma 4. *In an arbitrary acute triangle, at most one of the inner vertexes, i.e., M', K', or N', exists.*

Proof. According to the proof of Lemma 2, if two vertexes from M', K', and N' exist, two of the angles $\angle AOB$, $\angle BOC$, and $\angle AOC$ are smaller than $\frac{\pi}{2}$. Given $\angle AOB + \angle BOC + \angle AOC = 2\pi$, the rest one angle is larger than π, which means the point O is outside the triangle. All of $\angle AOB$, $\angle BOC$, and $\angle AOC$ cannot be smaller than $\frac{\pi}{2}$ as well. Therefore, at most one vertex from M', K', and N' exists. □

6.3.2 Problem Formulation

In an arbitrary acute triangle formed by three existing service nodes, we want to minimize the service distance of the worst point by optimizing the location of one newly deployed service node. Therefore, the original problem can be transformed to a Min-Max problem of the service distance. Given Lemma 1, 2 and 3 in Sec. 6.3.1, the optimization problem is given by

$$\min_{O} \ \max\left(||OZ_i^{(1)}||, ||OZ_j^{(2)}||, ||OZ_k^{(3)}||\right), \tag{6.1}$$
$$i = 1, ..., \Gamma^{(1)}, j = 1, ..., \Gamma^{(2)}, k = 1, ..., \Gamma^{(3)},$$

$$\textbf{s.t.} \ \ O \in \triangle ABC,$$

$$[Z_1^{(1)}, ..., Z_{\Gamma^{(1)}}^{(1)}] = \begin{cases} M', & \angle AOB \leq \dfrac{\pi}{2}, \\ [M, L], & \angle AOB > \dfrac{\pi}{2}, \end{cases}$$

$$[Z_1^{(2)}, ..., Z_{\Gamma^{(2)}}^{(2)}] = \begin{cases} K', & \angle BOC \leq \dfrac{\pi}{2}, \\ [K, P], & \angle BOC > \dfrac{\pi}{2}, \end{cases}$$

$$[Z_1^{(3)}, ..., Z_{\Gamma^{(3)}}^{(3)}] = \begin{cases} N', & \angle AOC \leq \dfrac{\pi}{2}, \\ [N, Q], & \angle AOC > \dfrac{\pi}{2}, \end{cases}$$

6.4 Optimal Algorithm Design

6.4.1 Isosceles Acute Triangle

As a starting point, we assume $\triangle ABC$ is an acute isosceles triangle, i.e. $||AB|| = ||AC||$, and thus $\alpha = \beta$.

Lemma 5. *If K' exists in the acute isosceles triangle $\triangle ABC$ where $||AB|| = ||AC||$, it is always on BC's mid-perpendicular line.*

Proof. Given K' is the circumcenter of $\triangle BOC$, $||K'B|| = ||K'C||$. Therefore, K' is on BC's mid-perpendicular line. □

Theorem 2. *Given an arbitrary acute isosceles triangle $\triangle ABC$ where $||AB|| = ||AC||$, the optimal location of a new service node to minimize the longest service distance is on BC's mid-perpendicular line.*

Proof. See Sec. 9.2. □

Lemma 6. *Given an acute isosceles triangle $\triangle ABC$ where $||AB|| = ||AC||$, if the location of a new service node O is on BC's mid-perpendicular line and inside $\triangle ABC$, inner vertexes M' and N' do not exist.*

Proof. Given O is on BC's mid-perpendicular line, it is easy to prove $\angle AOB > \frac{\pi}{2}$ and $\angle AOC > \frac{\pi}{2}$. Therefore, the circumcenters of $\triangle AOB$ and $\triangle AOC$ are outside the triangles, and thus there is neither M' nor N' inside $\triangle ABC$. □

Given Theorem 2 and Lemma 6, $\phi_1 = 0$, $||OM|| = ||ON||$, $||OL|| = ||OQ||$, $||OK|| = ||OP||$ and only the inner vertex K' possibly exists in an acute isosceles triangle. We only focus on $||OM||$, $||ON||$, $||OK||$, and $||OK'||$ in this section. The service distances to these four candidates of the worst point are derived as follows,

$$||OM|| = \frac{||AB|| \cos(\alpha + \phi_2)}{2\cos(\phi_2)\cos(\alpha)}, \tag{6.2}$$

$$||OL|| = \frac{||AB|| \sin(\alpha)}{2\cos(\phi_2)\sin(\alpha + \phi_2)}, \tag{6.3}$$

$$||OK|| = \frac{||AB|| \sin(\alpha)}{2\cos^2(\phi_2)}, \tag{6.4}$$

$$||OK'|| = \frac{||AB|| \sin(\alpha)}{\sin(2\phi_2)}. \tag{6.5}$$

Theorem 3. *Given an arbitrary acute isosceles triangle $\triangle ABC$ where $||AB|| = ||AC||$ and $\angle OBC = \phi_2$, the optimal location of a new service node (the global mini-*

mizer of ϕ_2) and the corresponding minimized service distance to the worst point are summarized in Table 6.1.

Table 6.1: Global minimizer

$\alpha \in$	Global minimizer
$[0, 0.106\pi]$	$\|\|OM\|\|^*_{\phi_2=A'(\alpha)} = \|\|OK'\|\|^*_{\phi_2=A'(\alpha)} = \dfrac{\|\|AB\|\| \cos(\alpha + A'(\alpha))}{2\cos(A'(\alpha))\cos(\alpha)}$;
$[0.106\pi, \frac{\pi}{6}]$	$\|\|OM\|\|^*_{\phi_2=A(\alpha)} = \|\|OK\|\|^*_{\phi_2=A(\alpha)} = \dfrac{\|\|AB\|\| \cos(\alpha + A(\alpha))}{2\cos(A(\alpha))\cos(\alpha)}$;
$[\frac{\pi}{6}, \frac{\pi}{4}]$	$\|\|OL\|\|^*_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}} = \|\|OK\|\|^*_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}} = \dfrac{\|\|AB\|\| \sin(\alpha)}{1+\sin(\alpha)}$;

Proof. See Sec. 9.3. □

It can be observed from Theorem 3 that ϕ_2 is the bisector of $\angle ABC$ when $\alpha \geq \frac{\pi}{6}$. Therefore, given the isosceles triangle, the optimal location is consistent with the incenter of $\triangle ABC$ if $\alpha \geq \frac{\pi}{6}$. For $\alpha < \frac{\pi}{6}$, the optimal location is not any of the existing centers, e.g., incenter, centroid, and circumcenter. Besides, if $\alpha \leq 0.106\pi$, there are 3 worst points with the longest service distance, i.e., M, N, and K'. Otherwise there are 4 worst points. This observation is generalized by Theorem 4 in the following section.

6.4.2 Non-isosceles Acute Triangle

For a non-isosceles acute triangle, we have to consider all of the nine possible vertexes discussed in Sec. 6.3.1. Given a arbitrary location of O in $\triangle ABC$, i.e., given ϕ_1 and

ϕ_2, the service distances to the nine vertexes are derived as follows,

$$||OM|| = \frac{||AB||\cos(\alpha+\phi_2)}{2\cos(\phi_1+\phi_2)\cos(\alpha-\phi_1)}, \tag{6.6}$$

$$||ON|| = \frac{||AC||\cos(\beta+\phi_3)}{2\cos(\phi_3-\phi_1)\cos(\beta+\phi_1)}, \tag{6.7}$$

$$||OL|| = \frac{||AB||\sin(\alpha-\phi_1)}{2\cos(\phi_1+\phi_2)\sin(\alpha+\phi_2)}, \tag{6.8}$$

$$||OK|| = \frac{||AB||\sin(\alpha-\phi_1)}{2\cos(\phi_1+\phi_2)\cos(\phi_2)}, \tag{6.9}$$

$$||OP|| = \frac{||AC||\sin(\beta+\phi_1)}{2\cos(\phi_1-\phi_3)\cos(\phi_3)}, \tag{6.10}$$

$$||OQ|| = \frac{||AC||\sin(\beta+\phi_1)}{2\cos(\phi_1-\phi_3)\sin(\beta+\phi_3)}, \tag{6.11}$$

$$||OK'|| = \frac{||AB||\sin(\alpha+\beta)}{2\sin(\phi_2+\phi_3)\cos(\beta)}, \tag{6.12}$$

$$||OM'|| = \frac{||AB||}{2\cos(\phi_1+\phi_2)}, \tag{6.13}$$

$$||ON'|| = \frac{||AC||}{2\cos(\phi_1-\phi_3)}, \tag{6.14}$$

where

$$||AC|| = ||AB||\frac{\cos(\alpha)}{\cos(\beta)},$$
$$||OA|| = ||AB||\frac{\cos(\alpha+\phi_2)}{\cos(\phi_1+\phi_2)},$$

and ϕ_3 is $\angle OCB$ and is given by

$$\phi_3 = \arctan\left[\frac{||AC||\cos(\beta) - ||OA||\cos(\phi_1)}{||OA||\sin(\phi_1) + ||AC||\sin(\beta)}\right].$$

Theorem 4. *Given non-isosceles triangle $\triangle ABC$, the service distances from the optimal location of a new service node to at least three different voronoi vertexes are*

equal to the longest service distance in $\triangle ABC$.

Proof. See Sec. 9.4. □

Given Theorem 4, we can pick up three different service distances from (6.6)-(6.14), make them equal to each other and obtain the result by solving the equation system. A candidate for the longest distance is obtained if the selected distance is the longest compared with other possible distances. After comparing the candidates of all possible combinations, the minimum of the longest distance and the corresponding optimal location (ϕ_1, ϕ_2) can be identified.

According to Lemma 4, at most one service distance from (6.12)-(6.14) exists. We propose a equal service distance (ESD) algorithm shown in Algorithm 2 to search the optimal location. The ESD is optimal because all the possible combinations are analyzed.

Algorithm 2 The ESD searching algorithm

```
for Every case that choose three from (6.6)-(6.11) do
    Solve the equation system
    if The selected distance is the longest among other possible distances then
        Record the result as a candidate
    end if
end for
for Everyone from (6.12)-(6.14) do
    Remove the distances to the on-boundary vertexes in the same vertex set defined in Sec. 6.3.1 from (6.6)-(6.11), denote the rest as set 𝔻
    for Every case that choose two from 𝔻 do
        Combine with the selected one from (6.12)-(6.14)
        Solve the equation system
        if The selected distance is the longest among other possible distances then
            Record the result as a candidate
        end if
    end for
end for
Compare all the candidates and find out the minimum of the longest service distance and identify the corresponding optimal location
```

Because of the difficulty to derive the closed-form solution of the equation system consisting of three equations from (6.6)-(6.14), the numerical method should be adopted, which contributes the major part of the computational load. According to Algorithm 2, we need $\binom{6}{3} = 20$ iterations for on-boundary vertexes and an additional $\binom{4}{2} \times 3 = 18$ iterations to further consider inner vertexes. Overall, 38 equation systems need to be solved to find the optimal location, which greatly reduces the complexity comparing with searching all the points exhaustively inside a triangle.

6.5 Simulation and Verification

In this section, we first compare the numerical results with the exhaustive search and other existing triangular centers for isosceles and non-isosceles cases, respectively. Then, we conduct the simulations of a UAV mesh network to reveal the impacts of the processing time and UAV's speed.

6.5.1 Isosceles Case

The exhaustive search results have been used to verify the optimal location in an acute isosceles triangle, i.e., Theorem 3. We put BC on x-axis and let its middle point be the origin point. The height of $\triangle ABC$ is fixed to 100, and we try to find the optimal location given different α. Because the optimal location is always on BC's mid-perpendicular line, which means the x-axis of the optimal location is 0. Thus, we only show the results of y-axis in Fig. 6.5. As shown in the figure, results applying Theorem 3 perfectly match the exhaustive search results.

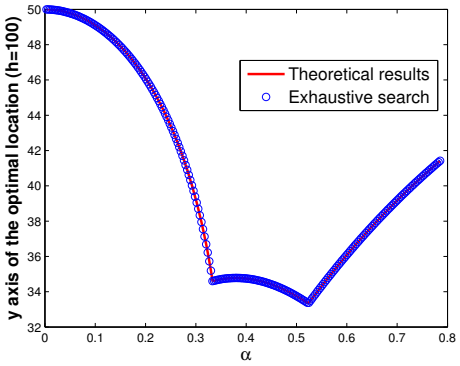
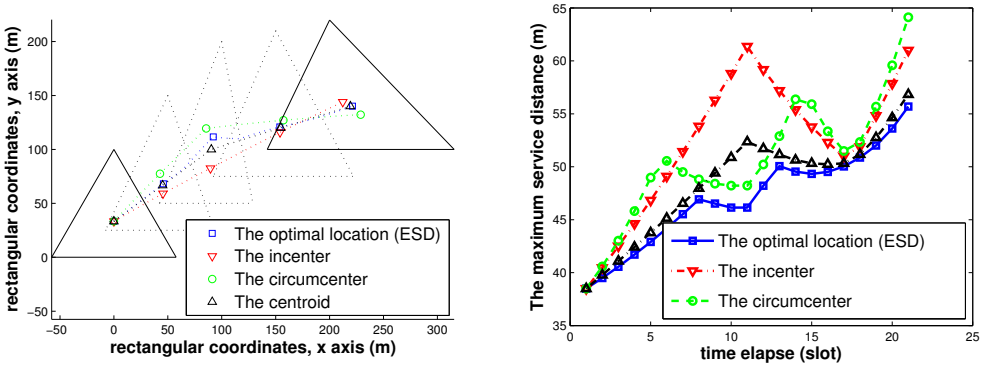


Figure 6.5: A triangle formed by three existing service nodes.

6.5.2 Non-isosceles Case

For the non-isosceles case, we examine the performance of the proposed scheme in a mobile UAV mesh backbone, i.e., a dynamic triangular warning area. The changing triangle and the trajectory of the optimal location to add the FR are shown in Fig. 6.6 (a). For comparison, the trajectories of other existing centers of a triangle, including the incenter, circumcenter, and centroid, are shown as well. We start from the equilateral triangle where all the existing centers consist with the optimal location. With the changing of the triangle, they deviate from each other. We divide the whole time duration of the procedure in Fig. 6.6 (a) into 20 equal slots, and record the maximum service distance of each slot in Fig. 6.6 (b). It can be observed that the proposed scheme can always achieve a better performance than other existing centers. On average, it can reduce the maximum service distance by 9.52%, 6.79%, and 2.86% compared to the incenter, circumcenter, and centroid, respectively.

In order to have a more comprehensive investigation, we compare the maximum service distance achieved by the proposed method with that of other existing centers exhaustively. α and β are adjusted to cover all the possible cases. The height of $\triangle ABC$, i.e., BC's perpendicular line passing through service node A, is fixed to 100 meters. The maximum service distances according to different locations of the newly

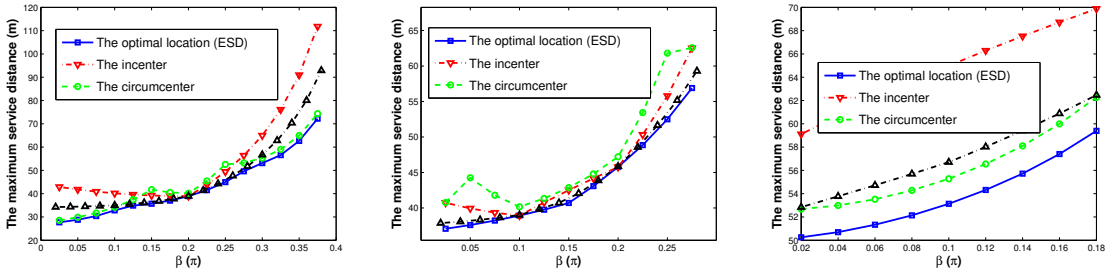


(a) The trace of the optimal location

(b) Comparisons with existing centers of a triangle

Figure 6.6: Simulation of a dynamic warning area.

added service node are summarized in Figs. 6.7. α is equal to 0.1π, 0.2π, and 0.3π in Fig. 6.7 (a), (b), and (c), respectively. It can be observed that the maximum service distance achieved by the proposed scheme is always the lower-bound, and thus verifies the optimality. The proposed ESD algorithm can reduce the maximum service distance by up to 35.71%, 15.91%, and 21.74% compared to the incenter, circumcenter, and centroid, respectively.



(a) $\alpha = 0.1\pi$ (b) $\alpha = 0.2\pi$ (c) $\alpha = 0.3\pi$

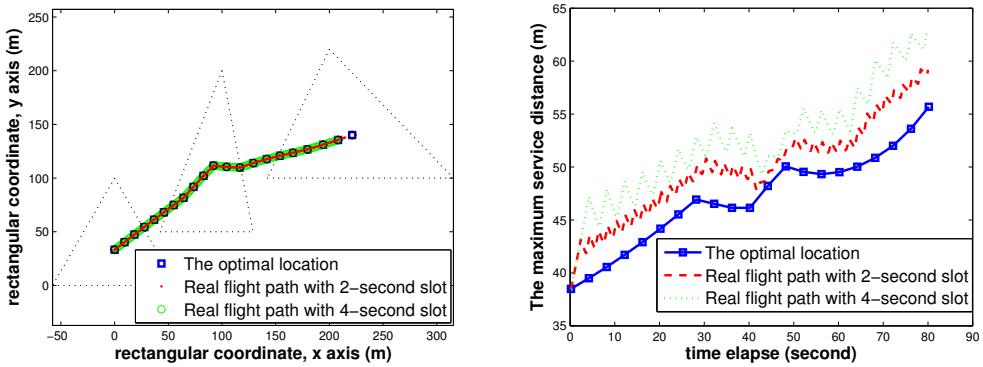
Figure 6.7: Comparisons with existing centers of a triangle.

6.5.3 Performance Evaluation Regarding to UAV's Flight Path

In this subsection, we examine the maximum service distance along the UAV's flight path. The results based on different slot duration, i.e., how long the UAV updates

its location, are shown in Fig. 6.8. We apply the same changing mesh backbone in Fig. 6.6 (a) and assume the procedure lasts for 80 seconds. The maximum actual speeds of three existing service nodes, i.e., three vertexes of the changing triangle, are 3.54m/s, 2.80m/s, and 4.83m/s, respectively. A maximum speed limit of 5m/s is assumed for the newly added UAV which always updates the location based on the latest information of the mesh.

The flight paths with 2-second and 4-second slot durations are almost the same with the trace strung by optimal locations, as shown in Fig. 6.8 (a). The maximum service distances along their flight paths are quite different as shown in Fig. 6.8 (b). Compared with 4-second slot, 2-second slot reduces the average gap to the optimal result from 6.15m to 3.03m. It is because a smaller slot duration results in a smaller deviation to the optimal location before the UAV's next update.

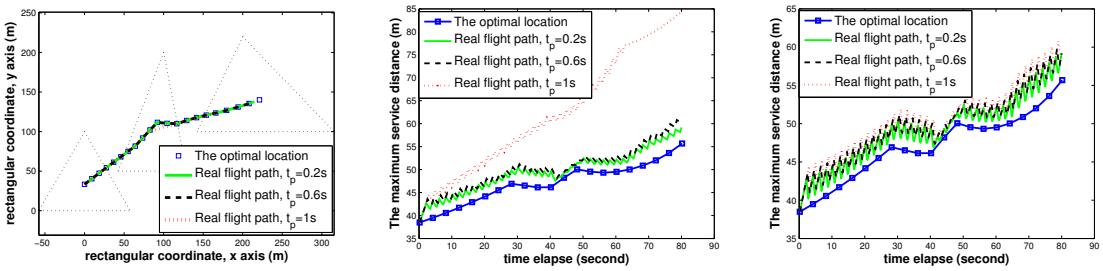


(a) The trace of the real flight path

(b) Comparisons between flight paths with different slot length

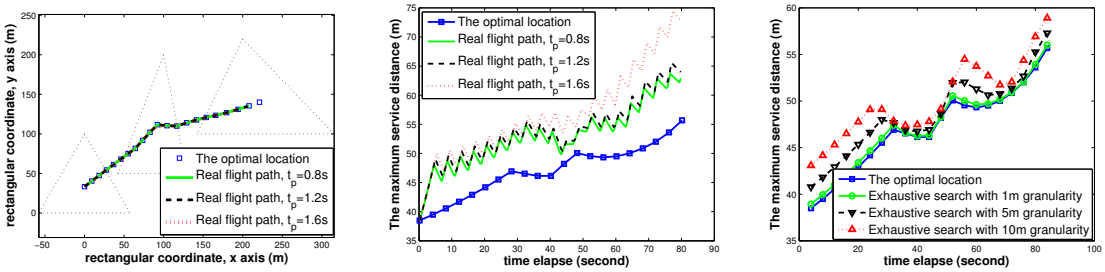
Figure 6.8: Simulation of flight paths given different slot length (t_p=0.4s).

However, the slot duration is not always the smaller the better. The processing time t_p, i.e., the time needed by a UAV to calculate the updated location, also plays an important role. We use $t_p = 0.4$s in Fig. 6.8 and the effects of different t_p are revealed in Fig. 6.9. A smaller t_p means a longer time for moving, and thus the UAV has a higher chance to reach the newly updated location. Given a fixed t_p, a small



(a) The trace of the real flight path, $v = 5$m/s (b) Comparisons between flight paths with different t_p, $v = 5$m/s (c) Comparisons between flight paths with different t_p, $v = 10$m/s

Figure 6.9: Simulation of flight paths given different t_p (slot length is 2s).



(a) The trace of the real flight path (b) Comparisons between flight paths with different t_p (c) Comparisons with exhaustive search given different searching granularities

Figure 6.10: Simulation of flight paths given different t_p (slot length is 4s) and exhaustive search given different searching granularities.

slot duration means a shorter moving time, the UAV may be unable to catch up the speed of the optimal location, e.g., $t_p = 1$s in Figs. 6.8 (a) and (b).

Given $t_p = 1$s, one way to improve the performance is to increase the UAV's speed. In Fig. 6.8 (c) it shows that the UAV keeps pace with the updated location well if the speed is increased to $v = 10$m/s. Considering mechanical issues, however, it may be difficult to further enhance the speed given the cost of a UAV. So, the slot duration needs to be prolonged to reduce the proportion of processing time. In Figs. 6.10 (a) and (b), the system can support up to $t_p = 1.2$s if the slot length is 4s. But the average gap of the maximum service distance to the optimal result is enlarged to 6.49m. This observation justify the necessity of the proposed low-complexity ESD

algorithm, by which a smaller slot duration can be used to achieve a better performance. Exhaustive search results with different granularities are shown in Fig. 6.10 (c). A coarser granularity reduces the complexity but significantly increases the gap between the updated locations and the optimal ones. Otherwise, it results in a large t_p and is not suitable for real-time updating especially when the size of the warning area, i.e., the triangle, is large.

6.6 Conclusion

For the UAV mesh network, we have proposed a second stage in addition to the existing seamless coverage (the first stage), where a supplementary UAV is introduced to further support URLLC in the warning area. A min-max optimization problem has been formulated to minimize the longest service distance in an arbitrary acute triangle. For the newly added service node, the closed-form expression of the optimal location within an isosceles triangle has been theoretically derived. An ESD algorithm with low complexity has been proposed to find the optimal location of the new service node within a non-isosceles triangle as well. The numerical comparisons with the exhaustive search results and triangle's existing centers, including incenter, circumcenter, and centroid, have been conducted, according to which the optimality of the proposed method has been validated. The impacts of UAV's speed and processing time on the performance have been studied.

This work can be applied to other fields where a new facility needs to be added to improve the experience of the worst user. The worst user is the customer with the lowest QoS, which can be the slowest response time in police, ambulance, or fire services, the longest travel distance to a supermarket, a postal office, or a recreation center. The worst possible experience is also important for a service provider because

it determines the guaranteed QoS advertised.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this dissertation, we have studied the combination of NC and D2D for MTC in the cellular system, introduced UAVs to assist the communication and support various types of traffic generated by MTC, MI, and URLLC. The following outlines the contributions we have achieved.

- In Chapter 2, based on the MTC characteristics described above, we have proposed an efficient scheme combined with NC and D2D communication for the multiple unicast scenario, which can substantially reduce the cellular resource consumption while the total UE energy consumption can also be reduced. A feasible system design including the protocol stack has been given, which is backward-compatible with the current LTE/LTE-A system and easy to implement. The eNB can fully control the transmission, and the security in RAN will not be affected. The closed-form probability mass functions of transmission times in both the cellular and D2D phases have been derived. The design and analysis can be extended to multicast scenarios. The error rates in both the cellular and D2D transmissions have been taken into account. The feedback load has also been analyzed.

Numerical and simulation results corresponding to different channel settings have been given, which can be used as references for the eNB to optimize the configuration of modulation and coding. Numerical and simulation results show that substantial reduction of cellular transmission times can be achieved comparing with the legacy LTE system. Meanwhile, the consumption of either the UE transmit power or the system transmit power can also be reduced. We also observe that with a higher BLER in the cellular transmissions, we can achieve a higher gain, so the proposed solution is more promising in the situation of poor channel conditions.

- In Chapter 3, we have proposed an efficient semi-centralized cooperative control method for the uplink transmissions in cellular systems. Two specific cooperative schemes based on D2D have been proposed: one is the random UE relay scheme and the other further applies NC. The proposed schemes can substantially increase the transmission efficiency while reducing the overhead, scheduling queue length, and devices' power consumption. A feasible system design including the protocol stack has been given, which is backward-compatible with the current LTE/LTE-A system and easy to implement. The system has been modeled based on two distinct traffic models, i.e., the MTC small-data traffic and the full-buffer traffic. The D2D interference is considered and modeled applying stochastic geometry.

  Extensive simulations have been conducted for the MTC small-data traffic in various scenarios to identify the performance gain and compared the two cooperative schemes. Finally, the performance with full-buffer traffic has been analyzed theoretically. The closed-form results have been obtained, according to which extensive numerical evaluations are then performed. These results provide important guidelines for the eNB, such as when each of cooperative scheme

is preferable, and how to determine the optimal density of the UE relays.

- In Chapter 4, considering the high penetration loss/deep shadowing channels of the MTC devices, we introduced the FR into the cellular system to improve the transmission efficiency and maximize the system throughput. Given the capacity limit of the FR's back-haul link and the maximum transmission power of each user, an optimization problem has been formulated to maximize the system throughput. The optimal power allocation strategy has been derived, and two effective online FR placement algorithms, i.e., the WCA for the unpredictable network and the A-WCA targeting on the predictable network, have been proposed. Two comparative off-line schemes, i.e., the direct and the SDP methods, have also been described.

  Extensive simulations have been conducted. The results show that the WCA algorithm can reach the near-optimal system throughput given the homogeneous shadowing and can find the global optimal location within a certain range for most of the identical shadowing cases. Collecting the statistic information and predicting the network's changing pattern is desirable in the independent shadowing case, where the WCA is easily misled by local maximizers and the A-WCA shows a significant advantage thanks to its adaptive step length in the searching process.

  More simulations focusing on the WCA algorithm have been included. When the homogeneous shadowing is applied, the convergence rate of the WCA algorithm is close to the upper-bound and its accumulated throughput along the flight path is as good as or even better than the offline schemes. However, the WCA's performance is worse in the independent shadowing case, which provides an insight on how the performance varies according to the ruggedness of the maximum system throughput over a 3-D area. It can be considered as the

reference to determine the necessity of introducing the network prediction and the A-WCA algorithm.

- In Chapter 5, we have proposed the UABS to solve the problems brought by the increasing traffic volume of MI and by serving MTC devices with special traffic characteristics and locations. It utilizes UAV based FR to enable heterogeneous deployment of additional FR-cells inside the macro cell and achieves dynamic and adaptive coverage. We focused on how to deploy the FR-cells and the associated issues. Comprehensive analyses on FR-cells deployment including frequency reuse, interference, backhaul resource allocation, and coverage were given.

- In Chapter 6, we have proposed a second stage in addition to the existing seamless UAV mesh network, where a supplementary UAV is introduced to further support URLLC in the warning area. A min-max optimization problem has been formulated to minimize the longest service distance in an arbitrary acute triangle. For the newly added service node, the closed-form expression of the optimal location within an isosceles triangle has been theoretically derived. An ESD algorithm with low complexity has been proposed to find the optimal location of the new service node within a non-isosceles triangle as well. The numerical comparisons with the exhaustive search results and triangle's existing centers, including incenter, circumcenter, and centroid, have been conducted, according to which the optimality of the proposed method has been validated. The impacts of UAV's speed and processing time on the performance have been studied.

  This work can be applied to other fields where a new facility needs to be added to improve the experience of the worst user. The worst user is the customer with

the lowest QoS, which can be the slowest response time in police, ambulance, or fire services, the longest travel distance to a supermarket, a postal office, or a recreation center. The worst possible experience is also important for a service provider because it determines the guaranteed QoS advertised.

7.2 Future Work

For the future work that plans beyond this dissertation, there are still many open issues of importance.

- There are many possibilities of applying the solution proposed by Chapter 2 in scenarios other than MTC transmissions. If the dynamic scheduling is applied, how to adjust the MCS and the target receiving BLER to improve the overall throughput is worth further research. Besides that, given certain QoS requirements, how the eNB decides the mature UE number target is another interesting problem. The coordination/scheduling mechanism between the cellular and D2D transmissions has not been deeply studied, which requires further investigation. It is possible to schedule concurrent D2D transmissions, if the mature UEs in the same group are far away from each other, to achieve the spatial multiplexing gain, while the signaling messages exchanged between UEs may increase. The trade-off of them needs further research.

- In Chapter 3, the current grouping method based on randomly selected D2D agents is efficient but not optimal. How to tune parameter a according to UE-specific features, such as the UE's receiving power of the eNB's downlink pilot, to further optimize the grouping requires further investigation. In that way, the D2D agents may no longer follow a PPP, and some of the theoretical results in this chapter need to be revisited. For the proposed schemes, the advantage

of the control overhead reduction has not been quantified yet, which is worth studying in the future. Due to the heavy control/feedback overhead and high complexity of advanced multi-antenna techniques, we do not consider them in this work. However, the trade-off between the high spectrum efficiency and the overhead may be an interesting further research issue. In this work, the minimum performance gain based on extreme cases has be achieved, but the exact performance gain that can be obtained in reality is still an open issue. This part needs to consider the detailed scheduling algorithm design for the concurrent D2D transmissions, which plays an important role in the overall system performance and beckons for further investigation.

- In Chapter 4, many open research issues are beckoning further studies. In this chapter, we tried to find the optimal location for the FR to maximize the system throughput. The fairness among users was not considered, which is an important issue and given the channel dynamics, there exists a trade-off between overall throughput and fairness. Fairness can be considered by defining a utility function as the product of each user's throughput, or as the summation of a log function of each user's throughput. The proposed approach can be applied to maximize the utility to make a trade-off between the total throughput and fairness. Another issue is the traffic model. In this chapter, we use the sum of the Shannon capacities of all the UEs' front-haul links and serving links to approximate the system throughput. The capacity may be achieved only if the traffic is the full-buffer type. The users' buffer status should be further considered. The system can also be optimized based on the delay tolerance of the traffic with the constraint of the FR's speed limit. The energy consumption combining with the solar-powered UAV in another important further research issue. The FR will not sit on one single optimal location. Instead, it should

cruise along an optimized closed route. Given the solar charging rate varies in different areas, the proportion of the route within the high charging rate area should be above a threshold to guarantee the sustainability. With this new constraint, an optimization problem can be formulated. Finally, we assumed that the FR is connected to a fixed serving BS in this chapter. An optimized handover triggering mechanism enabled by the cooperation between existing stationary BSs will further improve the system performance.

- In Chapter 5, instead of using the interference model of the extreme case, given a certain distribution of the macro users locations, the average interference coming from the whole macro area can be obtained for an alternative solution, which may result in a more precise performance gain. A constant spectrum efficiency of the macro cell, η_M, is assumed in the procedure of finding the optimal bandwidth for each FR-backhaul. When the number of the macro users is not large enough, or the deduction of the FR-cell covered users greatly affects the distribution of the macro users locations, the spectrum efficiency of the macro cell will be changed. How to model this change and solve the optimization problem needs further investigation. Besides the two-layer coverage extension method in 2-D (2-dimension) plane, the 3-D coverage extension, which is applicable for the case of skyscrapers, needs more in-depth analyses. When the FR extends the coverage, an optimized non-constant cruising speed and the energy consumption model can be applied to further increase the average FR-connection probability while optimizing FRs battery life.

- In Chapter 6, we focus on the acute triangle, which may impose restrictions on UAVs' deployment for the first-stage coverage, i.e., ensuring all the warning areas are acute triangles. Obtuse triangles can be studied and thus provide a higher flexibility. Homogeneous service nodes have been assumed in this chapter

which means all the nodes have the same capability, e.g., transmitting power in wireless communication, and the service quality only depends on the service distance. Heterogeneous service nodes and differentiated areas, e.g., shadowing in wireless communication, are worth for research. Based on a newly defined optimizing objective combining the service node's capability, the area features, and the service distance, the solution will be applicable to more scenarios. Besides, only one new service node in a 2-D plane has been considered in this chapter. How to coordinate multiple supplementary service nodes in a 3-D space for the same purpose beckons for further investigation.

Chapter 8

Publications

1. **Yue Li**, Hamed Mosavat-Jahromi, Lin Cai, and Jianping Pan, "Prediction and Modeling of Spectrum Occupancy for Cognitive Radio Communication Systems," submitted to INFOCOM 2019.

2. **Yue Li**, Yongmin Zhang and Lin Cai, "Placement of Supplementary Node in UAV Surveillance System," ready to submit.

3. **Yue Li**, Guangsheng Feng, Mohammad Ghasemiahmadi, and Lin Cai, "Power Allocation and Flight Path Planning for Floating Relay Supporting MTC Traffic in Cellular Systems," accepted by *IEEE Transactions on Mobile Computing*, 2018.

4. Zhe Wei, **Yue Li**, and Lin Cai, "Electric Vehicle Charging Scheme for a Park-and-Charge System Considering Battery Degradation Costs," accepted by *IEEE Transactions on Intelligent Vehicles*, 2018.

5. **Yue Li** and Lin Cai, "Cooperative Device-to-Device Communication for Uplink Transmission in Cellular System," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3903-3917, Jan. 2018.

6. Zhe Wei, **Yue Li**, Yongmin Zhang, and Lin Cai, "Intelligent Parking Garage EV Charging Scheduling Considering Battery Charging Characteristic," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2806-2816, 2018.

7. **Yue Li**, Kai Sun, and Lin Cai, "Cooperative Device-to-Device Communication With Network Coding for Machine Type Communication Devices," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 296-309, Jan. 2018.

8. **Yue Li** and Lin Cai, "UAV-assisted Dynamic Coverage in a Heterogeneous Cellular System," *IEEE Network*, vol. 31, no. 4, pp. 56-61, 2017.

9. Mohammad Ghasemiahmadi, **Yue Li**, and Lin Cai, "RSS-based Grouping Strategy for Avoiding Hidden Terminals with GS-DCF MAC Protocol," in *Proc. of IEEE WCNC*, 2017, pp. 1-5.

10. **Yue Li**, Mohammad Ghasemiahmadi, and Lin Cai, "Performance Analysis of Semi-Centralized Controlled Uplink Cooperative Transmission," in *Proc. of IEEE GLOBECOM*, 2016, pp. 1-6.

11. **Yue Li**, Mohammad Ghasemiahmadi, and Lin Cai, "Uplink Cooperative Transmission for Machine-Type-Communication Traffic in Cellular System," in *Proc. of IEEE VTC-Fall*, 2016, pp. 1-5.

12. Guangsheng Feng, **Yue Li**, Qian Zhao, Huiqiang Wang, Hongwu Lv and Junyu Lin, "Optimizing Broadcast Duration for Layered Video Streams in Cellular Networks," *Peer-to-Peer Networking and Applications*, pp. 1-15, 2016.

13. **Yue Li**, Kai Sun, and Lin Cai, "Small-Cell Planning Based on Uplink Interference and Traffic in Two-Tier Cellular System," in *Proc. of IEEE WCSP*, 2015, pp. 1-6.

14. **Yue Li**, Song Zhu, and Xiaolong Guo, "System Design for Multiple Users Cooperative Communication in LTE," in *Proc. of IEEE VTC-Fall*, 2013, pp. 1-5.

The work reported in Chapter 2 has been published in 7; The work reported in Chapter 3 will be soon published in 5; The work reported in Chapter 4 will be soon published in 3; The work reported in Chapter 5 has been published in 8; The work reported in Chapter 6 has been included in 2.

Bibliography

[1] Rapeepat Ratasuk, Athul Prasad, Zexian Li, Amitava Ghosh, and Mikko Uusitalo. Recent advancements in M2M communications in 4G networks and evolution towards 5G. In *Proc. of IEEE ICIN*, pages 52–57, 2015.

[2] 3GPP. Study on Enhancements to Machine-Type Communications (MTC) and other Mobile Data Applications; Release 12. TR 37.869, 3rd Generation Partnership Project (3GPP), 2013.

[3] 3GPP. Service requirements for Machine-Type Communications (MTC); Stage 1; Release 13. TS 22.368, 3rd Generation Partnership Project (3GPP), 2014.

[4] 3GPP. Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE; Release 12. TR 36.888, 3rd Generation Partnership Project (3GPP), 2013.

[5] RP-130396. Study on RAN aspects of Machine-Type and other mobile data applications Communications enhancements. Technical report, RAN-59, 2013.

[6] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures; Release 12. TS 36.213, 3rd Generation Partnership Project (3GPP), 2015.

[7] R1-130462. Feasibility of coverage extension of physical channels for MTC devices. Technical report, Alcatel-Lucent, Alcatel-Lucent Shanghai Bell, 2013.

[8] R1-132970. Simplification of MTC UE operation in coverage enhancement. Technical report, Alcatel-Lucent, Alcatel-Lucent Shanghai Bell, 2013.

[9] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding; Release 12. TS 36.212, 3rd Generation Partnership Project (3GPP), 2015.

[10] Afif Osseiran, Federico Boccardi, Volker Braun, Katsutoshi Kusume, Patrick Marsch, Michal Maternia, Olav Queseth, Malte Schellmann, Hans Schotten, Hidekazu Taoka, et al. Scenarios for 5G mobile and wireless communications: the vision of the metis project. *IEEE Commun. Mag.*, 52(5):26–35, 2014.

[11] 3GPP. Proximity-Based Services (ProSe); Stage 2; Release 12. TS 23.303, 3rd Generation Partnership Project (3GPP), 2014.

[12] Lei Lei, Zhangdui Zhong, Chuang Lin, and Xuemin Shen. Operator controlled device-to-device communications in LTE-advanced networks. *IEEE Wireless Commun.*, 19(3), 2012.

[13] Amila Tharaperiya Gamage, Hao Liang, Ran Zhang, and Xuemin Shen. Device-to-Device communication underlaying converged heterogeneous networks. *IEEE Wireless Commun.*, 21(6):98–107, 2014.

[14] David Astely, Erik Dahlman, Gabor Fodor, Stefan Parkvall, and Joachim Sachs. LTE release 12 and beyond. *IEEE Commun. Mag.*, 51(7):154–160, 2013.

[15] Serveh Shalmashi and Slimane Ben Slimane. Cooperative device-to-device communications in the downlink of cellular networks. In *Proc. of IEEE WCNC*, pages 2265–2270, 2014.

[16] Jiajia Liu, Yasutaka Kawamoto, Hiroki Nishiyama, Nei Kato, and Naoto Kadowaki. Device-to-Device communications achieve efficient load balancing in LTE-advanced networks. *IEEE Wireless Commun.*, 21(2):57–65, 2014.

[17] Bo Hu, H Vicky Zhao, and Hai Jiang. Wireless multicast using relays: Incentive mechanism and analysis. *IEEE Trans. Veh. Technol.*, 62(5):2204–2219, 2013.

[18] Tuan Ta, John S Baras, and Chenxi Zhu. Improving smartphone battery life utilizing device-to-device cooperative relays underlaying LTE networks. In *Proc. of IEEE ICC*, pages 5263–5268, 2014.

[19] Nuno K Pratas and Petar Popovski. Low-rate machine-type communication via wireless device-to-device (D2D) links. *arXiv preprint arXiv:1305.6783*, 2013.

[20] Xiao Lu, Hai Jiang, Dusit Niyato, Dong In Kim, and Zhu Han. Wireless-powered device-to-device communications with ambient backscattering: Performance modeling and analysis. *IEEE Trans. Wireless Commun.*, 17(3):1528–1544, 2018.

[21] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2; (Release 13). TS 36.300, 3rd Generation Partnership Project (3GPP), 2015.

[22] Jay Kumar Sundararajan, Devavrat Shah, Muriel Médard, Szymon Jakubczak, Michael Mitzenmacher, and Joao Barros. Network coding meets TCP: Theory and implementation. *Proceedings of the IEEE*, 99(3):490–512, 2011.

[23] Rudolf Ahlswede, Ning Cai, S-YR Li, and Raymond W Yeung. Network information flow. *IEEE Trans. Inf. Theory*, 46(4):1204–1216, 2000.

[24] S-YR Li, Raymond W Yeung, and Ning Cai. Linear network coding. *IEEE Trans. Inf. Theory*, 49(2):371–381, 2003.

[25] Jay Kumar Sundararajan, Devavrat Shah, and Muriel Médard. ARQ for network coding. In *Proc. of IEEE ISIT*, pages 1651–1655, 2008.

[26] Ting Zhou, Bin Xu, Tianheng Xu, Honglin Hu, and Lei Xiong. User-specific link adaptation scheme for device-to-device network coding multicast. *IET Commun.*, 9(3):367–374, 2015.

[27] Salim El Rouayheb, Alex Sprintson, and Parastoo Sadeghi. On coding for cooperative data exchange. In *Proc. of IEEE ITW*, pages 1–5, 2010.

[28] Nestor J Hernandez Marcano, Janus Heide, Daniel E Lucani, and Frank HP Fitzek. On the throughput and energy benefits of network coded cooperation. In *Proc. of IEEE CloudNet*, pages 138–142, 2014.

[29] Andrea Tassi, Francesco Chiti, Romano Fantacci, and Fabio Schoen. An energy efficient resource allocation scheme for RLNC-based heterogeneous multicast communications. *IEEE Commun. Letters*, 18(8):1399–1402, 2014.

[30] Shahriar Etemadi Tajbakhsh and Parastoo Sadeghi. Coded cooperative data exchange for multiple unicasts. In *Proc. of IEEE ITW*, pages 587–591, 2012.

[31] Afif Osseiran, Klaus Doppler, Cassio Ribeiro, Ming Xiao, Mikael Skoglund, and Jawad Manssour. Advances in device-to-device communications and network coding for IMT-advanced. *ICT Mobile Summit*, 2009.

[32] Yue Li, Mohammad Ghasemiahmadi, and Lin Cai. Uplink cooperative transmission for machine-type communication traffic in cellular system. In *Proc. of IEEE VTC FALL*, pages 1–5, 2016.

[33] Yue Li, Mohammad Ghasemiahmadi, and Lin Cai. Performance analysis of semi-centralized controlled uplink cooperative transmission. In *Proc. of IEEE GLOBECOM*, pages 1–6, 2016.

[34] Andrea Abrardo, Gabor Fodor, and Besmir Tola. Network coding schemes for device-to-device communications based relaying for cellular coverage extension. In *Proc. of IEEE SPAWC*, pages 670–674, 2015.

[35] S-YR Li, Raymond W Yeung, and Ning Cai. Linear network coding. *IEEE Trans. Inf. Theory*, 49(2):371–381, 2003.

[36] David JC MacKay. Fountain codes. *IEE Proceedings-Communications*, 152(6):1062–1068, 2005.

[37] Jalaluddin Qureshi. Random linear fountain code with improved decoding success probability. In *Proc. of IEEE APCC*, pages 6–11, 2016.

[38] 3GPP. User Equipment (UE) radio access capabilities; Release 12. TS 36.306, 3rd Generation Partnership Project (3GPP), 2015.

[39] Christos Gkantsidis, John Miller, and Pablo Rodriguez. Comprehensive view of a live network coding P2P system. In *Proc. of ACM SIGCOMM*, pages 177–188, 2006.

[40] SP-150149. 5G timeline in 3GPP. Technical report, 2015.

[41] Yue Li, Song Zhu, and Xiaolong Guo. System design for multiple users cooperative communication in LTE. In *Proc. of IEEE VTC FALL*, pages 1–5, 2013.

[42] Gábor Fodor, Erik Dahlman, Gunnar Mildh, Stefan Parkvall, Norbert Reider, György Miklós, and Zoltán Turányi. Design aspects of network assisted device-to-device communications. *IEEE Commun. Mag.*, 50(3):170–177, 2012.

[43] Minming Ni, Jianping Pan, and Lin Cai. Geometrical-based throughput analysis of device-to-device communications in a sector-partitioned cell. *IEEE Trans. Wireless Commun.*, 14(4):2232–2244, 2015.

[44] Minming Ni, Lei Zheng, Fei Tong, Jianping Pan, and Lin Cai. A geometrical-based throughput bound analysis for device-to-device communications in cellular networks. *IEEE J. Sel. Areas Commun.*, 33(1):100–110, 2015.

[45] 3GPP. Study on LTE Device to Device Proximity Services; Radio Aspects; Release 12. TR 36.843, 3rd Generation Partnership Project (3GPP), 2014.

[46] Chen Xu, Lingyang Song, and Zhu Han. *Resource management for device-to-device underlay communication*. Springer, 2014.

[47] Rose Qingyang Hu and Yi Qian. *Resource Management for Heterogeneous Networks in LTE Systems*. Springer, 2014.

[48] Ye Li, Bertan Bakkaloglu, and Chaitali Chakrabarti. A system level energy model and energy-quality evaluation for integrated transceiver front-ends. *IEEE Trans. VLSI Systems*, 15(1):90–103, 2007.

[49] ITU. Link-level Simulation Results for IMT.EVAL. Technical report, TTA PG707.

[50] Kiran Vanganuru, Matthew Puzio, Gregory Sternberg, Kandarp Shah, and Samian Kaur. Uplink system capacity of a cellular network with cooperative mobile relay. In *Proc. of IEEE WTS*, pages 1–7, 2011.

[51] Lefei Wang, Tao Peng, Yufeng Yang, and Wenbo Wang. Interference constrained relay selection of D2D communication for relay purpose underlaying cellular networks. In *Proc. of IEEE WiCOM*, pages 1–5, 2012.

[52] Seung Min Yu and Seong-Lyun Kim. Downlink capacity and base station density in cellular networks. In *Proc. of IEEE WiOpt*, pages 119–124, 2013.

[53] Pramod Viswanath, David NC Tse, and Rajiv Laroia. Opportunistic beamforming using dumb antennas. *IEEE Trans. Inf. Theory*, 48(6):1277–1294, 2002.

[54] Mohammad Torabi, Wessam Ajib, and David Haccoun. Performance analysis of rate-adaptive scheduling in MIMO systems with antenna selection. In *Proc. of IEEE PIMRC*, pages 1–6, 2008.

[55] Lin Yang and Mohamed-Slim Alouini. Performance analysis of multiuser selection diversity. *IEEE Trans. Veh. Technol.*, 55(6):1848–1861, 2006.

[56] Yue Li. Cooperative device-to-device communication for uplink transmission in cellular system. `http://web.uvic.ca/~liyue331/3.pdf`, 2016.

[57] Lu Ding, Jinwei He, and Zhimin Liu. A novel open loop power control method in LTE uplink. In *Proc. of IEEE WiCOM*, pages 1–4, 2012.

[58] Bhaskar Saha, Edwin Koshimoto, Cuong C Quach, Edward F Hogge, Thomas H Strom, Boyd L Hill, Sixto L Vazquez, and Kai Goebel. Battery health management system for electric uavs. In *Aerospace Conference, 2011 IEEE*, pages 1–9.

[59] Charles D Gavrilovich. Mobile communication system with moving base station, March 17 1998. US Patent 5,729,826.

[60] Google. Project loon, 2016.

[61] MARK ZUCKERBERG. The technology behind aquila, 2016.

[62] Ruslan Kirichek, Alexandr Paramonov, and Andrey Koucheryavy. Flying ubiquitous sensor networks as a queueing system. In *Proc. of IEEE ICACT*, pages 127–132, 2015.

[63] Abhijeet Alkesh, Ashutosh Kumar Singh, and N Purohit. A moving base station strategy using fuzzy logic for lifetime enhancement in wireless sensor network. In *Communication Systems and Network Technologies (CSNT), 2011 International Conference on*, pages 198–202. IEEE, 2011.

[64] YoungSang Yun and Ye Xia. Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications. *IEEE Transactions on mobile computing*, 9(9):1308–1318, 2010.

[65] Jun Luo, Jacques Panchard, Michał Piórkowski, Matthias Grossglauser, and Jean-Pierre Hubaux. Mobiroute: Routing towards a mobile sink for improving lifetime in sensor networks. In *International Conference on Distributed Computing in Sensor Systems*, pages 480–497. Springer, 2006.

[66] Xing Xu, Ji Luo, and Qian Zhang. Delay tolerant event collection in sensor networks with mobile sink. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[67] Mohammed Abo-Zahhad, Sabah M Ahmed, Nabil Sabor, and Shigenobu Sasaki. Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks. *IEEE sensors journal*, 15(8):4576–4586, 2015.

[68] Ioannis Chatzigiannakis, Athanasios Kinalis, and Sotiris Nikoletseas. Efficient data propagation strategies in wireless sensor networks using a single mobile sink. *Computer Communications*, 31(5):896–914, 2008.

[69] Mariam Akbar, Nadeem Javaid, Ayesha Hussain Khan, Muhammad Imran, Muhammad Shoaib, and Athanasios Vasilakos. Efficient data gathering in 3d linear underwater wireless sensor networks using sink mobility. *Sensors*, 16(3):404, 2016.

[70] Ilker Bekmezci, Ozgur Koray Sahingoz, and Şamil Temel. Flying ad-hoc networks (fanets): A survey. *Ad Hoc Networks*, 11(3):1254–1270, 2013.

[71] Shuai Gao, Hongke Zhang, and Sajal K Das. Efficient data collection in wireless sensor networks with path-constrained mobile sinks. *IEEE Transactions on Mobile Computing*, 10(4):592–608, 2011.

[72] Alberto Jimenez-Pacheco, Denia Bouhired, Yannick Gasser, Jean-Christophe Zufferey, Dario Floreano, and Bixio Rimoldi. Implementation of a wireless mesh network of ultra light mavs with dynamic routing. In *2012 IEEE Globecom Workshops*, pages 1591–1596. IEEE, 2012.

[73] SenseFly. ebee, 2016.

[74] David Shallcross and Eric van den Berg. Autonomous location of mobile wireless relay nodes. In *Unmanned Aircraft Systems (ICUAS), 2016 International Conference on*, pages 15–24. IEEE, 2016.

[75] Edison Pignaton de Freitas, Tales Heimfarth, Ivayr Farah Netto, Carlos Eduardo Lino, Carlos Eduardo Pereira, Armando Morado Ferreira, Flávio Rech Wagner, and Tony Larsson. UAV relay network to support WSN connectiv-

ity. In *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2010 International Congress on*, pages 309–314. IEEE, 2010.

[76] Constantine Mukasa, Valentine A Aalo, and George Efthymoglou. On the performance of a dual-hop network with a mobile relay in a nakagami fading environment. In *Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2016 IEEE 21st International Workshop on*, pages 43–47. IEEE, 2016.

[77] Jiangbin Lyu, Yong Zeng, and Rui Zhang. Cyclical multiple access in uav-aided communications: A throughput-delay tradeoff. *IEEE Wireless Communications Letters*, 5(6):600–603, 2016.

[78] Yong Zeng, Rui Zhang, and Teng Joon Lim. Wireless communications with unmanned aerial vehicles: opportunities and challenges. *IEEE Communications Magazine*, 54(5):36–42, 2016.

[79] 3GPP. Further advancements for E-UTRA physical layer aspects; Release 9. TR 36.814, 3rd Generation Partnership Project (3GPP), 2017.

[80] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[81] Peyman MohajerinEsfahani, Debasish Chatterjee, and John Lygeros. Motion planning for continuous time stochastic processes: A dynamic programming approach. 2012.

[82] Ingo Forkel, Marc Schinnenburg, and Markus Ang. Generation of two-dimensional correlated shadowing for mobile radio network simulation. *WPMC, sep*, 21:43, 2004.

[83] John Hoadley and Payam Maveddat. Enabling small cell deployment with HetNet. *IEEE Wireless Commun.*, 19(2):4–5, 2012.

[84] Imad Jawhar, Nader Mohamed, Jameela Al-Jaroodi, and Sheng Zhang. A framework for using unmanned aerial vehicles for data collection in linear wireless sensor networks. *J. Intell. Rob. Syst.*, 74(1-2):437, 2014.

[85] Marco AM Marinho, Edison Pignaton De Freitas, João Paulo C Lustosa da Costa, André Lima F de Almeida, and Rafael Timóteo de Sousa. Using cooperative MIMO techniques and UAV relay networks to support connectivity in sparse wireless sensor networks. In *IEEE ComManTel*, pages 49–54, 2013.

[86] Feng Jiang and A Lee Swindlehurst. Optimization of UAV heading for the ground-to-air uplink. *IEEE J. Sel. Areas Commun.*, 30(5):993–1005, 2012.

[87] Pengcheng Zhan, Kai Yu, and A Lee Swindlehurst. Wireless relay communications with unmanned aerial vehicles: Performance and optimization. *IEEE Trans. Aerosp. Electron. Syst.*, 47(3):2068–2085, 2011.

[88] Oleg Burdakov, Patrick Doherty, Kaj Holmberg, and Per-Magnus Olsson. Optimal placement of UV-based communications relay nodes. *J. Global Optim.*, 48(4):511–531, 2010.

[89] Zubair Md Fadlullah, Daisuke Takaishi, Hiroki Nishiyama, Nei Kato, and Ryu Miura. A dynamic trajectory control algorithm for improving the communication throughput and delay in UAV-aided networks. *IEEE Network*, 30(1):100–105, 2016.

[90] Weisi Guo, Conor Devine, and Siyi Wang. Performance analysis of micro unmanned airborne communication relays for cellular networks. In *IEEE CSNDSP*, pages 658–663, 2014.

[91] Arvind Merwaday and Ismail Guvenc. UAV assisted heterogeneous networks for public safety communications. In *IEEE WCNCW*, pages 329–334, 2015.

[92] Vishal Sharma, Mehdi Bennis, and Rajesh Kumar. UAV-assisted heterogeneous networks for capacity enhancement. *IEEE Commun. Lett.*, 20(6):1207–1210, 2016.

[93] Jiajia Liu, Yuichi Kawamoto, Hiroki Nishiyama, Nei Kato, and Naoto Kadowaki. Device-to-device communications achieve efficient load balancing in LTE-advanced networks. *IEEE Wireless Commun.*, 21(2):57–65, 2014.

[94] Jiajia Liu, Shangwei Zhang, Nei Kato, Hirotaka Ujikawa, and Kenichi Suzuki. Device-to-device communications for enhancing quality of experience in software defined multi-tier LTE-A networks. *IEEE Network*, 29(4):46–52, 2015.

[95] Kai Daniel and Christian Wietfeld. Using public network infrastructures for UAV remote sensing in civilian security operations. Technical report, DORTMUND UNIV (GERMANY FR), 2011.

[96] Wei Li, Changxin Huang, Kaimin Chen, and Songchen Han. A quantitative evaluation method of surveillance coverage of uavs swarm. In *International Conference on Cloud Computing and Security*, pages 772–782. Springer, 2017.

[97] Nikhil Nigam, Stefan Bieniawski, Ilan Kroo, and John Vian. Control of multiple uavs for persistent surveillance: algorithm and flight test results. *IEEE Trans. Control Syst. Technol.*, 20(5):1236–1251, 2012.

[98] Martin Saska, Vojtěch Vonásek, Jan Chudoba, Justin Thomas, Giuseppe Loianno, and Vijay Kumar. Swarm distribution and deployment for cooperative surveillance by micro-aerial vehicles. *J. Intell. Rob. Syst.*, 84(1-4):469–492, 2016.

[99] Jianping Pan, Lin Cai, Yiwei Thomas Hou, Yi Shi, and Sherman X Shen. Optimal base-station locations in two-tiered wireless sensor networks. *IEEE Trans. Mob. Comput.*, 4(5):458–473, 2005.

[100] Yue Li and Lin Cai. UAV-assisted dynamic coverage in a heterogeneous cellular system. *IEEE Network*, 31(4):56–61, 2017.

[101] Mohamed Alzenad, Amr El-Keyi, Faraj Lagum, and Halim Yanikomeroglu. 3d placement of an unmanned aerial vehicle base station (uav-bs) for energy-efficient maximal coverage. *IEEE Wireless Commun. Lett.*, 2017.

[102] R Irem Bor-Yaliniz, Amr El-Keyi, and Halim Yanikomeroglu. Efficient 3-d placement of an aerial base station in next generation cellular networks. In *Proc. of IEEE ICC*, pages 1–5, 2016.

[103] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage. *IEEE Commun. Lett.*, 20(8):1647–1650, 2016.

[104] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs. *IEEE Trans. Wireless Commun.*, 15(6):3949–3963, 2016.

[105] Elham Kalantari, Muhammad Zeeshan Shakir, Halim Yanikomeroglu, and Abbas Yongacoglu. Backhaul-aware robust 3d drone placement in 5g+ wireless networks. *arXiv preprint arXiv:1702.08395*, 2017.

[106] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Mobile internet of things: Can uavs provide an energy-efficient mobile architecture? In *Proc. of IEEE GLOBECOM*, pages 1–6, 2016.

[107] Mohammad Mozaffari, Walid Saad, Mehdi Bennis, and Mérouane Debbah. Mobile unmanned aerial vehicles (uavs) for energy-efficient internet of things communications. *arXiv preprint arXiv:1703.05401*, 2017.

[108] Luc-Normand Tellier. The weber problem: solution and interpretation. *Geographical Analysis*, 4(3):215–233, 1972.

[109] Alfred Weber. *Theory of the Location of Industries*. University of Chicago Press, 1929.

[110] Luc-Normand Tellier. *Économie spatiale: rationalité économique de l'espace habité*. Boucherville, Québec: G. Morin, 1993.

[111] Pitu B Mirchandani and Richard L Francis. *Discrete location theory*. 1990.

[112] Bernard Gendron, Paul-Virak Khuong, and Frédéric Semet. Comparison of formulations for the two-level uncapacitated facility location problem with single assignment constraints. *Computers & Operations Research*, 86:86–93, 2017.

[113] Ravishankar Krishnaswamy and Maxim Sviridenko. Inapproximability of the multilevel uncapacitated facility location problem. *ACM Trans. Algorithms*, 13(1):1, 2016.

[114] Hyung-Chan An, Mohit Singh, and Ola Svensson. Lp-based algorithms for capacitated facility location. *SIAM J. Comput.*, 46(1):272–306, 2017.

[115] Irina Harris, Christine L Mumford, and Mohamed M Naim. A hybrid multi-objective approach to capacitated facility location with flexible store allocation for green logistics modeling. *Transp. Res. Part E: Logistics and Transportation Review*, 66:1–22, 2014.

[116] FAA. Docket FAA-2015-0150, Amdt. 107-1, 81 FR 42209. 2016.

Chapter 9

Appendices

9.1 Proof of (3.7)

Proof. The CDF of R_2 can be obtained as follows,

$$F_{R_2}(r) = \Pr[R_2 < r] = \sum_{n=0}^{\infty} \Pr[M = n] \Pr[x_1 < r, x_2 < r, ..., x_n < r], \tag{9.1}$$

where $x_1, x_2, ..., x_n$ are the distances from capillary UEs in one Voronoi cell to their D2D agent. According to [52], the PMF of M is given by

$$\Pr[M = n] = \frac{3.5^{3.5}\Gamma(n + 3.5)(\lambda - \lambda_D/\lambda_D)^n}{n!\Gamma(3.5)(\lambda - \lambda_D/\lambda_D + 3.5)^{n+3.5}}. \tag{9.2}$$

In (9.1), according to the Bayes' theorem, we can obtain the following result,

$$\Pr[x_1 < r, x_2 < r, ..., x_n < r] = \Pr[x_1 < r]\Pr[x_2 < r \mid x_1 < r]...\Pr[x_n < r \mid x_1 < r, ..., x_{n-1} < r]. \tag{9.3}$$

We take $\Pr[x_2 < r \mid x_1 < r]$ as an example. As shown in Fig. 9.1, there are two

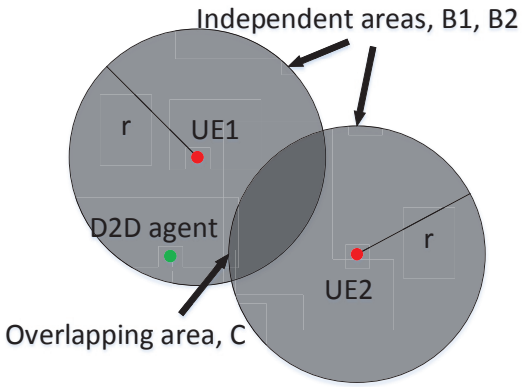


Figure 9.1: An example of two capillary UEs case.

capillary UEs (UE1 and UE2) and one D2D agent. $B1 = B2$, $B1+C = B2+C = r^2\pi$. Given that the distance between the D2D agent and the UE1, x_1, is smaller than r, we can obtain,

$$\begin{aligned}\Pr[x_2 < r \mid x_1 < r] =& 1 - \Pr[x_2 > r \mid x_1 < r] \\ =& 1 - \Pr[\text{no D2D agent in C} \mid x_1 < r]\Pr[\text{no D2D agent in B2} \mid x_1 < r].\end{aligned} \tag{9.4}$$

Because there is at least one D2D agent in the area of $B1 + C$, so that if there is no D2D agent in C, at least one D2D agent must be in $B1$. Thus,

$$\begin{aligned}&\Pr[\text{no D2D agent in C} \mid x_1 < r] \\ =& \Pr[\text{no D2D agent in C}]\frac{1 - \Pr[\text{no D2D agent in B1}]}{1 - \Pr[\text{no D2D agent in B1 + C}]} \\ <& \Pr[\text{no D2D agent in C}].\end{aligned} \tag{9.5}$$

Due to the independence, $\Pr[\text{no D2D agent in B2} \mid x_1 < r] = \Pr[\text{no D2D agent in B2}]$.

Therefore, by substituting (9.5) into (9.4), we have

$$\begin{aligned}\Pr[x_2 < r \mid x_1 < r] >& 1 - \Pr[\text{no D2D agent in C}]\Pr[\text{no D2D agent in B2}] \\ =& 1 - \Pr[\text{no D2D agent in B2} + \text{C}] = F_{R_1}(r). \end{aligned} \tag{9.6}$$

The overlapping part C cannot be avoided because all the UEs are connected to the same D2D agent. When the event $x_1 < r, x_2 < r, ..., x_n < r$ is true, the D2D agent will be in the overlapping area of all the UEs. We can easily extend our conclusion in (9.6) to the general case. Because of the existence of the overlapping area and the fact that there is at least one D2D agent existing in other UEs's circles (including the overlapping area), the probability of the event that there is no D2D agent existing in the overlapping area is reduced comparing with the independent case. Thus, we have $\Pr[x_n < r \mid x_1 < r, ..., x_{n-1} < r] > F_{R_1}(r),\ n \geq 2$.

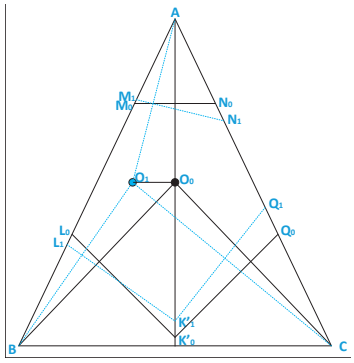
The lower-bound of the CDF of R_2 can be found as follows,

$$\begin{aligned} F_{R_2}(r) =& \sum_{n=0}^{\infty} \Pr[M = n]\Pr[x_1 < r]\Pr[x_2 < r \mid x_1 < r]...\Pr[x_n < r \mid x_1 < r, ..., x_{n-1} < r] \\ >& \sum_{n=0}^{\infty} \Pr[M = n] F_{R_1}(r)^n. \end{aligned} \tag{9.7}$$

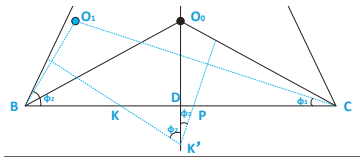
□

9.2 Proof of Theorem 2

As shown in Fig. 9.2 (a), we assume a point O_1 in the left side of BC's mid-perpendicular line, and five vertexes are marked as M_1, L_1, K_1', Q_1, and N_1. If O_1 moves horizontally to O_0, which is on BC's mid-perpendicular line, the vertexes change to M_1, L_1, K_1', Q_1, and N_1. We consider K' in this figure, and the case of K



(a) The case of K'



(b) The case of K and P

Figure 9.2: Changes of the vertexes when O moves horizontally.

and P is shown in Fig. 9.2 (b).

Given the property of voronoi cells, M_1N_1, M_0N_0, and AO_1 intersect at the same point, It is easy to get $||AM_1|| < ||AN_1||$, $||AM_1|| < ||AM_0||$, and $||AN_1|| > ||AN_0||$. Because $||M_1O_1|| = ||AM_1||$, $||N_1O_1|| = ||AN_1||$, $||M_0O_0|| = ||AM_0||$, $||N_0O_0|| = ||AN_0||$, we have $||M_1O_1|| < ||N_1O_1||$, $||M_1O_1|| < ||M_0O_0||$, and $||N_1O_1|| > ||N_0O_0||$. Therefore, we can conclude that once point O locates in the left side of BC's mid-perpendicular line, $||ON||$ is always longer than $||OM||$, and when O moves horizontally toward the mid-perpendicular line, the difference between $||ON||$ and $||OM||$ is reduced.

$||LK||$ and $||QP||$ will intersect at K' which locates on the mid-perpendicular line or its extension line (outside $\triangle ABC$). Given the topology in Fig. 9.2 (a), $\angle L_1K'_1O_0 > \angle Q_1K'_1O_0$. Thus $||BL_1|| < ||CQ_1||$ and $||O_1L_1|| < ||O_1Q_1||$. If O_1 moves to O_0, it is easy to prove $||BL_1|| < ||BL_0||$ and $||CQ_1|| > ||CQ_0||$, so that $||O_1L_1|| < ||O_0L_0||$ and $||O_1Q_1|| > ||O_0Q_0||$. Therefore, when O is in the left side of BC's mid-perpendicular line, $||OQ||$ is always longer than $||OL||$, and the difference decreases while O moving

from O_1 to O_0.

We use Fig. 9.2 (b) to discuss the case of K and P. It is easy to find out $||OK|| = \frac{||BC||}{2} - ||DK||$, $||OP|| = \frac{||BC||}{2} - ||DP||$, $\angle KK'D = \phi_2$, and $\angle PK'D = \phi_3$. Given that O is in the left side of BC's mid-perpendicular line, such as O_1 in the figure, $\phi_2 > \phi_3$. Thus, $||KD|| > ||DP||$, and $||OK|| < ||OP||$. If O moves horizontally from O_1 to O_0, ϕ_2 gets close to ϕ_3, so do $||OK||$ and $||OP||$.

For the case of K' shown in Fig. 9.2 (a), K' is always the circumcenter of $\triangle OBC$. Given Lemma 5, if O moves from O_0 to O_1, vertex K' shall move upward from K'_0 to K'_1, in order to satisfy $||K'B|| = ||K'C|| = ||OK'||$. Therefore, $||O_1K'_1|| > ||O_0K'_0||$. We can conclude that when O moves horizontally toward the mid-perpendicular line, $||OK'||$ is decreasing. Similar approach can be applied to prove that $||ON'||$ holds the same conclusion. (There is no M' inside the triangle when O is in the left side because $\angle AOB > \frac{\pi}{2}$.)

In summary, when there is no any vertex in M', K', and N' exists and O is in the left side of BC's mid-perpendicular line, $\max(||OM||, ||OL||, ||OK||) < \max(||ON||, ||OQ||, ||OP||)$. The longest service distance is one of $||ON||, ||OQ||, ||OP||$, and it is decreasing while O moving toward mid-perpendicular line. If inner vertex exists (according to Lemma 4) at most one of them exists), such as K' or N', the corresponding service distance is also decreasing while O moving toward mid-perpendicular line. It is easy to proof the same conclusions if O is in the right side of the mid-perpendicular line. It can be concluded that the maximum service distance is decreasing while the location of a new service node O moving horizontally toward mid-perpendicular line. Therefore, the optimal location of a new service node is on BC's mid-perpendicular line.

9.3 Proof of Theorem 3

When $||OK'|| \leq ||OK||$, K' will be used instead of K as a valid voronoi vertex in the triangle.

$$\begin{aligned} & ||OK'|| \leq ||OK|| \\ \Longrightarrow & \frac{||AB|| \sin(\alpha)}{\sin(2\phi_2)} \leq \frac{||AB|| \sin(\alpha)}{2\cos(\phi_2)^2} \\ \Longrightarrow & \sin(\phi_2) \geq \cos(\phi_2). \end{aligned} \tag{9.8}$$

Given (9.8), when $0 \leq \phi_2 \leq \frac{\pi}{4}$, $||OK||$ is used in the following analysis. When $\frac{\pi}{4} \leq \phi_2 \leq \frac{\pi}{2}$, $||OK'||$ will be applied. $||OK'|| = ||OK||$ when $\phi_2 = \frac{\pi}{4}$

9.3.1 If $||OL||$ is the maximum distance

Assuming that $||OL||$ is the longest transmission distance, we have

$$\begin{aligned} & ||OL|| \geq ||OK'|| \\ \Longrightarrow & \sin(\phi_2) \geq \sin(\alpha + \phi_2). \end{aligned} \tag{9.9}$$

Given $\phi_2 \leq \frac{\pi}{2}, \alpha + \phi_2 \leq \frac{\pi}{2}$, the inequality above is impossible, which means $||OL||$ cannot be larger than $||OK'||$. Therefore, $||OK'||$ does not exist when $||OL||$ is the longest distance.

When $0 \leq \phi_2 \leq \frac{\pi}{4}$, $||OK||$ is used and we have

$$\begin{aligned} & ||OL|| \geq ||OK|| \\ \Longrightarrow & \sin(\alpha + \phi_2) \leq \cos(\phi_2) \\ \Longrightarrow & \phi_2 \in \left[0, \frac{\pi}{4} - \frac{\alpha}{2}\right]. \end{aligned} \tag{9.10}$$

$$\begin{aligned} & ||OL|| \geq ||OM|| \\ \Longrightarrow & \sin(2\alpha) \geq \sin(2\alpha + 2\phi_2) \\ \Longrightarrow & \phi_2 \in \left[\max\{\frac{\pi}{2} - 2\alpha, 0\}, \frac{\pi}{2} - \alpha\right]. \end{aligned} \tag{9.11}$$

Taking the intersection of the two sets above and considering the fact that $0 \leq \phi_2 \leq \frac{\pi}{4}$ and $0 \leq \alpha \leq \frac{\pi}{4}$, we have

$$\phi_2 \in \left[0, \frac{\pi}{4} - \frac{\alpha}{2}\right] \bigcap \left[\max\{\frac{\pi}{2} - 2\alpha, 0\}, \frac{\pi}{2} - \alpha\right] \tag{9.12}$$

$$= \begin{cases} \emptyset, & \alpha \in \left[0, \frac{\pi}{6}\right), \\ \left[\frac{\pi}{2} - 2\alpha, \frac{\pi}{4} - \frac{\alpha}{2}\right], & \alpha \in \left[\frac{\pi}{6}, \frac{\pi}{4}\right]. \end{cases}$$

We calculate the derivative of $||OL||$ w.r.t. ϕ_2 as follows,

$$\begin{aligned}\frac{\partial||OL||}{\partial\phi_2} =& \frac{||AB||\sin(\alpha)}{2} \times \frac{\sin(\phi_2)\sin(\alpha+\phi_2)-\cos(\phi_2)\cos(\alpha+\phi_2)}{\left[\cos(\phi_2)\sin(\alpha+\phi_2)\right]^2} \\ =& -\frac{||AB||\sin(\alpha)\cos(2\phi_2+\alpha)}{2\left[\cos(\phi_2)\sin(\alpha+\phi_2)\right]^2},\end{aligned} \tag{9.13}$$

where only $\cos(2\phi_2+\alpha)$ needs to be discussed, and other items are obviously positive. Let $\cos(2\phi_2+\alpha)=0$, we can obtain that $\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}$, which is the upper bound of ϕ_2 in (9.12) given that $||OL||$ is the longest distance ($\alpha \geq \frac{\pi}{6}$). Considering $\frac{\partial||OL||}{\partial\phi_2} \leq 0$ when $\phi_2 \leq \frac{\pi}{4}-\frac{\alpha}{2}$, and $\frac{\partial||OL||}{\partial\phi_2} \geq 0$ when $\phi_2 \geq \frac{\pi}{4}-\frac{\alpha}{2}$, $\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}$ is the minimizer of $||OL||$.

Therefore, we can conclude that if $||OL||$ is the longest distance ($\alpha \geq \frac{\pi}{6}$), the minimum of $||OL||$ can be obtained as follows,

$$\begin{aligned}||OL||^* =& ||OL||_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}} \\ =& \frac{||AB||\sin(\alpha)}{2\cos(\frac{\pi}{4}-\frac{\alpha}{2})\sin(\frac{\pi}{4}+\frac{\alpha}{2})} \\ =& \frac{||AB||\sin(\alpha)}{1+\sin(\alpha)}.\end{aligned} \tag{9.14}$$

9.3.2 If $||OM||$ is the maximum distance

Assuming that $||OM||$ is the longest transmission distance, we first compare it with $||OK||$ and $||OL||$, respectively.

$$\begin{aligned}& ||OM|| \geq ||OK|| \\ \Longrightarrow & \frac{\cos(\alpha+\phi_2)}{\cos(\alpha)} \geq \frac{\sin(\alpha)}{\cos(\phi_2)} \\ \Longrightarrow & \cos(\alpha+2\phi_2)+\cos(\alpha) \geq \sin(2\alpha).\end{aligned} \tag{9.15}$$

Because of $0 \leq \alpha + 2\phi_2 \leq \pi$ and the fact that cosine function is monotonically decreasing in this range, we can obtain the following range of ϕ_2 to satisfy (9.15).

$$\phi_2 \leq \frac{\arccos\left[\sin(2\alpha) - \cos(\alpha)\right] - \alpha}{2} = A(\alpha) \tag{9.16}$$

Similarly,

$$\begin{aligned} & ||OM|| \geq ||OL|| \\ \Longrightarrow & \frac{\cos(\alpha + \phi_2)}{\cos(\alpha)} \geq \frac{\sin(\alpha)}{\sin(\alpha + \phi_2)} \\ \Longrightarrow & \sin(2\alpha + 2\phi_2) \geq \sin(2\alpha) \\ \Longrightarrow & \phi_2 \in \left[0, \max\left(\frac{\pi}{2} - 2\alpha, 0\right)\right]. \end{aligned} \tag{9.17}$$

From (9.17), it can be observed that if $\alpha \geq \frac{\pi}{4}$ (obtuse triangle), $||OM||$ cannot be the longest distance.

It is easy to find out that $A(\alpha) \leq \frac{\pi}{2} - 2\alpha$ when $\alpha \in [0, \frac{\pi}{6}]$, and $A(\alpha) \geq \frac{\pi}{2} - 2\alpha$ when $\alpha \in [\frac{\pi}{6}, \frac{\pi}{4}]$. Taking the intersection of two ranges derived in (9.17) and (9.16), we have

$$\begin{aligned} \phi_2 \in & [0, A(\alpha)] \bigcap \left[0, \max\{\frac{\pi}{2} - 2\alpha, 0\}\right] \\ = & \begin{cases} [0, A(\alpha)], & \alpha \in \left[0, \dfrac{\pi}{6}\right], \\ \left[0, \dfrac{\pi}{2} - 2\alpha\right], & \alpha \in \left[\dfrac{\pi}{6}, \dfrac{\pi}{4}\right]. \end{cases} \end{aligned} \tag{9.18}$$

The derivative of $||OM||$ w.r.t. ϕ_2 is given as follows,

$$\begin{aligned}\frac{\partial||OM||}{\partial\phi_2} =& \frac{||AB||}{2\cos(\alpha)} \times \frac{-\sin(\alpha+\phi_2)\cos(\phi_2)+\cos(\alpha+\phi_2)\sin(\phi_2)}{[\cos(\phi_2)]^2} \\ =& -\frac{||AB||\sin(\alpha)}{2\cos(\alpha)\left[\cos(\phi_2)\right]^2},\end{aligned} \tag{9.19}$$

It can be observed that given $\alpha \leq \frac{\pi}{2}$, $\frac{\partial||OM||}{\partial\phi_2} \leq 0$, so that the upper bound of ϕ_2 is a minimizer of $||OM||$. It should be noticed that $||OK||$ is considered only if $0 \leq \phi_2 \leq \frac{\pi}{4}$, otherwise $||OK'||$ will be applied.

Therefore, if $||OM||$ is the longest distance, the minimum of $||OM||$ is given by

$$||OM||^* = ||OM||_{\phi_2=A(\alpha)} = \frac{||AB||\cos(\alpha+A(\alpha))}{2\cos(A(\alpha))\cos(\alpha)}, \tag{9.20}$$

when

$$\begin{aligned}\left[A(\alpha) \leq \frac{\pi}{4}\right] \bigcap \left[\alpha \leq \frac{\pi}{6}\right] &= \left[\arccos\left[\sin(2\alpha)-\cos(\alpha)\right] \leq \frac{\pi}{2}+\alpha\right] \bigcap \left[\alpha \leq \frac{\pi}{6}\right] \\ &= \left[\sin(2\alpha)-\cos(\alpha)+\sin(\alpha) \geq 0\right] \bigcap \left[\alpha \leq \frac{\pi}{6}\right] \\ &\Longrightarrow \alpha \in \left[0.106\pi, \frac{\pi}{6}\right],\end{aligned} \tag{9.21}$$

and

$$\begin{aligned}||OM||^* =& ||OM||_{\phi_2=\frac{\pi}{2}-2\alpha} \\ =& \frac{||AB||\sin(\alpha)}{2\sin(2\alpha)\cos(\alpha)} \\ =& \frac{||AB||}{4\cos^2(\alpha)}.\end{aligned} \tag{9.22}$$

when

$$\begin{aligned} &\left[\frac{\pi}{2} - 2\alpha \leq \frac{\pi}{4}\right] \bigcap \left[\frac{\pi}{6} \leq \alpha \leq \frac{\pi}{4}\right] \\ \Longrightarrow &\alpha \in \left[\frac{\pi}{6}, \frac{\pi}{4}\right]. \end{aligned} \tag{9.23}$$

If $||OK'||$, instead of $||OK||$, is applied, we have

$$\begin{aligned} &||OM|| \geq ||OK'|| \\ \Longrightarrow &\frac{\cos(\alpha + \phi_2)}{\cos(\alpha)} \geq \frac{\sin(\alpha)}{\sin(\phi_2)} \\ \Longrightarrow &\sin(\alpha + 2\phi_2) \geq \sin(2\alpha) + \sin(\alpha). \end{aligned} \tag{9.24}$$

When $||OK'||$ is applied, $\phi_2 \geq \frac{\pi}{4}$. Thus, $\frac{\pi}{2} \leq \alpha + 2\phi_2 \leq \pi$, and sine function is monotonically decreasing in this range. Therefore, we have

$$\phi_2 \leq \frac{\pi - \arcsin\left[\sin(2\alpha) + \sin(\alpha)\right] - \alpha}{2} = A'(\alpha) \tag{9.25}$$

Also, we can observe that if $\sin(2\alpha) + \sin(\alpha) \geq 1$, (9.25) does not hold, and thus $||OM||$ is not the longest distance. The valid range can be calculated as follows,

$$\begin{aligned} &\sin(2\alpha) + \sin(\alpha) \leq 1 \\ \Longrightarrow &2\sin(\alpha)\sqrt{1 - \sin^2(\alpha)} + \sin(\alpha) \leq 1 \end{aligned} \tag{9.26}$$

Let $x = \sin(\alpha)$, it can be rewrote as

$$\begin{aligned} &4x^4 - 3x^2 - 2x + 1 \geq 0 \\ \Longrightarrow &(x-1)(4x^3 + 4x^2 + x - 1) \geq 0 \end{aligned} \tag{9.27}$$

Because of $\phi_2 \geq \frac{\pi}{4}$, $\alpha \leq \frac{\pi}{4}$ and thus $x - 1 = \sin(\alpha) - 1 \leq 0$,

$$\Longrightarrow (4x^3 + 4x^2 + x - 1) \leq 0 \tag{9.28}$$

Based on numerical calculation, we find that (9.28) holds when $x \leq 0.348$. Therefore, (9.25) holds when $\alpha \leq \arcsin(0.348)$. It is easy to find that when $\alpha \leq \arcsin(0.348)$, $\max(\frac{\pi}{2} - 2\alpha, 0)$ is always larger than $A'(\alpha)$. The intersection of the ranges in (9.25) and (9.18), i.e., the condition of $||OM|| \geq ||OL||$, is given by

$$\begin{aligned} \phi_2 \in [0, A'(\alpha)] \bigcap & \left[0, \max\left(\frac{\pi}{2} - 2\alpha, 0 \right) \right] \\ = & \left[0, A'(\alpha) \right], \alpha \in [0, \arcsin(0.348)]. \end{aligned} \tag{9.29}$$

It has been shown in (9.19) that $||OM||$ is monotonically decreasing with the increasing of ϕ_2. Therefore, $\phi_2 = A'(\alpha)$ is a minimizer of $||OM||$. In order to satisfy the existence of $||OK'||$, we have

$$\begin{aligned} & A'(\alpha) \geq \frac{\pi}{4} \\ \Longrightarrow & \cos(\alpha) - \sin(2\alpha) - \sin(\alpha) \geq 0 \\ \Longrightarrow & \alpha \leq 0.106\pi. \end{aligned} \tag{9.30}$$

It is easy to find that $[0, 0.106\pi] \cap [0, \arcsin(0.348)] = [0, 0.106\pi]$, so we have

$$\begin{aligned} ||OM||^* = & ||OM||_{\phi_2 = A'(\alpha)} \\ = & \frac{||AB|| \cos(\alpha + A'(\alpha))}{2\cos(A'(\alpha))\cos(\alpha)}, \quad \alpha \in [0, 0.106\pi]. \end{aligned} \tag{9.31}$$

9.3.3 If $||OK||$ is the maximum distance

First, let $\phi_2 \leq \frac{\pi}{4}$ to ensure the existence of $||OK||$. Similar with the results we obtained from Sec. 9.3.1 & 9.3.2, we have

$$\begin{aligned} &||OL|| \leq ||OK|| \\ \Longrightarrow &\sin(\alpha+\phi_2) \geq \cos(\phi_2) \\ \Longrightarrow &\phi_2 \in \left[\frac{\pi}{4} - \frac{\alpha}{2}, \frac{\pi}{4}\right], \end{aligned} \tag{9.32}$$

and

$$\begin{aligned} &||OM|| \leq ||OK|| \\ \Longrightarrow &\cos(\alpha+2\phi_2) + \cos(\alpha) \leq \sin(2\alpha) \\ \Longrightarrow &\phi_2 \in \left[A(\alpha), \frac{\pi}{4}\right]. \end{aligned} \tag{9.33}$$

If $A(\alpha) > \frac{\pi}{4} \Longrightarrow \alpha < 0.106\pi$, there is no feasible solution. Considering $\alpha \geq 0.106\pi$, the valid range of ϕ_2 is given by

$$\begin{aligned} \phi_2 &\in \left[\frac{\pi}{4} - \frac{\alpha}{2}, \frac{\pi}{4}\right] \bigcap \left[A(\alpha), \frac{\pi}{4}\right] \\ &= \begin{cases} \left[\frac{\pi}{4} - \frac{\alpha}{2}, \frac{\pi}{4}\right], & \alpha \in \left[\frac{\pi}{6}, \frac{\pi}{2}\right], \\ \left[A(\alpha), \frac{\pi}{4}\right], & \alpha \in \left[0.106\pi, \frac{\pi}{6}\right]. \end{cases} \end{aligned} \tag{9.34}$$

The derivative of $||OK||$ w.r.t. ϕ_2 can be obtained as follows,

$$\begin{aligned}\frac{\partial||OK||}{\partial\phi_2} =&\frac{||AB||\sin(\alpha)}{2}\frac{2\cos(\phi_2)\sin(\phi_2)}{[\cos(\phi_2)]^4}\\ =&\frac{||AB||\sin(\alpha)\sin(\phi_2)}{[\cos(\phi_2)]^3}\geq 0.\end{aligned} \tag{9.35}$$

Therefore, when $||OK||$ is the longest distance, the minimum of $||OK||$ is given by

$$\begin{aligned}||OK||^* =&||OK||_{\phi_2=A(\alpha)}\\ =&\frac{||AB||\sin(\alpha)}{2\cos^2(A(\alpha))},\quad \alpha\in\left[0.106\pi,\frac{\pi}{6}\right].\end{aligned} \tag{9.36}$$

and

$$\begin{aligned}||OK||^* =&||OK||_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}}\\ =&\frac{||AB||\sin(\alpha)}{1+\sin(\alpha)},\quad \alpha\in\left[\frac{\pi}{6},\frac{\pi}{2}\right].\end{aligned} \tag{9.37}$$

9.3.4 If $||OK'||$ is the maximum distance

If $||OK'||$ is applied, $\phi_2\geq\frac{\pi}{4}$ and $\alpha\leq\frac{\pi}{4}$. Similarly, we have

$$\begin{aligned}&||OL||\leq||OK'||\\ \Longrightarrow&\sin(\phi_2)\leq\sin(\alpha+\phi_2).\end{aligned} \tag{9.38}$$

and

$$\begin{aligned}&||OM||\leq||OK'||\\ \Longrightarrow\phi_2\geq&\frac{\pi-\arcsin\left[\sin(2\alpha)+\sin(\alpha)\right]-\alpha}{2}=A'(\alpha)\end{aligned} \tag{9.39}$$

It is obvious that (9.38) is always true given $\alpha + \phi_2 \leq \frac{\pi}{2}$. So the valid range of ϕ_2 is given by

$$\phi_2 \in \begin{cases} \left[\frac{\pi}{4}, \frac{\pi}{2}\right], & \alpha \in \left[0.106\pi, \frac{\pi}{4}\right], \\ \left[A'(\alpha), \frac{\pi}{2}\right], & \alpha \in [0, 0.106\pi]. \end{cases} \tag{9.40}$$

The derivative of $||OK'||$ w.r.t. ϕ_2 can be given as follows,

$$\frac{\partial ||OK'||}{\partial \phi_2} = -||AB|| \sin(\alpha) \frac{2\cos(2\phi_2)}{[\sin(2\phi_2)]^2} \tag{9.41}$$

Because of $\phi_2 \geq \frac{\pi}{4}$, $\cos(2\phi_2) \leq 0$ and thus $\frac{\partial ||OK'||}{\partial \phi_2} \geq 0$. Therefore, if $||OK'||$ is the longest distance, the minimum of $||OK'||$ is given by

$$\begin{aligned} ||OK'||^* =& ||OK'||_{\phi_2 = A'(\alpha)} \\ =& \frac{||AB|| \sin(\alpha)}{\sin(2A'(\alpha))}, \quad \alpha \in [0, 0.106\pi]. \end{aligned} \tag{9.42}$$

and

$$\begin{aligned} ||OK'||^* =& ||OK'||_{\phi_2 = \frac{\pi}{4}} \\ =& ||AB|| \sin(\alpha), \quad \alpha \in \left[0.106\pi, \frac{\pi}{4}\right]. \end{aligned} \tag{9.43}$$

9.3.5 Minimize the maximum distance

All the results in Sec. 9.3.1, 9.3.2, 9.3.3, and 9.3.4 are summarized in Table 9.1. By comparing the candidate minimizers in each range of α, the global minimizers are picked up and summarized in Table 2.

Table 9.1: Candidate minimizers

$\alpha \in$	Candidate minimizers and the longest distances
$[0, 0.106\pi]$	$\|\|OM\|\|^*_{\phi_2=A'(\alpha)} = \dfrac{\|\|AB\|\| \cos(\alpha + A'(\alpha))}{2\cos(A'(\alpha))\cos(\alpha)}$, $\|\|OK'\|\|^*_{\phi_2=A'(\alpha)} = \dfrac{\|\|AB\|\| \sin(\alpha)}{\sin(2A'(\alpha))}$;
$[0.106\pi, \frac{\pi}{6}]$	$\|\|OM\|\|^*_{\phi_2=A(\alpha)} = \dfrac{\|\|AB\|\| \cos(\alpha + A(\alpha))}{2\cos(A(\alpha))\cos(\alpha)}$, $\|\|OK\|\|^*_{\phi_2=A(\alpha)} = \dfrac{\|\|AB\|\| \sin(\alpha)}{2\cos^2(A(\alpha))}$, $\|\|OK'\|\|^*_{\phi_2=\frac{\pi}{4}} = \|\|AB\|\| \sin(\alpha)$;
$[\frac{\pi}{6}, \frac{\pi}{4}]$	$\|\|OL\|\|^*_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}} = \dfrac{\|\|AB\|\| \sin(\alpha)}{1+\sin(\alpha)}$, $\|\|OM\|\|^*_{\phi_2=\frac{\pi}{2}-2\alpha} = \dfrac{\|\|AB\|\|}{4\cos^2(\alpha)}$, $\|\|OK\|\|^*_{\phi_2=\frac{\pi}{4}-\frac{\alpha}{2}} = \dfrac{\|\|AB\|\| \sin(\alpha)}{1+\sin(\alpha)}$, $\|\|OK'\|\|^*_{\phi_2=\frac{\pi}{4}} = \|\|AB\|\| \sin(\alpha)$;

9.4 Proof of Theorem 4

Assume the distance from the location of the new service node, O, to only one vertex, v^*, equals the longest service distance, denoted by X. We define the set $\mathbb{V}^{(r)}$ including the distances from O to other 4 or 5 vertexes exist in this triangle. We have

$$X > \max\left(\mathbb{V}^{(r)}\right). \tag{9.44}$$

We can always find a direction to move O and reduce $||Ov^*||$ by δ_1, and increase $\max\left(\mathbb{V}^{(r)}\right)$ by δ_2. If we further ensure

$$\delta_2 < X - \max\left(\mathbb{V}^{(r)}\right) - \delta_1, \tag{9.45}$$

then the following inequality holds.

$$X - \delta_1 > \max\left(\mathbb{V}^{(r)}\right) + \delta_2. \tag{9.46}$$

Given δ_1 and δ_2 are generated by O's movement, $\delta_2 = 0$ when $\delta_1 = 0$. Considering $X - \max\left(\mathbb{V}^{(r)}\right)$ is positive, we can always satisfy (9.45) by finding a small enough δ_1. Therefore, there is always a movement of O to satisfy (9.46) and further reduce the longest service distance, which means when O is on the optimal location, it cannot have only one vertex as the worst point.

Next, we discuss the case of two vertexes leading to the same longest service distance. There may be a special case where the two worst vertexes and O are in a line. Any movement of O cannot reduce the longest distance. Let us first verify whether it is a possible scenario. Taking Fig. 2 as an example, it is obvious that any two adjacent vertexes cannot be the two worst points. We assume $\angle LOQ = \pi$ and $||LO|| = ||OQ||$. It is also easy to find out $||BL|| = ||QC||$. Given $\angle ABC \neq \angle ACB, ||AB|| \neq ||AC||$, either $\angle ALQ < \angle ABC$ or $\angle AQL < \angle ACB$ after $||AB||$ and $||AC||$ are deducted by the same length. For example, $\angle ALQ < \angle ABC$ because of $\alpha > \beta$ in Fig. 2, thus $2\angle LBO < \angle LBO + \phi_2$ and $\angle LBO < \phi_2$. Therefore, we can conclude that, in this case, $||OK|| = ||BK|| > ||BL|| = ||OL||$ and L, Q cannot be the only two worst points. If we want to make $||OL|| = ||ON||$ and $\angle LON = \pi$, O should move upward because N is above Q, and thus $||OK||$ (or possible $||OK'||$) further increases. Therefore, L, N cannot be the only two worst points while minimizing the longest service distance. The same method can be applied for M, Q to obtain the same conclusion.

We can easily verify other cases by rotating the triangle. For example, if AC is on the x-axis, by applying the same method above, we can prove that M and P, L and P, M and K cannot be the only two worst points when they are in a line with

O, respectively.

If the two worst vertexes and O are not in a line, similarly, we denote the longest service distance as X, the distances from O to the rest vertexes as $\mathbb{V}^{(r)}$. We can always find a direction to move O and reduce the longest service distance by δ_1, and increase $\max\left(\mathbb{V}^{(r)}\right)$ by δ_2. Similar to (9.44)-(9.46) and following analysis, we can prove that if the two worst vertexes and O are not in a line and O is on the optimal location, these two vertexes cannot be the only two worst points.

In summary, any two vertexes cannot be the only two worst points, and thus Theorem 3 is proved.