QoS/QoE Provisioning in Emerging 5G Wireless Networks

by

Zhifeng He

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 10, 2016

Keywords: Cognitive Radio, Millimeter Wave, Resource Allocation

Approved by

Shiwen Mao, Chair, Samuel Ginn Endowed Professor of Electrical and Computer Engineering
Mark Nelms, Professor and Chair of Electrical and Computer Engineering
Jitendra Tugnait, James B Davis Professor of Electrical and Computer Engineering
Fa Foster Dai, Ed and Peggy Reynolds Family Endowed Professor of Electrical and Computer Engineering

Abstract

In recent years, there has been a drastic growth in wireless traffic, and in the near future, a majority of wireless traffic will be video-related. How to support the increasing demand of radio resources from the bandwidth-hungry services has attracted intensive research interests from both academia and industrial areas. Milli-Meter Wave (mmWave) technology and Cognitive Radio (CR) technology was recently proposed to enhance the wireless network capacity.

However, some challenges are needed to be addressed before one can apply these techniques. For example, how to adjust coding scheme to the changing wireless network environment to against the problem of uncertainty of channel condition. Besides, in Cognitive Radio Network (CRN), it is important to coordinate the transmissions of the primary users and that of the secondary users so that Cognitive Users (CUs) will not cause unacceptable interference to Primary Users (PUs) while their utility can be maximized, which also depends on the radio resource allocation for the CUs. What's more, in the mmWave network, due to the dynamic channel conditions, how to optimally coordinate the concurrent transmissions of neighboring links based on their possible channel conditions, so that the network throughput is can be improved.

In this dissertation work we study how to use adaptive video coding to combat channel uncertainty in mmWave network and investigate how to optimize the channel allocation in video streaming over cognitive radio networks.

The first part of this dissertation investigates the problem of streaming uncompressed HD video over mmWave wireless networks so that the expected mean square error of the reconstructed video quality is minimized. An adaptive coding scheme that can dynamically adjust to the changing channel conditions is proposed so that error rate is reduced, and a dynamic interleaving based transmission strategy is incorporated to avoid busty errors in transmission. Efficient algorithm with low computational complexity is proposed to solve for the optimal setup.

The second part of this dissertation investigates the problem of video streaming over CRN. Spectrum sensing and spectrum allocation are optimized such that so that Quality of Experience (QoE) of CUs are maximized. Due to the non-linearity of the QoE model, the spectrum sensing problem and the spectrum accessing problem are solved separately and some performance may be lost. We discuss the case when each CU can sense only one channel each time and the case when each CU can sense multiple channels each time.

The third part of this dissertation investigates the problem of video streaming over CRN, where the spectrum sensing, spectrum allocation, and transmit power allocation are jointly optimized such that so that Quality of Service (QoS) of CUs are maximized, which is significantly different with the second part. We show that the formulated Mixed Integer NonLinear Programming problem can be decomposed into two sub-problems without sacrificing optimality, and with a much lower computational complexity. We analyze the proposed iterative algorithm with respect to complexity and time efficiency, and derive a performance upper bound.

The fourth part of this dissertation investigates the problem of relay and link selection in a dual-hop mmWave network aiming at minimizing the Maximum Expected Delivery Time among all Tx-Rx pairs, while exploiting reflected mmWave transmissions and considering link blockage dynamics. Due to the NP-hardness of the formulated problem, we develop a Decomposition Principle to transform this problem into two sub-problems, one for link selection and the other for relay assignment. We prove that the proposed scheme can achieve an optimality gap of just 1 time slot at greatly reduced complexity. The proposed schemes are validated with simulations with their superior performance observed.

The fifth part of this dissertation investigates the problem of user scheduling in multiple hop mmWave networks, so that the number of time slots needed to serve all user' traffic demand is minimized. Channel condition changes over time and at each time slot, given the possible channel states, the PNC decides the optimal routing path and which users should access the channel at current time slot, aiming at maximizing the long term utility of the whole network. We propose a heuristic algorithm with greatly reduced complexity to solve this problem, which first fix the

optimal routing path for a long term and then maximizes the instant network throughput. A simlilar problem in singel hop mmWave network is also studied and an effective algorithm is proposed. The performance of the heuristic algorithms is validated with simulations.

In summary, this dissertation aims to improving the QoS/QoE provisioning in emerging wireless networks by addressing the resource allocation and user scheduling problems. In-depth analysis and comprehensive results are also provided. Some of the findings may shed light on how to put emerging techniques into real applications.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my major advisor Dr. Shiwen Mao for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. In addition, I want to thank Dr. Mao's wife, Dr. Yihan Li for her hospitality.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Mark Nelms, Dr. Jitendra Tugnait, and Dr. Fa Foster Dai, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also goes to Dr. Ash Abebe, who supervised my Master in Probabilities and Statistics, Dr. Peng Zeng and Dr. Xiaoyu Li, who served as my committee members for my Master in Probabilities and Statistics.

I thank my fellow labmates in for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four and a half years.

Last but not the least, I would like to thank my family: my parents and to my brothers and sister for supporting me spiritually throughout writing this thesis and my life in general.

At last, this dissertation work is supported in part by the US National Science Foundation under Grant CNS-1320664, and Grant CNS-0953513, and through the NSF Broadband Wireless Access and Applications Center (BWAC) site and the Wireless Engineering Research and Education Center (WEREC) at Auburn University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Auburn Auburn University

LoA List of Abbreviations

Chapter 1

Introduction

Due to the rapid development of the mobile devices such as smartphone and laptops in the last decade, wireless data traffic has experienced a explosive growth, where video-related services is one of the wireless services that enjoyed the fastest growth in recent years, according to a images conducted by Cisco [6]. The drastic expansion of wireless traffic poses great challenge to the capacity of existing wireless networks, and efficient technologies, such as mmWave and CR, are proposed recently to magnify the wireless network capacity to accommodate the increasing wireless traffic.

To resist the channel uncertainty issue in wireless networks and the sequential packet loss problem, adaptive video coding is proposed to adjust to the changing channel conditions to ensure received video quality. When channel conditions are good, we may choose aggressive video coding algorithms so that more effective data can be transmitted without corrupted by transmission errors to improve the received video quality, and while channel conditions are poor, conservative video coding algorithms are chosen so that the most important information of the video is well-protected in transmission and received by the receiver so that a basic video quality is guaranteed at the receiver.

Moreover, to expand nowadays wireless network capacity and to support High Definition (HD) video streaming, mmWave frequency is exploited to accommodate the explosion of wireless data traffic, especially for bandwidth-hungry services such as video streaming. However, mmWave channel is highly directional and susceptible to blockage of the link-of-sight path caused by obstacles, due to its high frequency which leads to serious signal power attenuation and a poor capability of penetrating obstacles. We apply interleaving techniques to the packets of multiple consecutive

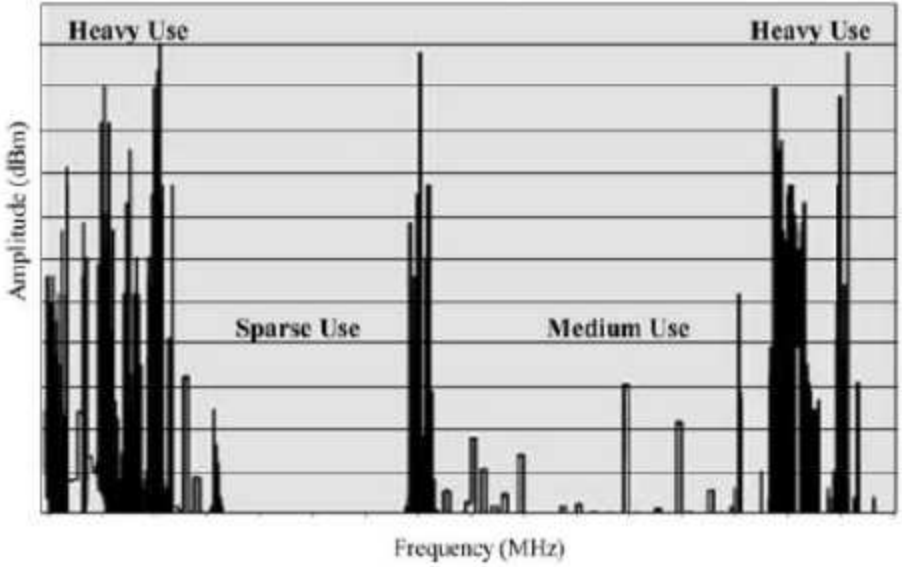

Figure 1.1: Frequency usage measurement

frames to reduce the bursty error rate in mmWave network, which brings in a greater probability of recovering received video with higher quality.

Cognitive radio is also proven to have great potential to magnify the wireless network capacity and has attracted intensive attention of researchers from both academy and industry. According to the FCC report [2], while some licensed bands are overcrowded, many others are underutilized, as in Fig. 1.1. Under traditional fixed spectrum allocation policy, when PUs are not active, the channels assigned to them are wasted (termed as spectrum opportunities). Cognitive radio networks (CRN) are proposed as a new wireless paradigm for exploiting such spectrum opportunities, to enable flexible and efficient access to radio spectrum. In CRN, unlicensed users (or, CUs) are allowed to access the licensed band opportunistically, while PUs gain by collecting revenue for spectrum leasing. Such a CR paradigm has been shown to have high potentials to enhance spectrum efficiency.

However, channel quality may be highly diverse in CRNs due to heterogeneous network, and a channel may have different utility (e.g., SNR) to different users, therefore it is important to optimize the channel allocation to users so that the overall utility of all users can be maximized. Besides, a reliable spectrum sensing result should not only discover the available idle spectrum for a Cognitive radio User (CU), but also not to misidentify the unavailable spectrum as available spectrum for CUs access, in order not to cause interference to PUs. If a CU spends more time on spectrum sensing to guarantee a good sensing result, then the data transmission time will be

shorten, which lowers the CU's throughput. And users may have different sensing accuracy for the spectrum.

Power allocation is also important to the performance of CUs. One way to ensure the desired SINR at the receiver is to control the transmit power of the target session and that of other video sessions in order to guarantee the strength of the desired signal and limit the interference to an acceptable level. When there are concurrent video sessions in CRN, it is necessary to control the power of each transmission in order to constraint the interference caused to other sessions, and meanwhile achieve the required SINR for each transmission.

For comprehensive evaluation of video quality, we study both the Quality of Service (QoS) and Quality of Experience (QoE) measurements in our dissertation, where QoS is an objective assessment method of the performance of network, while QoE is a subjective assessment method of a user's experiences with a service. In other words, QoE is the perceptual QoS from the users perspective. While monitoring and controlling QoS parameters of the video transmission system is important for achieving high video quality, it is more crucial to evaluate video quality from the users perspective. In this work, we study how to improve the QoS and QoE based on the developed QoS and QoE models by jointly optimization of spectrum sensing, channel allocation, and power allocation.

As mentioned above, although the mm Wave band is attractive for its capability of supporting high data rates, many challenges should be addressed to make mm Wave networks applicable [143]. The wireless signal propagating in mm Wave channels suffers from an attenuation that is much higher than that in 5GHz channels [147]. To overcome the high attenuation, beamforming should be used to increase the signal's effective power, while the small wavelength allows for integration of many antenna elements with a small form factor. It has been shown in [146] that the highly directional links, especially in the outdoor environment, can be treated as "pseudowired", i.e., the probability of collision even in a dense mm Wave network could be small. Although the pseudowired feature is attractive from the perspective of spatial reuse, extremely narrow transmission beamwidths will make it hard for network coordination and control [147], which all require

the nodes in a neighborhood to be able to hear from each other. In indoor mm Wave networks, the beamwidth is usually wider than that in outdoor networks due to the smaller transmission distance. The interference among neighboring links should be considered in this case. In our work we study both of the indoor transmission and outdoor transmission cases.

In addition, mmWave signals in the mm Wave band usually do not easily diffract around or penetrate obstacles. A line-of-sight (LOS) path between the transmitter and receiver is required for a successful transmission. When the LOS path is blocked (e.g., by a human body), relay nodes will be needed to forward data for a hidden receiver or wall reflections can be utilized [148]. The blockage may appear or disappear occasionally due to the movement of obstacles between the transmitter and receiver or the movement of the transmitter or receiver themselves [150, 151]. A flexible link model that considers both narrow and wide beamwidths and dynamic blockage will be desirable for the design of mm Wave network protocols.

In the work we investigate the problem of link scheduling and relay assignment in mm Wave networks, where a two-state Markov chain model is adopted to capture the dynamic blockage behavior of mm Wave channel. Based on the possible channel states (blocked or unblocked) at each time slot, we study the problem of optimally assigning relays to Tx-Rx pairs, and which Tx-Rx pairs should access channels at each time slot, so that the interference they cause to each other is refrained under an acceptable level, and the network performance in terms of throughput is maximized. Besides, unlike prior works on relay selection in mm Wave networks, we consider the LOS and multiple reflected NLOS links between transmitting and receiving nodes. Although using reflections will cause additional power loss and reduce power efficiency, it offers additional choices for increasing network coverage and improving network throughput.

The contributions of our work is summarized as follows:

1. We investigated how the bits in different positions in a pixel have different levels of significance on recovering the original video frame, and propose to combat the lossy mmWave channel with Multiple Description Coding (MDC). Specifically, bits in a pixel are optimally

divided into multiple segments according to their respective importance and channel condition, with different level of protection. Besides, packet interleaving are used to further reduce busty error rate in mmWave networks. We formulate the problem of minimizing the Mean Square Error (MSE) of the received video by optimal segmentation of the pixel bits and optimal interleaving depth. A heuristic algorithm is proposed to solve this problem with greatly reduced complexity.

2. We study the problem of maximizing the QoE of all the CUs by jointly optimizing spectrum sensing and channel assignment. Due to the non-linearity of the QoE model, the problem is formulated as an Nonlinear Integer Programming problem (NIP). Under the assumption that all the spectrum sensors work at the same operating point (i.e., with the same probability of detection and the same probability of false alarm), we show that this challenging problem can be solved with a two-step approach: First, the spectrum sensing scheduling problem is solved with a greedy algorithm; Second, the channel allocation problem, which is a Maximum Weight Matching problem and can be solved optimally with the Hungarian Method. We prove that the two-step solution algorithm is indeed optimal: decomposing the original problem into two sub-problems and solving them sequentially do not sacrifice the optimality of the solution.

3. We investigate the problem of QoS maximization video streaming over CRN. Considering the frequency diversity among users and the transmission power limit, the problem is formulated as a joint-optimization problem that spectrum sensing, channel allocation and transmission power allocation for users are jointly optimized, which is an NIP. We decompose the problem into two sub-problems that can be solved with greatly reduced complexity and without sacrificing optimality. To further reduce spectrum sensing overhead for users, we proposed a heuristic spectrum sensing algorithm with very competitive performance validated by simulation results.

4. We investigate the scheduling problem in centralized dual-hop 60GHz networks in the outdoor environment. We adopt the “pseudowired” assumption and thus the interference between different links can be ignored, and the objective is to minimize the Maximum Expected Delivery Time among all Tx-Rx pairs by jointly optimizing relay and link selection in a two-hop wireless network. We develop a Decomposition Principle to transform this problem into two sub-problems, one for link selection and the other for relay assignment. We prove that the proposed scheme can achieve an optimality gap of just 1 time slot at greatly reduced complexity. NLOS links are also utilized to further improve the coverage and throughput of the network.

5. We investigate the problem of link scheduling in both of one hop and multi-hop mmWave networks in the indoor environment, and the interference among neighboring links is considered in this case. The objective is to minimize the time duration needed to satisfy the traffic demand of all the links. The optimization consists of two major parts, one is to choose the optimal relays for the transmitters and receivers which are out of the transmission range of each other (for the multi-hop case), and the other is to determine which Tx-Rx pairs should access the channels based on the possible channel conditions at each time slot, so that the interference they cause to each other will not be too high to affect their transmissions, while the network throughput can be maximized at the same time. The formulated problem is NP hard, and we propose heuristic algorithms with greatly reduced complexity, which first finds the optimal streaming path for each Tx-Rx pair (for the multi-hop case)and then maximizes the instant network throughput by optimizing the link scheduling at each time slot.

Chapter 2

Adaptive Multiple Description Coding and Transmission of Uncompressed Video over mmWave Networks

2.1 Introduction

With the dramatic advances in wireless networking technology, there is an exponentially increasing demand for wireless data service. In particular, mobile video is predicted to grow at a compound annual growth rate (CAGR) of 90% from 2011 to 2016 [6]. Cognitive radio (CR) has been recognized as an important technology for enhancing spectrum efficiency, while many dynamic spectrum access techniques are developed to better utilize the allocated spectrum [72]. On the other hand, millimeter wave (mm-wave) communications in the mmWave band has gained considerable interest from academia, industry, and standards bodies [16]. There is huge unlicensed bandwidth (i.e., up to 7 GHz) in the mmWave band that is available in most parts of the world. In addition to indoor use, FCC recently updated the rules for the unlicensed mmWave band, which will allow higher emission limits for mmWave devices that operate outdoors, thus enabling broader deployment of point-to-point broadband systems [7]. The massive unlicensed bandwidth provides great potential to meet the surging wireless video demand, as well as supporting new bandwidth demanding applications [8, 20].

Recently, the problem of transmitting uncompressed High Definition (HD) videos via mmWave channels has attracted intensive interest. The immense bandwidth of the mmWave channel enables streaming of uncompressed HD videos with high data rates [9], such as the 1080p video, which has 1920×1080 pixels (each of which has 24 bits) per video frame and the frame size is about 6 MBs. Such high speed wireless links can not only replace the cumbersome HDMI cables, but also relieve the computational burden for video decoding at the display devices (e.g., projectors or HDTVs).

Although the larger bandwidth is a great advantage of mmWave networks, the mmWave transmissions are highly directional to overcome the high attenuation, making it susceptible to blockage of the line-of-sight path by obstacles or pedestrians. Thus the memory of the channel poses a negative effect on the performance of the system, and channel conditions such as packet loss rate may change over time [16]. How to adjust the operation parameters of the transmission system is critical to maintain a certain level of QoS under such conditions.

In this chapter, we investigate the problem of streaming uncompressed HD video over mmWave networks. Considering the fact that bits of different positions in a pixel have different levels of significance on recovering the original video frame [10], it is desirable to improve the quality of reconstructed video frames by offering more protection for the more important bits. We adopt multiple description (MD) coding (MDC) to combat the lossy mmWave channel. In MDC, a video is encoded into multiple descriptions, each can guarantee a low but acceptable video quality. More important, the more descriptions received, the better the video quality. MDC has been used mostly for compressed videos in lossy wireless or wireline networks in the literature [12, 13].

Here we adopt MD coding for uncompressed videos, exploiting the different significance of the different bits in a pixel. In particular, in order to ensure the quality of the reconstructed video, we divide the bits in a pixel into multiple segments according to their respective significance in enhancing video quality, and provide higher protection to bits of greater importance by assigning more forward error correction (FEC) symbols to them. This way, the decoder has a higher probability of recovering the most important bits, which affect the quality of the reconstructed image to a larger extent. Such an MD coder is termed MD-FEC. How to partition the bits, i.e., how many segments the bits should be divided into, and how many bits a particular segment should have, affects the significance of a particular segment and the probability of decoding the segment.

To combat the bursty errors while transmitting over the mmWave channel, interleaving among multiple video frames is applied in our scheme. We formulate a nonlinear integer programming problem, which can be solved to find the optimal partition of the pixel bits in the MD coder, as well as the optimal interleaving depth, for given channel conditions. To reduce computational

Table 2.1: Notation

Symbol	*Description*
M	Number of rows of the image matrix
N	Number of columns of the image matrix
C	Number of rows of the sub-image matrix
D	Number of columns of the sub-image matrix
Q	Number of sub-images of an image
S	Number of segments the 8 bits are divided into
m_i	Length of the i-th segment, $1 \leq i \leq S$
k	Number of bits constituting a symbol
g_i	Number of groups the data symbols of stream i are divided
L	Code word length of the RS code
$P_e(G)$	Probability of an error occurring in the good state
$P_e(B)$	Probability of an error occurring in the bad state
$P_{g\|b}$	Channel state transition probability from good to bad
$P_{b\|g}$	Channel state transition probability from bad to good
H	Number of images interleaved together
Δ	Maximum tolerable time of recovering the image
Λ_l	Minimum packet size
Λ_u	Maximum packet size

complexity, we solve the problem with a heuristic approach for sub-optimal solutions. The proposed scheme is adaptive to the dynamic mmWave link conditions for enhanced video quality. It is evaluated with simulations and is shown to outperform an existing scheme reported in [19] with considerable gains.

The remainder of this chapter is organized as follows. Related work is reviewed in Section 6.2. We present the proposed MD coding and transmission scheme in Section 2.3 and simulation results in Section 3.6. Section 6.8 concludes this chapter. The notation used in this chapter is summarized in Table 2.1.

2.2 Related Work

In this section, we briefly review related work on video transmission over mmWave channels, and two MD-FEC techniques and Reed-Solomon coding, which are the basis of our proposed scheme.

Considering the similarity in the most important bits of the neighboring pixels in an image, the authors of [10] propose a technique to correct the error in the most important bits in a pixel by comparing those bits of the pixel with the bits at the same positions in its neighboring pixels. Shao et al. in [18] also develop a method to recover the bits of a pixel using the neighboring pixels. With this scheme, the neighboring pixels are encapsulated into different packets that are transmitted separately, so that a lost pixel can still be approximated by other received packets containing the neighboring pixels. However, there is no measures on providing better protection for the most important bits during transmission, which affects the quality of reconstructed video frames.

In [19], an Uncompressed Video streaming over Wireless (UVoW) system is proposed for transmission of uncompressed video on mmWave channels. In this system, the 8 bits representing a pixel are divided into two 4-bit parts according to their significance in recovering the video: MSB (most significant bits) and LSB (least significant bits). The MSB portions are better protected than the LSB portions by being transmitted through a channel of better condition. The encoder retransmits the MSB part if it has an error, but corrupted LSB parts will not be retransmitted. However, there is no evidence that the system achieves the best performance by cutting the 8 bits into two 4-bit segments, and there is no investigation on, when channel condition changes, how to adjust the operational parameters to make it adaptive to the changing channel conditions.

In [5], Bosco develops a cross-layer adaptive scheme for HD videos streaming over mmWave channels, where the sender adjusts the modulation and coding schemes (MCS) in response to channel variations to maintain good video quality. A method of cooperation with relay for transmission of HD videos over mmWave channels is developed in [9]. The source and relay use different error correcting codes so that error correcting performance can be enhanced, and spatial diversity gain is

obtained by using the relay. As in [19], the message is also cut into two fixed parts and the optimal segmentation of the message is out of concern in this related work [9].

MD video coding has been shown to be highly effective for video over lossy networks [4, 11, 13, 14, 21]. Among the many MD coding techniques, MD-FEC is a popular one based on Priority Encoding Transmission (PET) [3], which assigns different levels of protection to data segments according to their respective priority defined by the user [21]. The higher the priority, the smaller number of packets the decoder needs to decode it, so that the higher priority segments have a greater probability of being decoded. The basis of this algorithm is Reed-Solomon (RS) code, which is a sub-class of the Bose-Chadhuri-Hocquenghem (BCH) codes. RS coding provides an effective way of recovering lost data symbols from a portion of data symbols received [17].

2.3 System Model, Problem Statement and Solution Procedure

We consider streaming uncompressed HD video over a mmWave wireless link (e.g., from a storage device to an HDTV or a video projector). The MD-FEC coding scheme for uncompressed HD video is introduced in Section 2.3.1. The mmWave channel model and the interleaving based transmission strategy is presented in Section 2.3.2. Then we present our problem formulation in Section 2.3.3 and solution procedure in Section 2.3.4.

2.3.1 MD-FEC Coding of Uncompressed Video

Consider an uncompressed video frame of $M \times N$ pixels. Each pixel consists of $3 \times 8 = 24$ bits, while each 8-bit block corresponding to one of the R/G/B color components as shown in Fig. 2.1. The j-th bit in each of the 8-bit block represents a value of 2^{8-j}. Obviously, the j-th bit is more important than the $(j+i)$-th bit, for all $i \geq 1$, because it has a greater influence on the color depth of the R/G/B component of the pixel.

Next we divide the frame evenly into Q *sub-images*, each of which has $C \times D$ pixels, i.e., $Q = \frac{MN}{CD}$, for the purpose of:

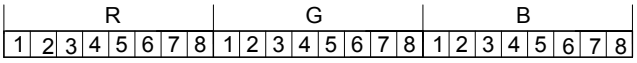

Figure 2.1: Structure of a pixel in the uncompressed video.

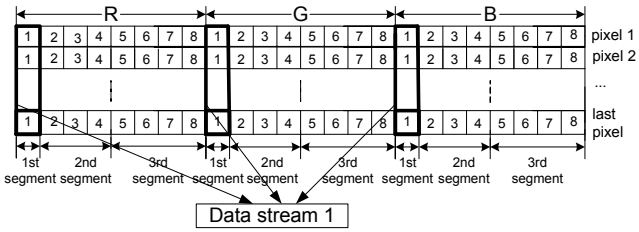

Figure 2.2: Structure of a data stream, assuming $S = 3$.

- the packet size will not become too large so that the packet transmission time can be fitted in the unit time slot in our channel model;
- the number of data symbols of a data stream will not become too large so that RS codes can be applied to the data symbols.

Now let's take the r-th sub-image as an example; each of the following procedures will be performed on other sub-images as well. First, each of the 8 bits representing the R/G/B color in a pixel is divided into S *segments*, $1 \leq S \leq 8$, and the length of (i.e., the number of bits in) segment i, is denoted as m_i, $1 \leq i \leq S$. Rearrange all the i-th segments of the R/G/B color of all the pixels in a sub-image into a new segment called *data stream* i, which has $\frac{3m_iMN}{Q}$ bits. We define the *significance* of data stream i of sub-image r as the summation of the values of all the data bits in segment i of sub-image r, and denote it as w_i^r.

$$w_i^r = \frac{3MN}{Q} \sum_{l=1}^{m_i-1} 2^{8-\sum_{j=1}^{i-1} m_j - l}. \tag{2.1}$$

Due to uncompressed video, there is no reason to prefer one pixel over another when semantic content is not considered. We thus have $w_i^r = w_i^q$ for any $r \neq q$, $1 \leq r \leq Q$, $1 \leq q \leq Q$. Fig. 2.2 shows an example of cutting each of the 8 bits of a pixel in a sub-image into 3 segments and combining all the first segments into data stream 1 of the sub-image.

From (2.1), we can see that $w_i^r \geq w_{i+1}^r$, for all $1 \leq r \leq Q$, $1 \leq i \leq S-1$, which implies that recovering data stream i is more important than recovering data stream $(i+1)$ for the purpose of enhancing the quality of reconstructed frames at the receiver. Therefore, we assign higher protection to data stream i than data stream $(i+1)$ to make the probability of recovering data stream i higher than that of recovering data stream $(i+1)$. We illustrate how to achieve this goal below.

First, we let k bits in a data steam form a *data symbol*. For example, we can take $k = 8$ for the convenience of computer processing. Then data stream i can be represented by $\frac{3m_iMN}{Qk}$ data symbols. These $\frac{3m_iMN}{Qk}$ symbols of data stream i is further divided into g_i groups, each of which has $\frac{3m_iMN}{Qkg_i}$ data symbols. We apply RS code to each group so that the total number of symbols (including both data and FEC symbols) is L, $L \leq 2^k - 1$, which is a constraint of RS coding.

This way, the number of FEC symbols assigned to each group of data stream i is $L - \frac{3m_iMN}{Qkg_i}$. The j-th symbol of all the groups of all the data streams compose the j-th *description* of the frame, and is transmitted in one packet. This procedure is illustrated in Fig. 2.3. To assign higher protection to data stream i than data stream $(i+1)$, the error correcting capacity, or the amount of FEC symbols for data stream i should be greater than that of data stream $(i+1)$, i.e., $L - \frac{3m_iMN}{Qkg_i} \geq L - \frac{3m_{i+1}MN}{Qkg_{i+1}}$, which implies that

$$\frac{m_i}{g_i} \leq \frac{m_{i+1}}{g_{i+1}}. \tag{2.2}$$

As shown in Fig. 2.3, each description (or packet) consists of video data bits and FEC bits from all the data streams, and is equally important for reconstructing the frame. The quality of the reconstructed frame is proportional to the number of descriptions received. If $\frac{3m_1MN}{Qkg_1}$ or more error-free descriptions are received, then all the groups of data stream 1 can be decoded, which guarantees a basic quality for this video frame. In general, if the decoder receives at least $\frac{3m_iMN}{Qkg_i}$ error-free descriptions, then all the data streams up to i can be decoded, resulting in an improved

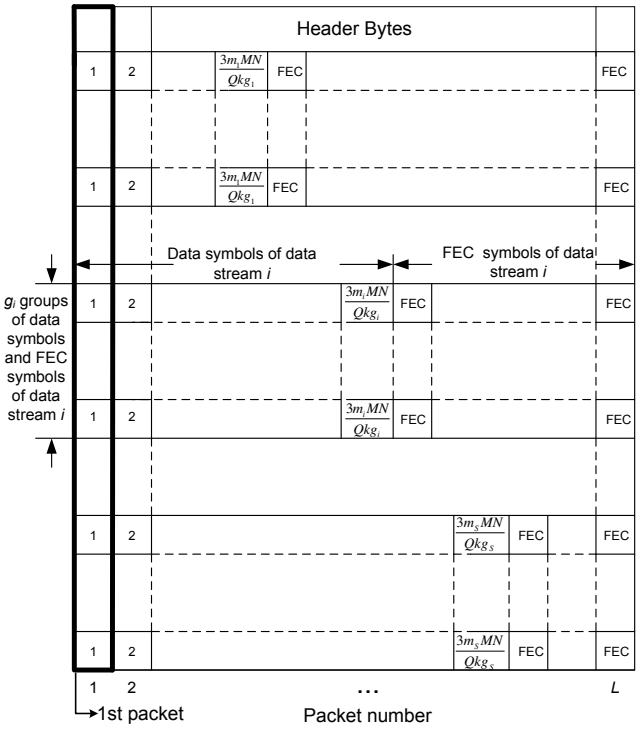

Figure 2.3: Structure of the packets of a sub-image.

quality for the frame. When more than $\frac{3m_S MN}{Qkg_S} \leq L$ error-free descriptions are received, the video frame can be reconstructed without any error.

2.3.2 mmWave Channel and Transmission Schedule

After MD-FEC coding, the video packets are then transmitted over the mmWave link from the sender (e.g., a storage device) to the receiver (e.g., an HDTV or video projector). As discussed, the mmWave channel is highly directional and susceptible to blockage of the link-of-sight path. Therefore, we model the mmWave link with the Gilbert-Elliot (GE) model, which is a discrete-time two-state Markov chain as shown in Fig. 2.4 [22]. The two states are: good state (denoted as G in the figure) and bad state (denoted as B). The probabilities that an error occurring at the good state and bad state are denoted as $P_e(G)$ and $P_e(B)$, respectively. The transition probabilities from G to B and from B to G are $P_{g|b}$ and $P_{b|g}$, respectively. Because the Gilbert-Elliot channel is a discrete-time model instead of a continuous-time model, we assume time is slotted and the

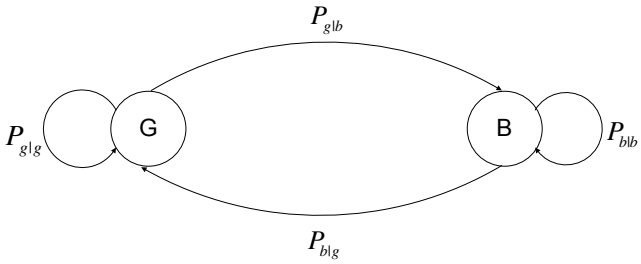

Figure 2.4: The Gilbert-Elliot link model.

transmission time of a packet over the mmWave link is less than or equal to the duration of one time slot [4, 13].

When multiple beam paths are available between the transmitter and receiver (e.g., a direct path and a second path due to reflection on the wall), the multipath transport technique can be used for the MD video packets [11]. In this chapter, we focus on the single beam path case, while the proposed technique can be extended to the case of multiple beam paths. Recall that MD-FEC coding can ensure a minimum video quality if at least $\frac{3m_1MN}{Qkg_1}$ error-free packets are received for a frame. However, if most of the packets are transmitted within a bad period, it is possible that less than $\frac{3m_1MN}{Qkg_1}$ packets can be received and the video frame cannot be decoded. In order to reduce the bursty error rate, we adopt *interleaving* to the packets of multiple consecutive frames.

The interleaving scheme is illustrated in Fig. 2.5. The *interleaving depth* is the time interval between the transmissions of two consecutive packets of the same sub-image. As shown in the figure, the interleaving depth is set to QH, while H is a non-negative integer. We can adjust H to find the best interleaving depth for different channel conditions in terms of $P_e(G)$, $P_e(B)$, and the transition probabilities. when the average bad period T_{off} is large, we can increase QH by increasing H. The probability that multiple consecutive packets of a sub-image are transmitted within the bad period can be reduced. Since RS coding is applied in our scheme, the probability of failing to decode data stream i of sub-image X of image Y is the probability that the decoder receives less than $\frac{3m_iMN}{Qkg_i}$ error-free data packets, which can be expressed as [22]

$$P_i^{XY} = \sum_{d=0}^{Z} P_z(d) \left\{ \sum_{e_b=0}^{d} \left[C_d^{e_b} P_e^{e_b}(B) \left(1 - P_e(B)\right)^{d-e_b} \right. \right.$$

$$\left.\left.\sum_{e_g=\max\{0,t+1-e_b\}}^{Z-d} C_{Z-d}^{e_g} P_e^{e_g}(G)\left(1-P_e(G)\right)^{Z-d-e_g}\right]\right\}, \tag{2.3}$$

where $P_Z(d)$ is the probability that given the GE channel is observed at Z continuous time slots, the probability that the channel is in the bad state for d time slots; Z is the block code length of the error correcting code being used and thus $Z = L = 2^k - 1$ in our example; e_b and e_g denote the numbers of packets in error when the channel is the bad state and good state, respectively; t is the error correcting capability of the error correcting code for data stream i; and $P_i^{XY} = P_i^{X'Y'} = P_i$, for any $X \neq X'$ and $Y \neq Y'$, due to uncompressed videos. Since the sequence numbers of all the data symbols are known, the RS code can correct up to $t = L - \frac{3m_iMN}{Qkg_i}$ known erasures as an erasure code [15].

For brevity, retransmission is not explicitly implemented in our system. Therefore, the transmission time for all the packets of a data stream i is LHQ time slots. A retransmission scheme can be translated to reduced $P_e(G)$ and $P_e(B)$, but the transmission times for the data streams could be longer [19].

2.3.3 Problem Statement

Now we are ready to formulate the problem of MD video over mmWave networks. Due to uncompressed video, the range of PSNR is from 0 (all lost) to positive infinity (error free), which makes no practical sense. On the other hand, the Mean Squared Error (MSE) of the reconstructed image reflects the distortion of the reconstructed image compared with the original image. Therefore our goal is to minimize the expected Mean Squared Error (i.e., MSE or $\sqrt{\text{MSE}}$) of the reconstructed video frames, which is a function of the various design factors as follows.

$$\text{MSE} = \frac{1}{3MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{\chi=1}^{3}\left[I(i,j,\chi) - I'(i,j,\chi)\right]^2. \tag{2.4}$$

In (2.4), $I(i,j,\chi)$ is a pixel in the original video frame at location (i,j) and $I'(i,j,\chi)$ is the pixel at the same location in the reconstructed video frame. The index χ indicates the color components

(i.e., R, G or B). Since it is assumed that the three color components are equally important for the image quality, we don't introduce different weights for the three color components.[1] It follows that

$$
\begin{aligned}
&\mathbb{E}\{\sqrt{\text{MSE}}\} \qquad (2.5)\\
&= \mathbb{E}\left\{\sqrt{\frac{1}{3MN}\sum_{q=1}^{Q}\left[\sum_{i=1}^{C}\sum_{j=1}^{D}\sum_{\chi=1}^{3}\left(\sum_{r=T_q}^{S_q}\frac{Q}{3MN}w_q^r\right)^2\right]}\right\}\\
&= \mathbb{E}\left\{\sqrt{\frac{1}{3MN}\sum_{q=1}^{Q}\left[3CD\left(\sum_{r=T_q}^{S_q}\frac{Q}{3MN}w_q^r\right)^2\right]}\right\}\\
&= \mathbb{E}\left\{\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\sum_{r=T_q}^{S_q}\frac{Q}{3MN}w_q^r\right)^2}\right\}\\
&= \sum_{T_1=1}^{S_1}\cdots\sum_{T_Q=1}^{S_Q}\left\{\left[\prod_{q=1}^{Q}P(x_{T_q-1}\le X_q<x_{T_q})\right]\times\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\sum_{r=T_q}^{S_q}\frac{Q}{3MN}w_q^r\right)^2}\right\}\\
&= \sum_{T_1=1}^{S_1}\cdots\sum_{T_Q=1}^{S_Q}\left\{\left[\prod_{q=1}^{Q}\left(P_{T_q}-P_{T_q-1}\right)\right]\times\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\sum_{r=T_q}^{S_q}\frac{Q}{3MN}w_q^r\right)^2}\right\}
\end{aligned}
$$

Where S_q, $1 \le q \le Q$ is the total number of segments that each pixel of the sub-image q is divided into; T_q is the first segment of sub-image q that cannot be decoded; X_q is the number of packets of sub-image q that are received and not corrupted by error; x_{T_q} is the number of data symbols required to decode T_q (and denote $x_0 = 0$); $P(X_q < x_{T_q})$ is the probability that X_q is less than x_{T_q}, such that $P(X_q < x_{T_q}) = P_{T_q}$, where P_{T_q} is the probability that data T_q cannot be decoded, as given in (2.3). Note that we assume if a bit cannot be decoded then it is discarded at the decoder, meaning that (2.5) is indeed the worst-case expected $\sqrt{\text{MSE}}$.

[1]The same scheme can be applied to videos stored in the YUV format, where the weights for the Y, U and V components may vary and three optimization problems need to be solved.

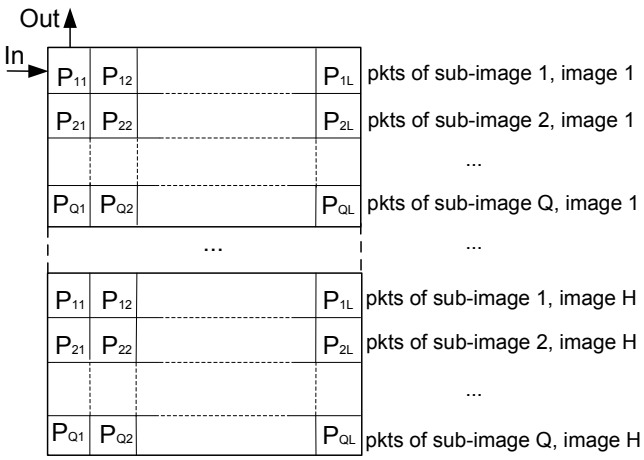

Figure 2.5: Structure of the interleaver.

We then formulate the following nonlinear integer programming (NLIP) problem.

$$\text{minimize} \quad \mathbb{E}\{\sqrt{\text{MSE}}\} \tag{2.6}$$

$$\text{subject to:} \quad \frac{m_i}{g_i} \leq \frac{m_{i+1}}{g_{i+1}}, \text{for all } i \tag{2.7}$$

$$\frac{3m_{S_q}MN}{Qkg_{S_q}} \leq L, \text{for all } q \tag{2.8}$$

$$L \leq 2^k - 1 \tag{2.9}$$

$$\sum_{i=1}^{S_q} m_i = 8, \text{for all } q \tag{2.10}$$

$$LHQ \leq \Delta \tag{2.11}$$

$$\Lambda_l \leq \sum_{i=1}^{S_q} \frac{g_i k}{8} \leq \Lambda_u, \text{for all } q. \tag{2.12}$$

Constraint (3.27) is due to the fact that the transmission time of all the packets of an image must not be larger than a tolerable transmission delay, i.e., the delay bound Δ. Constraint (2.12) is for the purpose of making the size of a packet adequate for transmission after adding header bytes to the packets, i.e., within the range of the minimum packet size Λ_l and the maximum packet size Λ_u.

2.3.4 Solution Strategy

The formulated problem aims to minimize the expected $\sqrt{\text{MSE}}$ given in (2.5). This goal is achieved by tuning the control parameters S_q, $\{m_1, m_2, \cdots, m_{S_q}\}$, $\{g_1, g_2, \cdots, g_{S_q}\}$, for all q, and H to find the optimal partition of the pixel bits for MD coding and the optimal interleaving depth for transmission. And the problem is solved once when the channel condition is changed.

Checking out the search space for the optimal solution, it can be seen that: (i) the feasible values for S_q is between 1 and 8; (ii) due to constraint (3.26), the m_i's are all small integers; and (iii) H is also limited by the maximum delay constraint (3.27). The parameters that dominate the search space of the problem are $\{g_1, g_2, \cdots, g_{S_q}\}$, for all q, since the packet size can be a value between hundreds to thousands of bytes. This motivates us to reduce the search space by "taking out" $\{g_1, g_2, \cdots, g_{S_q}\}$, for all q, such that the search space of the problem will be small enough to apply an exhaustive search for the remaining parameters. This way, the problem can be reduced to a sub-optimal problem having S_q, $\{m_1, m_2, \cdots, m_{S_q}\}$, and H as its optimization parameters.

In our solution strategy, we reduce constraint (2.12) into $\sum_{i=1}^{S_q} \frac{g_i k}{8} = \Lambda$, where Λ is a constant, for all q. For a particular S_q, for all q, we have $g_i = \frac{8\Lambda}{kS_q}$, for all i. Furthermore, constraint (3.21) can be transformed to $m_i \leq m_j$, for all $i < j$. Then the problem is reduced to a sub-optimal problem of finding the optimal $\{m_1, m_2, \cdots, m_{S_q}\}$ and H under a specific S_q value. The search space of the sub-optimal problem is now much smaller than that of the original problem, so that exhaustive search can be applied. The ideas of finding the best bits-partition scheme and interleaving depth are still maintained in the sub-optimal problem, thus the proposed scheme still being adaptive to channel dynamics. Note that for the case where $\frac{8\Lambda}{kS_q}$ is not an integer, we take $g_i = \lfloor \frac{8\Lambda}{kS_q} \rfloor$, for all $1 \leq i \leq S_q$ and the remaining space in the packet left by the data symbols are padded by redundancy.

In addition, we also aim to reduce the number of loops caused by Q, since a large number of loops will seriously drag down the speed of the search algorithm, as can be seen from (2.5). We cut the HD video image into many small images (different from the sub-images mentioned above) and take each of the small images as an intact image, and then interleaving is applied to these small

images. Thus $Q = 1$ in the exhaustive search algorithm and our following simulations. And we omit the subscript q for brevity in the following. This way, the number of loops is reduced and simulation speed is greatly improved.

After these simplifications, the computation overhead is dominated by the computation of (2.3), which is the probability of failing to decode a data stream, and by searching for the optimal pixel bit allocation and interleaving depth within a limited searching space. The computation can be further sped up by pre-computing the probabilities (2.3) under various cases and store them in a table for future use. The performance of this solution strategy will be demonstrated in the performance evaluation section.

2.4 Simulation Study

2.4.1 Simulation Setup

In this section, we evaluate the performance of the proposed scheme. The coding and transmission schemes are implemented and simulated using Matlab. In the simulations, we use 1920 $\times$ 1080 HD uncompressed images with 24 bits per pixel (i.e., 8 bits for each R/G/B component), instead of a real HD video stream. The simulation parameters are shown in Table 6.1. We compare the proposed scheme with UVoW presented in [19], as discussed in Section 6.2. Note that ARQ is not employed in both schemes for a fair comparison. The channel conditions are assumed to be known in advance, e.g., through a proper channel measurement/feedback scheme. In the simulations, we assume a time slot duration of 1 μs.

2.4.2 Simulation Results and Analysis

We first examine the operation of the proposed scheme to reveal its properties. Figure 2.6 visually shows the different significance of bits at different positions for reducing $\sqrt{\text{MSE}}$. In this simulation, the segmentation of the eight bits of a color component in a pixel is 1-1-1-2-3, which means the eight bits are divided unevenly into five segments, with 1, 1, 1, 2, and 3 bits, respectively.

Table 2.2: Simulation Parameters

Parameters	*Value*
M	1080
N	1920
C	60
D	80
k	6
Δ	10^6 time slots
L	51
H	$1 \sim 5$
Λ_l	800 Bytes
Λ_u	2000 Bytes
$\sum_{i=1}^{S} \frac{g_i k}{8}$	900 Bytes

The significance of the segments are 128, 64, 32, 24, and 7, respectively. When less than 10 packets are received, the video cannot be decoded, resulting in a $\sqrt{\text{MSE}}$ of 255. When more than 10 but less than 20 packets are received, the most important bits in the pixels can be decoded, resulting in greatly reduced $\sqrt{\text{MSE}}$. The $\sqrt{\text{MSE}}$ curve has a staircase shape and the gap is getting smaller as more and more packets are received.

We learn from the figure that recovering the bits of larger values has a greater effect on lowering the $\sqrt{\text{MSE}}$ than recovering the bits of smaller values, which is the guideline for the proposed algorithm. Recall that the total number of FEC symbols is $L \sum_{i=1}^{S} g_i - \frac{3 \cdot 8 \cdot M \cdot N}{k}$, where the first term is the total number of symbols and the second term is the total number of data symbols in a sub-image. Therefore, the amount of FEC symbols in a sub-image is fixed. If we want to provide stronger protection to the bits of greater values by assigning more FEC symbols to them, then the bits of smaller values will be less protected, and vice versa. If the channel is bad and the packet loss rate is high, then it is more urgent to offer higher level of protection to the bits of greater values than to the bits of smaller values considering their respective level of significance for lowering the $\sqrt{\text{MSE}}$. When channel is good and the packet loss rate is low, then the probability of recovering the bits of greater values will be high enough even if not many FEC symbols are assigned to them,

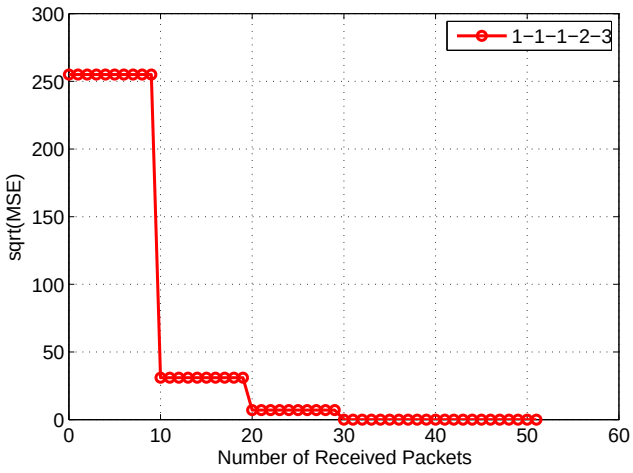

Figure 2.6: Significance of different bits at different positions to $\sqrt{\text{MSE}}$.

thus enough FEC symbols may be assigned to the bits of smaller values to enhance the probability of decoding them to further lower the $\sqrt{\text{MSE}}$.

This reasoning has been validated when we checked with the segmentation of the 8 bits under different channel conditions in simulations. Another question is how effective is the MDC component of the proposed scheme in improving video quality. To answer this question, we compare the performance between our proposed scheme and another scheme with interleaving only but without MDC. The results is presented in Fig. 2.7, under a channel condition of $P_{b|g}$ changes from 0.06 to 0.3, $P_{g|b} = 0.1$, and $P_e(G) = 0.01$. Note that according to the channel model, the average off duration is $T_{off} = 1/P_{b|g}$. It can be seen that when the channel condition is good (with short off periods), the gain achieved by MDC is not obvious. However, when the channel gets worse with increased off periods, the gap between the two curves are quickly increasing. Considerable reduction in $\sqrt{\text{MSE}}$ by MDC is observed under bad channel conditions.

The relationship between the packet loss trace and $\sqrt{\text{MSE}}$ trace is plotted in Fig. 2.8 under a channel condition of $P_{b|g} = 0.1$, $P_{g|b} = 0.1$, and $P_e(G) = 0.01$. The packet loss trace is plotted in dashed line using the left-hand-side y-axis, and the corresponding $\sqrt{\text{MSE}}$ is plotted in solid line using the right-hand-side y-axis. The corresponding optimal segmentation of the 8 bits is 2-3-3, and the optimal interleaving depth is 5. We can see from the above figure that when packet loss exceeds a certain threshold and thus the last two segments cannot be decoded, the $\sqrt{\text{MSE}}$ of the

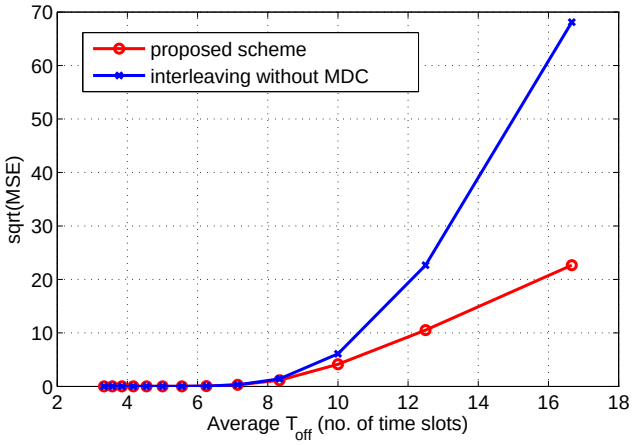

Figure 2.7: Impact of the MDC component under various channel conditions.

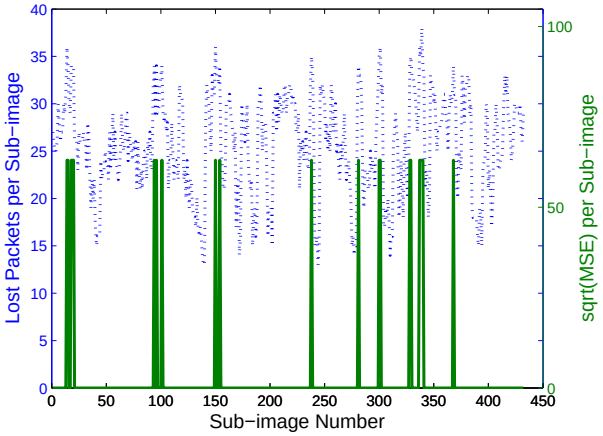

Figure 2.8: $\sqrt{\text{MSE}}$ vs. packet loss trace.

sub-image is about 60; otherwise, the $\sqrt{\text{MSE}}$ is 0, which means all the three segments are decoded. The statistical average $\sqrt{\text{MSE}}$ of all the tested sub-images is 3.0652 while the expected $\sqrt{\text{MSE}}$ calculated from the minimization problem is 4.1472. The difference may be explained by a limited number of tested sub-images here, and the dependence of the channel condition for the consecutive sub-images which is not considered on the minimization problem which tends to minimize the expected $\sqrt{\text{MSE}}$ of a sub-image and doesn't consider the channel condition correlation problem of consecutive sub-images.

We next focus on the impact of channel dynamics to show how the proposed scheme adapt to mmWave channels. We compare the expected $\sqrt{\text{MSE}}$ achieved by the proposed scheme and UVoW, by varying the average duration of the channel bad period T_{off}. In the simulations, $P_{g|b}$ is

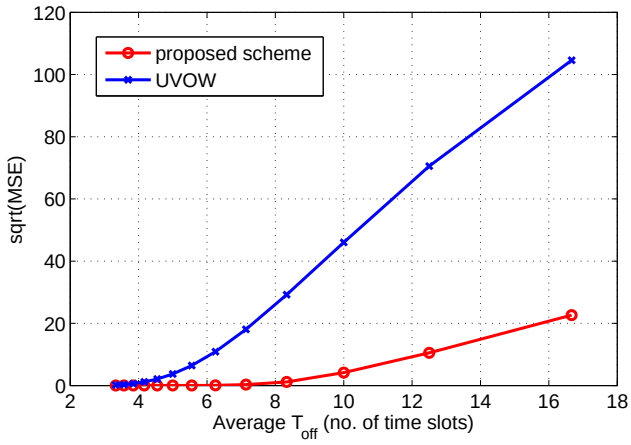

Figure 2.9: $\sqrt{\text{MSE}}$ vs. average off period of the mmWave channel T_{off}.

fixed at 0.1 and $P_e(G)$ is fixed at 0.01 for both schemes. The results are plotted in Fig. 2.9. Note that using our proposed scheme, the best combination of $\{S, m_1, m_2, \cdots, m_S, H\}$, which leads to the lowest expected $\sqrt{\text{MSE}}$ under the given channel condition, can be found for each different average T_{off} value. We can see from the figure that the expected $\sqrt{\text{MSE}}$ of the proposed scheme is lower than that of UVoW for the entire range. When channel condition is good (i.e., the average T_{off} is less than 4), the performance gap between the two schemes is not remarkable. However, as the channel condition degrades (i.e., average T_{off} is increased beyond 4), the performance gap quickly grows, for the reason that the packet loss rate grows thus it is of greater importance to provide stronger protection to the most important bits. We observed from the result that when T_{off} is 4, the optimal segmentation is 4-4 (as in UVoW) and the optimal interleaving depth is 5; when average T_{off} is 10, the optimal segmentation is 2-3-3 and optimal interleaving depth is 5, which means that more FEC symbols are assigned to the first 2 bits, and interleaving is applied to overcome the channel burstiness. Thus the proposed scheme can adapt to changing channel conditions, to strengthen the protection for the most important bits and increase the interleaving depth to combat the temporarily blocked channel. On average, the proposed scheme achieves 19.6055 reduction in $\sqrt{\text{MSE}}$ than UVoW.

Fig. 2.10 demonstrates the relationship between the expected $\sqrt{\text{MSE}}$ and $P_e(G)$, the packet loss probability when the channel is in the good state. In this simulation, $P_{b|g}$ and $P_{g|b}$ are fixed at

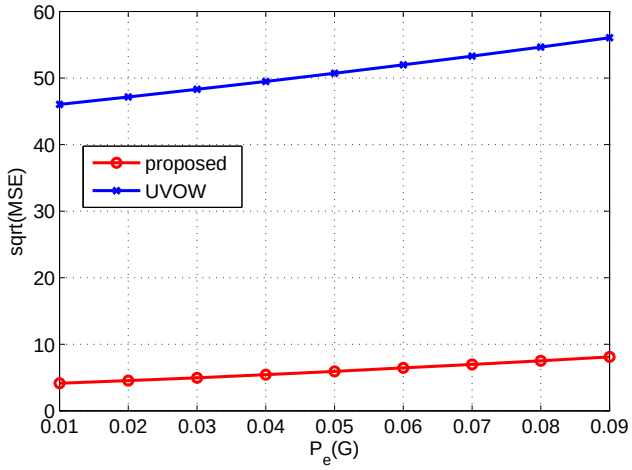

Figure 2.10: $\sqrt{\text{MSE}}$ vs. packet loss rate at the good state $P_e(G)$.

0.1 for both schemes. As $P_e(G)$ is increased, the probability of a packet being lost in the good state is increased, which means the error correcting capability of the bits of the highest priorities should be enhanced by re-segmentation of the bits. As Fig. 2.10 shows, the characteristic of adaptivity enables the proposed scheme outperform UVoW in terms of $\sqrt{\text{MSE}}$ for the entire range of $P_e(G)$ examined in the study. The average $\sqrt{\text{MSE}}$ of the proposed scheme is 44.8486 lower than that of UVoW. This result also indicates that our proposed scheme is extremely suitable for the case that $P_e(G)$ is comparatively high during communication.

We look into the optimal segmentation result obtained by solving the minimization problem (4.7). We find that, in general, when channel is good, more descriptions are created as the bits are divided into more segments, and the length of the segments of the more significant bits are shorter. This implies that the number of data symbols constituted by the bits of greater values are smaller compared with the number of FEC symbols assigned to protected them, so that these bits have a higher level of protection.

Fig. 2.11 and 2.12 plot the relationship between $\sqrt{\text{MSE}}$, $P_{b|g}$ and transmission delay tolerance Δ. Recall that the relationship between transmission delay tolerance (short for *delay* hereafter) and interleaving depth H is given in (3.27). Under a specific delay, the expected $\sqrt{\text{MSE}}$ will decrease as $P_{b|g}$ increases; this pattern also applies to the relationship between delay and expected $\sqrt{\text{MSE}}$ under a specific $P_{b|g}$. Besides, in the entire ranges of delay and $P_{b|g}$ as shown in this figure,

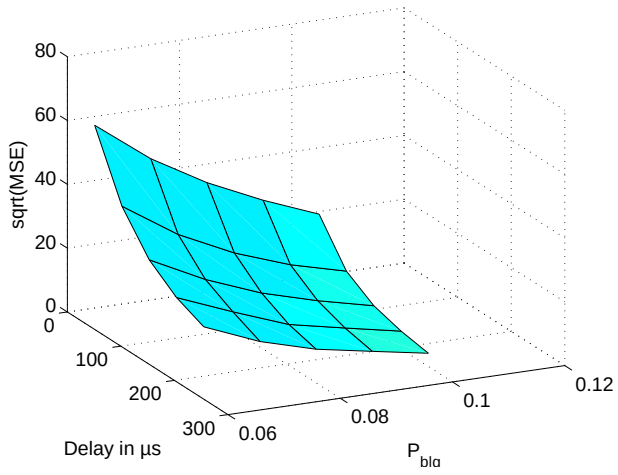

Figure 2.11: $\sqrt{\text{MSE}}$ vs. transmission delay tolerance under good channel condition.

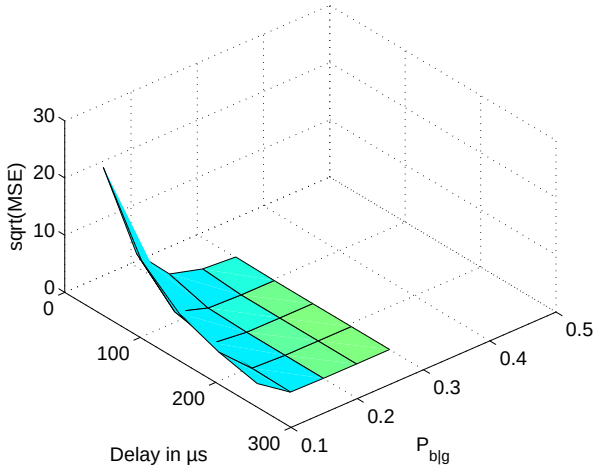

Figure 2.12: $\sqrt{\text{MSE}}$ vs. transmission delay tolerance under bad channel condition.

the effect of increased delay and that of increased $P_{b|g}$ on lowering the expected $\sqrt{\text{MSE}}$ do not conflict with each other, which means choosing a larger interleaving depth under a specific channel condition will be always good for improved video performance, especially when channel condition is bad (i.e., $P_{b|g}$ ranges from 0.06 to 0.1). This is because the reduction of $\sqrt{\text{MSE}}$ is much more significant than that achieved when channel condition is good (i.e., $P_{b|g}$ ranges from 0.1 to 0.3), as plotted in Fig. 2.12. This phenomenon motives us to carefully design the system parameters under certain channel conditions in order to achieve a balance between delay and desired video performance.

Finally, we examine the visual quality of reconstructed HD video. In the next simulation, example video images are used in our performance test. About 10 distinct images are evaluated using both the proposed scheme and UVOW. The image which reflects the simulation results of most tested images are selected to undergo about 100 distinct evaluations using both schemes in each test. We choose the results that are representative of the 100 tests to present in the following. The optimal solutions of $\{S, m_1, \cdots, m_S, H\}$ corresponding to the lowest expected $\sqrt{\text{MSE}}$ are recalculated when channel condition changes, and then applied to the transmission of each sub-image. Since we take $Q = 1$ to reduce the complexity of solving the optimization problem, we tend to obtain the optimal solution on sub-image basis, and the optimal solution is the same for each sub-image in our simulations under the same channel.

In the simulations, the entire image is divided into smaller sub-images and we show only 1/16 of the sub-images (termed partial image in the following). This is because the size of an entire HD image will be too large to fit into this chapter, and the quality of the partial images is representative since all the partial images are treated equally in uncompressed video streaming. The partial images as shown in Fig. 2.13. The partial image in Fig. 2.13(a), (b), and (c) are produced by the proposed scheme and the partial images in Figs. 2.13(d), (e), and (f) are produced by UVoW. Fig. 2.13(a) and (d) are obtained under a good channel condition; Fig. 2.13(b) and (e) are obtained under a bad channel condition; Fig. 2.13(c) and (f) are obtained under a severe channel condition. The specific parameters are as follows: (a)(d): $P_{b|g} = 0.3$; (b)(e): $P_{b|g} = 0.1$; (c)(f): $P_{b|g} = 0.06$. $P_{g|b} = 0.1$ and $P_e(G) = 0.01$ are fixed in all the simulations.

It can be seen that when the channel condition is good, both schemes work fine by producing high quality received videos. When the channel is in a bad condition (i.e., $P_{b|g} = 0.1$), the partial images produced by our scheme are still good with a comparatively much high quality, while a portion of the partial images using UVoW are black (meaning that those pixels are completely lost). Under such bad channels, the most important bits can still be received and decoded in our scheme, but are completely lost in UVoW. This is because the protection to the most important

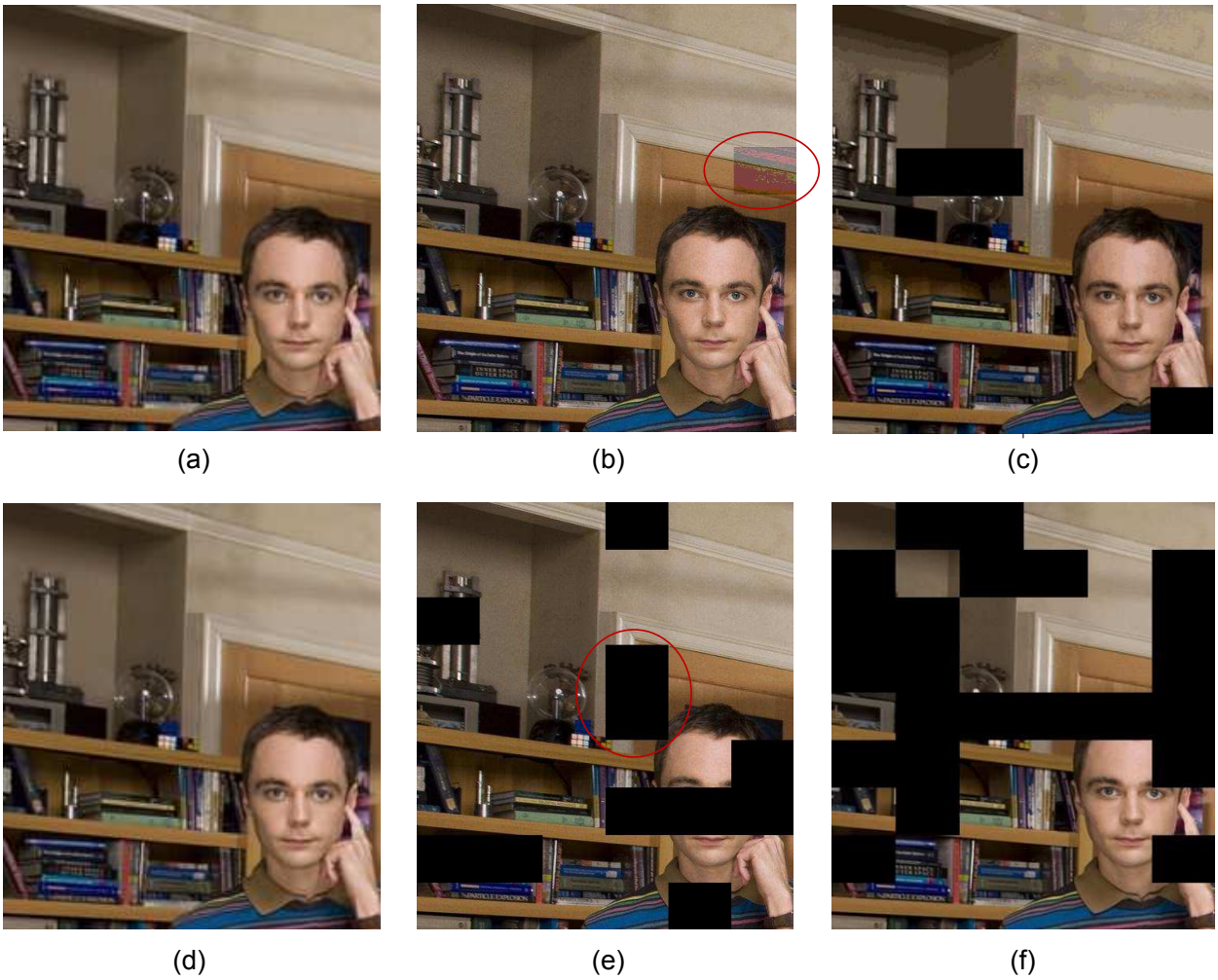

Figure 2.13: Visual quality comparison of the two schemes under different channel conditions.

bits is adaptively enhanced after re-segmentation of the pixel bits to adapt to the degraded channel condition.

We also observed that the optimal segmentation is 2-3-3 and the optimal interleaving depth is 5 under the bad channel condition. That is, the two most significant bits are better protected. The change in the color depth in Fig. 2.13(b) (indicated by the circle) is caused by the loss of the 6 least significant bits during transmission over the mm Wave channel. The data stream reconstituted by the last 6 bits cannot be decoded due to the transmission loss, but the first 2 most significant bits are correctly decoded since more FEC symbols are assigned to them to provide a higher level of protection, and consequently a basic visual quality is still guaranteed. Under the same channel

condition, the UVoW partial image in Fig. 2.13(e) has a few black blocks, which are caused by the loss of all the 8 bits in those pixels.

When the channel condition is even worse (i.e., $P_{b|g} = 0.06$), the visual quality of the reconstructed image using our proposed scheme (in Fig. 2.13(c)) is still much better than that using the UVoW (in Fig. 2.13(f)). The optimal segmentation is now 1-1-1-1-4 and the optimal interleaving depth is 5 under such channel condition. As can be seen from Figs. 2.13(c), although the color depth has a little change since the last four bits are lost (e.g., see the up right corner), and there are 3 black blocks, the visual quality is still acceptable compared with the UVoW image in Fig. 2.13(f), which has much more black blocks.

These results clearly demonstrate the advantages of being adaptive both at the MD video coder and at the transmission scheduler to the varying channel conditions.

2.5 Conclusion

In this chapter, we investigated the problem of streaming uncompressed HD video over mmWave wireless networks. We developed an MD-FEC coding scheme that partitions the pixel bits and an interleaving based transmission strategy to minimize the expected $\sqrt{\text{MSE}}$ of the reconstructed video quality. The main idea is that important bits should have higher level of protection in transmission. We formulated an Nonlinear Integer Programming problem for the optimal partition of the pixel bits and interleaving of packets, which is NP-hard, and derived a sub-optimal solution for this problem with much lower computational complexity. The performance of the proposed scheme was evaluated with simulations and shown to outperform an existing scheme with considerable gains.

Chapter 3

QoE Driven Multi-user Video Streaming in Cellular CRNs

3.1 Introduction

The Cisco Visual Network Index report predicts a drastic increase in mobile data and a dominant part of video related data in the near future [62] [45]. calls for a more flexible management of radio resources in today's and future wireless networks, in order to unlock the wireless network capacity by promoting more efficient use of spectrum. To this end, the Cognitive Radio (CR) technology has been widely recognized as an effective solution for efficient and flexible access to the radio spectrum. CR is an evolutionary technology for more efficient and flexible access to the radio spectrum. In a cognitive radio network (CRN), Cognitive Users (CUs) search for the unoccupied licensed spectrum of the Primary User (PU) network and then opportunistically access detected spectrum holes in an unobtrusive manner [26] [43] citeXiao16Globe [46]. Bandwidth-demanding and elastic mobile services, such as wireless video, will benefit enormously from this new wireless networking paradigm [27].

In this chapter, we address the challenging problem of downlink multi-user video streaming in cellular CRNs. We consider a CRN consisting of one cognitive base station (CBS) and multiple CUs. Without loss of generality, we assume each CU can access one channel at a time (i.e., with a single antenna). The CUs cooperatively sense PU signals on licensed channels and the CBS infers the channel states based on collected CU sensing results with an OR fusion rule. Once the idle channels are detected, the CBS then assigns them to active CUs for downlink multi-user video streaming. We incorporate the video assessment model proposed in [66, 71], aiming to maximize the CU QoE by optimal designs of spectrum sensing and access policies.

It is obviously a challenging problem to jointly design the spectrum sensing and access polices for QoE-aware multi-user video streaming, due to the large number of design factors and the

complex interactions that should be modeled in a cross-layer optimization framework. We first consider the case where each CU can sense and access at most one channel at a time slot. To make the problem tractable, we take a divide-and-conquer approach to break it into two sub-problems: (i) Optimal Assignment Sub-problem for Spectrum Sensing (OAPSS): to discover a sufficient amount of idle channels reliably and quickly to meet the bandwidth demand of the CUs; and (ii) Optimal Assignment Sub-problem for Video Transmission (OAPVT): to allocate available channels to CUs according to their respective QoE requirements and network status. We propose a distributed Greedy Poly-matching Algorithm that can compute optimal solution to the channel sensing sub-problem, and using the Hungarian Method to compute optimal solution to the channel assignment sub-problem.

Furthermore, we examine the more general case where each CU can sense multiple channels (e.g., with multiple spectrum sensors) but can still access only one channel at a time slot. We formulate an integrated problem that maximizes the QoE of all the CUs by *jointly* optimizing spectrum sensing and access policies. Under the assumption that all the spectrum sensors work at the same operating point (i.e., with the same probability of detection and the same probability of false alarm), we show that this challenging problem can be solved with a two-step approach: First, the spectrum sensing scheduling problem is solved with a greedy algorithm; Second, the channel allocation problem, which is a Maximum Weight Matching problem and can be solved optimally with the Hungarian Method. We prove that the two-step solution algorithm is indeed optimal: decomposing the original problem into two sub-problems and solving them sequentially do not sacrifice the optimality of the solution.

It is worth noting that if we also assume identical operating points for the spectrum sensors, the single-channel sensing scenario is a special case of the multi-channel sensing scenario, to which the optimal solution approach also applies. We validate the proposed schemes with simulations, and the simulation results demonstrate their superior performance in terms of the MOS that CUs can achieve under various network scenarios, when compared with benchmark schemes.

The remainder of this chapter is organized as follows. The system model is presented in Section 3.3. The problem for the case of single channel sensing is formulated and solved in Section 3.4. The problem for the case of multi-channel sensing is formulated and solved in Section 3.5. Simulation results are discussed in Section 3.6. Section 6.2 reviews related work and Section 6.8 concludes the paper.

3.2 Related Work

In this section we briefly review the prior work on Quality of Service (QoS) and QoE provisioning and video streaming over CRNs.

CR research has been largely focused on the aspects of spectrum sensing and dynamic spectrum access [72] [48]. In [73], the authors study the sensing-throughput tradeoff problem that optimizes the spectrum sensing time so that the CU's throughput can be maximized with restricted interference to the PUs. Unlike [73], the protocol proposed in [74] also considers the problem of which channel to sense, in addition to sensing parameters and access strategy optimization. Moreover, it is shown that the design of sensing strategy is independent to sensing parameters design and the access strategy, as specified in a *principle of separation* [74]. These works focus on the optimization of sensing parameters only, and there is no collaboration between CUs. Considering the fact that different CUs may have different spectrum sensing performance, the algorithm proposed in [75] forms groups of CUs for cooperative sensing, aiming to find the best grouping scheme to discover most idle channels. Moreover, the problem of sensing parameter optimization in addition to optimal sensor selection is investigated in [76], with the objective to achieve a trade-off between detection performance and sensing overhead.

The problem of video streaming over CRNs has been studied in a few prior works. The transmission of multimedia over CRN is first proposed by J. Mitola in [77]. In [63] the quality optimization problem is formulated as an mixed integer nonlinear programming (MINLP) problem and solved with effective algorithms. Authors of [78] develop an auction game model to deliver

content-aware multimedia. The authors in [79] consider the scenario where multimedia transmission is scheduled in CRN and a QoE Driven channel allocation scheme is proposed to optimize the multimedia transmission of priority-based CUs, where the MOS model proposed in [71] is used. Specifically, each CU has different QoE requirements and thus has different priority in utilizing the idle channels of the PU system. Upon the re-appearance of an active PU on the idle licensed channel, each CU utilizing the idle licensed channels will evacuate from the current channel it is using to avoid conflict with the active PU.

The authors of [93] propose a a learning-based QoE-driven spectrum handoff scheme for multimedia transmissions over CR networks. Reinforcement learning is applied to spectrum handoff scheme to maximize the QoE of video transmissions in the long term. The proposed learning scheme is asymptotically optimal, model-free, and can adaptively perform spectrum handoff for the changing channel conditions and traffic load. To extend the video streaming time for the CUs, the authors of [94]propose a flexible sensing scheme to reducing the unnecessary channel sensings. Besides, the network abstraction layer units in the SVC video are assigned utilities which accurately reflect their contributions to the video quality, and different layers are streamed over different channels based on their contributions to maximize the total utility of the received video. In order to comprehensively evaluate the utility of the CUs in video streaming, the authors of [95] propose to not only consider the video quality of each CU, but also consider the the number of satisfied CUs. A 3-dimensional scalable quality of the H.264/SVC video transmission problem is formulated and solved with an suboptimal solution. In [96], the authors consider the case that the future Internet network may become highly heterogeneous, and therefore an efficient cognitive network management is proposed for the optimization of network operations like management of resources, mobility or QoS in order to ensure smooth network operation and high user satisfaction.

3.3 System Model

We consider a primary network operating on N_1 orthogonal licensed channels. The primary network is co-located with a CR network, which consists of a CBS supporting M CUs. The CUs

Algorithm 1: Spectrum Sensing and Access for QoE-driven Multi-user Video Streaming

1. **Phase 1**: The CBS determines for each CU which channel to sense based on SNR feedback, and broadcasts the sensing schedule to the CUs ;
2. **Phase 2**: Each CU follows the sensing schedule to sense the channel to which it is assigned, and reports the sensing result to the CBS ;
3. **Phase 3**: The CBS makes two decisions: (i) channel availability at the current time slot, based on the sensing results and the fusion rule; and (ii) channel assignment to CUs for multi-user video transmission at the current time slot, based on channel availability, channel condition, Content Type (denoted as CT) of each CU, and other information. Then the CBS broadcasts the channel access schedule to the CUs ;
4. **Phase 4**: The CBS uses the assigned channels to transmit video data; each CU follows the channel access schedule to receive video data from its assigned channel.

sense the PUs' usage of the licensed channels and access the licensed channels in an opportunistic manner. As in prior work [75, 80], we assume the CUs, when they are not receiving data, measure the SNRs of PU transmissions over all the licensed channels and report the measured SNRs to the CBS through some feedback mechanism. Based on such feedback, the CBS then assigns those CUs with good channel conditions to sense each licensed channel, in order to achieve a good sensing performance.

Here we consider the downlink multi-user video streaming scenario, where the CBS streams a video to each active CU using the license channels that are detected idle. We assume time is divided into a series of non-overlapping GOP (Group of Pictures) windows, each consisting of T time slots. Each time slot can be further divided into four phases for spectrum sensing and access for multi-user video streaming, as shown in Algorithm 1.

Note that at the very beginning of the first GOP window, the SNR information used in Phase 1 may not be available yet. However, such information can be obtained via estimation or learning techniques, or by simply letting CUs probe the channels when they are idle [76].

3.4 The Case of Single Channel Sensing

In this section, we consider the case that each CU can only sense a single channel and access a single channel during a time slot. We consider the case that each spectrum sensor has its own

operating point, which may be different from that of other spectrum sensors. This turns out to be an IP (Integer Programming) problem, which is NP-hard in general. We then take a divide-and-conquer approach to break down the problem into an optimal assignment sub-problem for spectrum sensing (termed OAPSS) and an optimal assignment sub-problem for video transmission (termed OAPVT). We develop effective solution algorithms for each sub-problem and prove their optimality to each sub-problem. However, the overall solution is near-optimal due to the divide-and-conquer approach.

3.4.1 Problem Formulation

Optimal Assignment Sub-problem for Spectrum Sensing (OAPSS)

In a practical wireless network scenario, CUs are located at different geographical positions with different channel gains to primary transmitters. Thus their performance on detecting primary signals on a particular licensed channel would be different, e.g., a CU with better channel gains to a primary transmitter may have a higher probability of detecting the PU singles (if the single indeed exist). By selecting a group of CUs which has better channel gain to detect the PU, then the PUs signal will have a higher probability of being detected, if the PUs signal really exist, and then the probability of causing interference to PU transmissions can be reduced [80, 81].

Usually cooperative sensing is used to improve the detection performance by fusing the sensing results from multiple CUs [82], and a certain fusion rule is required to combine these results. In this chapter, the OR fusion rule is used at the CBS to determine the presence or absence of PU signal on a particular channel. With the OR rule, if any of the CUs reports the presence of a PU signal then the CBS decides that the channel is busy; Otherwise, the CBS decides that the channel is idle. We use an $M \times N_1$ matrix $\mathbf{X}$ to denote the sensing task assignment at time slot t, while the entry located at the i-th row and j-th column position is defined as

$$x_{ij}^t = \begin{cases} 1, & \text{CU } i \text{ senses channel } j \text{ in time slot } t \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

A useful metric to evaluate the performance of detecting a PU signal is *probability of detection*, which is the probability that a CU successfully detects the existence of an existing PU signal. Let $P^t_{d_{ij}}$ denote the probability of detection on channel j by CU i at time slot t. For an energy detector, we have [81]

$$P^t_{d_{ij}} = \frac{1}{2}\text{erfc}\left(\left(\frac{\lambda^t_{ij}}{\sigma^2_n} - \varsigma^t_{ij} - 1\right)\sqrt{\frac{K}{2\left(2\varsigma^t_{ij}+1\right)}}\right), \tag{3.2}$$

where λ^t_{ij} is the threshold of energy detection on channel j by CU i at time slot t, σ^2_n is the power of the i.i.d. Additive White Gaussian Noise (AWGN) at the CU, ς^t_{ij} is the SNR of PU's signal on channel j at CU i, K is the number of samples on channel j by energy detection. In (3.2), $\text{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty \text{e}^{-u^2}du$ is the complementary error function, and let $\text{erfc}^{-1}(\cdot)$ denote the inverse function of $\text{erfc}(\cdot)$.

According to the OR fusion rule, the probability of detection on channel j at time slot t is

$$P^t_{d_j} = 1 - \prod_{i=1}^{M}\left(1 - P^t_{d_{ij}}\right)^{x^t_{ij}}. \tag{3.3}$$

In order to guarantee the protection of the PUs, we set $P^t_{d_{ij}} = \bar{P}_d$ by tuning λ^t_{ij} for all i, j. Thus the probability of detection of the activity of a PU will be greater than $\bar{P}_d$ if the channel is sensed by some CUs (according to (3.3)). In the case that a channel is not sensed by any of the CUs, it will not be used for video streaming.

Under the assumptions that the PU signal is complex valued phase-shift keying (PSK) and the noise is circularly symmetric complex Gaussian (CSCG), then CU i's *probability of false alarm* on channel j, denoted by $P^t_{f_{ij}}$, can be expressed as [81]

$$P^t_{f_{ij}} = \frac{1}{2}\text{erfc}\left(\left(\frac{\lambda^t_{ij}}{\sigma^2_n} - 1\right)\sqrt{\frac{K}{2}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{2\varsigma^t_{ij}+1}\,\text{erfc}^{-1}\left(2\bar{P}_d\right) + \sqrt{\frac{K}{2}}\varsigma^t_{ij}\right). \tag{3.4}$$

The objective of sensing task assignment is to maximize the probability of detecting all the idle channels at time slot t, while maintaining fairness among the probabilities of detection of the

N_1 licensed channels. It has been shown that proportional fairness can be achieved by maximizing the sum of logarithmic functions. The optimal sensing task assignment problem is to maximize the following objective function.

$$\sum_{j=1}^{N_1} \log\left(1 - P_{f_j}^t\right) = \sum_{j=1}^{N_1} \log \prod_{i=1}^{M} \left(1 - P_{f_{ij}}^t\right)^{x_{ij}^t}$$
$$= \sum_{j=1}^{N_1} \sum_{i=1}^{M} \log\left(1 - P_{f_{ij}}^t\right) x_{ij}^t = \sum_{j=1}^{N_1} \sum_{i=1}^{M} \varphi_{ij}^t \cdot x_{ij}^t, \tag{3.5}$$

where $\varphi_{ij}^t \doteq \log\left(1 - P_{f_{ij}}^t\right)$, and $P_{f_{ij}}^t, P_{d_{ij}}^t$ are defined as in Chapter 3. We assume that each CU can sense one channel at each time slot, and the number of CUs that can be assigned to sense a channel i at each time slot is unrestricted. Therefore, the optimal sensing task assignment problem is formulated as

$$\underline{\textbf{OAPSS}}: \max \quad \sum_{j=1}^{N_1} \sum_{i=1}^{M} \varphi_{ij}^t \cdot x_{ij}^t \tag{3.6}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_1} x_{ij}^t = 1, \text{ for all } i \tag{3.7}$$

$$x_{ij}^t \in \{0, 1\}, \text{ for all } i, j. \tag{3.8}$$

Optimal Assignment Problem for Video Transmission (OAPVT)

For video quality assessment, we adopt the QoE model named Mean Score Opinion (MOS) that was proposed in [71], where the MOS of CU i using channel j during time slot t, denoted by Ψ_{ij}^t, can be expressed as

$$\Psi_{ij}^t = \alpha + CT_i\gamma + (\beta + CT_i\delta) \ln\left(SBR_{ij}^t\right) \tag{3.9}$$
$$= \alpha + CT_i\gamma + (\beta + CT_i\delta) \ln\left(B\log_2\left(1 + SNR_{ij}^t\right)\right),$$

where $\alpha = 3.9860$, $\beta = 0.0919$, $\gamma = -5.8497$, and $\delta = 0.9844$ are constants, CT_i is the Content Type of the video sequences required by CU i, B is the bandwidth of a channel in kbps, and SNR_{ij}^t is the SNR of the video signal using channel j measured at CU i at time slot t [71].

Remarks on QoE model (3.9): The QoE model in [71] is applied in the Universal Mobile Telecommunication Systems. In multimedia applications such as video conferences over the CR network, QoE that directly measures the satisfaction of the CUs cannot be easily realized due to the limited spectrum resources. Therefore, it is important to study how to allocate frequency or spectrum resources to SUs according to their QoE requirements. The authors of [79] where video is streamed over CR network also adopts the same QoE model from [71] as we do. For the achievable (or the maximum) bit rate based on Shannon Theorem, the channel capacity by definition is the maximum bit rate that can be transmitted across the channel with a specific bit error ratio [71]. As pointed out in [71] and [79], the SBR should be adjusted according to the changing channel conditions, e.g., channel data rate, therefore, SBR is an adjustable parameter here in our problem. The effect of channel degradation on SBR adjustment is already taken into account in calculating the channel capacity, and loss rate and channel congestion are the only two factors that affects SBR adjustment, and a gradual increase in SBR is allowed when the bandwidth is available and there is no/reduced congestion, whereas, quick action is taken to reduce the SBR in case of severe congestion, as pointed out in [90], and the adaptive video scheme gracefully adapted the SBR to the available network downlink bandwidth, as pointed out in [71], [91], and [90]. Since our simulation is not conducted over the real network which is the same as [71], [91], and [90], the network congestion is not considered in our work, therefore the SBR is set to the value of available channel data rate so that the channel bandwidth can support the SBR without wasting bandwidth resource.

We assume that N_2 channels are determined to be idle after the sensing phase, where $N_2 \leq N_1$. We consider a general case where not all the CUs have data to receive at all times. Instead, the probability of a CU has data to receive at each GOP window is $0 \leq \xi \leq 1$. The number of CUs that have data to receive in a GOP window (called active CUs) is denoted as M_1, where $M_1 \leq M$.

An $M_1 \times N_2$ matrix $\mathbf{Z}$ is used to represent channel access assignments on time slot t, while the entry located at the i-th row and j-th column position is

$$z_{ij}^t = \begin{cases} 1, \text{ assign channel } j \text{ to CU } i \text{ in time slot } t \\ 0, \text{ otherwise.} \end{cases} \tag{3.10}$$

We consider the case where each CU can use at most one channel at each time slot due to hardware constraints, and each channel can be used by at most one CU at each time slot. We aim to maximize the expected average MOS of all the CUs during a GOP window by assigning the available channels.

$$\max: \ \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M_1} \mathrm{E}\left[\Psi_i^t\right] \tag{3.11}$$

where Ψ_i^t is the MOS of CU i at time slot t. The above objective function can be maximized if we maximize the expected MOS increment of the M_1 CUs during each time slot [63], which can be written as

$$\begin{aligned} &\sum_{i=1}^{M_1} \mathrm{E}\left[\Psi_i^t\right] = \sum_{i=1}^{M_1}\sum_{j=1}^{N_2} \mathrm{E}\left[\Psi_{ij}^t\right] \cdot z_{ij}^t \\ =& \sum_{i=1}^{M_1}\sum_{j=1}^{N_2} [P\left(r_j^t = 0 | s_j^t = 0\right)\phi_{ij}^t + P\left(r_j^t = 1 | s_j^t = 0\right)\theta_{ij}^t] z_{ij}^t, \end{aligned} \tag{3.12}$$

where $s_j^t = 0$ indicates the channel is sensed idle; $P(r_j^t = 0)$ and $P(r_j^t = 1)$ are the probability of channel j to be idle or busy at time slot t, respectively; $P(r_j^t = 0 | s_j^t = 0)$ and $P(r_j^t = 1 | s_j^t = 0)$ (denoted as $P_{00}^{j,t}$ and $P_{10}^{j,t}$, respectively) are the conditional probability for channel j to be idle or busy conditioned on the sensing result, respectively. It follows that

$$P_{00}^{j,t} = P(r_j^t = 0 | s_j^t = 0) = \frac{(1 - P_{f_j}^t)P(r_j^t = 0)}{(1 - P_{f_j}^t)P(r_j^t = 0) + (1 - P_{d_j}^t)P(r_j^t = 1)} \tag{3.13}$$

$$P_{10}^{j,t} = P(r_j^t = 1 | s_j^t = 0) = 1 - P(r_j^t = 0 | s_j^t = 0). \tag{3.14}$$

In (3.12), ϕ_{ij}^t and θ_{ij}^t are the effective data rate of the received video sequence at CU i using channel j which is indeed idle or busy at time slot t, respectively. Denote μ_{ij}^t and ν_{ij}^t as the received SNR at CU i using channel j which is indeed idle or busy at time slot t, respectively. We then have

$$\begin{aligned}
\mu_{ij}^t &= \frac{\Gamma g_i}{n_0 B} \\
\nu_{ij}^t &= \frac{\Gamma g_i}{n_0 B(1+\varsigma_{ij}^t)} \\
\phi_{ij}^t &= \alpha + CT_i\gamma + (\beta + CT_i\delta)\ln\left(B_j \log_2\left(1+\mu_{ij}^t\right)\right) \\
\theta_{ij}^t &= \alpha + CT_i\gamma + (\beta + CT_i\delta)\ln\left(B_j \log_2\left(1+\nu_{ij}^t\right)\right),
\end{aligned}$$

where Γ is the transmit power of the CBS on channel j, for all j.

Define ϖ_{ij}^t as

$$\varpi_{ij}^t = P_{00}^{j,t} \cdot \phi_{ij}^t + P_{10}^{j,t} \cdot \theta_{ij}^t. \tag{3.15}$$

The optimal channel access problem is formulated as

$$\underline{\textbf{OAPVT}}: \max \quad \sum_{i=1}^{M_1}\sum_{j=1}^{N_2} \varpi_{ij}^t \cdot z_{ij}^t \tag{3.16}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_2} z_{ij}^t \le 1,\ i \in \{1,\cdots,M_1\}. \tag{3.17}$$

$$\sum_{i=1}^{M_1} z_{ij}^t \le 1,\ j \in \{1,\cdots,N_2\} \tag{3.18}$$

$$z_{ij}^t \in \{0,1\},\ \text{for all } i,j. \tag{3.19}$$

OAPVT considering fairness among CUs

Now we consider achieving fairness among CUs for the channel allocation problem. Considering the fact that our objective is to maximize the expected average MOS of all the CUs during a GOP window by assigning the available channels, we propose to achieve a long term fairness among CUs.

In order to achieve long term fairness among CUs, we propose that, fairness among CUs should be achieved by allocating channels to different CUs in different time slots. For example, when there is only one available channel and two CUs, A and B, if A is scheduled for video streaming in the previous time slot, then at current time slot, B should be scheduled for video streaming. Generally speaking, at current time slot, when the number of idle channels available for video streaming, e.g., N_2, is less than the number of CUs requesting video streaming, e.g., M_1,, the CUs that have not been scheduled in previous time slots will have a higher priority of being scheduling than the CUs that have been scheduled in previous time slots, and the objective is still to maximizing the MOS sum of all CUs. Here we assume that all CUs request video streaming at each time slot, i.e., $M_1 = M$.

To be more specific, we consider the following cases:

1. At current time slot, $N_2 \geq M_1$. In this case, all CUs can be scheduled for video streaming at current time slot, and the problem formulation will be the same as problem **OAPVT**.

2. $N_2 < M_1$. For the easy of presentation, denote Θ as the whole set of CUs requesting video streaming at current time slot. Denote Θ_k as a subset of Θ, where Θ_k is the set of CUs whose times of being scheduled in previous time slots is $k, k = 0, 1, 2, 3, \ldots$, and denote Θ_{k+1} as a subset of Θ, where Θ_{k+1} is the set of CUs whose times of being scheduled in previous time slots is $k+1$. Denote $\|\cdot\|$ as the number of elements in a set. Then $\|\Theta\| = M_1$, $\Theta_k \cup \Theta_{k+1} = \Theta$, $\|\Theta_k\| + \|\Theta_{k+1}\| = M_1$ (which we will prove in the following lemma).

   - If $N_2 \leq \|\Theta_k\| \leq M_1$, then the optimization problem is to maximize the MOS sum for the CUs in Θ_k , by choosing N_2 CUs from Θ_k and allocate the N_2 available channels to the N_2 CUs;

$$\textbf{P1}: \max \quad \sum_{i \in \Theta_k} \sum_{j=1}^{N_2} \varpi_{ij}^t \cdot z_{ij}^t \tag{3.20}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_2} z_{ij}^t \leq 1, \ i \in \Theta_k. \tag{3.21}$$

$$\sum_{i\in\Theta_k} z_{ij}^t \leq 1,\ j \in \{1,\cdots,N_2\} \tag{3.22}$$

$$z_{ij}^t \in \{0,1\},\ \text{for all } i \in \Theta_k, 1 \leq j \leq N_2. \tag{3.23}$$

- If $\|\Theta\| > N_2 > \|\Theta_k\|$, then all the CUs in Θ_k would be scheduled, and $N_2 - \|\Theta_k\|$ CUs would be chosen from set Θ_{k+1}, and then N_2 channels is allocated to the $\|\Theta_k\| + (N_2 - \|\Theta_k\|)$ CUs to maximize the MOS sum.

$$\underline{\textbf{P2}}: \max \quad \sum_{i\in\Theta}\sum_{j=1}^{N_2} \varpi_{ij}^t \cdot z_{ij}^t \tag{3.24}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_2} z_{ij}^t = 1,\ i \in \Theta_k. \tag{3.25}$$

$$\sum_{j=1}^{N_2} z_{ij}^t \leq 1,\ i \in \Theta_{k+1}. \tag{3.26}$$

$$\sum_{i\in\Theta} z_{ij}^t \leq 1,\ j \in \{1,\cdots,N_2\} \tag{3.27}$$

$$z_{ij}^t \in \{0,1\},\ \text{for all } i \in \Theta, 1 \leq j \leq N_2. \tag{3.28}$$

Now we are going to prove that $\Theta_k \cup \Theta_{k+1} = \Theta$. Once we have $\Theta_k \cup \Theta_{k+1} = \Theta$, then it is easy to get to $\|\Theta_k\| + \|\Theta_{k+1}\| = M_1$.

Lemma 3.1. *According to the definition of Θ, Θ_k, and Θ_{k+1}, we have $\Theta_k \cup \Theta_{k+1} = \Theta$, where $k = 0, 1, 2, \ldots$ at the beginning of all time slots.*

Proof. At the very beginning of the $1st$ time slot, each CU has not been scheduled, therefore $k = 0$, $\Theta_k = \Theta$, and $\Theta_{k+1} = \oslash$. So the lemma stands for the $1st$ time slot. Assume that at the beginning of time slot t, $t \geq 1$, $\Theta_k \cup \Theta_{k+1} = \Theta$.

1. Assume that problem P1 is solved at time slot t. Then N_2 CUs in set Θ_k at the beginning of t will be moved to set Θ_{k+1} at the end of t. Therefore $\Theta_k \cup \Theta_{k+1} = \Theta$ at the end of t.

2. Assume that problem P2 is solved at time slot t. Then CBS will choose $N_2 - \|\Theta_k\|$ CUs from set Θ_{k+1} to schedule, and the number of times of being scheduled at the end of t is $k+2$, then the set formed by these CUs is denoted as Θ_{k+2}. And all the CUs in set Θ_k at the beginning of t will be moved to set Θ_{k+1} at the end of t, and forms a new set Θ'_{k+1}. Then we have the following:

$$\begin{aligned}
&\Theta_{k+2} \cup \Theta'_{k+1} = \Theta_{k+2} \cup ((\Theta_{k+1} - \Theta_{k+2}) \cup \Theta_k) \\
&= \Theta_{k+2} \cup ((\Theta_{k+1} \cup \Theta_k) - (\Theta_{k+2} - \Theta_k)) \\
&= \Theta_{k+2} \cup (\Theta - (\Theta_{k+2} \cap \Theta_{k+1})) \\
&= \Theta_{k+2} \cup (\Theta - \Theta_{k+2}) \\
&= \Theta_{k+2} \cup \Theta - (\Theta_{k+2} - \Theta_{k+2}) \\
&= \Theta - \oslash = \Theta. \qquad (3.29)
\end{aligned}$$

   And from the definition we know that $\Theta_{k+2} \cap \Theta'_{k+1} = \oslash$. Therefore the lemma stands at the end of t.

3. Assume that OAPVT is solved at time slot t. Then all CUs in set Θ_{k+1} and all CUs in set Θ_k will be scheduled, which means that Θ_{k+1} becomes Θ_{k+2} and Θ_k becomes Θ_{k+1}. Therefore $\Theta_{k+1} \cup \Theta_{k+2} = \Theta_k \cup \Theta_{k+2} = \Theta$ at the end of t. Therefore the lemma stands at the end of t.

From the above discussions we know that the lemma stands at the end of any time slot t, i.e., at the beginning of time slot $t+1$, $t = 0, 1, 2, \ldots$, which completes our proof. □

3.4.2 Solution Algorithms and Analysis

Poly-matching Based Solution to OAPSS

We can see that the OAPSS problem is formulated as the well-known General Assignment Problem (GAP), which is NP-hard in general. However, since there is no constraint on how many CUs can be assigned to a channel, the problem is actually a Maximum Weight Poly-Matching

Algorithm 2: Greedy Poly-Matching Algorithm

```
1 for i = 1 → M do
2 |   for j = 1 → N_1 do
3 |   |   x_ij^t = 0 ;
4 |   end
5 |   j* = arg max_{j∈{1,···,N_1}} {φ_ij^t} ;
6 |   x_ij*^t = 1 ;
7 end
```

(MWPM) problem on a bipartite graph that matches CUs to licensed channels with edge weights defined as φ_{ij}^t. Furthermore, a channel can be matched to multiple CUs. It can be solved by the following greedy strategy presented in Algorithm 2 [83].

With this algorithm, each CU selects the channel with the largest weight, regardless whether the selected channel has been chosen by other CUs or not [83]. This greedy strategy has a time complexity of $\mathcal{O}(MN_1)$. In fact this is a distributed algorithm, since each CU can choose its best channels to sense and there is no need to involve the CBS in this phase. Since the CUs can launch their searching procedures in Line 5 in parallel, this distributed strategy has a time complexity of $\mathcal{O}(N_1)$.

In the following theorem, we also show that the Greedy Poly-matching Algorithm is optimal.

Theorem 3.1. *The Greedy Poly-matching Algorithm 2 achieves the optimal solution to the OAPSS problem.*

Proof. Exchanging the summation order, the objective function of the OAPSS sub-problem (3.6) becomes $\sum_{i=1}^{M}\left(\sum_{j=1}^{N_1}\varphi_{ij}^t\cdot x_{ij}^t\right)$, where $\sum_{j=1}^{N_1}\varphi_{ij}^t\cdot x_{ij}^t$ is the utility that CU i can achieve under the two constraints (3.7) and (3.8). Since each CU can have at most one channel, the maximum utility CU i can achieve is $\max_j\left\{\varphi_{ij}^t\cdot x_{ij}^t\right\}$, which is accomplished in Line 5 of Algorithm 2. Since the optimal strategies of the CUs do not conflict with each other and thus are independent to each other, the maximum utility of the CUs are also independent of each other. It follows that $\max\sum_{i=1}^{M}\left(\sum_{j=1}^{N_1}\varphi_{ij}^t\cdot x_{ij}^t\right)=\sum_{i=1}^{M}\left(\max_j\left\{\varphi_{ij}^t\cdot x_{ij}^t\right\}\right)$, and Algorithm 2 is optimal. □

Solution to Channel Accessing Problem

The three problems of OAPVT, P1, and P2 are all Integer Programming problems which is NP-hard in general. However, an interesting characteristic of the three problems is that the coefficients of the constraint matrix in these problems are either 0 or 1, such that the unimodularity property [24] is satisfied in these problems. As a result, both these problems have the optimal solution with their LP relaxations, and thus they can be solved with the Simplex method [15], [25], which has a a polynomial-time average-case complexity.

3.5 The Case of Multi-Channel Sensing

In this section, we consider the general case that a CU can sense multiple channels but can still access one channel at a time (e.g., each CU is equipped with multiple spectrum sensors but with only one transceiver). To make the problem tractable, we assume that all the spectrum sensors are tuned to have the same probability of detection and the same probability of false alarm. Under this assumption, we present a problem formulation that integrates both spectrum sensing and access for QoE driven video streaming. We then develop a two-step algorithm with proven optimality. Note that if this assumption is made for the problem examined in Section 3.4, then the single channel sensing problem becomes a special case of the multi-channel sensing problem, which can be solved with the optimal solution algorithms developed in this section. For brevity, we omit the superscript t on all the relevant symbols in the rest of this section.

3.5.1 Problem Formulation

We assume that there are M CUs and N licensed channels. CU i can sense at most C channels and access at most one channel at a time slot. Furthermore, each channel must have Λ CUs to sense it to guarantee that the cooperative probability of detection on a channel satisfies $P_d \geq 1 - \left(1 - \bar{P}_d\right)^{\Lambda}$. We also have $MC < N\Lambda$, which means that only parts of the N channels can be sensed at a time slot. As discussed, in addition to $P_{d_{ij}} = \bar{P}_d$, we also have $P_{f_{ij}} = \bar{P}_f$, for

all i, j. This can be achieved by solving the system of these two equations for the threshold of energy detection λ_{ij} and the number of samples K_{ij} for each spectrum sensor i on channel j with a different SNR value ς_{ij}. We have

$$K_{ij} = 2\left(\frac{\text{erfc}^{-1}(2\bar{P}_f) - \sqrt{2\varsigma_{ij}+1}\text{erfc}^{-1}(2\bar{P}_d)}{\varsigma_{ij}}\right)^2$$
$$\lambda_{ij} = \sigma_n^2\left(1 + \frac{\varsigma_{ij}\text{erfc}^{-1}(2\bar{P}_f)}{\text{erfc}^{-1}(2\bar{P}_f) - \sqrt{2\varsigma_{ij}+1}\text{erfc}^{-1}(2\bar{P}_d)}\right).$$

Let $I_{(\sum_i x_{ij}=\Lambda)}$ be an indicator function defined as

$$I_{(\sum_i x_{ij}=\Lambda)} = \begin{cases} 1, & \text{if } \sum_i x_{ij} = \Lambda \\ 0, & \text{otherwise.} \end{cases} \tag{3.30}$$

We then have

$$P(s_j = 0) = \left\{P(r_j = 0)(1 - P_{f_j}) + P(r_j = 1)(1 - P_{d_j})\right\} \cdot I_{(\sum_i x_{ij}=\Lambda)} \tag{3.31}$$
$$P(s_j = 1) = 1 - P(s_j = 0) \tag{3.32}$$

Let $\vec{\mathbf{S}} = \{s_j, j = 1, 2, \cdots, N\}$ represents the cooperative sensing results on the N licensed channels. There are 2^N possible outcomes for $\vec{\mathbf{S}}$ in total. Let $\vec{\mathbf{S}}_h$ be the h-th outcome, $0 \leq h \leq 2^N - 1$. Define $\Gamma_j(h)$ to be the j-th element in $\vec{\mathbf{S}}_h$, $j = 1, 2, \cdots, N$. Assuming independent channel states, the probability of getting outcome $\vec{\mathbf{S}}_h$ as a sensing result is

$$P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) = \prod_{j=1}^{N} P(s_j = \Gamma_j(h))$$
$$= \prod_{j=1}^{N} \left[(1 - \Gamma_j(h))P(s_j = 0) + \Gamma_j(h)P(s_j = 1)\right]. \tag{3.33}$$

For a sensing outcome $\vec{\mathbf{S}}_h$, let $\Phi_h = \{j : \Gamma_j(h) = 0, j = 1, 2, \cdots, N\}$ be the set of channels that are sensed idle. Let $\mathbf{Y}_h = \left[y_{ij}^h\right], 1 \leq i \leq M$, $j \in \Phi_h$, be the channel assignment matrix,

where $0 \leq y_{ij}^h \leq 1$ is the amount of time that CBS transmits to CU i on channel j in a time slot, when the sensing outcome is $\vec{\mathbf{S}}_h$. The channel assignment strategy can be expressed as $\mathbf{Y} = [\mathbf{Y}_0, \mathbf{Y}_1, \cdots \mathbf{Y}_{2^N-1}]$.

According to conditional expectation, the expected overall MOS can be derived as

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^{M} \Psi_i\right) &= \sum_{i=1}^{M} \sum_{h=0}^{2^N-1} \mathrm{E}(\Psi_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \\ &= \sum_{h=0}^{2^N-1} \sum_{i=1}^{M} \mathrm{E}(\Psi_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h). \end{aligned} \tag{3.34}$$

With the MOS model used in Section 3.4, we have

$$\mathrm{E}\left[\Psi_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h\right] = \mathrm{E}\left[\sum_{j=1}^{N} \Psi_{ij} | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h\right] = \sum_{j=1}^{N} \left(P_{00}^j \cdot \phi_{ij} + P_{10}^j \cdot \theta_{ij}\right) \cdot y_{ij}^h, \tag{3.35}$$

where

$$\begin{aligned} P_{00}^j &= P(r_j = 0 | s_j = 0) \\ &= \begin{cases} \frac{(1-P_f)P(r_j=0)}{(1-P_f)P(r_j=0)+(1-P_d)P(r_j=1)}, & \text{if } \sum_i x_{ij} = \Lambda \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \tag{3.36}$$

and

$$P_{10}^j = P(r_j = 1 | s_j = 0) = \begin{cases} 1 - P_{00}^j, & \text{if } \sum_i x_{ij} = \Lambda \\ 0, & \text{otherwise}. \end{cases} \tag{3.37}$$

Define

$$w_{ij} = \begin{cases} P_{00}^j \cdot \phi_{ij} + P_{10}^j \cdot \theta_{ij}, & \text{if } \sum_i x_{ij} = \Lambda \text{ and channel} \\ & j \text{ is sensed idle} \\ 0, & \text{otherwise}. \end{cases} \tag{3.38}$$

Then the master problem of maximizing the total expected QoE of all the video sessions, denoted as MP, can be formulated as follows.

$$\underline{\textbf{MP}}: \max: \sum_{h=0}^{2^N-1} \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} \cdot y_{ij}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \tag{3.39}$$

$$\text{s.t.} \sum_{j=1}^{N} y_{ij}^h \leq 1, \text{for all } i, h \tag{3.40}$$

$$\sum_{i=1}^{M} y_{ij}^h \leq 1, \text{for all } j \tag{3.41}$$

$$\sum_{i=1}^{M} x_{ij} \leq \Lambda, \text{for all } j \tag{3.42}$$

$$\sum_{j=1}^{N} x_{ij} \leq C, \text{for all } i \tag{3.43}$$

Equation (3.38)

$$x_{ij} = \{0, 1\}, \text{for all } i, j \tag{3.44}$$

$$y_{ij}^h = \{0, 1\}, \text{for all } i, j, h. \tag{3.45}$$

It can be observed that the formulated problem MP is an Integer NonLinear Programming (INLP) problem, which is NP-hard in general, although a rigorous proof is not given in this chapter. We next show that problem MP can be decomposed into two sub-problems and solved with a two-step approach without sacrificing optimality.

3.5.2 Solution Algorithms

First, we use Algorithm 7 to solve the spectrum sensing sub-problem, denoted as SP1, i.e., to determine the sensing task assignment matrix $\mathbf{X}$. In Algorithm 7, we sort the N channels according to $P(r_j = 0)$, $j = 1, 2, \ldots, N$, in descending order. We then assign CUs to sense the sorted channels sequentially as follows. For the first channel in the remaining channel list, if there are no less than Λ CUs each of which can still sense some extra channels, choose Λ CUs to sense the channel; Otherwise, the channel is conservatively claimed to be busy in order to avoid potential

collision with PUs using this channel. Initially each CU can sense C channels, i.e., with sensing capability C. Each time a CU is assigned to sense a channel, its sensing capability will be reduced by 1.

Specifically, line 1 sorts the N channels, line 2 to line 4 initialize the sensing capacity of each CU, line 5 to line 26 assigns CUs to sense the N channels, where line 6 to line 17 is to choose λ CUs to sense a channel, and the sensing capacity of a CU is reduced by 1 at each time the CU is chosen to sense a channel. Line 18 to line 25 checks if there remains sufficient number of CUs to sense the next channel. If yes, then assign CUs to sense the next channel; otherwise, stop sensing the remaining channels which are not sensed by CUs. Line 27 to line 31 determine the channels that is not sensed by sufficient number of CUs and therefore is determined busy.

After obtaining the sensing task assignment matrix $\mathbf{X}$ from Algorithm (7), spectrum sensing is conducted by CUs following the assignments and sensing results are reported to the CBS. The CBS then solves the following sub-problem, denoted as SP2, to obtain the channel allocation matrix $\mathbf{Y}$, which will be broadcast to the CUs for channel access.

$$\underline{\textbf{SP2}}: \max: \sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij} \cdot y_{ij} \tag{3.46}$$

$$\text{s.t.} \sum_{j=1}^{N} y_{ij} \leq 1, \forall\, i \tag{3.47}$$

$$\sum_{i=1}^{M} y_{ij} \leq 1, \forall\, j \tag{3.48}$$

$$y_{ij} = \{0,1\}, \text{for all } i,j. \tag{3.49}$$

Clearly SP2 is also a Maximum Weight Matching problem and is the same as OAPVT. It can be solved with optimal solution using the Hungarian Method.

Algorithm 3: Greedy Spectrum Sensing Algorithm

Sort the N channels in descending order of $P(r_j = 0)$ and let the sorted channel set be Ξ ;
for $i = 1 : M$ **do**
    $C_i = C$;
end
for $j = 1 : N$ **do**
    Let $j' = \Xi(j)$;
    $\eta'_j = 0$;
    for $i = 1 : M$ **do**
        if $\eta'_j \geq \Lambda$ **then**
            Break;
        end
        if $C_i > 0$ **then**
            $x_{ij'} = 1$;
            $C_i = C_i - 1$;
            $\eta'_j = \eta'_j + 1$;
        end
    end
    for $i = 1 : M$ **do**
        if $C_i > 0$ **then**
            $\mu_i = 1$;
        end
    end
    if $\sum_i \mu_i < \Lambda$ **then**
        Break ;
    end
end
for $j = 1 : N$ **do**
    if $\eta_j < \Lambda$ **then**
        Channel j is determined to be busy ;
    end
end

3.5.3 Optimality Proof

Although problem MP is solved with the two-step approach in Section 3.5.2, we show that the solution is actually optimal in Theorem 3.2.

Theorem 3.2. *Let* $[\mathbf{X}^*, \mathbf{Y}^*]$ *denote the optimal solution to problem MP, where* $\mathbf{X}^*$ *is the optimal spectrum sensing strategy and* $\mathbf{Y}^*$ *is the optimal channel allocation strategy. Then* $\mathbf{X}^*$ *can be obtained by running Algorithm 7 and* $\mathbf{Y}^*$ *can be obtained by solving problem SP2.*

Proof. Let j' and j^* be the indexes of two licensed channels such that $P(r'_j = 0) \geq P(r_{j^*} = 0)$. Define $\mathbf{W}' = [\mathbf{w'_{11}}, \cdots, \mathbf{w'_{M1}}, \cdots, \mathbf{w'_{1j'}}, \cdots, \mathbf{w'_{Mj'}}, \cdots, w'_{1N}, \cdots, w'_{MN}]$ and $\mathbf{W}^* = [\mathbf{w^*_{11}}, \cdots, \mathbf{w^*_{M1}}, \cdots, \mathbf{w^*_{1j^*}}, \cdots, \mathbf{w^*_{Mj^*}}, \cdots, w^*_{1N}, \cdots, w^*_{MN}]$. Also denote $\mathbf{X}' = [\mathbf{x'_{11}}, \cdots, \mathbf{x'_{M1}}, \cdots, x'_{1j'}, \cdots, x'_{Mj'}, \cdots, x'_{1N}, \cdots, x'_{MN}]$ and $\mathbf{X}^* = [\mathbf{x^*_{11}}, \cdots, \mathbf{x^*_{M1}}, \cdots, x_{1j^*}, \cdots, x_{Mj^*}, \cdots, x_{1N}, \cdots, x_{MN}]$ as the feasible sensing task assignment matrices corresponding to $\mathbf{W}'$ and $\mathbf{W}^*$ respectively.

Let $x'_{ij^*} = 0$ in $\mathbf{X}'$, for all i, and $x^*_{ij'} = 0$ in $\mathbf{X}^*$, for all i. Then we have $w'_{ij^*} = 0$ in $\mathbf{W}'$, for all i, and $w^*_{ij'} = 0$ in $\mathbf{W}^*$, for all i. Let $x'_{ij} = x^*_{ij}$, for all $j \neq j', j \neq j$, for all i. It follows that $w'_{ij} = w^*_{ij}$ (denoted as w_{ij}), for all $j \neq j', j \neq j^*$, for all i. Let $\sum_i x'_{ij'} = \sum_i x^*_{ij^*} = \Lambda$. Then according to (3.38), we have $w'_{ij'} \geq w^*_{ij^*}$, for all i.

We first proof the following lemma, which will serve as a basis for the later part of the proof for Theorem 3.2.

Lemma 3.2. *Denote SP2' and SP2* as the channel allocation problem corresponding to $\mathbf{W}'$ and $\mathbf{W}^*$ as defined above, respectively. If there is a feasible solution $\mathbf{Y}^*$ for SP2*, then there is always a feasible solution, denoted as $\mathbf{Y}'$, for $\mathbf{W}'$, such that $\mathbf{W}'\mathbf{Y}'^{\mathbf{T}} \geq \mathbf{W}^*\mathbf{Y}^{*\mathbf{T}}$, where $(\cdot)^T$ denotes the matrix transpose operation.*

Proof. Let $\mathbf{Y}^*= [y^*_{11}, \cdots, y^*_{M1}, \cdots, y^*_{1j^*}, \cdots, y^*_{Mj^*}, \cdots, y^*_{1N}, \cdots, y^*_{MN}]$ be the feasible channel assignment matrix corresponding to $\mathbf{W}^*$. Let $y^*_{ij} = y_{ij}$, for all $j \neq j'$ or j^*, $y_{ij} = 0$ or 1, for all i, $y^*_{ij'} = u_i, u_i = 0$ or 1, for all i, $y^*_{\widehat{i}j^*} = v_{\widehat{i}}, v_{\widehat{i}} = 0$ or 1, for an $\widehat{i} \in I = \{1, \cdots, M\}$, and $y^*_{ij^*} = 0$, for all $i \neq \widehat{i}, i \in I = \{1, \cdots, M\}$.

Then $\mathbf{Y}'= [y'_{11}, \cdots, y'_{M1}, \cdots, y'_{1j'}, \cdots, y'_{Mj'}, \cdots, y'_{1N}, \cdots, y'_{MN}]$ with $y'_{ij} = y_{ij}$, for all $j \neq j'$ or j^*, for all i, $y'_{ij^*} = u_i$, for all i (recall that $w'_{ij^*} = 0$, for all i in $\mathbf{W}'$), $y'_{\widehat{i}j'} = v_{\widehat{i}}$, and $y'_{ij'} = 0$, for all $i \neq \widehat{i}, i \in I = \{1, \cdots, M\}$, will be a feasible solution to SP2'. This is because the constraints in SP2' are still satisfied as follows.

- The number of users on any channel $j \neq j'$ or j^* in solution $\mathbf{Y}'$ is the same as that in solution $\mathbf{Y}^*$, i.e., $\sum_i y'_{ij} = \sum_i y_{ij} = \sum_i y^*_{ij}$.

- The number of users on channel j^* (or j') in solution $\mathbf{Y}'$ is the same as that on channel j' (or j^*) in solution $\mathbf{Y}^*$, i.e., $\sum_i y'_{ij^*} = \sum_i u_i = \sum_i y^*_{ij'}$ (or $\sum_i y'_{ij'} = v_{\widehat{i}} = \sum_i y^*_{ij^*}$). Note that the constraint on the number of users on channel j^* is the same as that on channel j'.

- The number of antennas that CU $\widehat{i}$ uses in solution $\mathbf{Y}'$ is the same as that in solution $\mathbf{Y}^*$, i.e., $\sum_{j\neq j^*,j'}(y'_{\widehat{i}j} + y'_{\widehat{i}j^*} + y'_{\widehat{i}j'}) = \sum_{j\neq j^*,j'}(y_{\widehat{i}j} + u_{\widehat{i}} + v_{\widehat{i}}) = \sum_{j\neq j^*,j'}(y'_{\widehat{i}j} + y^*_{\widehat{i}j'} + y^*_{\widehat{i}j^*})$.

- The number of antennas that CU i, for all $i \neq \widehat{i}$, uses in solution $\mathbf{Y}'$ is the same as that in solution $\mathbf{Y}^*$, i.e., $\sum_{j\neq j^*,j'}(y'_{ij} + y'_{ij^*} + y'_{ij'}) = \sum_{j\neq j^*,j'}(y_{\widehat{i}j} + u_{\widehat{i}} + 0) = \sum_{j\neq j^*,j'}(y'_{\widehat{i}j} + y^*_{ij'} + y^*_{ij^*})$.

From the first two bullets above, it can be seen that in SP2', the constraint for each channel j is satisfied. From the third and fourth bullets above, we know that the constraint for each CU is also satisfied. Therefore, we conclude that **Y'** is also a feasible solution to problem SP2'.

It then follows that

$$
\begin{aligned}
&\mathbf{W'Y'^{T}} - \mathbf{W^*Y^{*T}} \\
&= \sum_i \sum_j w'_{ij} y'_{ij} - \sum_i \sum_j w^*_{ij} y^*_{ij} \\
&= \sum_i \sum_{j\neq j^*,j'} w'_{ij} y'_{ij} + \sum_i w'_{ij^*} y'_{ij^*} + \sum_i w'_{ij'} y'_{ij'} - \\
&\quad \left(\sum_i \sum_{j\neq j^*,j'} w^*_{ij} w^*_{ij} + \sum_i w^*_{ij^*} y^*_{ij^*} + \sum_i w^*_{ij'} y^*_{ij'} \right) \\
&= \sum_i w'_{ij'} y'_{ij'} - \sum_i w^*_{ij^*} y^*_{ij^*} (\text{recall } w'_{ij^*} = 0 \text{ and } w^*_{ij'} = 0, \forall i) \\
&= \sum_{i\neq\widehat{i}} w'_{ij'} y'_{ij'} + w'_{\widehat{i}j'} y'_{\widehat{i}j'} - \left(\sum_{i\neq\widehat{i}} w^*_{ij^*} y^*_{ij^*} + w^*_{\widehat{i}j^*} y^*_{\widehat{i}j^*} \right) \\
&= \sum_{i\neq\widehat{i}} w'_{ij'} \cdot 0 + w'_{\widehat{i}j'} y'_{\widehat{i}j'} - \left(\sum_{i\neq\widehat{i}} w^*_{ij^*} \cdot 0 + w^*_{\widehat{i}j^*} y^*_{\widehat{i}j^*} \right) \\
&= w'_{\widehat{i}j'} y'_{\widehat{i}j'} - w^*_{\widehat{i}j^*} y^*_{\widehat{i}j^*} \\
&= w'_{\widehat{i}j'} v_{\widehat{i}} - w^*_{\widehat{i}j^*} v_{\widehat{i}}
\end{aligned}
$$

$$\geq \left(w^*_{\hat{i}j^*} - w^*_{\hat{i}j^*}\right) v_{\hat{i}} \qquad \text{(recall } w'_{ij'} \geq w^*_{ij^*}, \forall i\text{)}$$
$$= 0.$$

Then the lemma immediately follows. □

Denote MP' and MP^* as the original joint-optimization problem with $\{\mathbf{W}', \mathbf{Y}'\}$, and $\{\mathbf{W}^*, \mathbf{Y}^*\}$ as defined above, respectively, and Δ' and Δ^* as the corresponding objective function value of MP' and MP^*, respectively. It follows Lemma (3.2) that

$$\Delta' - \Delta^* = \sum_{h=0}^{2^N-1} P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \left(\mathbf{W}'\mathbf{Y}'^{\mathbf{T}} - \mathbf{W}^*\mathbf{Y}^{*\mathbf{T}}\right) \geq 0. \tag{3.50}$$

Eq. (3.50) indicates that when we have limited spectrum sensing capability and cannot guarantee a satisfactory probability of detection to all the channels, in order to maximize the expected utility we can obtain from the possible sensing results and the corresponding optimal transmission strategy, we should assign the highest priority to the channel that has the highest probability of being idle, and allocate CUs that still have sensing capability to sense this channel. It would be suboptimal if we allocate CUs with extra sensing capability (if they do exist) to sense other channel(s) that has(have) a lower probability of being idle. This is exactly the same strategy used in Algorithm 7, i.e., assigning CUs to sense the channels in a decreasing order of their probabilities of being idle.

This concludes the proof of the theorem. □

3.6 Simulation Study

3.6.1 Simulation Setup

In this section, the performance of the proposed algorithms is validated with Matlab simulations. We consider a scenario in which the PUs and CUs are randomly distributed around a CBS within the service radius of the CBS. Table 6.1 presents the values of simulation parameters used in

Table 3.1: Simulation Parameters

Parameters	*Value*	*Parameters*	*Value*
M	30	μ_{ij}^t	-21 dB to -11 dB
N_1	30	ν_{ij}^t	-80 dB to -60 dB
K	10^4	ς_{ij}^t	-30 dB to -10 dB
f_s	10^6 Hz	$\max_j \left\{\Pr\left(H_{0j}^t\right)\right\}$	0.9
T	10	$\bar{P}_d$	0.95
B	10^6 Hz	$\bar{P}_f$	0.1
C	3	Λ	4

the simulations, where f_s is the sampling frequency at the CUs with energy detection. It is verified that the range of MOS is within 1 to 5 with the value of the parameters provided as in [71].

We first examine the performance of the proposed algorithms for the single-channel sensing case which solves the OAPSS and OAPVT problems separately. Here we term this algorithm "Proposed Scheme 1 (PS1)" in the simulations. And we compare PS1 with three benchmark schemes presented in [89] (termed *Benchmark 1* in simulations), [79] (termed *Benchmark 2*), and [92] (termed *Benchmark 3*), respectively.

Specifically, in [89], the authors assume that the QoE model is known but the parameters are unknown. The algorithm estimates the QoE models through the observation of the realized QoE sum, then it dynamically changes the channel allocation based on the estimated QoE parameters, in order to maximize the QoE sum of all users. However, since our QoE model is adopted from [71], where the authors do not consider the packet error rate (PER) while the authors of [89]consider the PER in the QoE model, we set the PER in the QoE model adopted in [89], as the authors of [79] do, in order for a fair comparison. In [79], the whole group of CUs are categorized into three different classes of priority. Each CU has a priority, and the priority of a CU is determined by the video sequence that it acquires from the CBS. The CUs acquiring the "Suzie" sequence have the highest priority, the CUs acquiring the "Carphone" sequence have the second highest priority, and the CUs acquiring the "Football" sequence have the lowest priority, where "Suzie", "Carphone", and "Football" are three video sequences of different content types we use in our simulations.

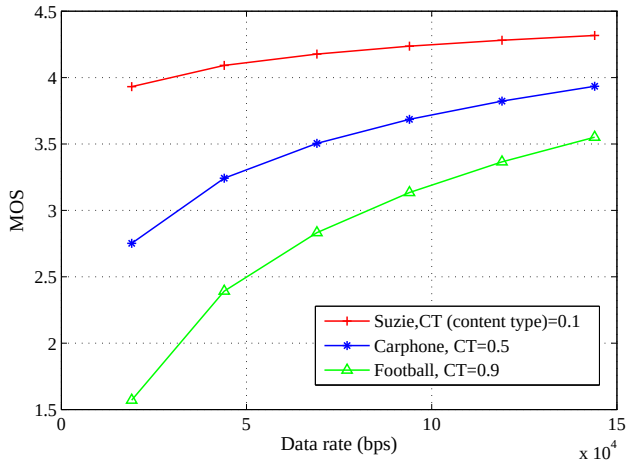

Figure 3.1: MOS and data rate relationship for three reference video sequences.

The CUs of a higher priority will have a higher priority of accessing a channel. However, this scheme doesn't consider the variability of channel gain among the CUs. We also compare the performance of our proposed algorithm with that of the algorithm proposed in [92], where the authors propose a set of novel acceptability-based QoE models, denoted as A-QoE, based on the results of comprehensive user studies on subjective quality acceptance assessments. The models are able to predict users acceptability and pleasantness in various mobile video usage scenarios.

3.6.2 Simulation Results and Analysis

As a basis for our simulations and discussions, Fig. 3.1 plots the relationship between MOS and data rate according to (3.9) for three widely used test video sequences with different content types, including Suzie, Carphone, and Football. The parameters are obtained from [71]. The results are as expected since generally for the same data rate, the MOS of a slow motion video sequence is higher than that of a high motion video sequence. We use these video sequences in the simulations presented in the rest of this section.

The effectiveness of the sensing algorithm component of PS1 is presented in Fig. 3.2. We increase the minimum channel idle probability $\min_j \left\{P\left(r_j^t = 0\right)\right\}$ from 0.1 to 0.47 and plot the real channel states and the sensed channel states. As a benchmark, we also present the simulation results with the random sensing scheme used in [85] and [88]. With random sensing, each CU

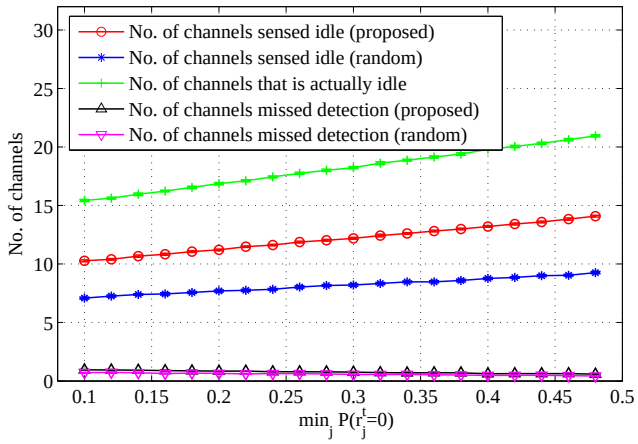

Figure 3.2: Performance of channel sensing vs. the minimum channel idle probability.

randomly and independently selects one of the N_1 channels to sense with equal probability. As utilization of the channels decreases, the number of idle channels increases. The proposed sensing algorithm can discover more idle channels for CUs to use. Moreover, the number of channels that miss detection is less than 0.5 on average, which is less than $N_1 \times (1 - P^t_{d_j}) = 1.5$. Recall that N_1 is the total number of channels and $P^t_{d_j}$ is the probability of detection. So $N_1 \times (1 - P^t_{d_j})$ is the expected number of channels that miss detection. The sensing algorithm offers an acceptable level of protection to the PUs, and is effective in detecting idle channels for the CUs.

We next compare the expected MOS of all the CUs at each time slot (denoted as Ψ_t) during an entire GOP window. In our simulations, each CU requests a video sequence of a certain content type (different CUs may request videos of different content type), and the request is sent to the BS. The BS decides the channel allocation based on the objective of maximizing the MOS sum of all CUs. In Fig. 3.3, we plot the achieved MOS sum of all the CUs achieved by PS1 and the Benchmark schemes. We set $\min_j \left\{P\left(r^t_j = 0\right)\right\} = 0.5$ and traffic load $\xi = 1$ in this simulation. Fig. 3.3 shows that the proposed QoE-aware scheme achieves a consistently high QoE sum than all the three benchmark schemes during the entire GOP window. The main reason is that Benchmark schemes 1 and 3 only consider channel gain diversity among the CUs while allocating channels, and Benchmark scheme 2 assigns channels to the CUs based on their respective priorities only and

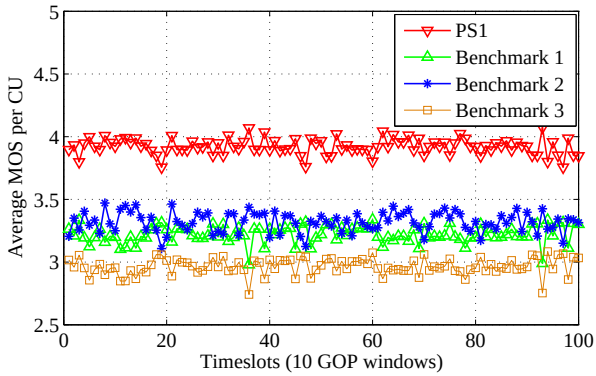

Figure 3.3: Instant MOS per CU over time during 10 GOP windows.

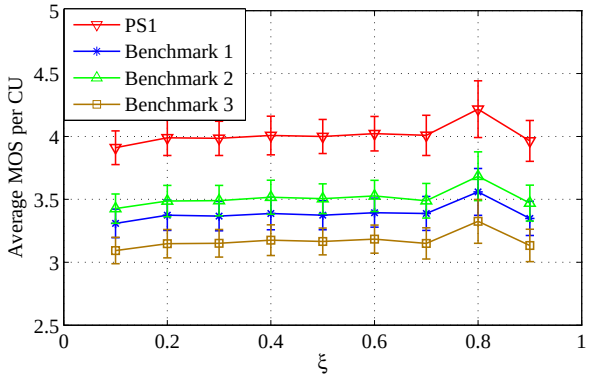

Figure 3.4: Average MOS per CU over time during 10 GOP windows for different traffic loads.

channel gain diversity is not considered among the CUs, which may result in a suboptimal strategy to the objective of maximizing the MOS of all the CUs.

Fig. 3.4 demonstrates how the CU video quality is affected by the traffic load of the CUs (i.e., ξ). The average MOS per CU during 10 GOP windows achieved by PS1 and the benchmark schemes are plotted, where 95% confidence intervals are plotted as error bars. As the CU traffic load is increased, more CUs need channels for video transmission. We can see that while the number of the really idle channels is greater than the number of active CUs, the average MOS per CU of all schemes increases with ξ, and the performance gap between our proposed scheme and the benchmark schemes grows larger. While the number of really idle channels is no greater than the number of active CUs, the average MOS sum of both schemes remain the same, since no more channel resource is available to satisfy the need of the extra CUs.

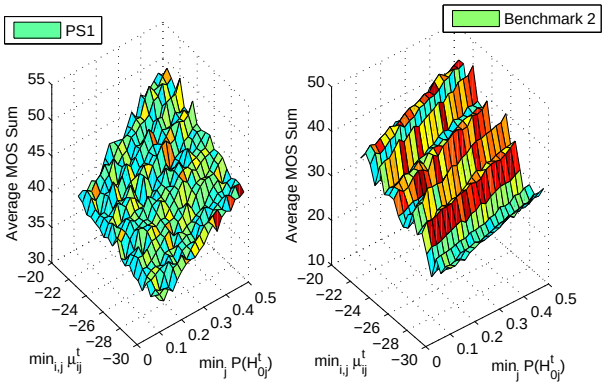

Figure 3.5: Average MOS sum of the CUs avg_Ψ, vs. the minimum channel idle probability, $\min_{i,j}\left\{\Pr\left(H_{0j}^t\right)\right\}$, and the minimum SNR of CUs, $\min_{i,j}\left\{\mu_{ij}^t\right\}$.

In Fig. 3.5, we examine the impact of PU channel utilization and the SNR at the CUs on CU video quality. In the 3-D plots, the x-axis is the minimum channel idle probability, i.e., $\min_j\left\{P\left(r_j^t=0\right)\right\}$, and the y-axis is the minimum SNR of CUs, i.e., $\min_{i,j}\left\{\mu_{ij}^t\right\}$. It can be observed from the figure that as channel utilization is decreased, a channel has a higher probability of being idle and there will be more channels available for CUs in the transmission phase. Thus the average MOS sum of the CUs is improved. Furthermore, it follows from (3.15) that

$$\begin{aligned}\varpi_{ij}^t &= P(r_j^t=0|s_j^t=0)\cdot\left(\phi_{ij}^t-\theta_{ij}^t\right)+\theta_{ij}^t \\ &= \frac{(1-P_{f_j}^t)P(r_j^t=0)\cdot\left(\phi_{ij}^t-\theta_{ij}^t\right)}{(1-P_{f_j}^t)P(r_j^t=0)+(1-P_{d_j}^t)P(r_j^t=1)}+\theta_{ij}^t.\end{aligned}$$

Since ϕ_{ij}^t and θ_{ij}^t are the MOS gain when channel j is idle and busy at time slot t, respectively, we have $(\phi_{ij}^t-\theta_{ij}^t)>0$. Therefore, w_{ij}^t is an increasing function of $P\left(r_j^t=0\right)$ and the overall MOS sum is improved with $P\left(r_j^t=0\right)$. On the other hand, an increased minimum SNR at the CUs leads to a higher data rate (i.e., a higher SBR in (3.9)), and results in a higher MOS value for the CUs according to the MOS model given in (3.9). We also find PS1 outperforms the Benchmark scheme for the entire range of $\min_{i,j}\left\{\mu_{ij}^t\right\}$ and $\min_j\left\{P\left(r_j^t=0\right)\right\}$ in terms of the average MOS sum over an GOP window in this simulation.

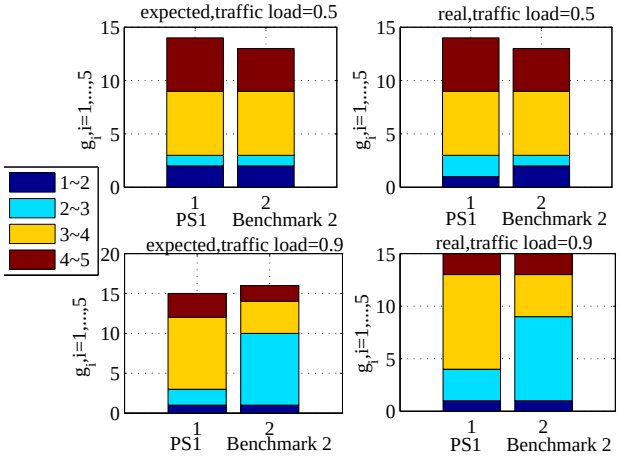

Figure 3.6: Distribution of the CU MOS values under different traffic loads.

We next show how the traffic load affects the performance of PS1 in Fig. 3.6. In particular, we simulated two traffic loads, i.e., $\xi = 0.5$ and $\xi = 0.9$, and plot the distribution of the CU MOS values. The entire MOS range (1 to 5) is evenly divided into 4 ranges with unit spans, and the number of CU MOS falling into each range is plotted in the stacked manner. We find that when the traffic load is light, most of the active CUs get the opportunity to receive video data, thus yielding a comparatively higher MOS value in this case. When the traffic load is heavy, the amount of idle channels becomes lower than the amount of active CUs, and thus some CUs are not scheduled for video streaming. The proposed scheme outperforms the benchmark schemes in both cases.

In the following we examine the performance of the proposed two-step approach for the multi-channel sensing phase, which is termed "PS2" in the simulations. In Figs. 3.7 and 3.8, we plot the number of idle channels detected and the achieved MOS values. We use a modified version of the Greedy Poly-matching Algorithm (GPA) that solves the OAPSS problem in Section 3.4.1 as a benchmark scheme where the channel allocation strategy is the same as that of our proposed two-step approach, since we only want to show the performance of the proposed two-step approach. We change the algorithm by letting the Right Hand Side (RHS) of constraint (3.7) be C, which is the spectrum sensing capacity of each CU, as $\sum_{j=1}^{N_1} x_{ij}^t = C$, for all i, and add a constraint $\sum_{i=1}^{M} x_{ij}^t = \Lambda$. This way, since all the P_{ij}^t's are identical, for all i, j, t, all the φ_{ij}^t's are also identical, for all i, j, t, and the algorithm will choose MC/Λ channels to sense.

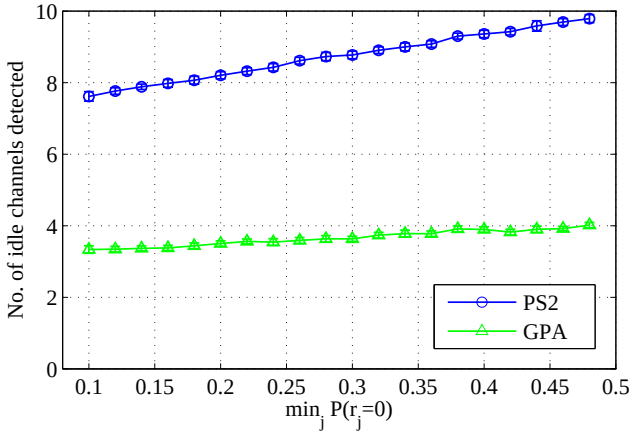

Figure 3.7: Sensing performance comparison.

In Fig. 3.7 we can see that the number of idle channels detected by PS2 is considerably greater than that by GPA. This is because in PS2, channels with a greater probability of being idle will have a higher probability of being sensed, while in GPA, all channels, regardless of the probability of being idle, have the same probability of being sensed. It is common sense that if a channel has a high probability of being idle, then it will have a high probability of being found idle by spectrum sensing. Note that in the multi-channel sensing case, at most a number of $\frac{MC}{\Lambda}$ channels will be sensed at a time slot.

In Fig. 3.8, we compare the MOS performance of PS2 and the channel assignment algorithm as in SP2 combined with GPA. Since the proposed scheme tends to find more idle channels than GPA does, more channels will be used for video streaming, leading to better QoE performance. This result also validate the fact that the proposed two-step approach which treats the spectrum-sensing-and-accessing-joint-optimization problem as an intact problem and solves for the jointly optimized sensing and accessing strategy, will achieve the optimality, the approach of decoupling the joint-optimization problem into two subproblems, as Section III does, will lost optimality in some extent.

Finally, we examine the fairness performance the channel allocation strategy considering fairness among CUs, i.e., P1 and P2. We adopt Jain's fairness index as in [112]: $f(e_1, e_2, \ldots, e_M) = \left(\sum_{i=1}^{M} e_i\right)^2 / \left(M \sum_{i=1}^{M} e_i^2\right)$, where e_i is the average MOS of CU i during a period of time (10

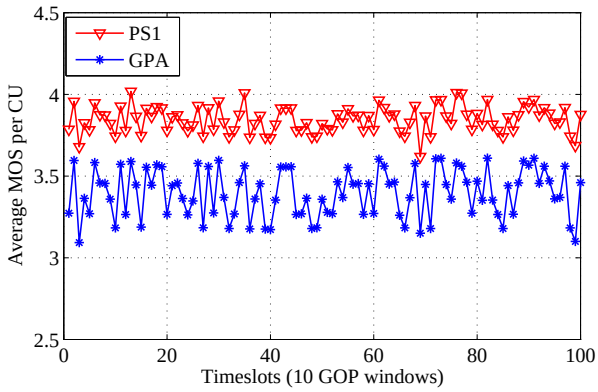

Figure 3.8: MOS performance comparison.

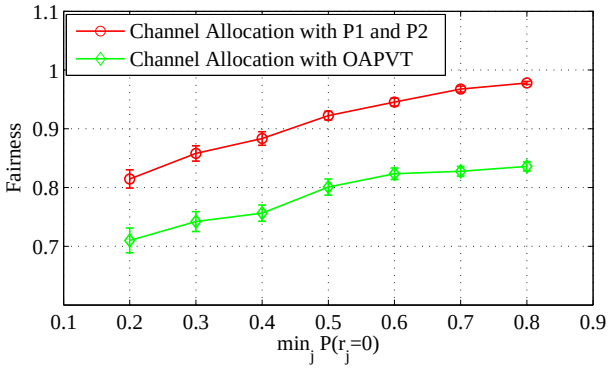

Figure 3.9: Fairness comparison between channel allocation strategy considering fairness among CUs and channel allocation strategy without considering fairness among CUs.

GOP windows in our simulation), $i = 1, 2, \ldots, M$. The fairness index ranges from 0 (worst) to 1 (best). The benchmark scheme is the channel allocation strategy without considering fairness among CUs, i.e., OAPVT.

3.7 Conclusion

In this chapter, we investigated the problem of QoE-aware video streaming over CRNs where each CU can access one channel at a time. For the case where each CU can sense and access at most one channel at a time, we formulated an IP problem on spectrum sensing and solved it with a optimal Greedy Poly-matching Algorithm. We then formulated a channel assignment problem and solved it with the Hungarian Method that is also optimal with respect to QoE of the multi-user videos. For the case where each CU can sense multiple channels but access only one channel, we

presented a more general, integrated formulation. Based on an assumption on the spectrum sensor configuration, we developed a two-step approach to solve the integrated problem and proved its optimality. The proposed schemes were shown to outperform several alternative schemes in the simulation study.

Chapter 4

A Decomposition Approach to QoS Driven Multi-user Video Streaming in Cellular CRNs

4.1 Introduction

In previous chapters, we discussed the subjective assessment method for the video quality, i.e., QoE. In this chapter, we are going to introduce the objective assessment method of video quality, i.e., QoS, which is to evaluate the video quality from the network performance perspective. QoS is an effective metric to evaluate the quality of multimedia applications. Various factors, such as throughput, delay, packet loss ration, and distortion, can be taken into consideration when developing QoS models. Besides, differnt from the pervious chapter where CUs can only access at most one channel each time, here we allow CUs to access multiple channels each time, so that CUs have higher data rate and thus the QoS is improved further. What's more, instead of fixed transmission power in previous chapter, here we also consider power adaptation for CUs so that transmission power can be dynamically adapted according to the channel conditions and a higher data rate for the channels can be achieved.

Although with great potential, the problem of video over CRNs brings about a whole level of technical challenges, particularly due to the extra dimension of dynamics on channel availability and the uncertainty from spectrum sensing and access. The manifold design trade-offs, multi-farious network dynamics, limited network resources and, on the other hand, video's stringent QoS constraints, necessitate a holistic cross-layer design approach to "squeeze" the most out of the CRN. Usually such cross-layer design results in a tremendously complex global optimization problem, where all the layers (i.e., the PHY, MAC, network, and application layers) and all the users (i.e., PUs and CUs) are tightly coupled [63] [52] [53]. A *decomposition principle* that helps to decouple the design of spectrum sensing, access, and application QoS provisioning would be crucial for making the problem manageable [32].

In this chapter, we consider the scenario of downlink multi-user video streaming in a cellular CRN, where each CU receives a video stream from the Cognitive Base Station (CBS). Each CU is able to sense (with multiple sensors or sequentially sensing the channels [33]) and access (e.g., with channel bonding/aggregation [85]) multiple channels. We adopt the H.264 Scalable Video Coding (SVC) (Quality Scalability) model from [36], and jointly design spectrum sensing, channel access, and power control for maximizing the QoS of all the CUs. There are two tightly coupled parts in this problem: the spectrum sensing problem (SP1) to determine which CU to sense which channel; and the channel assignment and power allocation problem (SP2) to allocate channels and transmit power to the CUs.

The formulated problem turns out to be a Mixed Integer NonLinear Programming (MINLP) problem, which is NP-hard in general [49]. However, as in [32], where a *separation principle* is established to decouple the design of sensing strategy from that of sensor and access policy, we show that our problem can also be decoupled into two relatively easier sub-problems with a *decomposition principle* and develop an effective Column Generation (CG) based solution algorithm [155].

The major contributions made in this chapter include:

1. A holistic *problem formulation* that jointly optimizes the spectrum sensing, channel assignment, and power allocation strategies for maximizing CU QoS.

2. A *decomposition principle* to decouple the original problem into a sensing strategy optimization problem SP1 and a resource allocation problem SP2, without sacrificing optimality under certain conditions, and *effective algorithms* to solve SP1 and SP2.

3. A *heuristic sensing scheme* that is less demanding on CU hardware than the optimal sensing strategy, but can achieve highly competitive sensing performance.

4. An *upper bound* for the performance of the CG-based distributed algorithm and an analysis of complexity and efficiency in terms of time savings.

5. *Simulation validation* to demonstrate the superior performance of our proposed algorithms in terms of sensing performance and the QoS achieved by CUs.

The remainder of this chapter is organized as follows. Section 6.2 reviews related work. The system model and problem formulation are presented in Section 4.3, while the decomposition principle and the two sub-problems SP1 and SP2 are presented in Section 4.4. The CG-based distributed algorithm to solve SP2 is developed in Section 4.5 and analyzed in Section 6.7.2. Section 5.6 presents the performance validation and Section 6.8 concludes this chapter.

4.2 Related Work

CR research has been largely focused on the aspects of spectrum sensing and dynamic spectrum access. In [73], the authors study the sensing-throughput tradeoff problem that optimizes the spectrum sensing time so that the CU's throughput can be maximized with restricted interference to the PUs. Unlike [73], the protocol proposed in [32] also considers the problem of which channel to sense, in addition to sensing parameters and access strategy optimization. Moreover, the design of sensing strategy is independent to sensing parameters design and the access strategy, as specified in a *principle of separation* [32]. These works focus on the optimization of sensing parameters only, and there is no collaboration among CUs. Considering the fact that different CUs may have different spectrum sensing performance, the authors in [75] propose an algorithm where groups of CUs are formed for cooperative sensing, aiming to find the best grouping scheme to discover most idle channels. Furthermore, the problem of sensing parameter optimization in addition to optimal sensor selection is addressed in [76], in order to achieve a trade-off between detection performance and sensing overhead.

Recently, cross layer design for video streaming over CRNs has attracted considerable interest. An auction game model is proposed in [78] to solve the problem of spectrum allocation in delay-sensitive content-aware multimedia delivering. Channel/path selection for multi-user video streaming is formulated as an MINLP problem in [63] to maximize the received video quality while restricting collisions with PUs. Packet scheduling is studied in [85] in which spectrum sensing at the PHY is integrated with packet scheduling at the MAC layer to improve delay-QoS provisioning over CRNs. The authors also analyze the throughput and delay performance with a Markov

chain and $M/G_Y/1$ queuing model. Beyond these, other cross layer factors such as Fine Grained Scalability (FGS) coding, error control, and modulation, are jointly considered in [27] to achieve the maximum QoS for CUs in a cellular CRN. Interestingly, cross layer optimization of streaming videos over a CR link can also be modeled as a POMDP (Partially Observable Markov Decision Process) as in [42], in which intra refreshing rate, a video codec parameter, along with spectrum sensing and access strategies are jointly designed.

Different from the above mentioned prior works that consider the physical layer factors, some other works treat the problem of video streaming over CRNs in the Medium Access Control (MAC) layer, Network layer, Transportation layer, and Application layer. The authors in [29] model the routing problem in video streaming over CRNs as a decision tree problem, where the quality of a multi-hop path is determined by the quality of the channels along this path, and the quality of the channels is inferred using prior distribution and posterior distribution. Then the routing scheme is optimized in order to maximize the PSNR of the received video sequence. Considering the fact that different CUs may have different channel data rate and different buffer storage size, which result in different abilities of tolerating network dynamics, the authors of [30] propose to allocate channels to CUs according to their buffer storages. Basically, the CU with a smaller buffer will have a higher priority of accessing channels to avoid playout outage. Reducing the playout speed to a certain extent when the buffered data at the receiver is low, is also a feasible solution to reduce the probability of playout outage, as the authors propose in [31]. The proposed scheme is based on the observation that varying the video playout rate by 25% to 50% can be unnoticeable by the viewers.

This chapter is motivated by these interesting prior works, and is mainly focused on the joint design of spectrum sensing and resource allocation strategies for streaming multi-user videos over the downlink of a cellular CRN with a novel decomposition principle, which is not well addressed in prior work but is essential for supporting the demand of large bandwidth for video applications and enhance the QoS of CUs.

4.3 System Model and Problem Statement

4.3.1 System Model

We consider a primary network operating on N_1 licensed orthogonal channels, while each channel j has bandwidth B_j. A CR network is co-located with the primary network, consisting of a CBS and M CUs. The CUs sense the PU activities on the licensed channels and access the channels in an opportunistic manner. As in prior work [33], we first assume that each CU is equipped with N_1 sensors so that it can sense all the channels simultaneously. This assumption is relaxed in Section 4.4.2, where each CU can only sense a few channels at a time. The CBS determines the status of the licensed channels based on the sensing results reported from the CUs.

We consider the scenario of downlink multi-user video streaming, where the CBS transmits different video streams to the CUs using the channels sensed idle. Once the channel states are estimated, the CBS and CUs determine the allocation of the idle channels, and the CBS selects a power level k, $k = 1, 2, \cdots, K$, for the video transmission to a CU on each allocated channel.

We assume that each CU and the CBS adopt the channel bonding/aggregation technique [35, 85], such that they can transmit on multiple assigned channels simultaneously to make use of all the available spectrum. To enforce a certain level of fairness among the CRs, we define an upper bound C_i on how much total time a CU i can access all the channels. If C_i is less than the total number of channels, this can limit how much channel time a CR can have at most. Otherwise, if C_i is equal to the number of channels, then there is no such fairness constraint and a CR can access all the channels for the entire time slot.

We assume time is divided into a series of non-overlapping Group of Pictures (GOP) windows, each consisting of $\mathbb{T}$ time slots. The operations of the CBS and CUs in each time slot, as discussed above, are summarized in Fig. 4.1.

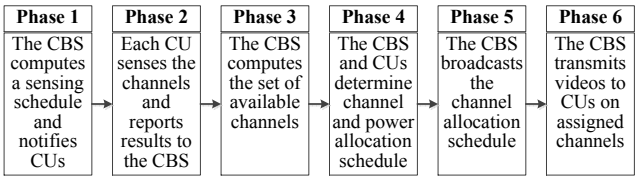

Figure 4.1: Operations of CBS and CUs in a time slot.

4.3.2 Problem Statement

Due to multipath fading and shadowing, different CUs usually experience different SNRs when detecting a PU signal, and thus may have different sensing performance. It is important to choose a suitable set of CUs to sense a licensed channel [75]. To achieve diversity gain, cooperative sensing is usually used to improve the detection performance by fusing the sensing results from multiple CUs, where a certain fusion rule is used to combine the CU sensing results. In this chapter, we consider cooperative sensing with the OR fusion rule: if any of the CUs reports the presence of a PU signal on a channel, the CBS will determine that the channel is busy; otherwise, the channel is considered to be idle.

We use an $M \times N_1$ matrix $\mathbf{X}$ to represent the assignment of sensing tasks, where each element x_{ij} is defined as

$$x_{ij} = \begin{cases} 1, & \text{CU } i \text{ is assigned to sense channel } j \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

As in previous chapter, here we adopt the energy detection method for spectrum sensing, and the probability of detection of PU signal on channel j by CU i, $P_{d_{ij}}$,the probability of false alarm on channel j by CU i, $P_{f_{ij}}$, and the probability of detection of PU signal on channel j, P_{d_j}, the probability of false alarm on channel j, P_{f_j} for cooperative sensing with the OR fusion rule,, are the same as in previous chapter.

To provide a graceful protection to PUs, we set $P_{d_j} = P_{req}$, where P_{req} is the maximum interference from the CU system that can be tolerated by the PU system, and $\sum_{i=1}^{M} x_{ij} = \Lambda_j$, for all $j = 1, 2, \cdots, N_1$, where Λ_j is the minimum number of CUs to sense channel j. If we set $P_{d_{ij}} = \bar{P}_{d_j} = 1 - (1 - P_{req})^{1/\Lambda_j}$, it follows (3.3) that $P_{d_j} = P_{req}$.

Let r_j denote the real state of channel j: $r_j = 0$ when channel j is idle, and $r_j = 1$ otherwise. Also let s_j be the cooperative sensing result on channel j: $s_j = 0$ if the channel is determined idle, and $s_j = 1$ otherwise. We have

$$\begin{cases} P(s_j=0) = P(r_j=0)(1-P_{f_j})+P(r_j=1)(1-P_{d_j}) \\ P(s_j=1) = P(r_j=0)P_{f_j}+P(r_j=1)P_{d_j}. \end{cases} \tag{4.2}$$

The cooperative sensing results on the N_1 channels can be represented as $\vec{\mathbf{S}} = \{s_j, j = 1, \cdots, N_1\}$. There are 2^{N_1} possible outcomes for $\vec{\mathbf{S}}$, and let $\vec{\mathbf{S}}_h$ be the h-th outcome, $0 \leq h \leq 2^{N_1} - 1$. To determine the j-th element in $\vec{\mathbf{S}}_h$, let $s_j = \Gamma_j(h), j = 1, 2, \cdots, N_1$, denote the relationship between $\vec{\mathbf{S}}_h$ and s_j. Assuming independent channel states, the probability of getting outcome $\vec{\mathbf{S}}_h$ can be written as

$$\begin{aligned} P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) &= \prod_{j=1}^{N_1} P(s_j = \Gamma_j(h)) \\ &= \prod_{j=1}^{N_1} \left[(1 - \Gamma_j(h))P(s_j = 0) + \Gamma_j(h)P(s_j = 1)\right]. \end{aligned} \tag{4.3}$$

We adopt the QoS model for H.264 SVC (Quality Scalability) from [36] as

$$\rho_i = \alpha_i + \beta_i \cdot R_i, \tag{4.4}$$

where ρ_i is the Y-PSNR (Peak Signal-to-Noise Ratio) of the received video at CU i, α_i and β_i are constants dependent on the content type of the video sequence, and R_i is the effective data rate of the video sequence. According to conditional expectation, the expected overall QoS can be derived as

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^{M} \rho_i\right) &= \sum_{i=1}^{M} \sum_{h=0}^{2^{N_1}-1} \mathrm{E}(\rho_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \\ &= \sum_{h=0}^{2^{N_1}-1} \sum_{i=1}^{M} \mathrm{E}(\rho_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h). \end{aligned} \tag{4.5}$$

For $\vec{\mathbf{S}} = \vec{\mathbf{S}}_h$, let $\Phi_h = \{j : \Gamma_j(h) = 0, j = 1, 2, \cdots, N_1\}$ be the set of channels sensed idle. Let $\mathbf{Y}_h = \left[y_{ijk}^h\right], 1 \leq i \leq M$, $j \in \Phi_h$, $1 \leq k \leq K$, be the channel assignment and power allocation matrix, where $0 \leq y_{ijk}^h \leq 1$ is the amount of time that CBS transmits to CU i with a power level of k on channel j in a time slot, when the sensing outcome is $\vec{\mathbf{S}}_h$. The channel assignment and power allocation strategy can be expressed as $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \cdots \mathbf{Y}_{2^{N_1}}]$.

Putting it all together, it follows that

$$\begin{aligned}
&\mathrm{E}(\rho_i|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) = \alpha_i + \beta_i \cdot \mathrm{E}(R_i|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \\
&= \alpha_i + \beta_i \cdot \mathrm{E}\left(\sum_{j\in\Phi_h}\sum_{k=1}^{K} R_{ijk} \cdot y_{ijk}^h|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h\right) \\
&= \alpha_i + \beta_i \sum_{j\in\Phi_h}\sum_{k=1}^{K}\left(P_{00}^j R_{00}^{ijk} + P_{10}^j R_{10}^{ijk}\right) \cdot y_{ijk}^h
\end{aligned} \tag{4.6}$$

where G_k is the power of level k, d_{ij} is the channel gain between the CBS and CU i on channel j, $P_{00}^j = P(r_j = 0|s_j = 0) = \frac{(1-P_{f_j})P(r_j=0)}{(1-P_{f_j})P(r_j=0)+(1-P_{d_j})P(r_j=1)}$, $P_{10}^j = P(r_j = 1|s_j = 0) = 1 - P_{00}^j$, $R_{00}^{ijk} = B_j \log_2(1 + G_k d_{ij}/(n_0 B_j))$, and $R_{10}^{ijk} = B_j \log_2(1 + G_k d_{ij}/(n_0 B_j(1 + \gamma_{ij})))$.

Define $w_{ijk} = \alpha_i + \beta_i B_j(P_{00}^j R_{00}^{ijk} + P_{10}^j R_{10}^{ijk})$. The master problem of maximizing the total expected QoS, denoted as P0, can be formulated as follows.

$$\underline{\mathbf{P0}}: \max: \sum_{h=0}^{2^{N_1}-1}\sum_{i=1}^{M}\sum_{j\in\Phi_h}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \tag{4.7}$$

$$\text{s.t.} \sum_{j\in\Phi_h}\sum_{k=1}^{K} y_{ijk}^h \leq C_i, \forall\, i, h \tag{4.8}$$

$$\sum_{i=1}^{M}\sum_{k=1}^{K} y_{ijk}^h \leq 1, \forall\, j \tag{4.9}$$

$$\sum_{i=1}^{M}\sum_{j\in\Phi_h}\sum_{k=1}^{K} y_{ijk}^h \cdot G_k \leq G_{total}, \forall\, h \tag{4.10}$$

$$\sum_{i=1}^{M} x_{ij} = \Lambda_j, \forall\, j \tag{4.11}$$

$$x_{ij} = \{0, 1\}, \forall\, i, j \tag{4.12}$$

$$y_{ijk}^h \begin{cases} \in [0,1], & \text{if } G_k d_{ij}/(n_0 B_j) \geq \bar{\gamma} \\ = 0, & \text{otherwise}, \end{cases} \quad \forall\, i,j,k. \tag{4.13}$$

In each time slot, constraint (6.15) enforces that the sum of time CU i spend on all channels is less than C_i.

According to [33, 35, 85], using discontinuous orthogonal frequency division multiplexing (D-OFDM), the CBS or a CU can aggregate multiple discontinuous orthogonal channels together to form an aggregated channel and then access the aggregated channel. This way, the CBS or a CU can aggregate and access all the channels assigned to it simultaneously. Therefore, after computing all the y_{ijk}^h's, the CBS transmits to all CUs sequentially according to the y_{ijk}^h's, and each CU is informed of this schedule before transmission so that each CU knows when to access which channel, then it is a feasible and optimal scheduling.

Constraint (6.16) enforces that the transmission time on each channel is within 1 time slot. constraint (4.10) enforces that the total transmission power of the CBS must not exceed the average power limit G_{total}; constraint (4.11) enforces that there are Λ_j CUs to sense each channel j; and constraint (4.13) enforces that the necessary condition for the CBS to transmit to CU i on channel j with power level k is that the resulting SNR must be greater than a predefined threshold $\bar{\gamma}$ such that CU i can successfully decode the received video.

Note that constraint (4.13) indicates $y_{ijk}^h \geq 0$. Combined with constraint (6.16), it follows that each $0 \leq y_{ijk}^h \leq 1$. Therefore, constraint (4.13) can be rewritten as

$$y_{ijk}^h \begin{cases} \geq 0, & \text{if } G_k d_{ij}/(n_0 B_j) \geq \bar{\gamma} \\ = 0, & \text{otherwise}, \end{cases} \quad \forall\, i,j,k. \tag{4.14}$$

The upper bound of 1 on the y_{ijk}^h's is thus removed and the problem can be solved more efficiently, since usually LP solvers solve an LP without upper bounds faster than LPs with upper bounds.

4.4 Problem Decomposition

4.4.1 Optimal Sensing Strategy

The formulated problem P0 is an MINLP, which is NP-hard. However, we observe that the optimal sensing strategy can be obtained by solving a relatively easier problem as follows.

We first introduce Lemma 4.1 as a basis for our later analysis.

Lemma 4.1. *The objective value of P0 is a decreasing function of P_{f_j}, for all $j = 1, 2, \cdots, N_1$.*

Proof. Define $F(\mathbf{X}, \mathbf{Y}_h) = \sum_{i=1}^{M} \sum_{j \in \Phi_h} \sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h)$ and $f(\mathbf{X}, \mathbf{Y}) = \sum_{h=0}^{2^{N_1}-1} F(\mathbf{X}, \mathbf{Y}_h)$. The partial derivative of $F(\mathbf{X}, \mathbf{Y}_h)$ with respect to P_{f_j} is

$$\frac{\partial F(\mathbf{X}, \mathbf{Y}_h)}{\partial P_{f_j}} = -y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \cdot \sum_{i=1}^{M} \sum_{k=1}^{K} \Big(\beta_i \cdot R_{00}^{ijk} \cdot$$
$$((1-P_{req})P(r_j = 0)P(r_j = 1) + (1-P_{f_j})P(r_j = 0)^2) +$$
$$\alpha_i P(r_j = 0)(R_{00}^{ijk} P_{00}^j + R_{10}^{ijk} P_{10}^j)/B_j \Big) \leq 0.$$

It follows that $\frac{\partial f(\mathbf{X},\mathbf{Y})}{\partial P_{f_j}} = \sum_{h=0}^{2^{N_1}-1} \frac{\partial F(\mathbf{X},\mathbf{Y}_h)}{\partial P_{f_j}} \leq 0$. □

Theorem 4.1. *The optimal spectrum sensing strategy to problem P0 can be obtained by solving the following problem SP1.*

$$\underline{\textbf{SP1}} : \forall\ j = 1, 2, \cdots, N_1$$
$$\min : P_{f_j} = 1 - \prod_{i=1}^{M} (1 - P_{f_{ij}})^{x_{ij}} \tag{4.15}$$
$$\text{s.t.} \quad \sum_{i=1}^{M} x_{ij} = \Lambda_j. \tag{4.16}$$

Proof. Let the optimal solution to the original problem P0 be $(\mathbf{X}', \mathbf{Y}')$ and the solution to SP1 be $\mathbf{X}^*$. Since $\mathbf{X}'$ is optimal to the maximization problem P0, we have $f(\mathbf{X}^*, \mathbf{Y}') \leq \mathbf{f}(\mathbf{X}', \mathbf{Y}')$. On the other hand, since $\mathbf{X}^*$ is optimal to the minimization problem SP1, it follows Lemma 4.1 that

$f(\mathbf{X}^*, \mathbf{Y}^{'}) \geq \mathbf{f}(\mathbf{X}^{'}, \mathbf{Y}^{'})$. Therefore we conclude that $f(\mathbf{X}^*, \mathbf{Y}^{'}) = \mathbf{f}(\mathbf{X}^{'}, \mathbf{Y}^{'})$ and $\mathbf{X}^*$ is optimal to problem P0.

After obtaining $\mathbf{X}^*$, we substitute it into problem P0 to solve for $\mathbf{Y}^*$. The $\mathbf{Y}^*$ obtained this way is also optimal to P0, i.e., we have $f(\mathbf{X}^*, \mathbf{Y}^*) \geq \mathbf{f}(\mathbf{X}^*, \mathbf{Y}) \geq \mathbf{f}(\mathbf{X}, \mathbf{Y})$, for all $\mathbf{X}, \mathbf{Y}$. The proof is completed. □

From the proof of Theorem 4.1, we have the following Decomposition Principle for the joint sensing, channel assignment and power allocation problem.

Corollary 4.1.1. *If there is no restriction on the sensing capability for each CU, or $\eta_i \leq \Theta_i$, for all i, then the problem P0 that jointly optimizes spectrum sensing, channel assignment, and power allocation can be decomposed into two sub-problems: one for the optimal spectrum sensing strategy, and the other for the optimal channel assignment and power allocation, without sacrificing optimality.*

Problem SP1 can be rewritten as the following problem SP1a,

$$\underline{\textbf{SP1a}} : \forall\ j = 1, 2, \cdots, N_1,$$

$$\max : \sum_{i=1}^{M} x_{ij} \cdot \log_2(1 - P_{f_{ij}}) \tag{4.17}$$

$$\text{s.t.} \ \sum_{i=1}^{M} x_{ij} = \Lambda_j, \tag{4.18}$$

which can be solved easily with an LP solver.

4.4.2 Optimal Solution to a More General Problem

Due to the time constraint (i.e., when the channels are sensed sequentially) or the hardware constraint, it may not be feasible for a CU to sense all the channels (although this may not be a problem for the CBS). In this section, we consider a more general case of spectrum sensing, where each CU i can only sense at most Θ_i channels simultaneously at a time slot. Therefore

the following additional constraint is added to Problem P0. Problem P0 is a special case when $\Theta_i = N_1$.

$$\sum_{j=1}^{N_1} x_{ij} \leq \Theta_i \leq N_1, \forall\, i. \tag{4.19}$$

Denote the more general problem as P0a. We still apply the same solution algorithm as in Theorem 4.1 to Problem P0a, and each channel j will select Λ_j users with the best sensing performance (smallest false alarm probability). There are two different cases. First, if a CU i is selected by more than Θ_i channels, then the new Constraint (4.19) is violated. Second, if CU i is selected by less than Θ_i channels, then Constraint (4.19) is still satisfied and Theorem 1 still holds true and the solution is optimal under the new constraint. Denote Π_j as the set of Λ_j CUs with the best sensing performance regarding to channel j. We can use the following procedure shown in Algorithm 4 to check if each CU i is selected by less than Θ_i channels.

Algorithm 4: Applicability of Theorem 4.1

```
for i = 1 : M do
    η_i = 0;
    for j = 1 : N_1 do
        if CU i ∈ Π_j then
            η_i = η_i + 1;
        end
    end
    if η_i > Θ_i then
        Theorem 1 is not applicable ;
        Break ;
    end
end
```

Algorithm 4 has a polynomial complexity of $\mathcal{O}(MN_1^2)$. We conjecture that if the PUs are widely separated, or when the channels are highly diverse, Theorem 4.1 will be more likely to hold true under the new practical constraint (4.19). However, if Theorem 4.1 is not applicable under (4.19), we can use a heuristic algorithm to obtain a near-optimal solution to P0a, which is presented in the following section.

Algorithm 5: Heuristic Spectrum Sensing Algorithm

1 Sort the N_1 channels in descending order of $P(r_j = 0)$ and let the sorted channel set be Ξ ;
2 **for** $j = 1 : N_1$ **do**
3 Let $j' = \Xi(j)$;
4 Solve problem SP1b and denote the solution as $\underline{\Theta}_{j'}$;
5 **if** $\underline{\Theta}_{j'} = \emptyset$ **then**
6 Channel j' is determined to be busy ;
7 **end**
8 **if** $x_{ij'} = 1$ **then**
9 $\Theta_i = \Theta_i - 1$;
10 **end**
11 **end**

4.4.3 Heuristic Spectrum Sensing Algorithm

The idea of heuristic algorithm to Problem P0a is to sort the N_1 channels according to $P(r_j = 0)$, for all $j = 1, 2, \ldots, N_1$, in the descending order, and then minimize P_{f_j}, for all $j = 1, \ldots, N_1$, sequentially. The heuristic spectrum sensing algorithm is presented in Algorithm 7. In Line 4, the following problem SP1b is solved.

$$\underline{\textbf{SP1a}} : \forall\ j = 1, 2, \cdots, N_1,$$

$$\min : P_{f_{j'}} = 1 - \prod_{i=1}^{M} (1 - P_{f_{ij'}})^{x_{ij'}} \tag{4.20}$$

$$\text{s.t. } \sum_{i=1}^{M} x_{ij'} = \Lambda_{j'} \tag{4.21}$$

$$x_{ij'} \leq \Theta_i. \tag{4.22}$$

In Lines 5∼7, if there is no feasible solution to problem SP1b, there is not a sufficient number of CUs to sense channel j', and we conservatively assume that channel j' is busy to avoid collision with PUs. Each time if CU i is assigned to sense a channel, Θ_i is decreased by 1 as in Lines 8∼10. When Θ_i reaches 0, constraint (4.22) will prevent CU i to be assigned to sense any more channels.

In Section 5.6, we will show that the performance of Algorithm 7 is very close to that of the optimal sensing strategy in terms of both sensing performance and the expected overall QoS, even when $\Theta_i \ll N_1$.

4.4.4 Optimal Chanel Assignment and Power Allocation Solution

After obtaining $\mathbf{X}^*$, cooperative sensing is conducted and the CBS determines the set of available channels based on sensing results, as shown in Fig. 4.1. From now on, we omit the subscript (or superscript) h in all the symbols, since the cooperative sensing results on the N_1 channels is already determined. Denote the number of channels sensed idle as N_2, and re-index the N_2 idle channels as $1, 2, \cdots, N_2$. Then the remaining channel assignment and power allocation problem SP2 can be written as follows.

$$\underline{\textbf{SP2}}: \max: \sum_{i=1}^{M}\sum_{j=1}^{N_2}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk} \tag{4.23}$$

$$\text{s.t.} \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \leq C_i, \forall\, i \tag{4.24}$$

$$\sum_{i=1}^{M}\sum_{k=1}^{K} y_{ijk} \leq 1, \forall\, j \tag{4.25}$$

$$\sum_{i=1}^{M}\sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \cdot G_k \leq G_{total} \tag{4.26}$$

Constraint (4.14).

In practice, there may be a large number of CUs and licensed channels, and the CBS also has a great flexibility to choose the power level for transmission on a channel. Therefore, the constraint matrix of SP2 could be huge and it may be hard to solve with an LP solver due to its size. In the next section, we propose to use the Column Generation (CG) method [155] to solve SP2 and derive a decentralized algorithm for better scalability. With the proposed CG method, the CBS and CUs solve different sub-problems, thus alleviating the computational burden on the CBS.

4.5 Column Generation Method for Sub-problem

4.5.1 Dantzig-Wolfe Decomposition

We first reformulate problem SP2 from the standard form into a *disaggregated formulation* by applying Dantzig-Wolfe decomposition of LP problems [54].

For $i = 1, 2, \cdots, M$, let $\overline{\Theta}_i = \left\{\chi_i^1, \chi_i^2, \cdots, \chi_i^{Q_i}\right\}$ denote the set of feasible channel assignment and power allocation schemes to CU i. Then

$$\chi_i^q = \left\{y_{ijk}^q, j = 1, 2, \cdots, N_2, k = 1, 2, \cdots, K\right\},$$

for $q = 1, 2, \cdots, Q_i$, is a feasible scheme satisfying all the constraints, where $y_{ijk}^q = 1$ if the CBS transmits to CU i on channel j at power level k, and $y_{ijk}^q = 0$ otherwise. Thus, the feasible schemes are indeed the extreme points of the the feasible region of SP2, which is the key for Dantzig-Wolfe decomposition [54].

Introduce a variable $0 \leq z_i^q \leq 1$ to denote the amount of time the CBS transmits using feasible scheme χ_i^q within a time slot. Let the "utility" gained by using χ_i^q for CU i as $\varpi_i^q = \sum_{j=1}^{N_2} \sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^q$. Then SP2 can be represented in a *set-partition form*, termed the Master Problem (MP), as

$$\underline{\textbf{MP}}: \max: \quad \sum_{i=1}^{M} \sum_{q=1}^{Q_i} \varpi_i^q \cdot z_i^q \tag{4.27}$$

$$\text{s.t.} \quad \sum_{q=1}^{Q_i} z_i^q \leq 1, \forall\, i, \tag{4.28}$$

$$\sum_{i=1}^{M} \sum_{q=1}^{Q_i} \left(\sum_{k=1}^{K} y_{ijk}^q \right) z_i^q \leq 1, \forall\, j, \tag{4.29}$$

$$\sum_{i=1}^{M} \sum_{q=1}^{Q_i} \left(\sum_{j=1}^{N_2} \sum_{k=1}^{K} y_{ijk}^q \cdot G_k \right) z_i^q \leq G_{total}, \tag{4.30}$$

$$0 \leq z_i^q \leq 1, \forall\, i, q. \tag{4.31}$$

Constraint (4.28) ensures that $0 \leq y_{ijk} = \sum_{q=1}^{Q_i} y_{ijk}^q z_i^q \leq 1$, for all i, j, k; constraints (4.29) and (4.30) correspond to constraints (4.25) and (4.26), respectively; and constraints (4.24) and (4.14) are specified in the INitialization Problem (INP) and Pricing Problem (PP) defined next in Section 4.5.2. For convenience of our later discussion, the problem containing a subset of the columns and cost coefficients (variables) of the MP is called Restricted MP (RMP).

4.5.2 Design of the Column Generation Method

Obviously, it is infeasible to solve the MP directly due to the exponential number of columns. However, usually most of the variables in the optimal solution to the MP are equal to zero, with only a small number of positive-valued variables. The MP solution can be re-optimized iteratively by finding the variables having the potential to improve the objective value at each iteration. This is done by iteratively solving the PP, which examines whether there exists a variable with a negative (in the case of a minimization problem) or positive (in the case of a maximization problem) reduced cost, and then generates the corresponding column to add it to the RMP.

The RMP contains only a small subset of all the feasible columns and variables of the MP and thus can be solved quickly. The simplex multiplier obtained from the RMP will be passed to the PP to identify a new column to enter the RMP again, until there is no variables whose reduced cost is negative (in the case of a minimization problem) or positive (in the case of a maximization problem). Thus an optimal feasible solution to the MP is found. The purpose of RMP is to generate the simplex multiplier for solving the PP.

The CG based Distributed Optimization Algorithm (CDOA) includes the following six steps.

Algorithm 2: CG Based Distributed Optimization Algorithm

Step 1: CU i solves the following i-th INP and reports its solution to the CBS, $i = 1, 2, \cdots, M$.

$$\underline{\textbf{INP}}: \max: \sum_{j=1}^{N_2} \sum_{k=1}^{K} w_{ijk} \cdot y_{ijk} \tag{4.32}$$

$$\text{s.t.} \sum_{j=1}^{N_2} \sum_{k=1}^{K} y_{ijk} \leq C_i, \tag{4.33}$$

$$\sum_{k=1}^{K} y_{ijk} \leq 1, \forall\, j, \tag{4.34}$$

$$y_{ijk} \begin{cases} \in \{0,1\}, & \text{if } G_k d_{ij}/(n_0 B_j) \geq \bar{\gamma} \\ = 0, & \text{otherwise}, \end{cases} \quad \forall\, j, k. \tag{4.35}$$

Each of the M solutions generates a feasible column of the MP. The CBS uses the M feasible columns and the corresponding cost coefficients to initiate the RMP, which has the same formulation with the MP, but with $Q_i = 1$, for all $i = 1, 2, \cdots, M$.

Step 2: The CBS solves the RMP, from which a vector of simplex multiplier $\Omega^T = (\nu^T, \mu^T, \varphi)$ is obtained, where $(\cdot)^T$ denotes the transpose of a vector, ν^T is a $1 \times M$ vector with the i-th entry ν_i corresponding to the i-th constraint in the RMP, μ^T is a $1 \times N_2$ vector with the j-th entry μ_j corresponding to the $(M + j)$-th constraint in the RMP, and φ is the simplex multiplier corresponding to the last constraint in the RMP. The objective value of the RMP is a *lower bound* to the MP.

Step 3: The CBS broadcasts Ω^T to all CUs and assigns CU i to solve the following i-th PP, to find the column and the corresponding variable with the most positive reduced cost [155] to enter the RMP to improve the objective value of the MP.

$$\underline{\textbf{PP}}: \max: \Delta_i = \sum_{j=1}^{N_2} \sum_{k=1}^{K} (w_{ijk} - \mu^j - \varphi G_k) \cdot y_{ijk} - \nu_i \tag{4.36}$$

$$\text{s.t.} \sum_{j=1}^{N_2} \sum_{k=1}^{K} y_{ijk} \leq C_i, \tag{4.37}$$

$$\sum_{k=1}^{K} y_{ijk} \leq 1, \forall\, j, \tag{4.38}$$

Constraint (4.35).

Step 4: Each CU i decides when to report its optimal solution to the i-th PP to the CBS according to a delay

$$\tau_i = \xi(\Delta_i), \tag{4.39}$$

where $\xi(\cdot)$ denotes a decreasing function of Δ_i. Define an index $a = \arg\max_{i=1,\cdots,M}\{\Delta_i\}$. In case that $\Delta_a > 0$, then the current optimal solution to the RMP is not optimal to the MP and CU a sends its solution in the earliest time τ_a (since it has the maximum value Δ_a). Other CUs overhearing CU a's message will not send their respective messages. In case that $\Delta_a \leq 0$, the current optimal solution to the RMP is also optimal to the MP, and no CU sends message to CBS.

Step 5: The CBS verifies the optimality of the current solution: if nothing is received from the CUs after a predefined period of time, the CBS concludes that $\Delta_a \leq 0$ and thus the CG method is terminated; otherwise, go to Step 6.

Step 6: For index $a = \arg\max_{i=1,\cdots,M}\{\Delta_i\}$, let $Q_a = Q_a + 1$ and generate the column

$$H_a^{Q_a} = \left[e_a, \sum_{k=1}^{K} y_{a1k}^{Q_a}, \cdots, \sum_{k=1}^{K} y_{aN_2k}^{Q_a}, \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ajk}^{Q_a} \cdot G_k \right]^T \tag{4.40}$$

with the solution to the a-th PP derived in Step 3, where e_a is a $1 \times M$ unit vector with the a-th entry being 1. Add the column and the corresponding variable $z_a^{Q_a}$ to the RMP and go to Step 2.

4.6 Upper Bound, Complexity and Time Efficiency

4.6.1 Upper Bound for the Master Problem

In the following, we derive an upper bound for the optimal objective value of the MP in each iteration of the CG method.

Theorem 4.2. *At each iteration, let Ω^T be the simplex multiplier vector of the RMP; $\Delta_a = \max_{i=1,\cdots,M}\{\Delta_i\}$ be the most positive reduced cost obtained from the PPs; $\vec{b}$ be a $(M + N_2 + 1) \times 1$ column vector with the i-th entry being the value of right hand side of the i-th constraint of the RMP, $i = 1, 2, \cdots, (M + N_2 + 1)$; $\vec{g}$ be a $(M + N_2 + 1) \times 1$ column vector as $g = (\underbrace{11\cdots1}_{M \text{ ones}} \quad \underbrace{00\cdots0}_{(N_2+1) \text{ zeros}} \quad)^T$. Then an upper bound for the MP can be derived as: $\overline{\Omega}^T\vec{b} = (\Omega^T + \Delta_a\vec{g})\vec{b}$.*

Proof. Let $\widehat{\Omega}^T$ be a feasible solution to the dual problem of the MP (termed DMP), according to the relationship between the dual and primal problems [155], we have

$$\widehat{\Omega}^T H_i^q \geq \varpi_i^q, \;\; i = 1, \cdots, M, \; q = 1, \cdots, Q_i, \tag{4.41}$$

where H_i^q is given in (4.40). As discussed, at each iteration we can obtain a simplex multiplier vector Ω^T by solving the RMP, as well as the most positive reduced cost by solving the PP.

$$\begin{aligned} &\Delta_a = \max_{i,q}\{\varpi_i^q - \Omega^T H_i^q\} \\ \Rightarrow\; &\Omega^T H_i^q \geq \varpi_i^q - \Delta_a, i = 1, \cdots, M, q = 1, \cdots, Q_i, \end{aligned} \tag{4.42}$$

where $\Delta_a > 0$. Denote $\overline{\Omega}^T = (\Omega^T + \Delta_a\vec{g})$ and multiply its both sides by H_i^q. We have

$$\begin{aligned} &\overline{\Omega}^T H_i^q = \Omega^T H_i^q + \Delta_a\vec{g}H_i^q \Rightarrow \overline{\Omega}^T H_i^q = \Omega^T H_i^q + \Delta_a \cdot 1 \\ \Rightarrow\; &\overline{\Omega}^T H_i^q - \Delta_a = \Omega^T H_i^q \Rightarrow \overline{\Omega}^T H_i^q - \Delta_a \geq \varpi_i^q - \Delta_a \\ \Rightarrow\; &\overline{\Omega}^T H_i^q \geq \varpi_i^q, \; i = 1, \cdots, M, \; q = 1, \cdots, Q_i. \end{aligned}$$

The first inequality is from (4.42). This means that $\overline{\Omega}^T$ is a feasible solution to the DMP. By duality, the corresponding dual LP of a maximization LP is a minimization LP [155]. So the DMP is a minimization LP, and $\overline{\Omega}^T$ is a feasible solution to the DMP.

Let the optimal solution to DMP be $\underline{\Omega}^T$. It follows that

$$\overline{\Omega}^T \vec{b} \geq \underline{\Omega}^T \vec{b} = \Upsilon^*, \tag{4.43}$$

where Υ^* is the optimal objective value of the minimization LP DMP. Due to *Strong Duality*, Υ^* is also the optimal objective value of the MP. It follows that $\overline{\Omega}^T \vec{b} = (\Omega^T + \Delta_a \vec{g})\vec{b}$ is an upper bound for the MP according to (4.43). □

4.6.2 Complexity and Optimality Analysis

In the general problems solved by the CG method, the INP and PP problems are at least as hard as the one dimensional 0-1 Knapsack problem, which is NP-hard [87].

However, an interesting characteristic of the INP and PP in our case is that the coefficients of the constraint matrix in the INP and PP are either 0 or 1, such that the *unimodularity property* [86] is satisfied in both problems. As a result, both the INP and PP have the optimal solution with their LP relaxations, and thus they can be solved with the Simplex method [87, 155]. Again, the upper bound of 1 on y_{ijk} can be removed as in P0.

Lemma 4.2. *The INP and PP are indeed LPs and thus can be solved with the Simplex method with a polynomial-time average-case complexity.*

Although the INP and PP can be solved with the Simplex method which has a average complexity of polynomial-time, it is possible that the Simplex method will require exponential time in extreme cases. Here we introduce a Greedy algorithm which solves the INP and PP in strongly polynomial time and still gets the optimal solution, to reduce the time complexity of CDOA.

The basic idea of our Greedy Algorithm is that, for the INP or PP of each CU, the decision variables corresponding to the combinations of channel and power level that having the highest ”utilities” among all possible combinations and still satisfy the constraints of the INP or PP, take value of 1, while the other decision variables take value of 0. In this way a feasible and optimal solution is guaranteed.

Note that Line 1 to Line 5 of Greedy Algorithm executes in Step 1 of CDOA to solve the INP and thus executes only once, in order to further reduce the computation complexity of our Greedy Algorithm, while Line 6 to Line 36 executes in Step 3 of CDOA to solve the PP and thus may executes multiple times during a whole operation of CDOA.

Theorem 4.3. *The Greedy Algorithm yields the optimal solution for the INP and PP.*

Proof. We can see from constraint (4.38) that each channel can be chosen at most once with one power level on it, and for each combination of channel and power level, there is an associated utility ξ_{ijk}.

Therefore, in order to maximize the total utility:

1. For each channel, we choose the combination of channel and power level having the greatest utility among all the K combinations;

2. If a combination has a negative utility, say, ξ_{ijk}, then the combination should not be used, i.e., $y_{ijk} = 0$ in the optimal solution. So the chosen combination should have a positive utility;

3. From constraint (4.35) we know that if the resulting SNR of a combination, say δ_{ijk}, is less than the threshold, then the combination should not be used. So the chosen combination should have a resulting SNR greater than the threshold.

We form a set using the combinations where each combination must satisfy the above three conditions.

From constraint (4.37) we know that each CU can use at most C_i channels. Therefor, from the above set, we choose ρ_i combinations having the greatest utilities among all the combinations, where $\rho_i = C_i$ if the number of elements in this set is greater or equal to C_i, and $\rho_i < C_i$ otherwise.

In this algorithm, all the constraints are used and thus the solution is feasible, and from the discussion we know that the solution is also optimal.

□

For the greedy algorithm that solves the INP at optimality, we just need to replace $\xi_{ijk} = w_{ijk} - \mu^j - \varphi G_k$ with $\xi_{ijk} = w_{ijk}$ at Line 6 of the above algorithm.

Lemma 4.3. *The Greedy Algorithm has a polynomial computational complexity of* $\mathcal{O}(N_2^2 + N_2 K)$.

Proof. From the algorithm we can see that Line 1 to Line 5 has a complexity of $N_2 K$, line 6 to line 15 has a complexity of $2N_2 K$, line 16 to line 19 has a complexity of $N_2(K+1)$, line 20 has a complexity of 1, line 21 to line 25 has a complexity less or equal to $2N_2$, line 26 has a complexity of $\mathcal{O}(\rho_i^2)$ which is less or equal to $\mathcal{O}(N_2^2)$, line 27 to line 36 has a complexity of $1 + 2\min\{\rho_i, C_i\}$ which is less or equal to $1 + 2N_2$. Summing up all these together, we conclude that the complexity of the algorithm is $\mathcal{O}(N_2^2 + N_2 K)$. □

4.6.3 Time Efficiency

Finally we analyze the time efficiency of CDOA. We compare the time needed to solve the MP with CODA (with a distributed parallel execution), t_1, and that without using CODA (with a centralized sequential execution), t_2.

Let the number of iterations of CDOA be L. Let τ_{ml} be the amount of time to process the RMP, τ_{pl} the amount of time to process a PP, τ_b the time for the CBS to broadcast the simplex multiplier to the CUs, τ_{rl} the time for CU a (where $a = \max_{i=1,\cdots,M}\{\Delta_i\}$) to report its solution to the a-th PP to the CBS, all in the l-th iteration of CDOA. Also assume that the time to broadcast the simplex multiplier to the CUs by the CBS is negligible. The *time saving* $s = t_2 - t_1$ achieved by the distributed, parallel execution can be approximated as

$$\begin{aligned} s &= \sum_{l=1}^{L}(\tau_{ml} + M \cdot \tau_{pl}) - \sum_{l=1}^{L}(\tau_{ml} + \tau_{pl} + \tau_b + \tau_{rl}) \\ &= \sum_{l=1}^{L}((M-1) \cdot \tau_{pl} - \tau_b - \tau_{rl}). \end{aligned} \tag{4.44}$$

Since the size of the PP does not change during each iteration, we assume that $\tau_{pl} = \tau_p$. Then (4.44) can be rewritten as

$$s = \tau_p \sum_{l=1}^{L} \left((M-1) - \frac{\tau_b + \tau_{rl}}{\tau_p} \right). \tag{4.45}$$

In (4.45), τ_b and τ_{rl} are usually negligible compared with τ_p for even a modest problem having hundreds or thousands of constraints and variables. When the number of simplex iterations is proportional to the number of constraints, the overall cost of the simplex method is $\mathcal{O}(m^4 + nm^2)$ arithmetic operations for a problem having m constraints and n variables [87]. Therefore, s can be further approximated by

$$s = \mathcal{O}\left(LM(N_2^4 + KN_2^3)\right). \tag{4.46}$$

It can be seen that s is an increasing function of L, M, N_2, and K, which represent the size of the problem. Such improvement in time efficiency demonstrate the advantages of the distributed CG-based algorithm.

4.7 Simulation Study

4.7.1 Simulation Setup

In this section, Matlab simulation results are used to demonstrate the performance of the proposed algorithms. Unless specified, the value of simulation parameters are as shown in Table 6.1. Each simulated point in the figures is obtained by repeating the simulation $r_p = 50$ times with different random seeds, and 95% confidence intervals are computed and plotted in the figures to guarantee credible results.

Table 4.1: Simulation Parameters

Parameter	*Value*	*Parameter*	*Value*
M	30	$\bar{\gamma}$	-25 dB
N_1	30	Λ_j	3
K	10	r_p	50
κ	10^4	Θ_i	3
d_{ij}	$-15 \sim -9$ dB	γ_{ij}	$-100 \sim 0$ dB
C_i	3	G_{total}	50
G_k	$10^{-\frac{k}{10}}$	P_{req}	0.99
f_s	10^6 H_z	$\min_j \{P(r_j=0)\}$	0.2
n_0	10^{-6}	$\max_j \{P(r_j=0)\}$	0.9
B_j	10^6 H_z		

4.7.2 Simulation Results and Analysis

As a basis for our simulations and discussions in the following, Fig. 4.2 plots the relationship between Y-PSNR and data rate according to (4.4) for three widely used test video sequences of different content types.

The three videos are in uncompressed YUV4MPEG format. They represent three levels of motions: slow (Suzie), medium (Carphone), and quick (Football). Videos of different motion levels are often used as tested sequences to demonstrate the performance of proposed algorithms [71, 79]. These and other video sequences can be downloaded from $https://media.xiph.org/video/derf/$

In the figure, the markers are obtained by experiment with the real video sequences, and the lines are obtained by linear regression. Video QoS parameters α_i and β_i for the three sequences are calculated based on their respective linear regression plots.

In Figs. 4.3 and 4.4 we compare the performance of the optimal sensing strategy and that of the Heuristic algorithm, in terms of sensing performance and the resulting overall Y-PSNR of the received videos, respectively, We consider four cases that the number of sensors a CU has in the Heuristic algorithm Θ_i = 3, 4, 5, and 6, where $\Theta_i \ll N_1 = 30$. Note that $M \cdot \Theta_i \geq N_1 \cdot \Lambda_j$ is a necessary condition to have any channel sensed by at lest Λ_j sensors. In Fig. 4.3, the legend 'Idle channels' means the number of channels cooperatively sensed idle; 'Missed channels' means

Algorithm 6: Greedy Algorithm

for $j = 1 : N_2$ **do**
  for $k = 1 : k$ **do**
    Compute $\delta_{ijk} = \frac{G_k d_{ij}}{n_0 B_j}$;
  end
end
for $j = 1 : N_2$ **do**
  for $k = 1 : K$ **do**
    if $\delta_{ijk} < \bar{\gamma}$ **then**
      Set $\xi_{ijk} = \sigma$, where $\sigma < 0$ and is a fixed constant;
      else
        Compute $\xi_{ijk} = w_{ijk} - \mu^j - \varphi G_k$;
      end
    end
  end
end
for $j = 1 : N_2$ **do**
  Find k' such that $\xi_{ijk'} \geq \xi_{ijk}, \forall k \neq k'$;
  Denote $\widehat{\xi}_{ij} = \xi_{ijk'}$;
end
Let $\rho_i = 0$;
for $j = 1 : N_2$ **do**
  if $\widehat{\xi}_{ij} > 0$ **then**
    $\rho_i = \rho_i + 1$;
  end
end
Sort the ρ_i channels in the decreasing order of $\widehat{\xi}_{ij}$ and let the sorted channel set be $\mathcal{A}$;
if $\rho_i \leq C_i$ **then**
  for $j = 1 : \rho_i$ **do**
    Set $y_{i\mathcal{A}(j)k'} = 1$;
  end
  else
    for $j = 1 : C_i$ **do**
      Set $y_{i\mathcal{A}(j)k'} = 1$;
    end
  end
end

the number of channels cooperatively sensed idle while these channels are actually busy, so the number of channels that are sensed idle and are actually idle is the difference between the two.

From these two figures we can see that the heuristic algorithm achieves a performance close to the optimal sensing strategy when $\Theta_i \geq 3$.

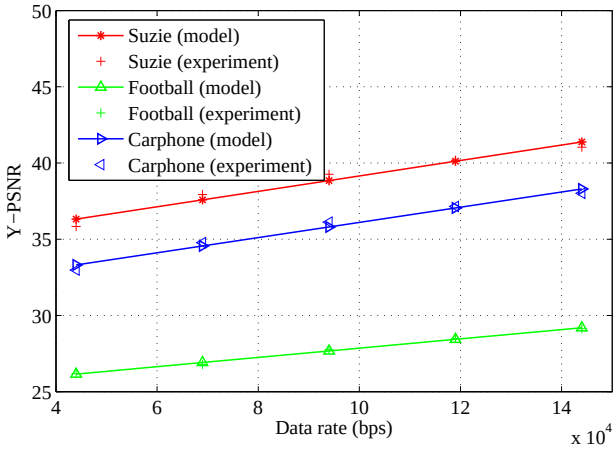

Figure 4.2: Y-PSNR versus data rate for three reference video sequences: model versus simulation.

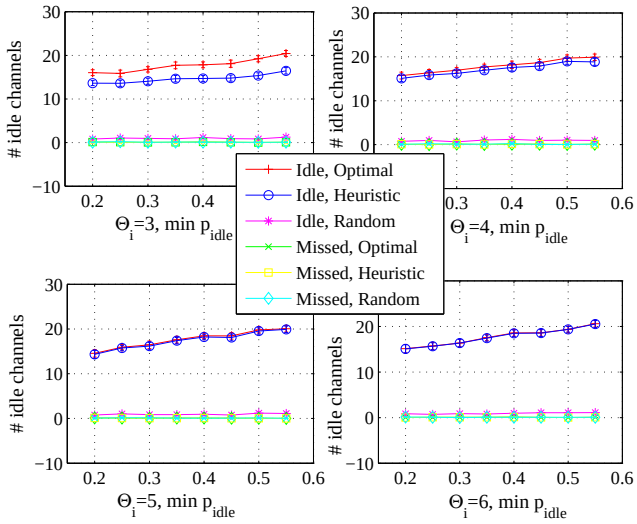

Figure 4.3: Optimal versus Heuristic Sensing in terms of sensing performance for $\Theta = 3, 4, 5, 6$.

In this case, we have $M \cdot \Theta_i = N_1 \cdot \Lambda_j$. A channel j, may not have 'good' CUs (i.e., with a relative small $P_{f_{ij}}$) to sense it, since these CUs may have already been assigned to another channel j', which has a higher $P(r_{j'} = 0)$ and thus has a higher priority of being optimized. Thus channel j may be discarded due to insufficient CUs to sense it, resulting in a lower value of SP2 due to fewer available channels for video streaming.

Recall that Θ is the number of channels that a CU can sense at a time slot. Once a CU is assigned to sense Θ channels, it is deleted from the remaining CUs list. Therefore, as Θ is increased, which means that a CU can sense more channels at a time slot, it is more likely that

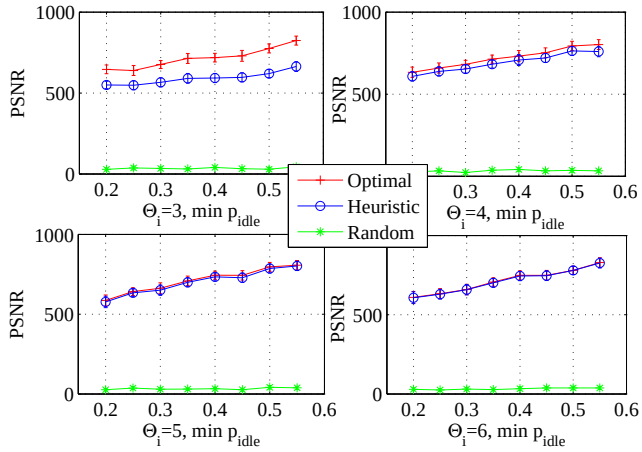

Figure 4.4: Optimal versus Heuristic Sensing in terms of overall Y-PSNR in dB for $\Theta = 3, 4, 5, 6$.

each channel can be sensed by the best CUs. Thus the sensing performance of each channel will be improved. We found that when Θ is selected from 3 to 6, the result is sufficiently good.

However, as the last three sub-figures shows, the Heuristic algorithm achieves almost the same performance as the optimal sensing strategy when Θ_i = 4, 5, and 6. Thus even the channels having a lower priority will have a higher chance to be sensed by 'good' CUs. Then P_{f_j}, $j = 1, 2, \ldots, N_2$ is more likely to be reduced, and the objective function value is improved.

Note that both the heuristic and the optimal sensing have the same computational complexity and workload for the 'good' cognitive users, since in both algorithms each channel will be sensed by the CUs with the best sensing performance. The only difference is that in the optimal sensing strategy, the sensing of each channel is optimized at the same time; while in the heuristic algorithm, the sensing of each channel is optimized sequentially, in the order of decreasing priority, which is determined by its probability of being idle.

Besides, we also compare the sensing performance and overall Y-PSNR performance of our proposed Heuristic algorithm with a Benchmark Algorithm called the Random Algorithm, as the authors in [85] did, which randomly assigns Λ_j CUs to sense channel j, $\forall j$. We can see that the Heuristic algorithm outperforms the Random Algorithm significantly, the main reason for which is that the Random Algorithm doesn't consider the sensing accuracy of different CUs for different channels. It is very likely that for a particular channel, the Random Algorithm will assign CUs

having a high false alarm probability to this channel to sense it, which results in a great probability of this channel being false alarmed. That's why the number of idle channels found by the Random Algorithm is very small, although the number of missed detected channels of the Random Algorithm is close with that of the Heuristic Algorithm and the optimal sensing strategy. Under such sensing performance, it is within our expectation that the overall Y-PSNR performance of the Random Algorithm falls behind that of the Heuristic Algorithm drastically. Therefore, we claim that Heuristic Algorithm provides a much better performance than Random Algorithm does.

Fig. 4.5 demonstrates the convergence of the CG method in two cases: $M = 9, N_1 = 18$ (the upper figure) and $M = 15, N_1 = 30$ (the lower figure). We set $C_i = 2$, for all $i = 1, 2, \ldots, M$ in both cases.

We have the following observations. (i) The number of iterations is positively correlated to the problem size, since as the number of CUs and channels grows, there may be more feasible schemes to improves the current objective value at a specific iteration. (ii) The most positive reduced cost $\max_i \{\Delta_i\}$ tends to decrease over iterations. This trend is the result of the greedy approach of the CG algorithm, which means that the algorithm chooses the feasible scheme having the most positive reduced cost (thus possibly having the greatest potential to improve the current objective value of the MP) from the remaining candidate feasible schemes, to enter the RMP at each iteration. (iii) The increment of the objective function at a specific iteration is positive correlated to $\max_i \{\Delta_i\}$. This follows directly from the above discussions. (iv) The upper bound to the optimal objective function value converges quickly, and is also positively correlated to $\max_i \{\Delta_i\}$. From Theorem 4.2, the upper bound at a certain iteration $\overline{\Omega}^T b$ is actually a positive function of $\max_i \{\Delta_i\}$ (Δ_a in Theorem 4.2) at this iteration. Since $\max_i \{\Delta_i\}$ drops quickly and become very close to 0, $\overline{\Omega}^T b$ drops and then converges to the optimal objective function value quickly. Note that $\overline{\Omega}^T b$ is not necessary decreasing as the iteration goes. There are two main reasons: (a) $\max_i \{\Delta_i\}$ is not necessary decreasing as the iteration goes although it shows a trend of decreasing. (b) $\overline{\Omega}^T b$ also depends on the simplex multiplier Ω_T, whose convergence is hard to analysis.

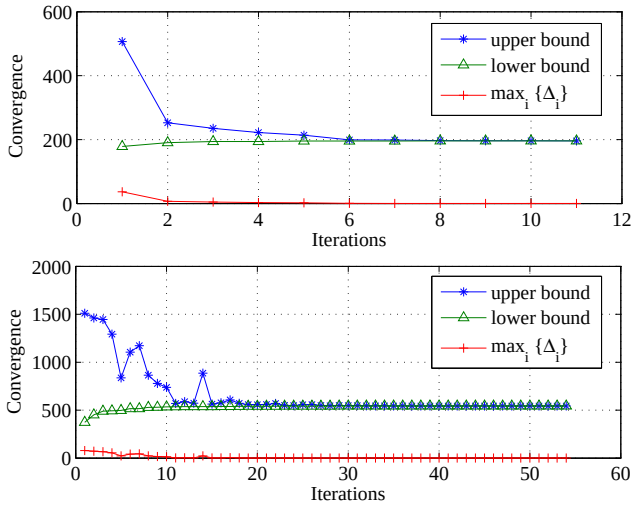

Figure 4.5: Convergence performance of Column Generation method

The convergence of the Y-PSNR of each type of received video under two cases: $M = 9, N_1 = 18$ (the upper figure) and $M = 15, N_1 = 30$ (the lower figure), where $C_i = 2$, for all $i = 1, 2, \ldots, M$ in both cases, is shown in Fig. 4.6.

It is observed that for the type 3 video, the Y-PSNR shows a tendency of increasing as the iteration goes. For the type 1 video, the Y-PSNR shows a trend of decreasing, especially in the lower plot in Fig. 4.6. The Y-PSNR of type 2 video is decreased in the upper plot and doesn't show much change in the lower plot. However, the overall Y-PSNR of the three types of videos increases along with the iterations. The main reason is that to improve the overall Y-PSNR, at each iteration, the CG algorithm preferentially assigns channel and power resources to the type of video which may have the greatest potential for improving the overall Y-PSNR. Recall that under the same data rate, the Y-PSNR of type 3 video is the greatest and the Y-PSNR of type 1 video is the smallest among the three types. Thus on the condition that the number of available channels is limited and is less than the number of video sessions, channel and power resources will be taken from what have been allocated to type 1 videos and then be assigned to type 3 videos first, then to type 2 video. Besides, the magnitude of changes of the three types of videos decreases quickly over iterations, which is in consistence with the change of the overall Y-PSNR, as shown in Fig. 4.5.

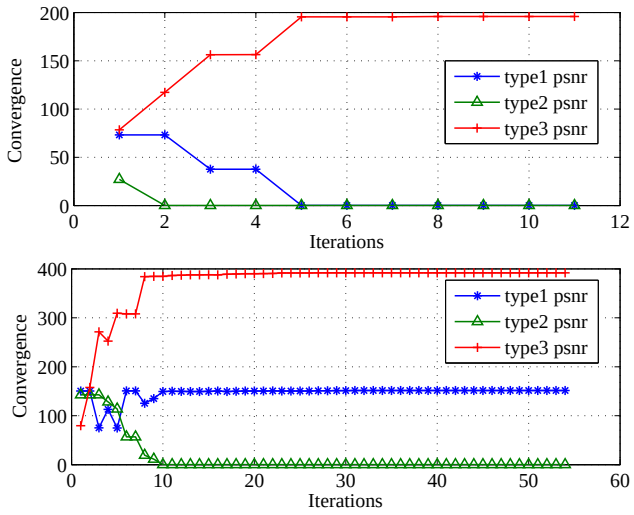

Figure 4.6: PSNR comparison for three types of videos achieved by Column Generation method.

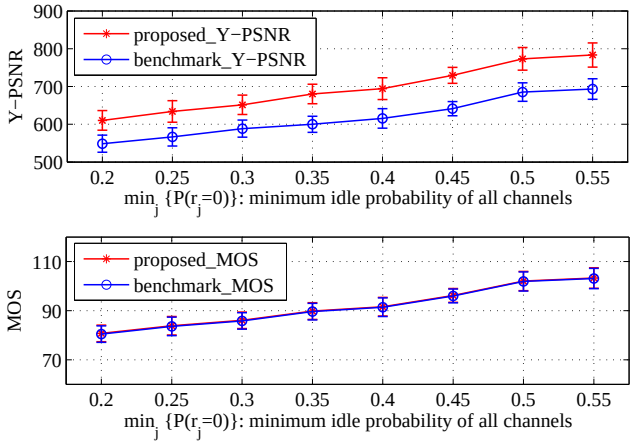

Figure 4.7: QoS and QoE comparison between the proposed scheme and the benchmark.

In Fig. 4.7, we compare the proposed algorithm with a benchmark scheme [79], with respect to both QoS (i.e., Y-PSNR) and QoE of the received videos. In the benchmark scheme, an idle channel, say channel j, with a higher $P(r_j = 0)$ is assigned to a CU, say CU i, that is less delay-tolerant and thus have a higher priority to use the available channels. In our simulation we randomly assign an integer priority level ranging from 1 to 3 to each CU. We use the MOS (Mean Opinion Score) to evaluate QoE and the MOS model is adopted from [71] as the benchmark scheme does [79].

It can be seen that when $\min_j\{P(r_j = 0)\} \geq 0.35$, the Y-PSNR achieved by the benchmark scheme is about 100 dB lower than that achieved by the proposed scheme, which means that each CU with our proposed scheme has an average Y-PSNR of 3 dB higher than that with the benchmark scheme. Both curves tends to increase as $\min_j \{P(r_j = 0)\}$ increases, since more channels are likely to be available for video transmission. Within the entire range of $\min_j \{P(r_j = 0)\}$, our proposed algorithm outperforms the benchmark scheme in terms of Y-PSNR, because the solution to the problem in [79] is not necessary optimal to our problem and thus yields a lower bound to our problem. Furthermore, we find that the overall MOS of the two schemes are almost the same. Thus our proposed algorithm achieves a considerably higher QoS and a comparable QoE performance as the benchmark scheme.

4.8 Conclusion

In this chapter, we investigated the problem of QoS-driven multi-user video streaming over cellular CRNs. We showed that there exists a *decomposition principle* in the optimal joint design of spectrum sensing, channel assignment, and power allocation that circumvents the *curse of dimensionality* in general MINLPs. The decomposed spectrum sensing problem was solved with an optimal algorithm, along with a heuristic algorithm that is much less demanding on CUs' hardware. A CG-based decentralized channel assignment and power allocation algorithm was next developed to relieve the computation burden on the CBS. And a Greedy Algorithm which solves the sub-problems generated during the CG based algorithm at optimality in polynomial time is proposed to reduce the computational complexity. We analyzed the complexity and time efficiency, and derive an upper bound for the CG-based algorithm, and validated its performance with simulations.

Chapter 5

A Decomposition Principle for Link and Relay Selection in Dual-hop mm Wave Networks

5.1 Introduction

mm Wave Millimeter wave (mmWave) communications has become a research hot spot recently. There is up to 7 GHz license-free spectrum in this band that is available in many countries, making mm Wave communications and networks a promising technique to meet the wireless data challenge, as well as a core technology for future 5G Wireless systems [112]. What's more, the authors of [113] propose to augment wired hybrid data center networks with highly directional 60GHz wireless links to provide flexible network connectivity, which reveals the great potential of 60GHz technology in other applications. However, to make mm Wave mmWave networks applicable, many research challenges should be addressed. The wireless signal attenuation in mm Wave channels is much serious than that in the 5 GHz or 2.5 GHz channels [148], making beamforming indispensable. The authors in [146] show that the highly directional links, especially in the outdoor environment, can be regarded as *pseudo-wired* with negligible collision probabilities. This model has been adopted in many works on mm Wave networks [112, 116–119].

Furthermore, mmWave signals usually do not diffract around or penetrate obstacles. A Line-Of-Sight (LOS) path between the transmitter and receiver is required for a successful transmission. However, in practical networks, an LOS path may not always exist, and it is possible that an LOS path is blocked (e.g., by a pedestrian or car) from time to time. First, relay nodes should be used to forward data for a distant or blocked receiver [120], by setting up an LOS path between the transmitter and relay, and that between the relay and receiver. Second, the blockage between two nodes may appear or disappear dynamically due to the movement of objects or the nodes themselves [150, 151]. A realistic mm Wave network protocol should consider the use of relay nodes and model the dynamic blockages of mm Wave links [50].

As a unique feature of mmWave communications, network connectivity can be enhanced by exploiting reflections from walls and other surfaces to steer around obstacles [123]. The authors in [124] use static reflectors to maintain the coverage of mm Wave networks. When the LOS path is blocked, the authors in [125] suggest to switch the beam path from an LOS link to a Non-Line-Of-Sight (NLOS) link. Although using reflections will cause additional power loss and reduce power efficiency, it offers additional choices for increasing network coverage and improving network throughput.

In this chapter we investigate the scheduling problem in a centralized dual-hop mm Wave network. The network here we consider consists of one PNC, which is the central coordinator, multiple SD pairs, and multiple relays. When a source and destination are unable to directly communicate with each other (e.g., out of range of each other, or permanently blocked by an obstacle), a relay will be used to forward their traffic. There are multiple links, including both the LOS link and NLOS links (e.g., reflected from a wall), from a source to a relay and a relay to a destination. We adopt a two-state Markov chain model to capture the dynamic blockage behavior of mm Wave links. At each time slot, the PNC decides the link and relay selection for each SD pair to minimize the Maximum Expected Delivery Time (MEDT) among all SD pairs, by jointly optimizing relay and link selection, while exploiting reflected mmWave transmissions and considering link blockage dynamics. We develop a nonlinear integer programming formulation of the link and relay selection problem, and then develop effective algorithms that can provide highly competitive solutions. Specifically, we develop a Decomposition Principle to transform this problem into two sub-problems, one for link selection and the other for relay assignment when there is enough replays. We prove that the proposed scheme can achieve an optimality gap of just 1 time slot at greatly reduced complexity. We also develop a heuristic scheme to handle the case when there is no enough relays. The proposed schemes are validated with simulations with their superior performance observed.

- Unlike prior works on relay selection in mm Wave networks [118,119,126–129], we consider the LOS and multiple reflected NLOS links between source, relay, and destination nodes, as

well as link blockage dynamics in our formulation, and provide a rigorous analysis of the joint link and relay selection problem.

- We develop a Decomposition Principle to break down the formulated NIP problem into a link selection sub-problem and a relay selection sub-problem. We prove that the two sub-problems together provides a sub-optimal solution to the NIP problem with greatly reduced complexity, and more important, the optimality gap is bounded by only 1 time slot, if there is a sufficient number of relays.

- When there is no enough relays, we propose a heuristic algorithm that can still achieve highly competitive solutions at a low complexity.

- We validate the proposed algorithms with extensive simulations and comparison with two scheduling algorithms for mm Wave networks. We find both the proposed Decomposition Principle and heuristic scheme outperform the two benchmarks in all the cases that we simulate, with respect to delay, MEDT, throughput, and fairness.

In the rest of this chapter, related work is reviewed in Section 6.2, the system model is presented in Section 6.3 and the problem formulation in Section 5.4. We develop the Decomposition Principle and the heuristic algorithm in Section 5.5 and evaluate their performance in Section 5.6. and Section 5.7 concludes the chapter.

5.2 Related Work

There have been some interesting work on link scheduling in mm Wave networks. The authors in [131] propose a Partially Observable Markov Decision Process (POMDP) framework to model the link status in mm Wave networks, and a greedy scheduling strategy that aims to maximize the instant throughput at each time slot. However, this strategy is only applicable for single-hop centralized networks, and the multiple potential links between a node pair is not explored in this chapter. A similar problem is studied in [129] with a special scenario of a single-transmitter.

To improve network throughput, the authors in [126] propose a fast relay selection algorithm to reduce the overhead of relay selection time, so that there will be more time for data transmission. The basic idea is to determine the sectors where the best relay may be located, and then find the best relays in the selected sector. However, the authors do not consider coordinating concurrent transmissions of multiple transmitters. It is possible that different transmitters may select the same relay and thus collision happens.

In [119], the authors consider the fact that different relays may have different path losses, and thus having different outage probabilities. A relay selection scheme is proposed to minimize the outage probability for a single transmitter. In both indoor and outdoor environments, the obstacles may change over time (e.g., pedestrians move) and thus the blockage of a mm Wave link is actually not static. Such dynamic channel condition is not considered in [119]. A network throughput maximization problem for a dual-hop network is studied in [118], where different relays may provide different capacities for a SD pair. Relays assignment for multiple SD pairs is optimized to maximize the network throughput. The path loss and blockage model considered in this work are also time-invariant, and thus the proposed algorithm may not be suitable for mm Wave networks with dynamic link conditions. To overcome the problem of link breakage and degradation in point-to-point 60GHz networks, the authors of [136] propose to use *repeaters* to provide alternate paths when the direct path between transmitter and receiver degrades. It is assumed that the nodes and repeaters can beamform in any direction and thus by tuning the transmitting and receiving antenna to the repeater, a new link between the transmitter and receiver can be established. However, it is worthy of noting that in complex environments, the placement and selection of repeaters is a non-trivial problem. In [137], to reinforce transmission efficiency and also reduce power consumption of 60GHz devices, the authors propose a fast beam-switching scheme, which employs an efficient beamforming training algorithm based on the direct numerical search so that only a small portion of beam-pairs will be sequentially tested while most other beam-pairs will never be probed, so that the search complexity can be significantly reduced.

In [112], the authors propose a heuristic scheduling scheme for given traffic demands under static channel conditions, aiming to minimize the time needed to clear all the traffic demands. The pseudo-wired mm Wave channel model is adopted in this work. The authors in [127] study the relationship between the collision probability of two concurrent transmissions on two links and the link distances. It is found that the collision probability is an increasing function of link distance. Based on this finding, the authors propose a hop selection metric based on link distance, to reduce the collision probability of concurrent transmissions. By replacing a single long hop with multiple short hops, the proposed scheme can improve the number of concurrent transmission flows while constraining the harmful interference below an acceptable level. However, the algorithm is heuristic and lacks consideration of multiple coexisting links. The time slot allocation problem in multi-hop mm Wave networks is investigated in [128], where the direct path shares time slots with the relay path. Different time slot allocation schemes may result in different system throughput, and the effective system throughput is optimized with time slot allocation. A sub-optimal solution is proposed to solve the formulated NP-hard problem. Besides time slot allocation, channel allocation is also significant to the improvement of network throughput. The authors of [139] investigate the problem of channel allocation in 60 GHz indoor WLANs in order to maximize throughput, and two SDMA (Spatial Division Multiple Access) algorithms are proposed, for the single-channel case and the multiple-channel case respectively, to exploit the peculiar propagation properties so that data rates to end users can be improved.

There are also some papers on designing MAC protocols for 60GHz wireless networks and performance analysis of MAC protocols. In [135], the authors claim that conventional directional CSMA/CA protocols do not work well at 60GHz networks due to the impaired carrier sensing at the transmitters. To overcome this difficulty, the authors propose a novel protocol which adopts virtual carrier sensing instead of physical carrier sensing, and relies on a central coordinator to distribute network allocation vector (NAV) information. The authors of [138] present an analytical model for computing the saturation throughput of a Medium- Transparent MAC protocol in 60GHz radio-over-fiber networks. Both of the contention at the optical and the wireless layer are considered. The

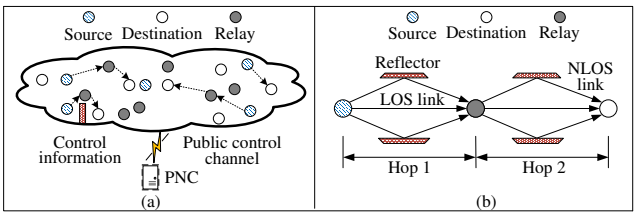

Figure 5.1: (a) Network model. (b) Two-hop relay path model.

authors derive the saturation throughput performance of the Medium Transparent MAC protocol under various scenarios. To provide a more comprehensive performance analysis of the Medium Transparent MAC protocol for 60GHz radio-over-fiber networks, the authors of [140] analysis the delay fairness performance of the Medium Transparent MAC protocol, and it is shown that delay equalization can be achieved even for highly varying user population patterns among the different antenna units when certain wavelength availability conditions are satisfied.

5.3 System Model

As shown in Fig. 5.1(a), we consider a centralized dual-hop 60GHz network consisting of multiple nodes and one PNC. Each node can be either a source node (S), a destination node (D), or a potential relay node (R). When the source and destination nodes are unable to directly communicate with each other (e.g., permanently blocked by an obstacle/wall, or out of range), a relay is used to forward their traffic. Due to point-to-point mm Wave links (unlike traditional broadcast-based relay networks), we assume that each SD pair can choose only one relay at a time. However, a relay may serve multiple SD pairs at different time slots (but not at the same time slot).

To overcome the deafness problem, which makes it highly challenging for coordination of the highly directional links, we assume a lower frequency public control channel (e.g., a WiFi channel) for all nodes and the PNC [130]. Due to the omnidirectional transmissions, better propagation, and larger coverage, the nodes on the control channel actually form a single-hop network. Network state and control information of the dual-hop mm Wave network can be effectively exchanged among the nodes on the control channel [118, 119, 130, 131], including the following at each time slot.

- The PNC collects network state information, such as traffic requests, link status, etc., from each node
- The PNC broadcasts link transmission schedules to all the nodes if it makes a new schedule at this time slot
- Receiving nodes inform the PNC whether the reception is successful or not, so that the PNC knows the link status.

As shown in Fig. 5.1(b), there may be multiple links between a pair of nodes within range of each other: there may be one LOS link, as well as other NLOS links formed by exploiting reflections from walls and other surfaces [116]. Due to the moving obstacles, the state of a link is either blocked or unblocked at each time slot. We assume the link state follows a discrete-time Markov process [129, 131], while the nodes learn the transition probabilities of their links and inform the PNC these parameters. Note that if the LOS link is more likely to be blocked, an NLOS link may be a better choice. A successful transmission on a link requires the link being unblocked.

Without loss of generality, we assume each node is equipped with an electronically steerable antenna array to beamform in the transmitting or receiving directions; so each node works in the *half-duplex* mode [132]. Both transmission and reception are directional with a very narrow beamwidth. The beamforming weights learned when receiving from a given node can then be used to transmit back to that node, assuming *channel reciprocity*. Some commercial mm Wave products can have a beamwidth of 1.4° or even as small as 0.6°. A probabilistic analysis is presented in [116] on the interference caused by uncoordinated transmissions in such highly directional mm Wave networks. The analysis shows that "interference can essentially be ignored in the MAC design" and the links can be regarded as *pseudo-wired* [116]. We adopt such a pseudo-wired link model in this chapter, as in prior works [116–119].

A relay can be in one of the three states at each time slot: idle, transmitting, or receiving. If a relay is selected for a source, it receive from the source in the first hop. Once finishing the reception, the relay transmit the received packet to the specified destination in the second hop.

Until the packet is successfully transmitted to the destination, the relay cannot receive more data from this or other sources due to the half-duplex operation. If a relay is not selected for any source, it stays in the idle state. This model is illustrated in Fig. 5.1(b) as in prior works [118, 119, 133].

5.4 Problem Formulation

5.4.1 Dynamic Link Blockage Model

For a link l, denote C_l^t as the event that link l is unblocked at time slot t, and $\bar{C}_l^t$ the opposite. Recall that link state follows a discrete-time two-state Markov chain. Let $0 \leq p_l, q_l \leq 1$ be the one-step transition probability from blocked to unblocked, and from unblocked to blocked, respectively. The one-step transition probability matrix of link l is

$$\mathbf{P}_l(1) = \begin{pmatrix} P(\bar{C}_l^{t+1}|\bar{C}_l^t) & P(C_l^{t+1}|\bar{C}_l^t) \\ P(\bar{C}_l^{t+1}|C_l^t) & P(C_l^{t+1}|C_l^t) \end{pmatrix} = \begin{pmatrix} 1-p_l & p_l \\ q_l & 1-q_l \end{pmatrix}.$$

The n-step transition probability matrix of link l is [134]

$$\begin{aligned} \mathbf{P}_l(n) &= \begin{pmatrix} P(\bar{C}_l^{t+n}|\bar{C}_l^t) & P(C_l^{t+n}|\bar{C}_l^t) \\ P(C_l^{t+n}|C_l^t) & P(\bar{C}_l^{t+n}|C_l^t) \end{pmatrix} \\ &= \frac{1}{p_l+q_l}\begin{pmatrix} q_l & p_l \\ q_l & p_l \end{pmatrix} + \frac{(1-p_l-q_l)^n}{p_l+q_l}\begin{pmatrix} p_l & -p_l \\ -q_l & q_l \end{pmatrix} \\ &= \begin{pmatrix} 1-p_l(n) & p_l(n) \\ q_l(n) & 1-q_l(n) \end{pmatrix}, \quad n = 1, 2, \ldots. \end{aligned} \tag{5.1}$$

5.4.2 Expected Delivery Time (EDT)

We consider two types of SD pairs. The first type, denoted as $\mathcal{S}_i$, is that the source and destination are within one-hop distance with each other and are not permanently blocked (e.g., by a wall). Hence the SD pair can either communication with each other directly, or use a relay if the

direct link is poor. The second type, denoted as $\mathcal{S}_j$, is that the SD pair are either out of range or blocked by a permanent obstacle between them. Thus a relay is needed for them to communicate with each other. Define $\mathcal{S}_{i\cup j} = \mathcal{S}_i \cup \mathcal{S}_j$. We next derive the expected delivery time (EDT) for the relay-assisted and direct transmission cases.

EDT via Relay

Let s denote a source with destination $d(s)$, and r be a relay that can communicate directly with both s and $d(s)$. Denote a link between s and r as l_{sr}, and the set of all s–r links as $\mathcal{L}_{sr}$. Similarly, we define $l_{rd(s)}$ as a link between r and $d(s)$ and $\mathcal{L}_{rd(s)}$ as the set of these links.

Let $T_{l_{sr}}$ be the delivery time from s to r when link l_{sr} is used in the first hop for a block of data no greater than the channel capacity (normalized to a time slot), and T be the current time slot, $T \geq 1$. The expectation $\mathbb{E}(T_{l_{sr}})$ is the average number of trials until the first successful transmission happens on l_{sr}, which can be expressed as

$$\begin{aligned}\mathbb{E}(T_{l_{sr}}) &= \sum_{t=1}^{\infty} tP(T_{l_{sr}} = t) = 1 \cdot P(C_{l_{sr}}^T) + \\ &\sum_{t=2}^{\infty} tP(\bar{C}_{l_{sr}}^T)(1 - p_{l_{sr}})^{(t-2)} p_{l_{sr}} = 1 + \frac{1 - P(C_{l_{sr}}^T)}{p_{l_{sr}}},\end{aligned} \tag{5.2}$$

where $P(C_{l_{sr}}^T)$ is the probability that l_{sr} is unblocked at T.

Now let $T_{l_{sr},l_{rd(s)}}$ be the delivery time from r to $d(s)$ when link l_{sr} is chosen in the first hop and link $l_{rd(s)}$ is used in the second hop. To derive $\mathbb{E}(T_{l_{sr},l_{rd(s)}})$, we first note that

$$\begin{aligned}&\mathbb{E}(T_{l_{sr},l_{rd(s)}}|T_{l_{sr}} = t) \\ &= \sum_{t'=1}^{\infty} t' P(T_{l_{sr},l_{rd(s)}} = t'|T_{l_{sr}} = t) = 1 + \frac{1 - P(C_{l_{rd(s)}}^{T+t})}{p_{l_{rd(s)}}}.\end{aligned} \tag{5.3}$$

According to the *law of total expectation*, we have

$$\mathbb{E}(T_{l_{sr},l_{rd(s)}}) = \mathbb{E}(\mathbb{E}(T_{l_{sr},l_{rd(s)}}|T_{l_{sr}} = t))$$

$$\sum_{t=1}^{\infty} P\left(C_{l_{rd(s)}}^{T+t}\right) P(T_{l_{sr}}=t)$$
$$= \sum_{t=1}^{\infty} \left(P\left(\bar{C}_{l_{rd(s)}}^{T_{l_{rd(s)}}}\right) p_{l_{rd(s)}}(T+t-T_{l_{rd(s)}}) + P\left(C_{l_{rd(s)}}^{T_{l_{rd(s)}}}\right)(1-q_{l_{rd(s)}}(T+t-T_{l_{rd(s)}}))\right) P(T_{l_{sr}}=t)$$
$$= \frac{P(\bar{C}_{l_{sr}}^{T})}{p_{l_{rd}}+q_{l_{rd(s)}}} + \frac{P(\bar{C}_{l_{sr}}^{T})(p_{l_{rd(s)}}+q_{l_{rd(s)}})p_{l_{sr}}}{p_{l_{rd(s)}}+q_{l_{rd(s)}}} \cdot \frac{(1-p_{l_{sr}}-q_{l_{rd(s)}})^{T-T_{l_{rd(s)}}+2}}{(1-(1-p_{l_{rd(s)}}-q_{l_{rd(s)}})(1-p_{l_{sr}}))}. \quad (5.6)$$

$$= \sum_{t=1}^{\infty} \mathbb{E}(T_{l_{sr},l_{rd(s)}}|T_{l_{sr}}=t)P(T_{l_{sr}}=t)$$
$$= 1 + \frac{1}{p_{l_{rd(s)}}} - \frac{1}{p_{l_{rd(s)}}} \sum_{t=1}^{\infty} P(C_{l_{rd(s)}}^{T+t})P(T_{l_{sr}}=t). \quad (5.4)$$

To calculate $\mathbb{E}(T_{l_{sr}})$ in (5.2) and $\mathbb{E}(T_{l_{sr},l_{rd(s)}})$ in (5.4), we need to derive $P(C_{l_{sr}}^{T})$ and $P(C_{l_{rd(s)}}^{T+t})$. Let t_l be the last time (before the current time T) that PNC knew the state of link l (being either blocked or unblocked). We have

$$P(C_l^T) = P(\bar{C}_l^{t_l})p_l(T-t_l) + P(C_l^{t_l})(1-q_l(T-t_l))$$
$$= \frac{p_l}{p_l+q_l} + \frac{(1-p_l-q_l)^{T-t_l}}{p_l+q_l}\left(q_l + P(C_l^{t_l})(p_l+q_l)\right). \quad (5.5)$$

Furthermore, we derive the summation term in (5.4), which is given at the top of next page. Substituting (5.5) and (5.6) into (5.2) and (5.4), we thus derive the closed-form expression for the EDT when link l_{sr} and link $l_{rd(s)}$ are chosen for the two-hop relay path, denoted as $\mathbb{E}(T_{srd(s)})$, which is given by

$$\mathbb{E}(T_{srd(s)}) = \mathbb{E}(T_{l_{sr}}) + \mathbb{E}(T_{l_{sr},l_{rd(s)}}). \quad (5.7)$$

EDT via Direct link

Consider the case when s and $d(s)$ use a direct link $l_{sd(s)}$ between them to communicate without using a relay. The EDT from s to $d(s)$ via link $l_{sd(s)}$, denoted as $\mathbb{E}(T_{sd(s)})$, can be derived as $\mathbb{E}(T_{sd(s)}) = 1 + \frac{1-P(C^T_{l_{sd(s)}})}{p_{l_{sd(s)}}}$.

5.4.3 Problem Formulation

Let $\mathcal{R}(s)$ be the set of relays that can communicate directly with both source s and its destination $d(s)$, and $\mathcal{S}(r)$ be the set of sources that can communicate directly with relay r. Denote all the relays and sources as $\mathcal{R}$ and $\mathcal{S}_{i\cup j}$, respectively. Let $\mathcal{L}_{sd(s)}$ be the set of all s-$d(s)$ links. We then define the following decision variables.

$$x_{l_{sr}} = \begin{cases} 1, & \text{source } s \text{ transmits on link } l_{sr} \text{ in hop } 1 \\ 0, & \text{otherwise,} \end{cases}$$

$$\forall\ s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), l_{sr} \in \mathcal{L}_{sr} \tag{5.8}$$

$$x_{l_{rd(s)}} = \begin{cases} 1, & \text{relay } r \text{ transmits on link } l_{rd(s)} \text{ in hop } 2 \\ 0, & \text{otherwise} \end{cases}$$

$$\forall\ s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), l_{rd(s)} \in \mathcal{L}_{rd(s)} \tag{5.9}$$

$$x_{l_{sd(s)}} = \begin{cases} 1, & \text{source } s \text{ transmits to its destination } d(s) \\ & \text{via direct link } l_{sd(s)} \\ 0, & \text{otherwise} \end{cases}$$

$$\forall\ s \in \mathcal{S}_{i\cup j}, l_{sd(s)} \in \mathcal{L}_{sd(s)}. \tag{5.10}$$

Since each relay r can be selected by at most 1 SD pair and only one link can be selected at each hop, we have

$$\sum_{s\in\mathcal{S}(r)} \sum_{l_{sr}\in\mathcal{L}_{sr}} x_{l_{sr}} \leq 1, \forall r \in \mathcal{R}. \tag{5.11}$$

Note that if relay r is selected by s in the hop 1, then r must also be selected by $d(s)$ in hop 2, i.e.,

$$\sum_{l_{sr}\in\mathcal{L}_{sr}} x_{l_{sr}} = \sum_{l_{rd(s)}\in\mathcal{L}_{rd(s)}} x_{l_{rd(s)}}, \forall s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s). \tag{5.12}$$

An SD pair can use either a relay or a direct link to communicate directly. So we have the following constraint.

$$\sum_{r\in\mathcal{R}(s)} \sum_{l_{sr}\in\mathcal{L}_{sr}} x_{l_{sr}} + \sum_{l_{sd(s)}\in\mathcal{L}_{sd(s)}} x_{l_{sd(s)}} = 1, \forall\ s \in \mathcal{S}_{i\cup j}. \tag{5.13}$$

Furthermore, a type $\mathcal{S}_j$ SD pair has to use a relay, i.e.,

$$\sum_{r\in\mathcal{R}(s)} \sum_{l_{sr}\in\mathcal{L}_{sr}} x_{l_{sr}} = 1, \forall\ s \in \mathcal{S}_j. \tag{5.14}$$

If a relay is selected for an SD pair with source s and destination $d(s)$, the EDT from s to $d(s)$, denoted as g_s, is

$$\begin{aligned} g_s &= \sum_{r\in\mathcal{R}(s)} \left(\sum_{l_{sr}\in\mathcal{L}_{sr}} \mathbb{E}(T_{l_{sr}}) + \left(\sum_{l_{rd(s)}\in\mathcal{L}_{rd(s)}} \mathbb{E}(T_{l_{sr},l_{rd(s)}}) x_{l_{rd(s)}} \right) x_{l_{sr}} \right) \\ &= \sum_{r\in\mathcal{R}(s)} \sum_{l_{sr}\in\mathcal{L}_{sr}} \mathbb{E}(T_{l_{sr}}) x_{l_{sr}} + \sum_{r\in\mathcal{R}(s)} \sum_{l_{sr}\in\mathcal{L}_{sr}} x_{l_{sr}} \sum_{l_{rd(s)}\in\mathcal{L}_{rd(s)}} \mathbb{E}(T_{l_{sr},l_{rd(s)}}) x_{l_{rd(s)}}. \end{aligned} \tag{5.15}$$

If a relay is not selected, the EDT from s to $d(s)$, denoted as u_s, is

$$u_s = \sum_{l_{sd(s)}\in\mathcal{L}_{sd(s)}} \mathbb{E}(T_{l_{sd(s)}}) x_{l_{sd(s)}}. \tag{5.16}$$

Determine If There Is Enough Relays

Before our problem formulation, we first need to determine whether each SD pair in $\mathcal{S}_j$ can have a relay. Let index variable $y_{rs} = 1$ denote that relay r is assigned to SD pair s, and $y_{rs} = 0$ otherwise. This problem can be formulated as follows.

$$\textbf{P0}: \max: \sum_{s\in\mathcal{S}_j}\sum_{r\in\mathcal{R}(s)} y_{rs} \tag{5.17}$$

$$\text{s.t.} \sum_{s\in\mathcal{S}(r)} y_{rs} \leq 1,\ \forall\ r \in \mathcal{R} \tag{5.18}$$

$$\sum_{r\in\mathcal{R}(s)} y_{rs} \leq 1,\ \forall\ s \in \mathcal{S}_j \tag{5.19}$$

$$Y \leq \|R\|\,. \tag{5.20}$$

where $\|R\|$ is the total number of relays in the network.

The constraints are due to the fact that each source can use up to one relay, and each relay can serve at most one source at a time. Let Y be the Objective Function Value (OFV) of problem **P0**. If $Y \geq \|\mathcal{S}_j\|$, where $\|\cdot\|$ denotes the cardinality of a set, each SD pair in $\mathcal{S}_j$ can be served by a relay; otherwise, there are some SD pairs in $\mathcal{S}_j$ that cannot have a relay. We have the following two cases.

When $Y \geq \|\mathcal{S}_j\|$

This is the case when each SD pair in $\mathcal{S}_j$ can have a relay. In this case, our objective is to minimize the **MEDT** among all the SD pairs. We thus have the following problem formulation.

$$\textbf{P1 :}\ \text{min: max}_{s\in\mathcal{S}_{i\cup j}}\{g_s + u_s\} \tag{5.21}$$

$$\text{s.t. } (5.8) - (5.14).$$

Note that when $Y \geq \|\mathcal{S}_j\|$, problem **P1** must have a solution. Although the constraints are linear, the objective function is not. Therefore problem **P1** is a nonlinear integer programming problem (NIP), which is generally NP-hard. In the next section, we propose a Decomposition Principle to solve this problem.

When $Y < \|\mathcal{S}_j\|$

In this case, problem **P1** is not applicable since a type $\mathcal{S}_j$ SD pair may not have a relay to forward its data, if all relays within range are assigned to other SD pairs. We develop a heuristic algorithm to address this case in Section 5.5.8.

5.5 Problem Decomposition and Solution

In this section, we present the Decomposition Principle for the case when each SD pair in $\mathcal{S}_j$ can have a relay, which breaks down problem **P1** into a subproblem for link selection and another subproblem for relay selection. The basic idea is to determine the link selection for each relay first, and then determine the relay selection based on the result of link selection. Moreover, the link selection sub-problem can be further decomposed into three sub-problems, one for link selection in hop 1, the second for link selection in hop 2, and the third for direct link selection. We develop effective algorithms to solve the decomposed problems, and more important, prove a tight *optimality bound* for the decomposition principle solution. In the case that there is no enough relays for the SD pairs in $\mathcal{S}_j$, we develop a heuristic algorithm that can still produce highly competitive solutions.

5.5.1 Optimal Choice and Greedy Choice

We first define an optimal choice, Optimal Choice 1 (OC1), and a greedy choice, Greedy Choice 1 (GC1), as follows.

- **Optimal Choice 1 (OC1)**: Given a link $l_{rd(s)}$ in hop 2, choose the hop 1 link as

$$l_{sr}^{*} = \arg\min_{l_{sr} \in \mathcal{L}_{sr}} \{\mathbb{E}(T_{l_{sr}}) + \mathbb{E}(T_{l_{sr}, l_{rd(s)}})\}. \tag{5.22}$$

That is, choose the hop 1 link that minimizes the EDT from s to $d(s)$ for a given hop 2 link.

- **Greedy Choice 1 (GC1)**: Given a link $l_{rd(s)}$ in hop 2, choose the hop 1 link as

$$l_{sr}^{+} = \arg\min_{l_{sr} \in \mathcal{L}_{sr}} \mathbb{E}(T_{l_{sr}}). \tag{5.23}$$

That is, choose the hop 1 link that minimizes the EDT from s to r for a given hop 2 link.

Obviously, the choice of l_{sr}^{*} depends on $l_{rd(s)}$ but that of l_{sr}^{+} does not. We have the following theorem for OC1 and GC1.

Theorem 5.1. *For a given relay r and a hop 2 link, GC1 can achieve an EDT from s to $d(s)$ via r that is at most 1 time slot greater than OC1 does.*

Proof. Let the hop 2 link be l, and recall that l_{sr}^{+} and l_{sr}^{*} are the links chosen by GC1 and OC1, respectively. For two time slots t_1 and t_2, denote $\Delta_t = t_2 - t_1$, which is an integer.

We consider the following four cases.

1. Case 1: $p_l + q_l \leq 1$ and $\Delta_t \geq 0$. From (5.3) and (5.5), we have (5.24) given on top of the next page. Since $0 < (1 - p_l - q_l)^{\Delta_t} < 1$, then $-1 < (1 - p_l - q_l)^{\Delta_t} - 1 < 0$. And since $-p_l \leq q_l - P(C_l^{t_l})(p_l + q_l) \leq q_l$, we have

$$-1 \leq -\frac{p_l}{p_l + q_l} \leq \frac{q_l - P(C_l^{t_l})(p_l + q_l)}{p_l + q_l} \leq \frac{q_l}{p_l + q_l} \leq 1.$$

Since $(1 - p_l - q_l)^{T + t_2 - t_l} < 1$, it follows (5.24) that

$$\left|\mathbb{E}(T_{l_{sr},l}^{+} | T_{l_{sr}^{+}} = t_1) - \mathbb{E}(T_{l_{sr},l}^{*} | T_{l_{sr}^{*}} = t_2)\right| \leq 1, \tag{5.26}$$

where $|\cdot|$ denotes the absolute value.

2. Case 2: $p_l + q_l \leq 1$ and $\Delta_t < 0$. A similar reasoning as the above yields the following inequality.

$$\left|\mathbb{E}(T_{l_{sr}^{*},l} | T_{l_{sr}^{*}} = t_2) - \mathbb{E}(T_{l_{sr}^{+},l} | T_{l_{sr}^{+}} = t_1)\right| \leq 1$$

$$\mathbb{E}\left(T_{l_{sr},l}^{+}|T_{l_{sr}^{+}}=t_1\right)-\mathbb{E}\left(T_{l_{sr}^{*},l}|T_{l_{sr}^{*}}=t_2\right)=\frac{P(C_{rd}^{T+t_2})-P(C_{rd}^{T+t_1})}{p_{l_{rd(s)}}}$$

$$=\frac{(1-P(C_l^{t_l}))q_l-P(C_l^{t_l})p_l}{p_l+q_l}((1-p_l-q_l)^{T+t_2-t_l}-(1-p_l-q_l)^{T+t_1-t_l})$$

$$=\frac{q_l-P(C_l^{t_l})(p_l+q_l)}{p_l+q_l}((1-p_l-q_l)^{T+t_2-t_l})((1-p_l-q_l)^{\Delta_t}-1). \tag{5.24}$$

$$\left|\frac{q_l-P(C_l^{t_l})(p_l+q_l)}{p_l+q_l}((1-p_l-q_l)^{T+t_2-t_l})((1-p_l-q_l)^{\Delta_t}-1)\right|$$

$$=\left|\frac{q_l-P(C_l^{t_l})(p_l+q_l)}{p_l+q_l}\right|\cdot\left|(1-p_l-q_l)^{T+t_2-t_l}\right|\cdot\left|(1-p_l-q_l)^{\Delta_t}-1\right|$$

$$\leq\left|q_l-P(C_l^{t_l})(p_l+q_l)\right|\cdot\frac{1}{p_l+q_l}\cdot\left|(1-p_l-q_l)^{T+t_2-t_l}\right|(p_l+q_l)$$

$$=\left|q_l-P(C_l^{t_l})(p_l+q_l)\right|\cdot\left|(1-p_l-q_l)^{T+t_2-t_l}\right|\leq 1. \tag{5.25}$$

$$\Rightarrow \text{Inequality } (5.26). \tag{5.27}$$

3. Case 3: $1<p_l+q_l\leq 2$ and $\Delta_t\geq 0$. It follows that

$$|(1-p_l-q_l)^{\Delta_t}-1|\leq|(1-p_l-q_l)^{\Delta_t}|+1\leq 1+$$

$$|1-p_l-q_l|=1-(1-p_l-q_l)=p_l+q_l. \tag{5.28}$$

Then we have (5.25) as shown on top of the next page, which implies inequality (5.26).

4. Case 4: $1<p_l+q_l\leq 2$ and $\Delta_t<0$. Again, a similar reasoning as the above yields (5.27).

From Cases 1,2,3 and 4, we conclude that for all $t_1\geq 1, t_2\geq 1$, and $0\leq p_l+q_l\leq 2$, inequality (5.26) holds.

According to (5.4), we can compute $\mathbb{E}(T_{l_{sr},l}^{+})-\mathbb{E}(T_{l_{sr}^{*},l})$ as inequality (5.29) on top of the next page. Recall $\mathbb{E}(T_{l_{sr}^{+}})=\min_{l_{sr}\in\mathcal{L}_{sr}}\mathbb{E}(T_{l_{sr}})$. Assume $\mathbb{E}(T_{l_{sr}^{+}})-\mathbb{E}(T_{l_{sr}^{*}})=P(C_{l_{sr}^{*}}^{T})/p_{l_{sr}^{*}}-P(C_{l_{sr}^{+}}^{T})/p_{l_{sr}^{+}}=\alpha\leq 0$. By (5.29) we have

$$-1+\alpha\leq\mathbb{E}(T_{l_{sr},l}^{+})+\mathbb{E}(T_{l_{sr}^{+}})-\mathbb{E}(T_{l_{sr}^{*},l})-\mathbb{E}(T_{l_{sr}^{*}})\leq 1+\alpha\leq 1. \tag{5.30}$$

$$
\begin{aligned}
&\mathbb{E}(T_{l_{sr}^+,l}) - \mathbb{E}(T_{l_{sr}^*,l}) = \mathbb{E}\left(\mathbb{E}(T_{l_{sr}^+,l} - T_{l_{sr}^*,l} \left| T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right)\right) \\
&= \sum_{t_1,t_2} \mathbb{E}\left(T_{l_{sr}^+,l} - T_{l_{sr}^*,l} \left| T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right.\right) \cdot P\left(T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right) \\
&= \sum_{t_1,t_2} \left(\mathbb{E}(T_{l_{sr}^+,l} \left| T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right.) - \mathbb{E}(T_{l_{sr}^*,l} \left| T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right.)\right) \cdot P\left(T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2\right) \\
&= \sum_{t_1,t_2} \left(\mathbb{E}(T_{l_{sr}^+,l} \left| T_{l_{sr}^+} = t_1\right.) - \mathbb{E}(T_{l_{sr}^*,l} \left| T_{l_{sr}^*} = t_2\right.)\right) \cdot P(T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2) \\
&\leq \sum_{t_1,t_2} 1 \cdot P(T_{l_{sr}^+} = t_1, T_{l_{sr}^*} = t_2) = 1. \qquad (5.29)
\end{aligned}
$$

Moreover, if $\alpha \leq -1$, then we have $\mathbb{E}(T_{l_{sr}^+,l}) + \mathbb{E}(T_{l_{sr}^+}) - \mathbb{E}(T_{l_{sr}^*,l}) - \mathbb{E}(T_{l_{sr}^*}) \leq 0$. Recall $\mathbb{E}(T_{l_{sr}^*}) + \mathbb{E}(T_{l_{sr}^*,l}) = \min_{l_{sr} \in \mathcal{L}_{sr}}\{\mathbb{E}(T_{l_{sr}}) + \mathbb{E}(T_{l_{sr},l})\}$. Therefore $\mathbb{E}(T_{l_{sr}^+,l}) + \mathbb{E}(T_{l_{sr}^+}) - \mathbb{E}(T_{l_{sr}^*,l}) - \mathbb{E}(T_{l_{sr}^*}) = 0$, which means GC1 equals OC1 in terms of EDT from s to $d(s)$.

Thus we conclude that theorem holds true. □

In the following, we show how to use GC1 to reduce problem **P1** into a simpler problem.

5.5.2 Link Selection in Hop 1

The problem is to minimize the MEDT among the SD pairs while there are plenty of relays. With GC1, we consider links l_{sr}^+, for all $s \in \mathcal{S}_{i\cup j}$, $r \in \mathcal{R}(s)$ in hop 1 of problem **P1**, as

$$
x_{l_{sr}^+} \in \{0,1\}, x_{l_{sr}} = 0, \forall\, l_{sr} \neq l_{sr}^+, l_{sr} \in \mathcal{L}_{sr}, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s). \qquad (5.31)
$$

Substitute constraint (5.31) into problem **P1** and then we have a reduced problem, termed problem **P2**, as follows.

$$
\begin{aligned}
\textbf{P2}:\ \min: \max_{s\in\mathcal{S}_{i\cup j}} \Bigg\{ & \sum_{r\in\mathcal{R}(s)} \mathbb{E}(T_{l_{sr}^+}) x_{l_{sr}^+} + \sum_{r\in\mathcal{R}(s)} x_{l_{sr}^+} \times \\
& \sum_{l_{rd(s)}\in\mathcal{L}_{rd(s)}} \mathbb{E}(T_{l_{sr}^+,l_{rd(s)}}) x_{l_{rd(s)}} +
\end{aligned}
$$

$$\sum_{l_{sd(s)} \in \mathcal{L}_{sd(s)}} \mathbb{E}(T_{l_{sd(s)}}) x_{l_{sd(s)}} \Bigg\} \tag{5.32}$$

$$\text{s.t.} \sum_{s \in \mathcal{S}(r)} x_{l_{sr}^+} \leq 1, \ \forall \ r \in \mathcal{R}, \tag{5.33}$$

$$x_{l_{sr}^+} = \sum_{l_{rd(s)} \in \mathcal{L}_{rd(s)}} x_{l_{rd(s)}}, \ \forall \ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s), \tag{5.34}$$

$$\sum_{r \in \mathcal{R}(s)} x_{l_{sr}}^+ + \sum_{l_{sd(s)} \in \mathcal{L}_{sd(s)}} x_{l_{sd(s)}} = 1, \ \forall \ s \in \mathcal{S}_{i \cup j} \tag{5.35}$$

$$\sum_{r \in \mathcal{R}(s)} x_{l_{sr}}^+ = 1, \forall \ s \in \mathcal{S}_j \tag{5.36}$$

$$x_{l_{sr}^+} = \{0, 1\}, \ \forall \ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s) \tag{5.37}$$

Constraints (5.9) and (5.10).

The number of decision variables of problem **P2** is much less than that of problem **P1**. We will prove below that the difference between the OFV of problem **P2** and that of problem **P1** is at most 1 time slot. We first introduce a lemma as a basis of the proof. For ease of presentation, let $\mathcal{S}_1$ denote the set of sources that are assigned with relays in the optimal solution to problem **P1**, i.e., $z_s = \mathbb{E}(T_{l_{sr}}) + \mathbb{E}(T_{l_{sr}, l_{rd(s)}})$, for all $s \in \mathcal{S}_1$, and $\mathcal{S}_2$ be the set of sources that are not assigned with relays and communicate with their destinations using a direct link, i.e., $z_s = \mathbb{E}(T_{l_{sd(s)}})$, for all $s \in \mathcal{S}_2$. Also denote $\mathcal{S}_{1 \cup 2} = \mathcal{S}_1 \cup \mathcal{S}_2$. Note that $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}_i \cup \mathcal{S}_j$.

Lemma 5.1. *Denote* $\phi^* = \{x_{l_{sr}^*} = 1, x_{l_{rd(s)}^*} = 1, \forall \ s \in \mathcal{S}_1,$ *and* $x_{l_{sd(s)}^*} = 1, \forall \ s \in \mathcal{S}_2\}$ *as the optimal solution to problem* **P1**. *For all* $s \in \mathcal{S}_1$, *set* $x_{l_{sr}^+} = 1$ *and then set* $x_{l_{sr}^*} = 0$. *Then* $\phi = \{x_{l_{sr}^+} = 1, x_{l_{rd(s)}^*} = 1, \forall \ s \in \mathcal{S}_1,$ *and* $x_{l_{sd(s)}^*} = 1, \forall \ s \in \mathcal{S}_2\}$ *is a feasible solution to problem* **P2**.

Proof. Comparing ϕ with ϕ^*, only the hop 1 link choice is different. Since for all s, we set $x_{l_{sr}^+} = 1$ and then set $x_{l_{sr}^*} = 0$, the link choice of hop 1 still satisfies all the constraints in problem **P2**. Hence ϕ is a feasible solution to problem **P2**. □

Theorem 5.2. *The OFV of problem* **P2** *is at most 1 time slot greater than that of problem* **P1**.

Proof. If ϕ^* is also the optimal solution to problem **P2**, denote the corresponding EDT of each SD pair as z_s^*, for all $s \in \mathcal{S}_{1\cup2}$, and the OFV of problem **P1** as $\mathbf{z} = \max_{s\in\mathcal{S}_{1\cup2}}\{z_s^*\}$. if ϕ is a feasible solution to problem **P2**, denote the corresponding EDT of each SD pair as z_s^+, for all $s \in \mathcal{S}_{1\cup2}$. Then the difference between the OFV of problem **P2** and that of problem **P1** can be written as

$$\max_{s\in\mathcal{S}_{1\cup2}} z_s^+ - \mathbf{z} = \max_{s\in\mathcal{S}_{1\cup2}} \{z_s^+ - \mathbf{z}\}. \tag{5.38}$$

Since links $\{l_{sr}^+, \forall\ s, r \in \mathcal{R}(s)\}$ are chosen by GC1, it follows Theorem 5.1 that $z_s^+ - z_s^* \leq 1$, for all $s \in \mathcal{S}_1$. Besides, we have $z_s^+ = z_s^* = \mathbb{E}(T_{l^*_{sd(s)}})$, for all $s \in \mathcal{S}_2$. Thus we have

$$z_s^+ - z_s^* \leq 1,\ \forall\ s \in \mathcal{S}_{1\cup2}. \tag{5.39}$$

Since $\mathbf{z} \geq z_s^*$, for all $s \in \mathcal{S}_{1\cup2}$, we have

$$\begin{aligned} z_s^+ \leq z_s^* + 1 \leq \mathbf{z} + 1 \Rightarrow z_s^+ - \mathbf{z} \leq 1, \forall s \in \mathcal{S}_{1\cup2}, \\ \Rightarrow \max_{s\in\mathcal{S}_{1\cup2}} \{z_s^+ - \mathbf{z}\} \leq 1 \Rightarrow \max_{s\in\mathcal{S}_{1\cup2}} \{z_s^+\} - \mathbf{z} \leq 1. \end{aligned}$$

Thus we conclude that Theorem 1 holds true. □

5.5.3 Link Selection in Hop 2

Lemma 5.1 indicates that $\phi = \{x_{l_{sr}^+} = 1, x_{l^*_{rd(s)}} = 1, \forall\ s \in \mathcal{S}_1$, and $x_{l^*_{sd(s)}} = 1, \forall\ s \in \mathcal{S}_2\}$ is a feasible, but not necessary optimal, solution to problem **P2**. Furthermore, $l^*_{rd(s)}, \forall\ s \in \mathcal{S}_1$ is hard to obtain because it requires computing the EDT of all possible links in hops 1 and 2. To obtain the optimal solution to problem **P2**, we first define another greedy choice, termed Greedy Choice 2 (GC2), as follows.

- **Greedy Choice 2 (GC2)**: Given a link l^+_{sr} obtained by GC1 in hop 1, choose the hop 2 link $l^+_{rd(s)}$ as

$$l^+_{rd(s)} = \underset{l_{rd(s)} \in \mathcal{L}_{rd(s)}}{\arg\min} \ \mathbb{E}(T_{l^+_{sr}, l_{rd(s)}}). \tag{5.40}$$

  That is, choose the hop 2 link that minimizes the EDT from r to $d(s)$ for given hop 1 link l^+_{sr}.

With GC2, we only consider links $l^+_{rd(s)}$, for all $s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s)$ in hop 2 for problem **P2**, which means

$$\begin{aligned} & x_{l^+_{rd(s)}} \in \{0,1\}, x_{l_{rd(s)}} = 0, \ \forall \ l_{rd(s)} \neq l^+_{rd(s)}, l_{rd(s)} \in \mathcal{L}_{rd(s)}, \\ & \forall \ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s). \end{aligned} \tag{5.41}$$

Then we have the following claims for the optimal solution to problem **P2**.

Lemma 5.2. *Denote* $\hat{\phi}^* = \{x_{l^+_{sr}} = 1, x_{l^*_{rd(s)}} = 1, \forall \ s \in \mathcal{S}_1, \ x_{l^*_{sd(s)}} = 1, \forall \ s \in \mathcal{S}_2\}$ *as the optimal solution to problem* **P2**. *For all* $s \in \mathcal{S}_1$, *set* $x_{l^+_{rd(s)}} = 1$ *and* $x_{l^*_{rd(s)}} = 0$. *Then* $\hat{\phi} = \{x_{l^+_{sr}} = 1, x_{l^+_{rd(s)}} = 1, \forall \ s \in \mathcal{S}_1, \ x_{l^*_{sd(s)}} = 1, \forall \ s \in \mathcal{S}_2\}$ *is a feasible solution to problem* **P2**.

Proof. Comparing $\hat{\phi}$ with $\hat{\phi}^*$, only the hop 2 link choice is different. Since for all s, we set $x_{l^+_{rd(s)}} = 1$ and then set $x_{l^*_{rd(s)}} = 0$, the link choice of hop 2 still satisfies all the constraints in problem **P2**. Hence $\hat{\phi}$ is a feasible solution to problem **P2**. □

Theorem 5.3. *In the optimal solution to problem* **P2**, *the link selection in hop 2 is* $\{x_{l^+_{rd(s)}} \in \{0,1\}$, $x_{l_{rd(s)}} = 0, \forall \ l_{rd(s)} \neq l^+_{rd(s)}, l_{rd(s)} \in \mathcal{L}_{rd(s)}, \forall \ s \in \mathcal{S}_1, r \in \mathcal{R}(s)\}$.

Proof. Recall $\hat{\phi}$ is a feasible solution to problem **P2**. With this solution, define $z_s^{++} = \mathbb{E}(T_{l^+_{sr}}) + \mathbb{E}(T_{l^+_{sr}, l^+_{rd(s)}})$, for all $s \in \mathcal{S}_1$, and $z'_s = \mathbb{E}(T_{l^+_{sr}}) + \mathbb{E}(T_{l^+_{sr}, l_{rd(s)}})$, for all $l_{rd(s)} \neq l^+_{rd(s)}, l_{rd(s)} \in \mathcal{L}_{rd(s)}$, for all $s \in \mathcal{S}_1$.

Since $\mathbb{E}(T_{l^+_{sr}, l^+_{rd(s)}}) = \min_{l_{rd(s)} \in \mathcal{L}_{rd(s)}} \mathbb{E}(T_{l^+_{sr}, l_{rd(s)}})$, we have $z_s^{++} \leq z'_s$, for all $s \in \mathcal{S}_1$. We also have $z_s^{++} = z'_s = \mathbb{E}(T_{l^*_{sd(s)}})$, for all $s \in \mathcal{S}_2$.

Let $\mathbf{z}' = \max_{s\in\mathcal{S}_{1\cup 2}}\{z'_s\}$. It follows that

$$z_s^{++} \leq z'_s \leq \mathbf{z}',\ \forall s\ \in \mathcal{S}_{1\cup 2}. \tag{5.42}$$

We thus have

$$\begin{aligned}
&\max_{s\in\mathcal{S}_{1\cup 2}}\{z_s^{++}\} - \max_{s\in\mathcal{S}_{1\cup 2}}\{z'_s\} = \max_{s\in\mathcal{S}_{1\cup 2}}\{z_s^{++}\} - \mathbf{z}' \\
=\ &\max_{s\in\mathcal{S}_{1\cup 2}}\{z_s^{++} - \mathbf{z}'\} \overset{Inequality\ (5.42)}{\leq} 0 \\
\Rightarrow\ &\max_{s\in\mathcal{S}_{1\cup 2}}\{z_s^{++}\} \leq \max_{s\in\mathcal{S}_{1\cup 2}}\{z'_s\}.
\end{aligned}$$

Note that the objective of problem **P2** is to minimize the MEDT among all SD pairs. So $\hat{\phi}$ is optimal to problem **P2**. The proof is completed. □

5.5.4 Link Selection in Direct Path

Link selection when the SD pair communicate directly in the optimal solution to problem **P2** can also be obtained with a greedy approach.

For all $s \in \mathcal{S}_j$, set $E(T_{l_{sd(s)}}), \forall\ l_{sd(s)} \in \mathcal{L}_{sd(s)}$, to an arbitrary constant, because the constraints will ensure that for all $s \in \mathcal{S}_j$, the direct link will not be selected. Define a greedy choice, termed Greedy Choice 3 (GC3), as follows.

- **Greedy Choice 3 (GC3)**: Choose the link in the direct SD path as $l^+_{sd(s)} = \arg\min_{l_{sd(s)}\in\mathcal{L}_{sd(s)}} \mathbb{E}(T_{l_{sd(s)}})$. That is, choose the link in the direct SD path that minimizes the EDT from s to $d(s)$.

Theorem 5.4. *In the optimal solution to problem* **P2**, *link selection in the direct path is* $\{x_{l^+_{sd(s)}} \in \{0,1\}, x_{l_{sd(s)}} = 0, \forall\ l_{sd(s)} \neq l^+_{sd(s)}, l_{sd(s)} \in \mathcal{L}_{sd(s)}, \forall s \in \mathcal{S}_{i\cup j}\}$.

Proof. Define $\hat{z}_s = \mathbb{E}(T_{l^+_{sd(s)}})$, for all $s \in \mathcal{S}_2$, and $\tilde{z}_s = \mathbb{E}(T_{l_{sd(s)}})$, for all $l_{sd(s)} \neq l^+_{sd(s)}, l_{sd(s)} \in \mathcal{L}_{sd(s)}$, for all $s \in \mathcal{S}_2$. Then we have $\hat{z}_s \leq \tilde{z}_s$. We also have $\hat{z}_s = \tilde{z}_s$, for all $s \in \mathcal{S}_1$.

Let $\bar{z} = \max_{s \in \mathcal{S}_{1\cup 2}}\{\tilde{z}_s\}$. It follows that

$$\hat{z}_s \leq \tilde{z}_s \leq \bar{z},\ \forall s\ \in \mathcal{S}_{1\cup 2}. \tag{5.43}$$

We thus have

$$\begin{aligned}
&\max_{s\in\mathcal{S}_{1\cup 2}}\{\hat{z}_s\} - \max_{s\in\mathcal{S}_{1\cup 2}}\{\tilde{z}_s\} = \max_{s\in\mathcal{S}_{1\cup 2}}\{\hat{z}_s\} - \bar{z} \\
= &\max_{s\in\mathcal{S}_{1\cup 2}}\{\hat{z}_s - \bar{z}\} \overset{Inequality\ (5.43)}{\leq} 0 \\
\Rightarrow &\max_{s\in\mathcal{S}_{1\cup 2}}\{\hat{z}_s\} \leq \max_{s\in\mathcal{S}_{1\cup 2}}\{\tilde{z}_s\}.
\end{aligned}$$

Note that the objective of problem **P2** is to minimize the MEDT among all SD pairs. So $\hat{z}_s$ is the optimal solution for direct link selection to problem **P2**. The proof is completed. □

5.5.5 Relay Assignment

Now that the hop 1, hop 2, and direct link selection sub-problems having been solve with GC1, GC2, and GC3, respectively, we next solve the remaining problem of relay assignment. Substituting the following into problem **P2**,

$$\begin{aligned}
&x_{l^+_{rd(s)}} = x_{l^+_{sr}}, x_{l_{sr}} = 0, x_{l_{rd(s)}} = 0, x_{l_{sd(s)}} = 0, \forall\ l_{sr} \neq l^+_{sr}, \\
&\forall\ l_{rd(s)} \neq l^+_{rd(s)}, \forall\ l_{sd(s)} \neq l^+_{sd(s)}, l_{sr} \in \mathcal{L}_{sr}, l_{rd(s)} \in \mathcal{L}_{rd(s)}, \\
&l_{sd(s)} \in \mathcal{L}_{sd(s)}, \forall\ s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s),
\end{aligned} \tag{5.44}$$

we obtain a reduced problem, termed **SP2**, as follows.

$$\begin{aligned}
\textbf{SP2 :}\ \min :\ \max_{s\in\mathcal{S}_{i\cup j}} \Bigg\{ &\sum_{r\in\mathcal{R}(s)} \left(\mathbb{E}(T_{l^+_{sr}}) + \mathbb{E}(T_{l^+_{sr},l^+_{rd(s)}})\right) x_{l^+_{sr}} + \\
&\sum_{l_{sd(s)}\in\mathcal{L}_{sd(s)}} \mathbb{E}(T_{l^+_{sd(s)}}) x_{l^+_{rd(s)}} \Bigg\}
\end{aligned}$$

$$\text{s.t.} \sum_{r \in \mathcal{R}(s)} x^+_{l_{sr}} + x_{l^+_{sd(s)}} = 1, \forall s \in \mathcal{S}_{i \cup j} \tag{5.45}$$

$$x_{l^+_{sd(s)}} = \{0, 1\}, \forall s \in \mathcal{S}_{i \cup j} \tag{5.46}$$

$$\text{Constraints } (5.33),\ (5.36),\ \text{and } (5.37).$$

Also the OFV of problem **SP2** equals to that of problem **P2**. According to Theorem (5.2), we have Theorem (5.5) as follows.

Theorem 5.5. *The OFV of problem* **SP2** *is at most 1 time slot greater than that of problem* **P1**.

5.5.6 Decomposition Principle and Problem Reformulation

With analysis in Sections 5.5.1 to 5.5.5, we are now able to present the following theorem on the Decomposition Principle.

Theorem 5.6. *Problem* **P1** *can be solved with the following four-step procedure, and the OFV of the solution is at most 1 time slot larger than that of the optimal solution.*

- *Step 1: Choose the set of links in hop 1, i.e.,* $\{l^+_{sr}\}$*, as* $l^+_{sr} = \arg\min_{l_{sr} \in \mathcal{L}_{sr}} \mathbb{E}(T_{l_{sr}}), \forall\ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s)$.
- *Step 2: With* $\{l^+_{sr}\}$*, choose the set of links in hop 2, i.e.,* $\{l^+_{rd(s)}\}$*, as* $l^+_{rd(s)} = \arg\min_{l_{rd(s)} \in \mathcal{L}_{rd(s)}} \mathbb{E}(T_{l^+_{sr}, l_{rd(s)}}), \forall\ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s)$.
- *Step 3: Choose the set of links in the direct path, i.e.,* $\{l^+_{sd(s)}\}$*, as* $l^+_{sd(s)} = \arg\min_{l_{sd(s)} \in \mathcal{L}_{sd(s)}} \mathbb{E}(T_{l_{sd(s)}}), \forall\ s \in \mathcal{S}_{i \cup j}, r \in \mathcal{R}(s)$.
- *Step 4: With (5.44) derived, solve problem* **SP2**.

Let the problem in Step 1, Step 2, and Step 3 of Theorem 5.6 be termed **SP1**. Note that problem **SP2** is not in the general Integer Linear Programming (ILP) form. To solve problem **SP2**, we reformulate it into a linear programming (LP) problem. Introducing a new variable $w =$

$\max_{s\in\mathcal{S}_{i\cup j}}\{\mathbb{E}(T_{l^+_{sd(s)}})x^+_{l_{sd(s)}} + \sum_{r\in\mathcal{R}(s)}(\mathbb{E}(T_{l^+_{sr}}) + \mathbb{E}(T_{l^+_{sr},l^+_{rd(s)}}))x_{l^+_{sr}}\}$, we have

$$w \geq \mathbb{E}(T_{l^+_{sd(s)}})x^+_{l_{sd(s)}} + \sum_{r\in\mathcal{R}(s)}(\mathbb{E}(T_{l^+_{sr}}) + \mathbb{E}(T_{l^+_{sr},l^+_{rd(s)}}))x_{l^+_{sr}},\ \forall\ s \in \mathcal{S}_{i\cup j}. \tag{5.47}$$

Then **SP2** can be rewritten as

SP2': $\min : w$

s.t. Constraints, $(5.33), (5.36), (5.37), (5.45), (5.46)$, and (5.47).

Problem **SP2'** is a mixed integer linear programming problem (MILP) and can be solved with an existing effective solver. Once the relay and link selection are completed, the PNC will inform the nodes to start transmission as scheduled. If and only if at least one of the following events happens, the PNC will reschedule the link selection and relay assignment for all the SD pairs based on feedback.

- Case 1: If a source had no traffic in the previous time slot but has traffic in the current time slot.
- Case 2: Whenever a relay finishes transmission to a destination and thus becomes available for source(s).

5.5.7 Complexity Analysis

Since problem **SP1** is easy to solve, we just compare the complexity of problem **P1** and problem **SP2'** from the following aspects.

- Problem **P1** is an NIP, while problem **SP2'** is an MILP. Currently there are existing efficient solvers for MILP, such as the Gorubi MIP solver and the Matlab Intlinprog function (implementing the Branch and Bound algorithm). Such kind of problems have been solved effectively in prior works [131, 133], especially when solution space is relatively small.

Algorithm 7: Heuristic Algorithm for Link and Relay Assignment When Some $\mathcal{S}_j$ SD Pairs Do Not Have Relays

```
Solve problem P0 ;
if Y ≥ ‖S_j‖ then
    Apply the Decomposition Principle to solve problem P1 ;
else
    for ∀s ∈ S_i do
        Choose direct link l^+_{sd(s)} to communicate with d(s) ;
    end
    Assign relays to type S_j SD pairs according to the solution to P0 ;
    Denote the set of type S_j SD pairs that have a relay as S'_j ;
    Find l^+_{sr}, l^+_{rd(s)}, for all s ∈ S'_j, r ∈ R(s) ;
end
```

- The number of decision variables of problem **P1** is $\sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{sr}\|+\sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\left\|\mathcal{L}_{rd(s)}\right\|+\sum_{s\in\mathcal{S}_i}\left\|\mathcal{L}_{sd(s)}\right\|$, while the number of decision variables of **SP2'** is $\sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\|l^+_{sr}\|+\sum_{s\in\mathcal{S}_{i\cup j}}\left\|l^+_{sd(s)}\right\| = \sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}1+\sum_{s\in\mathcal{S}_{i\cup j}}1=\sum_{s\in\mathcal{S}_{i\cup j}}\|\mathcal{R}(s)\|+\|\mathcal{S}_{i\cup j}\|$, considerably smaller than that of problem **P1**.

5.5.8 When $Y < \|\mathcal{S}_j\|$

If a type $\mathcal{S}_j$ SD pair cannot be served by a relay, its EDT cannot be defined as in (5.15) or (5.16). Thus we cannot directly employ the Decomposition Principle to solve the link and relay assignment problem in this case. We then propose a heuristic algorithm to solve the problem. The basic idea is to maximize the number of SD pairs that can transmit concurrently by relay assignment. We let each type $\mathcal{S}_i$ SD pair transmit via its direct link, and then assign relays to type $\mathcal{S}_j$ SD pairs to maximize the number of $\mathcal{S}_j$ SD pairs that can transmit concurrently. The more concurrent transmissions, the smaller the MEDT.

The Heuristic algorithm is presented in Algorithm 7.

5.6 Simulation Study

5.6.1 Simulation Setup

In this section we validate the performance of the proposed Decomposition Principle by Matlab simulations. Unless otherwise specified, the values of simulation parameters are as given in Table 6.1. Each simulated point in the figures is obtained by repeating the simulation 50 times with different random seeds, while 95% confidence intervals are computed and plotted as error bars in the figures.

We compare the performance of the proposed algorithm in Theorem (5.6) (termed *Proposed*) with two existing schemes designed for mmWave networks. The first one (termed *Benchmark 1*) is proposed in [129], where a source tries to maximize its throughput by choosing the optimal Access Points (APs), and the source-AP channels are modeled as Markov chains. A heuristic algorithm is used to solve the formulated NP-hard problem in [129]. The second one (termed *Benchmark 2*) is proposed in [128], where relay paths are determined for multiple SD pairs with a heuristic to maximize the total throughput under static channel conditions. Throughput fairness among multiple SD pairs is not considered in this scheme.

The performance metrics to evaluate the proposed algorithm are delay, MEDT among all SD pairs, and network throughput. The delay of a packet is the time it spends at the source queue plus the packet delivery time from source to destination. The traffic is generated with a Bernoulli process [112]. At each time slot, the source generates a number of packets with a predetermined probability, denoted as P_G, and the total volume of bits of the packets generated at each time slot does not exceed the channel capacity.

5.6.2 Simulation Results and Analysis

The performance of the proposed algorithm is demonstrated in Fig. 5.2 by comparing the OFV of problem **SP2'** with that of problem **P1** (i.e., the *Optimal*) under increasing channel transition probability q_l. Problem **P1** is an NIP whose solution takes a very long time to obtain using

Table 5.1: Simulation Parameters

Parameter	*Value*
$\|\mathcal{S}_{i\cup j}\|$	10
$\|\mathcal{R}\|$	10
$\|\mathcal{L}_{sr}\|, \forall s, r$	random $\in [3, 7]$
$\left\|\begin{matrix}\mathcal{L}_{rd(s)}\\ \mathcal{L}_{sd(s)}\end{matrix}\right\|, \forall s, r$	random $\in [3, 7]$
$\left\|\begin{matrix}\mathcal{L}_{rd(s)}\\ \mathcal{L}_{sd(s)}\end{matrix}\right\|, \forall s$	random $\in [0, 3]$
$\max_{\{l\in(\mathcal{L}_{sr}\cup\mathcal{L}_{rd(s)}\cup\mathcal{L}_{sd(s)}),\forall s,r\}} q_l$	0.9
$p_l, \forall l \in (\mathcal{L}_{sr} \cup \mathcal{L}_{rd(s)} \cup \mathcal{L}_{sd(s)}), \forall s, r$	random $\in [0.3, 0.7]$
Chanel capacity	1 Gbps
Time slot duration	1 s

exhaustive search even for a moderately-sized network. Therefore we simulate a relative small network with 2 SD pairs, 2 relays, 2 links in hop 1, 2 links in hop 2, and 1 link in the direct path, for each SD pair and relay, to obtain the optimal solution within a reasonable time. From Fig. 5.2, we fine the difference between the OFV of problem **SP2'** and that of problem **P1** is strictly within 1 time slot over the entire range of q_l. The gap is actually much smaller than 1 time slot, which suggests that 1 time slot is in fact the worst case upper bound. Furthermore, the gap increases as q_l grows, since a sub-optimal schedule may result in a relatively worse performance when channel conditions are bad, which means a greater MEDT.

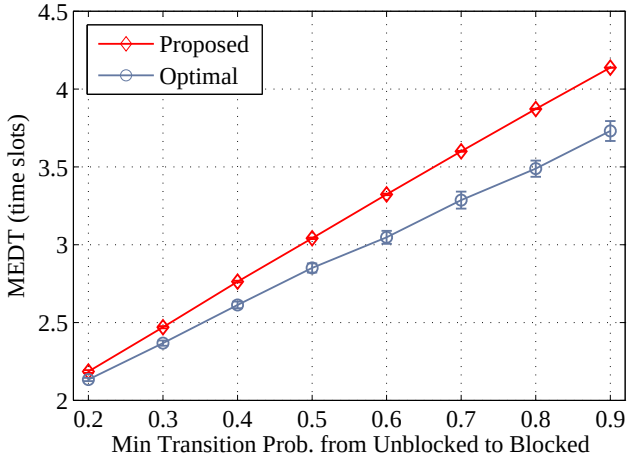

Figure 5.2: The OFV of the proposed decomposition principle and that of the optimal solution versus $\min_l\{q_l\}$, while $P_G = 0.8$.

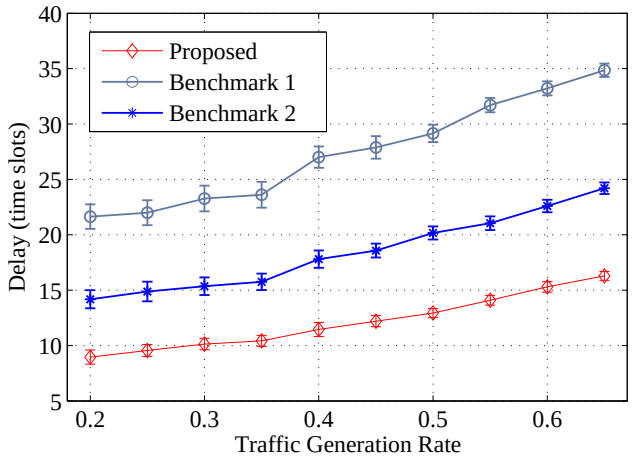

Figure 5.3: Delay versus minimum traffic generation probability P_G, while $\min_{\{l\}}\{eq_l\} = 0.2$.

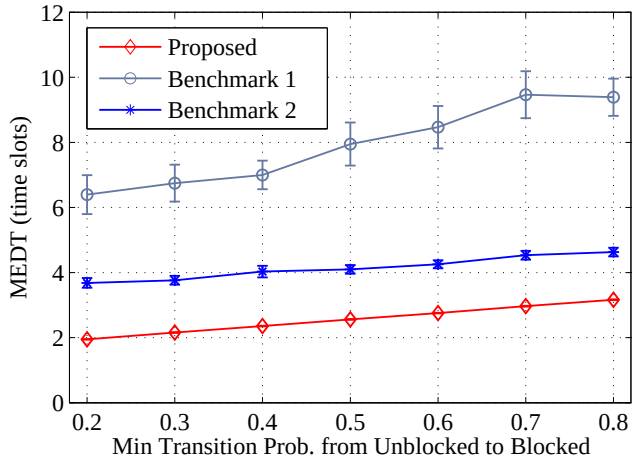

Figure 5.4: MEDT versus $\min_{\{l\}}\{q_l\}$, while $P_G = 0.8$.

We next compare the delay performance of the proposed scheme with that of the two benchmark schemes in Fig. 5.3 under various traffic generation rate P_G. As P_G is increased, the average delays of all the three schemes increase due to the increased traffic load, while the average delay of our proposed algorithm is always considerably lower than that of the two benchmark schemes. Benchmark 1 does not consider coordinating the concurrent transmissions among SD pairs. Therefore different SD pairs may select the same relay and thus collision happens, resulting in an increased delivery time. Benchmark 2 does not consider channel dynamics and thus its schedules may be sub-optimal. This comparison also demonstrates that traffic collision has a serious negative effect on delay performance.

Figure 5.4 shows the MEDT among all SD pairs under changing channel state transition probability q_l. The proposed scheme achieves the lowest MEDT among the three. The confidence interval of Benchmark 1 is greater than that of the other two schemes, indicating Benchmark 1 is less stable in terms of the number of trails until the first successful transmission is achieved. Benchmark 2 considers the channel conditions as static and it lacks adaptation to the channel dynamics, which certainly has an effect on the instantaneous scheduling decision for the current time slot.

The throughput performance achieved by the three schemes is presented in Fig. 5.5. The network throughput is defined as the total number of bits delivered for all the SD pairs per time slot, i.e., per second. As channel condition degrades, the number of links available for transmission is decreased at each time slot. So the number of bits that can be delivered at each time slot is reduced. For Benchmark 1, due to the possible collisions, the number of bits successfully delivered per time slot is less than that of the proposed algorithm. For Benchmark 2, due to lack of consideration of channel dynamics, although it tends to maximize the total expected throughput of all SD pairs, it still makes sub-optimal scheduling decisions under dynamic channel conditions, thus achieving a lower throughput.

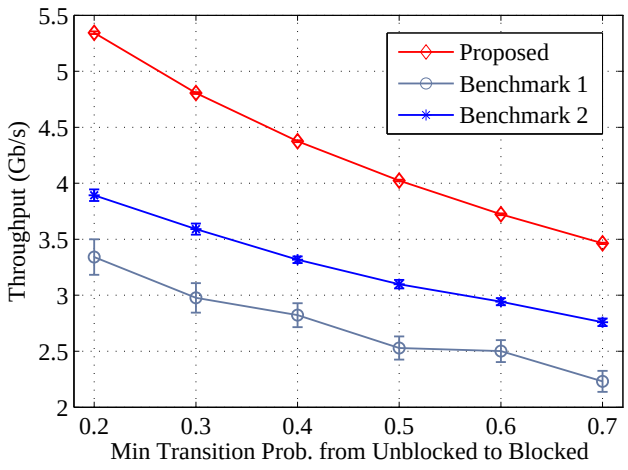

Figure 5.5: Throughput versus $\min_{\{l\}}\{q_l\}$, while $P_G = 0.8$.

In Fig. 5.6, we compare the MEDT of the Heuristic link and relay selection algorithm described in Algorithm (7) (termed *Heuristic*) with that of the two benchmarks under the condition

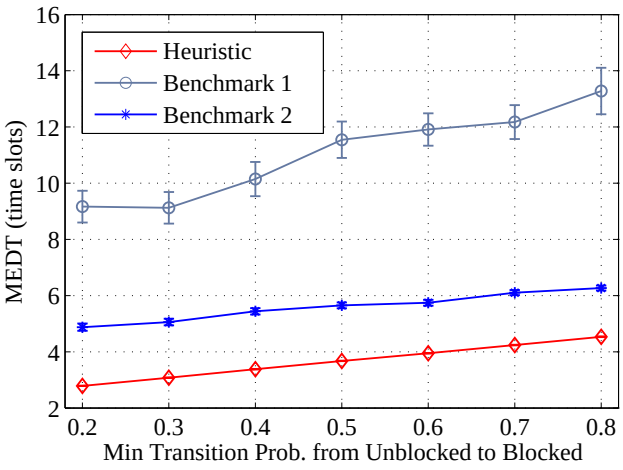

Figure 5.6: MEDT versus $\min_{\{l\}}\{q_l\}$ for the proposed Heuristic algorithm, while $P_G = 0.8$ and the number of relays is 6.

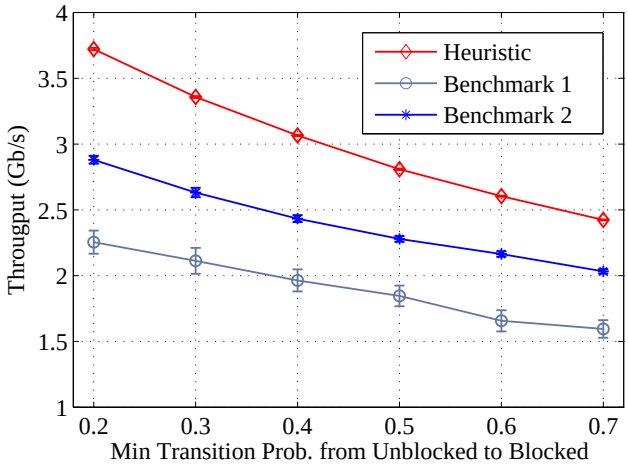

Figure 5.7: Throughput vs $\min_{\{l\}}\{q_l\}$ for the proposed Heuristic algorithm, while $P_G = 0.8$ and the number of relays is 6.

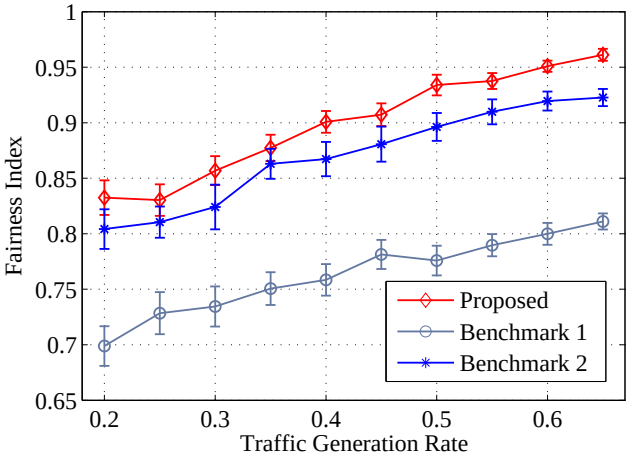

Figure 5.8: Fairness performance versus minimum traffic generation probability, for Proposed Scheme, while $\min_{\{l\}}\{q_l\} = 0.2$ and $P_G \leq 0.8$.

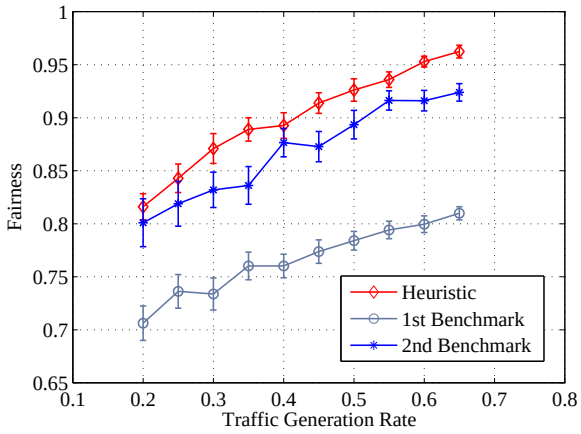

Figure 5.9: Fairness performance versus minimum traffic generation probability, for Heuristic Scheme, while $\min_{\{l\}}\{q_l\} = 0.2$ and $P_G \leq 0.8$.

that there is an insufficient number of relays to serve all the SD pairs. Here we set the number of relays equals to 6, while other parameters are as given in Table 6.1. Comparing to the results in Fig. 5.4, the MEDT of Heuristic is slightly higher, due to the insufficient number of realys to serve the 10 SD pairs. For example, when $\min_{\{l\}}\{q_l\} = 0.2$, the MEDT is increased from 2 to 2.7881. However, it can be observed that Heuristic still outperforms both benchmark schemes with considerable gains. This result makes sense since letting type $\mathcal{S}_i$ SD pairs to communicate using direct links, instead of using a relay, will save more relaying opportunity to Type $\mathcal{S}_j$ SD pairs, so these SD pairs may need less time to successfully delivery their packets. A similar reasoning can be applied to the comparison of throughput performance of Heuristic with that of the two benchmarks, as shown in Fig. 5.7.

Finally, we compare the fairness performance of the three schemes, in terms of average delay of the SD pairs. Fig. 5.8 shows the fairness performance comparison between the proposed scheme and the benchmark schemes. We adopt Jain's fairness index as in [112]: $f(e_1, e_2, \ldots, e_N) = \left(\sum_{n=1}^{N} e_n\right)^2 / \left(N \sum_{n=1}^{N} e_n^2\right)$, where e_n is the average delay of SD pair n, $n = 1, 2, \ldots, N$. The fairness index ranges from 0 (worst) to 1 (best). We can see that our proposed algorithm consistently achieves a higher fairness index than the other two schemes do, due to the *minimax* approach adopted in the problem formulation.

The fairness performance comparison between the heuristic scheme and the benchmark schemes are shown in Fig. 5.9. We can see that the fairness performance of the heuristic scheme is also consistently better than that of the benchmark schemes.

5.7 Conclusion

We developed a Decomposition Principle for the problem of link and relay selection in centralized dual-hop mm Wave networks. The objective was to minimize the MEDT, and the main idea was to decompose the original problem into a sub-problem for link selection, and the other for relay selection. When there are a sufficient amount of relays, we proved that the two sub-problems together can provide a sub-optimal solution to the original problem with an optimality gap bounded by 1 time slot, with greatly reduced complexity. We also developed a heuristic scheme to handle the case when there is not enough relays to serve the SD pairs. Through simulations, we showed that both proposed schemes outperformed two mm Wave network scheduling schemes with considerable gains.

Chapter 6

On Link Scheduling under Blockage and Interference in mmWave Ad Hoc Networks

6.1 Introduction

Before one can deploy the mmWave networks, many challenges need to be addressed. Wireless signal attenuation in mmWave (e.g., 60GHz) channels is a serious issue, much more than in the case of 5GHz or 2.5GHz channels [145], thereby making beamforming indispensable. Because of the small wavelength, large antenna arrays are feasible. The authors in [146] show that the highly directional links, especially in the outdoor environment, can be regarded as "pseudowired" with negligible collision probabilities. Although this characteristic is very attractive for spatial reuse, network coordination and scheduling become extremely challenging with such narrow beamwidths [147]. In the indoor environment, as mentioned above, the beamwidth is usually wider and thus the interference among neighboring links should be considered since the pseudowired assumption may not hold in this case.

Furthermore, mmWave signals usually do not diffract around or penetrate obstacles. A line-of-sight (LOS) path between the transmitter and receiver is required for successful transmission. However, in practical networks, a LOS path may not always exist; besides, it is possible that a LOS path can be blocked (e.g., by a human body) from time to time. In that case, relay nodes would be needed to forward data to a distant or blocked receiver [148], by setting up a LOS path between the transmitter and relay, and then between the relay and receiver. Second, the blockage between two nodes may appear or disappear intermittently due to the movement of objects between them or the movement of the nodes themselves [150, 151]. A realistic mmWave network protocol should consider the multi-hop path model and dynamic blockages between nodes.

In this chapter, we investigate the problem of link scheduling in mmWave Ad Hoc networks. We consider the case where each device (DEV) schedules its own transmission based on the traffic

demand (e.g., the amount of packets backlogged in its buffer for other DEVs, or the amount of traffic requested by the DEVs for the next scheduling period) and the status of the mmWave links. We adopt a directional link model from the literature, which incorporates the beamwidths as well as the beam directions to allow flexibly modeling interference among the directional links. Such a model is known as "directional beamforming," in contrast to today's cellular use of the term beamforming that merely looks to null out interfering users without maximizing gain in a particular direction [152]. We also model blockage of the LOS path with a discrete-time Markov chain model. A successful transmission requires an unblocked LOS path as well as a good Signal to Interference plus Noise Ratio (SINR). By tuning the parameters of the interference and blockage models, both indoor and outdoor mmWave links can be modeled.

We first consider the single-hop network case with N DEVs, where all the data transmissions are through one-hop, single links. That is, any pair of DEVs can communicate with each other directly. We formulate the link scheduling problem as a constrained Binary Integer Programming (BIP) problem, aiming to determine the minimum time length schedule, i.e., to minimize the time duration needed to satisfy the traffic demand of all the links. We develop a greedy algorithm (GA) to maximize the instantaneous throughput of each time slot to solve the formulated BIP problem heuristically.

We then consider the multi-hop case for N DEVs, where some DEVs are not within one-hop distance with each other, e.g., if their distance is longer than the transmission range or if the LOS path between them is blocked. In this case, intermediate DEVs are needed to relay the traffic for DEVs that are not within one-hop distance of each other. A multi-hop minimum time length scheduling problem is formulated incorporating routing and the flow conservation constraint. We then propose an effective algorithm to solve this problem. The performance of the proposed algorithms are validated with simulation and comparison with a benchmark scheme.

The remainder of this chapter is organized as follows. Related work is reviewed in Section 6.2. The system model and problem formulation for the single-hop network scenario are presented in Section 6.3. The proposed solution algorithms for the single-hop scenario are presented in

Section 6.4. The problem formulation for the multi-hope network case is described in Section 6.5 and the solution algorithm is presented in Section 6.6. The proposed algorithms are evaluated in Section 6.7. Section 6.8 concludes the paper.

6.2 Related Work

There have been considerable work on link scheduling in wireless networks. However, most of the prior work do not consider the specific properties of mmWave channels and thus may not be applied for mmWave networks. In [158, 160], the authors solve the problem of spatial TDMA scheduling in ad hoc networks, where each link has its traffic demand and the objective is to find the optimal scheduling of the links to minimize the time length needed to satisfy the traffic demand of all links. However, the channel state of each link is assumed to be static during the entire period, which may not be a valid assumption for mmWave networks where the channel state may change dramatically due to blockage of the LOS path.

The uncertainty of channel availability in Cognitive Radio Networks is considered in [161]. The problem of deciding which channel to sense and access in order to maximize the throughput of the secondary user is formulated as a Partially Observable Markov Decision Process (POMDP), and a separation principle is proposed to reveal the optimality of myopic spectrum sensing and accessing strategies. However, this chapter only considers the case where only one channel can be sensed and accessed at each time slot, and the interference between links of concurrent transmissions is not considered, which is obviously not the case when spatial reuse is considered.

There are also several interesting prior works on link scheduling and interference modeling in mmWave networks. For example, the authors in [146] find that the interference between links of concurrent transmission can be ignored in outdoor mmWave networks because of high attenuation of mmWave channel and the extremely small beamwidths of the directional transmissions. Motivated by this observation, a Graph Coloring method is proposed in [162, 163] as a scheduling algorithm to compute a schedule for given traffic demands, such that the total transmission time is

minimized for the mmWave network. These papers consider "pseudowired" mmWave links and do not take the potential co-channel interference (CCI) [51] into consideration.

On the other hand, some prior work rely on a low-rate omni-directional transmission to overcome the deafness problem [164, 165]. The authors of [166] propose the concept of *exclusive regions*, which is described by the relative geo-location and the antenna angle between the transmitter and receiver, to exclude certain concurrent transmissions in the mmWave network. To provide a more accurate attenuation model for the mmWave channel, the authors of [147] conduct extensive urban cellular and peer-to-peer RF wideband channel measurements and find that there are very few unique antenna angles for creating a link, i.e., a directional link is hard to find in mmWave networks. Motivated by the prior works, in this chapter we consider a more general interference and blockage model compared with the previous literature as described in Section 6.3 and develop effective link scheduling algorithms. In our previous work [167], we propose a link scheduling algorithm to minimize the required time length to serve a given data demand for all the nodes. However, the algorithms can only apply to single-hop centralized mmWave downlink networks where a PNC (Piconet coordinator) is required to coordinate the traffic and only downlink transmissions are considered.

6.3 System Model and Problem formulation: The Single-hop Case

6.3.1 System Model

We first consider a mmWave Ad Hoc network consisting of N DEVs with directional transmissions, where all DEVs are within one-hop distance with each other. We assume slotted time with unit length time slots and a common control channel (e.g., a WiFi channel) through which each DEV can broadcast its traffic demands and link statistics to other DEVs. Alternatively, some control time slots can be reserved for the DEVs to exchange traffic demand and link statistics with their neighbors through directional transmissions. A times lot consists of a data transmission phase and an acknowledgement (ACK) phase. Each DEV then executes a link scheduling algorithm to compute the link transmission schedule based on the common information. Finally, the DEVs

point to each other to transmit data according to the transmission schedule. Following the data transmission, the receiving DEV will send feedback (i.e., ACKs) to the sender in the ACK phase. From the feedback, the transmitting DEV can learn the state of the directional link by end of the current time slot and update the remaining traffic demands. The DEV will remain transmitting until it has reached its allocated time slot limit (which is assumed to be sufficiently large in this chapter) or until it has successfully transmitted its traffic demand. Any new traffic that arrives during the current transmission period will be saved in the DEV buffers. When the current traffic demands are all served, the buffer backlogs will be used as new traffic demand for the next scheduling phase, which will be served with a new link schedule during that phase, and so forth.

Denote the set of all links within the one-hop distance as $\mathcal{A}$. Link $l_{ij} \in \mathcal{A}$ is the LOS link from DEV i to DEV j, $1 \leq i, j \leq N$ and $i \neq j$. We assume 2D beam pattern for indoor or low antenna heights in this chapter, although this approach could be generalized to the 3D case. We use the directional antenna gain model $|h_{ij}|^2\Gamma(\theta)$, where $|h_{ij}|^2$ is the maximum gain, θ is the angle offset from the peak gain direction, and $\Gamma(\theta)$ is a non-negative, non-increasing function of θ in [0,1] with $\Gamma(0) = 1$. With the directional antenna gain model, the received signal power of l_{ij} is $P_t|h_{ij}|^2\Gamma^2(0) = P_t|h_{ij}|^2$, where P_t is the transmit power. The interference from transmitter i^* of link $l_{i^*j^*}$ to receiver j of link l_{ij} is $P_t|h_{i^*j}|^2\Gamma(\theta(l_{i^*j}, l_{i^*j^*}))\Gamma(\theta(l_{ji^*}, l_{ji}))$, where $\theta(l_{i^*j}, l_{i^*j^*})$ is the angle between link $l_{i^*j^*}$ and link l_{i^*j}, as shown in Fig. 6.1, and similar for $\theta(l_{ji^*}, l_{ji})$. Let $\Omega(i)$ denote the set of DEVs within one hop distance to node i, $1 \leq i \leq N$, excluding node i itself, and assume that $j \in \Omega(i)$ in the following. Note that $j \in \Omega(i)$ implies $i \in \Omega(j)$. Transmission on an unblocked link l_{ij} will be successful if and only if the SINR at receiver j exceeds a fixed threshold γ, which can be expressed as

$$\frac{|h_{ij}|^2 P_t}{\sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2\Gamma(\theta(l_{i^*j}, l_{i^*j^*}))\Gamma(\theta(l_{ji^*}, l_{ji}))P_t + \sigma^2} \geq \gamma, \quad \text{for all } l_{ij} \in \mathcal{A}. \tag{6.1}$$

where σ^2 is the noise power.

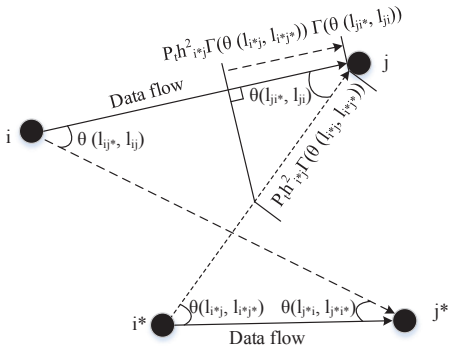

Figure 6.1: Interference between two directional links

We also model the dynamic blockage ((also called shadowing)) of a mmWave link with a discrete-time two-state Markov chain. For link l_{ij}, let G denote the good state (unblocked) and B denote the bad state (blocked); $\Pr(g|b)_{l_{ij}}$ and $\Pr(b|g)_{l_{ij}}$ are the transition probabilities from G to B and from B to G, respectively; $\Pr(g|g)_{l_{ij}} = 1 - \Pr(g|b)_{l_{ij}}$ and $\Pr(b|b)_{l_{ij}} = 1 - \Pr(b|g)_{l_{ij}}$.

6.3.2 Problem Formulation

Define the link state variable s_{ij} as

$$s_{ij} = \begin{cases} 1, & \text{link } l_{ij} \text{ is in the good state} \\ 0, & \text{otherwise,} \end{cases} \quad \forall\, l_{ij}. \tag{6.2}$$

In the ACK phase of each time slot, the receiver DEV returns an ACK to the transmitter DEV if the transmission is successful. Define index variable a as

$$a = \begin{cases} 1, & \text{an ACK is received} \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

Let $U^m(a|s_{ij})$ be the probability that when link l_{ij} is activated at time slot m, the ACK status is a conditioned on that the state of link l_{ij} is s_{ij}. Assuming error-free ACKs, we have

$$\begin{cases} U^m(1|s_{ij}) = s_{ij} \\ U^m(0|s_{ij}) = 1 - U^m(1|s_{ij}) = 1 - s_{ij}, \end{cases} \quad \forall\, m. \tag{6.4}$$

It can be easily seen that $U^m(a|s_{ij})$ is either 1 (i.e., ACK received) or 0 (ACK missing).

Define scheduling index variables x_{ij}^m as

$$x_{ij}^m = \begin{cases} 1, & \text{link } l_{ij} \text{ is activated in time slot } m \\ 0, & \text{otherwise}, \end{cases} \quad \forall\, l_{ij}, m. \tag{6.5}$$

If a DEV works in the Half-Duplex (HD) mode, it can only either transmit or receive at each time slot. We have the following capability constraint for a DEV.

$$\sum_{i \in \Omega(j)} x_{ij}^m + \sum_{z \in \Omega(j)} x_{jz}^m \leq 1,\ \forall\, j, m. \tag{6.6}$$

Since many antenna beam pointing positions are available, it is also possible to configure the DEV to work in the same-channel Full-Duplex (FD) mode [153], where the DEV can transmit to one direction and receive from another direction on the same channel simultaneously [154]. Assuming effective self-interference cancellation and negligible residual self-interference, we have the following capability constraint that holds true for a relay DEV.

$$\sum_{i \in \Omega(j)} x_{ij}^m \leq 1; \quad \sum_{z \in \Omega(j)} x_{jz}^m \leq 1,\ \forall\, j, m. \tag{6.7}$$

The probability that the state of link l_{ij} is s_{ij} at time slot $(m+1)$, denoted as $\lambda_{s_{ij}}^{m+1}$, can be derived as

$$\lambda_{s_{ij}}^{m+1} = x_{ij}^{m+1}(1 - a - s_{ij} + 2as_{ij})$$

$$+ (1 - x_{ij}^{m+1}) \sum_{s'_{ij}=0}^{1} \lambda_{s_{ij}}^{m} \Pr(s_{ij}|s'_{ij}), \tag{6.8}$$

where $\Pr(s_{ij}|s'_{ij})$ is the channel state transition probability. We set the channel state at time slot 0 according to the steady state distribution, as

$$\lambda_{s_{ij}}^{0} = \begin{cases} \frac{\Pr(b|g)_{l_{ij}}}{\Pr(b|g)_{l_{ij}}+\Pr(g|b)_{l_{ij}}}, & s_{ij} = 1 \\ \frac{\Pr(g|b)_{l_{ij}}}{\Pr(b|g)_{l_{ij}}+\Pr(g|b)_{l_{ij}}}, & s_{ij} = 0, \end{cases} \tag{6.9}$$

Define channel utilization index variable t^m for all time slots m as

$$t^m = \begin{cases} 1, \text{one or more links activated at time } m \\ 0, \text{otherwise.} \end{cases} \tag{6.10}$$

Recall that there are N DEVs and let M be a sufficiently large integer indicating a future time slot when all the traffic demands will be certainly transmitted. Let $V_{ij}^{m}(\lambda_{s_{ij}}^{m})$ be the expected amount of traffic delivered by link l_{ij} from time slot m to the future time slot M. We then have

$$\begin{aligned} V_{ij}^{m}(\lambda_{s_{ij}}^{m}) = & \sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^{m} \sum_{a=0}^{1} U^{m}(a|s_{ij})[a r_{ij} t^{m} + \\ & V_{ij}^{m+1}(\lambda_{s_{ij}}^{m+1})], \ \forall\ 1 \leq m \leq M-1, \end{aligned} \tag{6.11}$$

where r_{ij} is the number of packets delivered by link l_{ij} if the transmission on link l_{ij} is successful, i.e., the SINR of link l_{ij} is higher than a threshold and the link is not blocked.

Let c_{ij}^{m} be the expected amount of traffic delivered by link l_{ij}. Recall that x_{ij}^{m} indicates if link l_{ij} is activated or not in time slot m. The expected amount of traffic delivered by link l_{ij} from time slot 1 to time slot M can be derived as

$$V_{ij}^{1}(\lambda_{s_{ij}}^{1}) = \sum_{m=1}^{M} c_{ij}^{m} x_{ij}^{m}, \forall\ l_{ij}, \tag{6.12}$$

where c_{ij}^m can be derived from the link states and feedbacks from time slot 1 to time slot m as

$$c_{ij}^m = \sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^1 \sum_{a=0}^{1} U^1(a|s_{ij}) \left(\sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^2 \sum_{a=0}^{1} U^2(a|s_{ij}) \cdots \right.$$
$$\left. \cdots \left(\sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^m \sum_{a=0}^{1} U^m(a|s_{ij}) a r_{ij} \right) \right). \tag{6.13}$$

Denote the series of packet transmissions, i.e., a session, from DEV i to DEV j as w_{ij}. The total amount of packets to be transmitted is this session from DEV i to DEV j is D_{ij}. We aim to minimize the total amount of time slots, and thus the minimum amount of time, used to serve all the traffic demands under SINR and blockage constraints. The problem, denoted as **P1**, to solve the minimum time used (i.e., the smallest number of time slots) to serve all the traffic demands assuming the 2D directional antenna interference model and the dynamic link blockage model is formulated as

$$\textbf{(P1)} \min: \ \tau = \sum_{m=1}^{M} t^m \tag{6.14}$$

$$\text{s.t.} \ \ V_{ij}^1(\lambda_{s_{ij}}^1) \geq D_{ij}, \ \forall \ l_{ij} \tag{6.15}$$

$$t^m \geq t^{m+1}, \ \forall \ m \tag{6.16}$$

$$x_{ij}^m \in \{0, 1\}, \ \forall \ l_{ij}, m. \tag{6.17}$$

With this formulation, the traffic demands will be served in a certain amount of *consecutive* time slots (for which $t^m = 1$); when the traffic demands are all cleared, we have $t^m = 0$ for all the future time slots, as given in constraints (6.10) and (6.16). Note that the SINR constraint (6.1) is implicitly expressed in the formulated problem in that the expected number of packets transmitted in time slot m, i.e., c_{ij}^m's, are generated by the feasible x_{ij}^m set, for all l_{ij} and m that satisfy the SINR constraint. Furthermore, the feasible x_{ij}^m set should also satisfy the capability constraint (6.6). Note that M can be set to a sufficiently large value so that there is always a feasible solution (i.e., all the traffic demands can certainly be served within M time slots). As we will show later, with the

proposed algorithms, the value of M does not affect the solution and objective value as long as it is sufficiently large.

To formulate the problem when the DEVs can operate in the FD mode, we only need to replace constraint (6.6) (which is implicitly expressed in problem **P1**) with constraint (6.7), which will give us the problem formulation for the FD case, denoted as problem **P2**.

6.4 Solution Algorithms: The Single-hop Case

Problem **P1** is a Binary Integer Programming (BIP) problem. The coefficients of the constraint matrix all take on continuous values between $[0, a]$, which indicate that the BIP does not satisfy the property of *unimodularity* [155]. Thus the BIP cannot be reduced into a Linear Programming (LP) problem. It is, in fact, NP-hard [156].

Furthermore, it is infeasible to list all the columns of the constraint matrix for constraints (6.15) and (6.16)), since the number of all feasible columns in the constraint matrix is as large as M. An exhaustive search to construct the constraint matrix is impractical. Even if it is possible, a huge memory may be needed to store the constraint matrix. In this section, we introduce two effective algorithms to solve the BIP problem with greatly reduced complexity.

6.4.1 Greedy Algorithm

We first propose to solve the BIP problem with an iterative greedy algorithm (GA). The main idea is rather than to minimize $\tau = \sum_{m=1}^{M} t^m$, we instead maximize the instant throughput of the current time slot.

Denote H^m as the set of links whose traffic demands have not yet been served at the m-th iteration (i.e., time slot m).

$$\sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t x^m_{i^*j^*}$$

$$+ \left(\sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t + \sigma^2 - \gamma^{-1} |h_{ij}|^2 P_t \right) x^m_{ij}$$

$$\leq \sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t, \ \forall \ l_{ij} \in \mathcal{A}. \tag{6.20}$$

HD Case

The problem to be solved at the m-th iteration, denoted as **PH**^{m}, can be formulated as

$$\textbf{(PH}^m\textbf{)} \text{ max : } \sum_{l_{ij} \in H^m} c^m_{ij} x^m_{ij} \tag{6.18}$$

$$\text{s.t. constraints } (6.6) \text{ and } (6.20)$$

$$x^m_{ij} \in \{0, 1\}, \ \forall \ l_{ij} \in H^m, \tag{6.19}$$

where constraint (6.20) (given subsequently) ensures that the SINR requirement is satisfied on all of the active links. This problem can be solved with the Branch-and-Bound technique [157], which, for example, is implemented in the $Bintprog$ function in Matlab.

Once the x^m_{ij}'s are obtained, transmissions are scheduled for the current time slot m according to the x^m_{ij}'s. With feedback from DEV j, DEV i will know λ^{m+1}_{ij} and c^{m+1}_{ij}, for all l_{ij}, with which it can solve **PH**$^{m+1}$ to obtain x^{m+1}_{ij} for the next time slot $(m+1)$, for all l_{ij}, until H^m becomes empty or **PH**^{M} is solved.

The iterative greedy algorithm is presented above as in Algorithm 8, which is executed at each DEV. Denote the most recently completed time slot as m_{ij} in which link l_{ij} was activated. In line 11 of Algorithm 8, c^m_{ij} is updated as

$$c^m_{ij} = \sum_{s_{ij}=0}^{1} \lambda^{m_{ij}}_{s_{ij}} \sum_{a=0}^{1} U^{m_{ij}}(a|s_{ij}) \left(\sum_{s_{ij}=0}^{1} \lambda^{m_{ij}+1}_{s_{ij}} \right.$$

$$\sum_{a=0}^{1} U^{m_{ij}+1}(a|s_{ij}) \sum_{a=0}^{1} U^{m_{ij}+1}(a|s_{ij}) \ldots \left(\sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^{m} \sum_{a=0}^{1} U^{m}(a|s_{ij}) a r_{ij} \right) \Bigg), \ \forall \ l_{ij}. \tag{6.21}$$

Algorithm 8: Greedy Algorithm for Problem P1

```
m = 1 ;
while H^m ≠ ∅ do
    Solve PH^m to obtain x_ij^m, for all l_ij ;
    Schedule transmission according to x_ij^m, for all l_ij ;
    m ← m + 1 ;
    Update H^m and D_ij's according to feedback ;
    for l_ij ∈ A do
        for s_ij = 0 : 1 do
            Update λ_{s_ij}^m ;
        end
        Update c_ij^m as in (6.21) ;
    end
end
```

FD Case

For the FD case, we need to replace constraint (6.6) with constraint (6.7) in problem **PH**^{m}, which will give us a new problem for the FD case, denoted as **PF**^{m}. Problem **PF**^{m} can be solved with a similar algorithm as in Algorithm 8.

6.5 System Model and Problem Formulation: The Multi-hop Case

We now consider the case where there may not be an LOS path between any pair of DEVs in the network. Thus intermediate DEVs may be needed to relay traffic for a source DEV to its destination DEV. For a DEV i, recall that the set of DEVs that are within one-hop distance is denoted by $\Omega(i)$. AS before, we denote the session from DEV y to DEV z as w_{yz} and the set of all sessions as $\mathcal{W}$, where $w_{yz} \in \mathcal{W}$. Without loss of generality, we allow a source DEV to transmit multiple sessions, each to a different destination DEV, and a destination DEV to receive multiple

sessions, each from a different source DEV. We need to redefine the scheduling index variables for the multi-hop case. For session w_{yz}, define scheduling index variables $x_{ij}^{m,w_{yz}}$ as

$$x_{ij}^{m,w_{yz}} = \begin{cases} 1, & \text{link } l_{ij} \text{ is activated to transmit session} \\ & w_{yz} \text{in time slot } m \\ 0, & \text{otherwise.} \end{cases}$$

$$\forall\ l_{ij}, w_{yz}. \tag{6.22}$$

From the definitions of $x_{ij}^{m,w_{yz}}$ and t^m (which is the same as that in the single-hop case), we have that $t^m = 1$ if $\sum_{w_{yz}} \sum_{l_{ij}} x_{ij}^{m,w_{yz}} \geq 1$; $t^m = 0$ if $\sum_{w_{yz}} \sum_{l_{ij}} x_{ij}^{m,w_{yz}} = 0$.

The probability that the state of link l_{ij} is s_{ij} at time slot $(m+1)$, i.e., $\lambda_{s_{ij}}^{m+1}$, can now be derived as

$$\begin{aligned} \lambda_{s_{ij}}^{m+1} = & \left(\sum_{w_{yz}} x_{ij}^{m+1,w_{yz}} \right) (1 - a - s_{ij} + 2as_{ij}) + \\ & \left(1 - \sum_{w_{yz}} x_{ij}^{m+1,w_{yz}} \right) \sum_{s'_{ij}=0}^{1} \lambda_{s'_{ij}}^{m} \Pr(s_{ij} | s'_{ij}), \end{aligned} \tag{6.23}$$

Let $V_{ij}^{m,w_{yz}}(\lambda_{s_{ij}}^m)$ be the expected amount of traffic of session w_{yz} delivered by link l_{ij} from time slot m to a future time slot M. We have

$$\begin{aligned} V_{ij}^{m,w_{yz}}(\lambda_{s_{ij}}^m) = & \sum_{s_{ij}=0}^{1} \lambda_{s_{ij}}^m \sum_{a=0}^{1} U^m(a|s_{ij}) \left[ar_{ij} x_{ij}^{m,w_{yz}} + \right. \\ & \left. V_{ij}^{(m+1),w_{yz}}(\lambda_{s_{ij}}^{m+1}) \right], \text{ for } 1 \leq m \leq M-1. \end{aligned} \tag{6.24}$$

The expected amount of traffic delivered by link l_{ij} for session w_{yz} from time slot 1 to M can be derived as

$$V_{ij}^{1,w_{yz}}(\lambda_{s_{ij}}^1) = \sum_{m=1}^{M} c_{ij}^m x_{ij}^{m,w_{yz}}, \tag{6.25}$$

where c_{ij}^m is the same as that in the single-hop case.

Recall that DEV y is the source of session w_{yz} and $\Omega(y)$ is the set of one hop neighbors of node y. For DEV y we have

$$\sum_{m=1}^{M} \sum_{j \in \Omega(y)} c_{yj}^m x_{yj}^{m,w_{yz}} \geq D_{yz}, \ \forall \ 1 \leq y, z \leq N, y \neq z. \tag{6.26}$$

If DEV j relays traffic for session w_{yz} and m' is a time slot, then we have

$$\sum_{m=1}^{m'} \sum_{i \in \Omega(j)} c_{ij}^m x_{ij}^{m,w_{yz}} \geq \sum_{m=1}^{m'} \sum_{k \in \Omega(j)} c_{jk}^m x_{jk}^{m,w_{yz}},$$
$$\forall \ w_{yz}, 1 \leq j \leq N, j \neq z, 1 \leq m' \leq M, \tag{6.27}$$

which means that the traffic received by a DEV should be no less than the traffic it forwards for a flow.

In the HD mode, the following capability constraint holds for any DEV.

$$\sum_{i \in \Omega(j)} \sum_{w_{yz}} x_{ij}^{m,w_{yz}} + \sum_{k \in \Omega(j)} \sum_{w_{yz}} x_{jk}^{m,w_{yz}} \leq 1,$$
$$\forall \ 1 \leq j \leq N. \tag{6.28}$$

In the FD mode, we have the following capability constraint holds for any DEV.

$$\sum_{i \in \Omega(j)} \sum_{w_{yz}} x_{ij}^{m,w_{yz}} \leq 1; \quad \sum_{k \in \Omega(j)} \sum_{w_{yz}} x_{jk}^{m,w_{yz}} \leq 1,$$
$$\forall \ 1 \leq j \leq N. \tag{6.29}$$

DEV z is the destination of session w_{yz}. We have

$$\sum_{m=1}^{M} \sum_{j \in \Omega(z)} c_{jz}^m x_{jz}^{m,w_{yz}} \geq D_{yz}, \ \forall \ w_{yz}. \tag{6.30}$$

$$\sum_{w_{yz}} \sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t x_{i^*j^*}^{m,w_{yz}}$$

$$+ \sum_{w_{yz}} \left(\sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t + \sigma^2 - \gamma^{-1} |h_{ij}|^2 P_t \right) x_{ij}^{m,w_{yz}}$$

$$\leq \sum_{w_{yz}} \sum_{l_{i^*j^*} \neq l_{ij}} |h_{i^*j}|^2 \Gamma(\theta(l_{i^*j}, l_{i^*j^*})) \Gamma(\theta(l_{ji^*}, l_{ji})) P_t, \ \forall \ l_{ij} \in \mathcal{A}. \tag{6.31}$$

The SINR constraint (6.1) can be rewritten as (6.31) on top of the next page, which guarantees that the SINR of each activated link at each time slot be above the SINR threshold.

We aim to minimize the total amount of time slots used to serve the traffic demands under SINR and blockage constraints in the multi-hop network. The problem can be formulated as

$$\textbf{(P3)} \min : \ \tau = \sum_{m=1}^{M} t^m \tag{6.32}$$

$$\text{s.t. Constraints (6.26), (6.27), (6.28), (6.30), } (6.31)$$

$$t^m \begin{cases} = 1, & \text{if } \sum_{w_{yz}} \sum_{l_{ij}} x_{ij}^{m,w_{yz}} \geq 1 \\ = 0, & \text{otherwise,} \end{cases} \quad \forall \ m \tag{6.33}$$

$$t^m \geq t^{m+1}, \ \forall \ m \tag{6.34}$$

$$x_{ij}^{m,w_{yz}} \in \{0, 1\}, \ \forall \ l_{ij}, w_{yz}, m. \tag{6.35}$$

To formulate the problem when the DEVs can operate in the FD mode, we only need to replace constraint (6.28) with constraint (6.29) in Problem **P3**. The new multi-hop problem for the FD case is denoted as problem **P4**.

6.6 Solution Algorithms: The Multi-hop Case

The formulated problem **P3** of the multi-hop case is much more complicated than the single-hop problem **P1**. Therefore we develop a heuristic algorithm to solve problem **P3**. The basic idea is: first, it determines the optimal transmission path for each data flow based on a certain reliability

criterion; next, it maximizes the overall throughput of all the data flows by deciding the set of links to be activated at the current time slot based on the estimated link states and interference.

6.6.1 Path Selection Algorithm

To simplify our problem, we restrict that each data flow should be carried by one and only one path. Denote a path for session w_{yz} as $\mathcal{P}(yz)$ and the set of directional links that belong to $\mathcal{P}(yz)$ as $\mathcal{L}(\mathcal{P}(yz))$. We define the reliability of $\mathcal{P}(yz)$ as the joint reliability probability of all the links along $\mathcal{P}(yz)$, i.e.,

$$\pi(G_{\mathcal{P}(yz)}) = \prod_{l_{ij} \in \mathcal{L}(\mathcal{P}(yz))} \pi(G_{l_{ij}}) = \prod_{l_{ij} \in \mathcal{L}(\mathcal{P}(yz))} \frac{\Pr(g|b)_{l_{ij}}}{\Pr(g|b)_{l_{ij}} + \Pr(b|g)_{l_{ij}}}.$$

Naturally, if we define the link weight (for those valid links l_{ij}) as $\omega_{ij} = -\log(\pi(G_{l_{ij}}))$, then a shortest path routing algorithm will find the path $\mathcal{P}(yz)$ with the largest $\pi(G_{\mathcal{P}(yz)})$, i.e., the most reliable path. In this chapter, we adopt *Dijkstra's Algorithm* to find the shortest path between the source and destination DEV in the network, which has the largest reliability probability among all possible paths.

6.6.2 Link Scheduling Algorithm

After determining the path for each data flow, we decide which links to activate at each time slot. Let the current time slot be m', and denote as $H^{m'}$ the set of sessions whose traffic demands have not been satisfied yet at time slot m'. For a session $w_{yz} \in H^{m'}$, denote the number of its packets that link l_{ij} has delivered at time slot m as $f_{ij}^{m,w_{yz}}$, $1 \leq m \leq m'$. Let $\mathcal{L}(\mathcal{P}(yz))$ be the set of links l_{ij} such that is along path $\mathcal{P}(yz)$ and the traffic it has carried for the session up to time slot m' is less than the demand D_{yz}. Also denote the set of relay nodes along path $\mathcal{P}(yz)$ as $\mathcal{R}(\mathcal{P}(yz))$.

$$\sum_{w_{yz}\in H^{m'}}\sum_{l_{i^*j^*}\in\mathcal{L}(\mathcal{Q}(yz)),l_{i^*j^*}\neq l_{ij}}|h_{l_{i^*j}}|^2G_0^2\Gamma(\theta(l_{ij},l_{i^*j^*}))\Gamma(\theta(l_{ji^*},l_{ji}))P_t x_{l_{i^*j^*}}^{m',w_{yz}}$$
$$+\sum_{w_{yz}\in H^{m'}}\left(\sum_{l_{i^*j^*}\in\mathcal{L}(\mathcal{Q}(yz)),l_{i^*j^*}\neq l_{ij}}|h_{l_{i^*j}}|^2G_0^2\Gamma(\theta(l_{ij},l_{i^*j^*}))\Gamma(\theta(l_{ji^*},l_{ji}))P_t+\sigma^2-\gamma^{-1}|h_{l_{ij}}|^2G_0^2P_t\right)x_{l_{ij}}^{m',w_{yz}}$$
$$\leq\sum_{w_{yz}\in H^{m'}}\sum_{l_{i^*j^*}\in\mathcal{L}(\mathcal{Q}(yz)),l_{i^*j^*}\neq l_{ij}}|h_{l_{i^*j}}|^2G_0^2\Gamma(\theta(l_{ij},l_{i^*j^*}))\Gamma(\theta(l_{ji^*},l_{ji}))P_t,\text{ for all }l_{ij}\in\mathcal{L}(\mathcal{Q}(yz)). \tag{6.40}$$

For the HD case, we aim at maximizing the instant overall throughput of links in $\mathcal{L}(\mathcal{P}(yz))$, for all $w_{yz}\in H^{m'}$ at time slot m', by solving Problem $\mathbf{LS}_h^{m'}$ defined as follows.

$$\mathbf{LS}_h^{m'}\ \min:\ \sum_{w_{yz}\in H^{m'}}\sum_{l_{ij}\in\mathcal{L}(\mathcal{P}(yz))}c_{ij}^{m'}x_{ij}^{m',w_{yz}} \tag{6.36}$$

s.t. SINR Constraint (6.40) (at top of the next page)

$$\sum_{m=1}^{m'-1}f_{ij}^{m,w_{yz}}+c_{ij}^{m'}x_{ij}^{m',w_{yz}}\leq\sum_{m=1}^{m'-1}f_{jk}^{m,w_{yz}},\ \forall\, l_{ij},$$
$$l_{jk}\in\mathcal{L}(\mathcal{P}(yz)),j\in\mathcal{R}(\mathcal{P}(yz)),w_{yz}\in H^{m'} \tag{6.37}$$
$$\sum_{w_{yz}\in H^{m'}}\sum_{l_{ij}\in\mathcal{L}(\mathcal{P}(yz))}x_{ij}^{m',w_{yz}}+\sum_{w_{yz}\in H^{m'}}\sum_{l_{jk}\in\mathcal{L}(\mathcal{P}(yz))}x_{jk}^{m',w_{yz}}$$
$$\leq 1,\qquad\forall\, j\in\mathcal{R}(\mathcal{P}(yz)),w_{yz}\in H^{m'} \tag{6.38}$$
$$x_{ij}^{m,w_{yz}}\in\{0,1\},\ \forall\, l_{ij}\in\mathcal{L}(\mathcal{P}(yz)),w_{yz}\in H^{m'}. \tag{6.39}$$

Constraint (6.37) indicates the flow conservation condition for each relay node, i.e., the volume of transmitted data must not exceed the volume of received data. Constraint (6.38) corresponds to the HD constraint (6.28) in Problem **P3**. Problem $\mathbf{LS}_h^{m'}$ can be solved using the *Binary Integer Programming* technique [155]. The greedy algorithm for solving problem **P3** is given in Algorithm 9.

Algorithm 9: Greedy Algorithm for Problem P3

Set link weights to $-\log(\pi(G_{l_{ij}}))$ for all valid link l_{ij} ;
Use Dijkstra's Algorithm to find the shortest path $\mathcal{P}(yz)$ for each session w_{yz} ;
$m' = 1$;
while $H^{m'} \neq \emptyset$ **do**
    Solve $\mathbf{LS}_h^{m'}$ to obtain $x_{ij}^{m',w_{yz}}$, for all l_{ij} ;
    Schedule link transmissions according to $x_{ij}^{m',w_{yz}}$'s, for all l_{ij} ;
    $m' = m' + 1$;
    Updates $H^{m'}$ and $\mathcal{L}(\mathcal{P}(yz))$, for all $w_{yz} \in H^{m'}$;
    for $l_{ij} \in \mathcal{L}(\mathcal{P}(yz)), \forall\, w_{yz} \in H^{m'}$ **do**
        for $s_{ij} = 0 : 1$ **do**
            Update $\lambda_{s_{ij}}^{m'}$;
        end
        Update $c_{ij}^{m'}$;
    end
end

In the FD case, the flow conservation constraint for a relay node $j \in \mathcal{R}(\mathcal{P}(yz))$ becomes

$$\sum_{m=1}^{m'-1} f_{ij}^{m,w_{yz}} + c_{ij}^{m'z} x_{ij}^{m',w_{yz}} \leq \sum_{m=1}^{m'-1} f_{jk}^{m,w_{yz}} + c_{jk}^{m',w_{yz}} x_{jk}^{m',w_{yz}},$$
$$\forall\, l_{ij}, l_{jk} \in \mathcal{L}(\mathcal{P}(yz)), j \in \mathcal{R}(\mathcal{P}(yz)), w_{yz} \in H^{m'}. \tag{6.41}$$

The capability constraint for relay node $j \in \mathcal{R}(\mathcal{P}(yz))$ becomes

$$\sum_{z \in H(m')} \sum_{l_{ij} \in \mathcal{L}(\mathcal{P}(yz))} x_{ij}^{m',w_{yz}} \leq 1, \forall j \in \mathcal{R}(\mathcal{P}(yz)), w_{yz} \in H^{m'} \tag{6.42}$$

$$\sum_{z \in H(m')} \sum_{l_{jk} \in \mathcal{L}(\mathcal{P}(yz))} x_{jk}^{m',w_{yz}} \leq 1, \forall j \in \mathcal{R}(\mathcal{P}(yz)), w_{yz} \in H^{m'}. \tag{6.43}$$

Replacing constraints (6.37) and (6.38) in $\mathbf{LS}_h^{m'}$ with constraints (6.41)–(6.42) yields the instant throughput maximization problem for the FD case for the multi-hop scenario, denoted as problem $\mathbf{LS}_f^{m'}$. The FD problem $\mathbf{LS}_f^{m'}$ can be solved with a similar algorithm as in Algorithm 9.

Table 6.1: Simulation Parameters

Parameter	*Value*	*Parameter*	*Value*
P_t	1 watt	$\|h_{l_{ij}}\|^2\theta(l_{i^*j}, l_{i^*j^*})$	$[0,1]$, $\forall l_{ij}$
σ^2	0.1 watt	r_{ij}	10 pkts/slot, $\forall l_{ij}$
N	10	D_z	$[50,60]$ pkts, $\forall z$

6.7 Simulation Study

6.7.1 Simulation Setup

In this section we use Matlab simulations to validate the performance of the proposed algorithms. Unless otherwise specified, the values of simulation parameters are set as shown in Table 6.1. Each simulated point in the figures is obtained by repeating the simulation 50 times with different random seeds, and 95% confidence intervals are computed and plotted in the figures as error bars to guarantee credible results. We compare the performance of our proposed algorithms with that of a benchmark scheme proposed in [158]. The benchmark scheme does not consider link blockages when making routing selection and solving the subproblems with the greedy method. With the benchmark scheme, $c_{ij}^m = 1$, for all l_{ij}, and $\pi(G_{l_{ij}}) = 1$, for all l_{ij}, in routing selection. It randomly selects a route. Note that the capacity of a link is affected by the noise plus interference level in the benchmark scheme. We set the total number of sessions in the network equals to $\frac{N}{2}$, where N is the total number of DEVs. We allow a DEV to have multiple sessions.

6.7.2 Simulation Results and Analysis

The performance of the proposed algorithms under different SINR thresholds γ is shown in Fig. 6.2 and Fig. 6.3 for the HD mode single-hop and multi-hop case, respectively. All the algorithms have degraded performance when γ is increased. The reason is that a larger γ means that for a specific link, given a fixed channel gain and transmission power, a lower interference can be tolerated. Fewer concurrent transmissions can be accommodated in the system to leverage

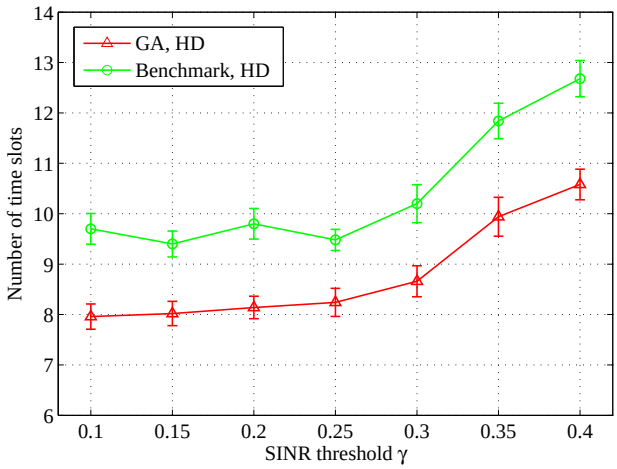

Figure 6.2: Number of time slots to serve the traffic demand under various SINR threshold γ: HD and single-hop case ($\Pr(g|b)_{l_{ij}} = \Pr(b|g)_{l_{ij}} = [0.3, 0.6]$, for all l_{ij}).

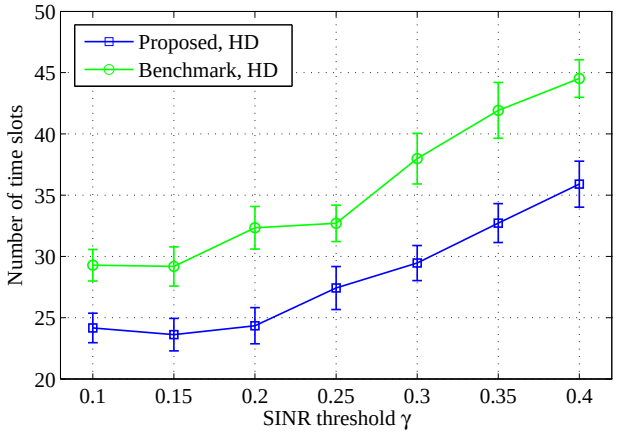

Figure 6.3: Number of time slots to serve the traffic demand under various SINR threshold γ: HD and multi-hop case ($\Pr(g|b)_{l_{ij}} = \Pr(b|g)_{l_{ij}} = [0.4, 0.7]$, for all l_{ij}).

spatial reuse and the throughput of each time slot is reduced. Therefore, the number of time slots needed to satisfy the traffic demand is increased as γ grows.

It would also be helpful to examine constraint (6.20) of problem **PH**^{m} in Section 6.4. It can be seen that if γ is increased, the value of the left hand side (LHS) of (6.20) will also increase, which means that constraint (6.20) will become tighter. Therefore, **PH**^{m} will have a smaller solution space and its optimal objective value may be reduced. The system throughput in each time slot may be reduced as a result and the time length needed to schedule all the traffic will be prolonged.

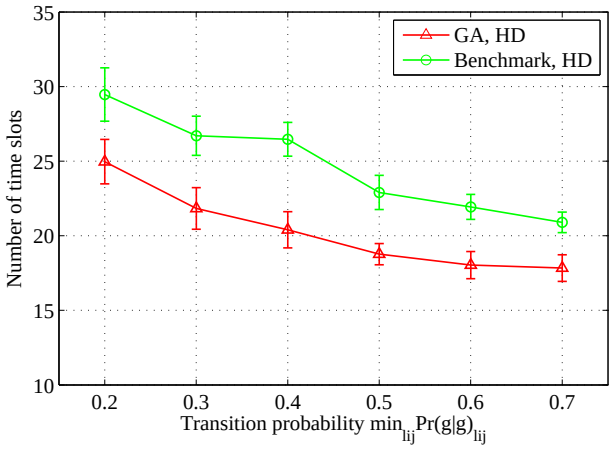

Figure 6.4: Number of time slots to serve the traffic demand under various transition probabilities for the mmWave channels, $\Pr(g|g)_{l_{ij}}$: HD single-hop case ($\Pr(g|b)_{l_{ij}} = [0.2, 0.4]$, and $\gamma = 0.4$).

A similar observation can be obtained for the multi-hop case if we check the SINR constraint in the subproblem $\mathbf{LS}_h^{m'}$.

The performance gap on scheduling time between our proposed scheme and the benchmark scheme under different minimum link state transition probabilities $\min_{l_{ij}}\{\Pr(g|g)_{l_{ij}}\}$ is shown in Figs. 6.4 and 6.5 for the HD single-hop case and HD multi-hop case, respectively. It can be seen from (6.9) that as $\Pr(g|g)_{l_{ij}} = 1 - \Pr(b|g)_{l_{ij}}$ is increased, the stationary probability of link l_{ij} in the good state is increased. Thus more links may have successful concurrent transmissions at each time slot, which improves the network throughput. It may take less time slots for a DEV to deliver the requested packets to its neighboring DEVs; for the DEVs, the transmission delay between two adjacent DEVs on a path will be shortened due to a greater probability of successful transmissions. These two factors are the main reasons for the shortened scheduling time when $\min_{l_{ij}}\{\Pr(g|g)_{l_{ij}}\}$ is increased.

However, the decrement cannot go indefinitely and the time length needed will converge to a certain threshold, regardless of the increase of $\min_{l_{ij}}\{\Pr(g|g)_{l_{ij}}\}$ once it goes beyond a specific value. The main reason is that since a certain level of SINR must be satisfied for a successful transmission, the maximum number of links that can transmit concurrently is limited. Even if $\pi(G_{l_{ij}})$ is large for all l_{ij}, i.e., each link has a high probability of being in good state, the maximum

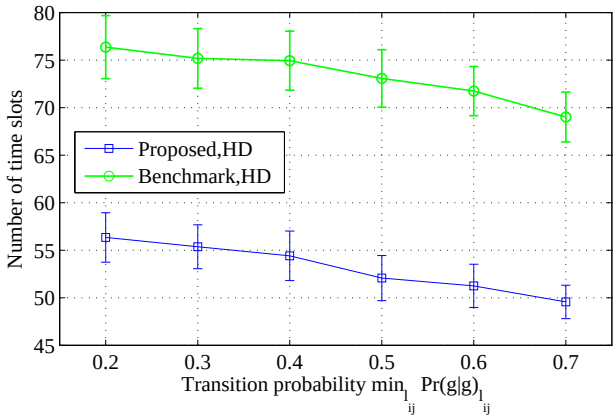

Figure 6.5: Number of time slots to serve the traffic demand under various transition probabilities for the mmWave channels, $\Pr(g|g)_{l_{ij}}$: HD multi-hop case ($\Pr(g|b)_{l_{ij}} = [0.3, 0.7]$, and $\gamma = 0.4$).

number of concurrent transmission links is still limited. Therefore we have a bounded network throughput at each time slot and thus a lower bound on the minimum time length to schedule the traffic demand of all the links.

Figures 6.6 and 6.7 present the performance comparison of the proposed and the benchmark algorithm under different SINR thresholds γ for the FD single-hop case and FD multi-hop case, respectively. We can see that for all the algorithms, FD achieves a better performance than HD does. The main reason is that FD allows more concurrent transmissions. The FD network is expected to achieve a higher throughput than the HD network, which leads to a lower scheduling time. The proposed scheme outperforms the benchmark in both FD and HD modes, since it considers the dynamic link states and always chooses the best set of links in each time slot.

Examining constraints (6.6) and (6.7) for the single-hop case will also help us to better understand the advantage of FD over HD. The HD constraint (6.6) is a subset of the FD constraint (6.7). The feasible solution region for the FD case is larger than that of the HD case. Thus a better solution may be found and the optimal objective value can be increased, which translates to a higher network throughput for FD. Similar observations can be made for the multi-hop case by checking constraints (6.38), (6.42), and (6.43).

The performance gap between the proposed algorithms and the benchmark under different link state transition probabilities $\min_{l_{ij}}\{\Pr(g|g)_{l_{ij}}\}$ for the FD single-hop case and FD multi-hop case

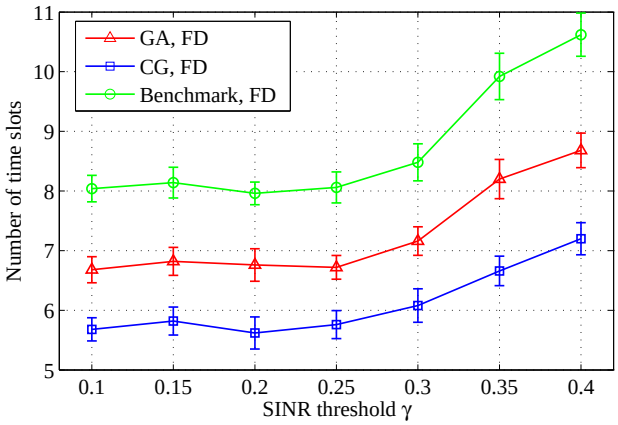

Figure 6.6: Number of time slots to serve the traffic demand under various SINR threshold γ: FD single-hop case ($\Pr(g|b)_{l_{ij}} = \Pr(b|g)_{l_{ij}} = [0.3, 0.6]$, for all l_{ij}).

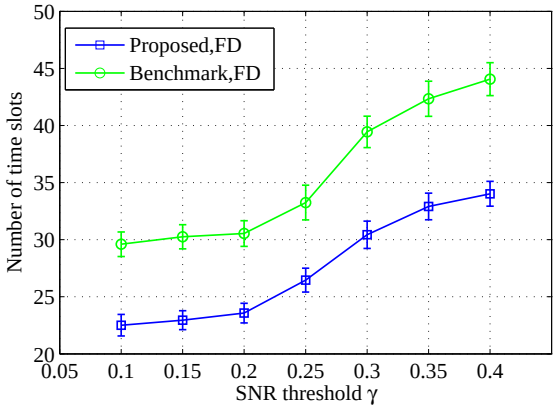

Figure 6.7: Number of time slots to serve the traffic demand under various SINR threshold γ: FD multi-hop case ($\Pr(g|b)_{l_{ij}} = \Pr(b|g)_{l_{ij}} = [0.4, 0.7]$, for all l_{ij}).

is shown in Figs. 6.8 and 6.9, respectively. The previous explanation for the changing γ scenario at the FD mode is also applicable for the simulation results of the changing $\min_{l_{ij}}\{\Pr(g|g)_{l_{ij}}\}$ scenario.

6.8 Conclusion

In this chapter, we investigated the problem of minimum time length scheduling in mmWave Ad Hoc networks under both traffic demand and SINR constraints. We considered both single-hop and multi-hop ad hoc networks and HD and FD transmission scenarios. We formulated the

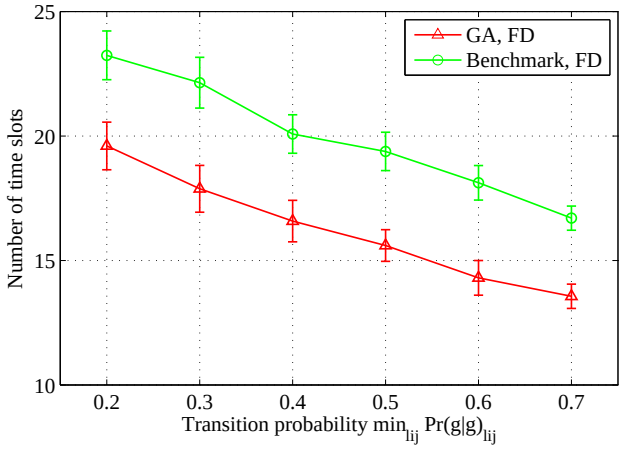

Figure 6.8: Number of time slots to serve the traffic demand under various transition probabilities for the mmWave channels $\Pr(g|g)_{l_{ij}}$: FD single-hop case ($\Pr(g|b)_{l_{ij}} = [0.2, 0.4]$, and $\gamma = 0.4$).

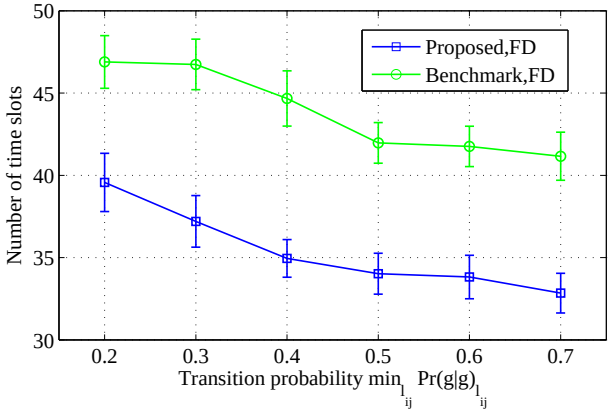

Figure 6.9: Number of time slots to serve the traffic demand under various transition probabilities for the mmWave channels $\Pr(g|g)_{l_{ij}}$: FD multi-hop case ($\Pr(g|b)_{l_{ij}} = [0.3, 0.7]$, and $\gamma = 0.4$).

minimum time length scheduling problems adopting a general directional interference model and a dynamic channel blocking model, and developed effective solution algorithms. Simulation results validated the performance of our proposed algorithms by comparison with a benchmark scheme.

Chapter 7

Conclusions

In this dissertation work, we investigate the problem of QoS/QoE provision in emerging wireless networks such as mmWave and networks. mmWave technology and CR technology was recently proposed to enhance the wireless network capacity, but both are suffered from channel condition uncertainty, therefore there are many technical challenges for incorporating these techniques to existing wireless networks. For instance, how to adjust the coding schemes to adapt to the changing channel conditions, how to coordinate the transmissions of the PUs and that of the CUs, and how to optimally assign radio resources and adopt transmission power for users, and how to optimally coordinate the concurrent transmissions of neighboring links, are crucial issues for the performance improvement of mmWave and CR networks. To address these challenges motivates this dissertation work.

In chapter 2, the problem of streaming uncompressed HD video over mmWave wireless networks is investigated. The objective is to minimize the expected mean square error of the reconstructed video quality. An MD-FEC coding scheme that partitions the pixel bits is proposed so that more important bits have higher level of protection thus have a higher probability of recovery at the receiver, an interleaving based transmission strategy to avoid busty errors in transmission. We formulated an Nonlinear Integer Programming problem for the optimal partition of the pixel bits and interleaving of packets, which is NP-hard, and derived a sub-optimal solution for this problem with much lower computational complexity. The performance of the proposed scheme was evaluated with simulations.

In chapter 3, the problem of video streaming over CRNs is investigated, aiming at maximizing the CU Quality of Experience (QoE) by optimal designs of spectrum sensing and access policies.

Due to the non-linearity of the QoE model, we solve the spectrum sensing problem and the spectrum accessing problem separately. The CUs chosen to sense a channel is optimized such that the false alarm probability of the channel is minimized , and the CUs allowed to access a channel is optimized such that the network throughput is maximized. For the case where each CU can sense and access at most one channel at a time, we derive simple algorithms to solve the channel sensing and accessing problem and prove its optimaility. For the case where each CU can sense multiple channels but access only one channel, we presented a more general, integrated formulation. Based on an assumption on the spectrum sensor configuration, we developed a two-step approach to solve the integrated problem and proved its optimality. The proposed schemes were shown to outperform several alternative schemes in the simulation study.

In chapter 4, the problem of video streaming over CRNs is investigated, aiming at maximizing the CU Quality of Service (QoS) by optimal designs of spectrum sensing, access and power allocation. Different with the previous work which solves the spectrum sensing problem and accessing problem separately, in this chapter the spectrum sensing, channel assignment, and power allocation strategies are jointly optimized to maximize the QoS for the CUs. What's more, in this chapter each CU can sense and access multiple channels at a timeWe show that the formulated Mixed Integer NonLinear Programming problem can be decomposed into two sub-problems: SP1 for the optimal spectrum sensing strategy, and SP2 for the optimal channel assignment and power allocation, without sacrificing optimality. We show that SP1 can be optimally solved under certain conditions and develop an algorithm to solve SP2 iteratively in a distributed manner. We also develop a heuristic algorithm for spectrum sensing with greatly reduced requirement on CU hardware, while still achieving a highly competitive sensing performance. We analyze the proposed algorithms with respect to complexity and time efficiency, and derive a performance upper bound. The proposed algorithms are validated with simulations.

In chapter 5, the relay and link selection problem in a dual-hop mmWave network is studied. The objective is to minimize the Maximum Expected Delivery Time among all Tx-Rx pairs by jointly optimizing relay and link selection, while exploiting reflected mmWave transmissions and

considering link blockage dynamics. We develop a Decomposition Principle to transform this problem into two sub-problems, one for link selection and the other for relay assignment when there is enough replays. We prove that the proposed scheme can achieve an optimality gap of just 1 time slot at greatly reduced complexity. We also develop a heuristic scheme to handle the case when there is no enough relays. The proposed schemes are validated with simulations with their superior performance observed.

In chapter 6, the problem of minimizing the scheduling time length to serve users traffic demand by user scheduling in one hop and multi-hop mmWave wireless networks is studied. Different with the previous chapter which considers the outdoor transmission environment and assumes the links are *psudo-wired*, in this chapter we consider the indoor transmission environment and there are interference between neighboring links and thus the *psudo-wired* assumption doesn't hold. Channel conditions change overtime, and at each time slot, the PNC decides which users should access the channel based on the possible channel state at current slot, aiming at maximizing the long term utility of the whole network. The formulated problem incorporates a flexible interference model for directional transmissions and a Markov chain based blockage model. We propose efficient algorithms with greatly reduced complexity to solve the one-hop problem and multi-hop problem, respectively. For the one-hop problem, the instant network throughput is maximized at each time slot; for the multi-hop problem, we first fix the optimal routing path for a long term and then maximizes the instant network throughput. In this way, we achieve a balance between long term utility and short term utility. The performance of the heuristic algorithm is validated with simulations.

Bibliography

[1] Z. He, S. Mao. Adaptive multiple description coding and transmission of uncompressed video over 60GHz networks. *ACM ACM Mobile Computing and Communications Review (MC2R)*, 18(1):14–24, Jan. 2014.

[2] Federal Communications Commission. Spectrum Policy Task Force. Rep. ET Docket no.02–135, Nov. 2002.

[3] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. *IEEE IEEE Trans. Inform. Theory*, 42(6):1737–1744, Nov. 1996.

[4] A. C. Begen, Y. Altunbasak, O. Ergun and M. H. Ammar. Multi-path selection for multiple description video streaming over overlay networks. *EURASIP Signal Processing: Image Commun.*, 20(1):39–60, Jan. 2005.

[5] B. Bosco. Adaptive cross layer design and implementation for Gigabit multimedia applications using 60 GHz wireless links. Master's thesis, Arizona State Univ., Phoenix, AZ, 2011.

[6] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. [online] Available: `http://www.cisco.com`. May 2012.

[7] Federal Communications Commission. FCC 13-112: In the Matter of Revision of Part 15 of the Commission's Rules Regarding Operation in the 57-64 GHz Band. *ET Docket No. 07-113*, Aug. 9, 2013.

[8] M. X. Gong, D. Akhmetov, R. Want, and S. Mao. Directional CSMA/CA protocol with spatial reuse for mmWave wireless networks. In *Proc. IEEE GLOBECOM 2010*, pages 1–5, Miami, FL, Dec. 2010.

[9] W. Lee, K. Noh, S. Kim, and J. Heo. Efficient cooperative transmission for wireless 3D HD video transmission in 60GHz channel. *IEEE Trans. Consum. Electron.*, 56(4):2481–2488, Nov. 2010.

[10] M. Manohara, R. Mudumbai, J. Gibson, and U. Madhow. Error correction scheme for uncompressed HD video over wireless. In *Proc. IEEE ICME 2009*, pages 802–805, New York, NY, June/July 2009.

[11] S. Mao, D. Bushmitch, S. Narayanan, and S. S. Panwar. MRTP: A multi-flow real-time transport protocol for ad hoc networks. *IEEE Trans. Multimedia*, 8(2):356–369, Apr. 2006.

[12] S. Mao, X. Cheng, Y. T. Hou, and H. D. Sherali. Multiple description video multicast in wireless ad hoc networks. *ACM/Springer Mobile Netw. Appl.*, 11(1):63–73, Jan. 2006.

[13] S. Mao, S. Lin, S. Panwar, Y. Wang, and E. Celebi. Video transport over ad hoc networks: Multistream coding with multipath transport. *IEEE J. Sel. Areas Commun.*, 21(10):1721–1737, Dec. 2003.

[14] S. Mao, S. Lin, Y. Wang, S. S. Panwar, and Y. Li. Multipath video transport over wireless ad hoc networks. *IEEE Wireless Commun.*, 12(4):42–49, Aug. 2005.

[15] A. E. Mohr, E. A. Riskin, and R. E. Lader. Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction. *IEEE J. Sel. Areas Commun.*, 18(6):819–828, Jun. 2000.

[16] T. S. Rappaport, J. N. Murdock, and F. Gutierrez. State of the art in 60-GHz integrated circuits and systems for wireless communications. *Proc. IEEE*, 99(8):1390–1436, Aug. 2011.

[17] L. Rizzo. Effective erasure codes for reliable computer communication protocols. *ACM Computer Commun. Rev.*, 27(2):24–36, Apr. 1997.

[18] H.-R. Shao and et al. Adaptive multi-beam transmission of uncompressed video over 60GHz wireless systems. In *Proc. Future Generation Commun. Netw. 2007*, pages 430–435, Jeju-Island, South Korea, Dec. 2007.

[19] H. Singh, X. Qin, H. Shao, C. Ngo, C. Kwon, and S. S. Kim. Support of uncompressed video streaming over 60GHz wireless networks. In *Proc. IEEE CCNC 2008*, pages 243–248, Las Vegas, NV, Jan. 2008.

[20] I.-K. Son, S. Mao, M. Gong, and Y. Li. On frame-based scheduling for directional mmWave WPANs. In *Proc. IEEE INFOCOM 2012*, pages 2149–2157, Orlando, FL, Mar. 2012.

[21] Y. Wang, A. Reibman, and S. Lin. Multiple description coding for video delivery. *Proc. IEEE*, 93(1):57–70, Jan. 2005.

[22] L. R. Wilhelmsson and L. B. Milstein. On the effect of imperfect interleaving for the Gilbert Elliott channel. *IEEE Trans. Commun.*, 47(5):681–688, May 1999.

[23] Y. Zhao, S. Mao, J. Neel, and J. H. Reed. Performance evaluation of cognitive radios: metrics, utility functions, and methodologies. *Proc. IEEE*, 97(4):642–659, Apr. 2009.

[24] Z. He and S. Mao, QoS driven multi-user video streaming in cellular CRNs: The case of multiple channel access. *Proc. IEEE MASS 2014*, Philadelphia, PA, Oct. 2014, pp.28–36.

[25] Cisco. Visual Networking Index (VNI). Feb. 2014. [Online]. Available: http://www.cisco.com/.

[26] Y. Zhao, S. Mao, J. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: metrics, utility functions, and methodologies," *Proc. IEEE*, vol.97, no.4, pp.642–659, Apr. 2009.

[27] D. Hu, S. Mao, Y. T. Hou, and J. H. Reed, "Scalable video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, Special Issue on Wireless Video Transmission, vol.29, no.3, pp.334–344, Apr. 2010.

[28] D. Hu, and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol.11, no.9, pp.3501–3511, Nov. 2011.

[29] S. Soltani, and M. Mutka, "Decision tree modeling for video routing in cognitive radio mesh networks," *Proc. IEEE WoWMoM'13*, Madrid, Spain, June. 2013, pp.1–9.

[30] S. Li, Tom. Luan, and X. Shen "Channel allocation for smooth video delivery over cognitive radio networks," *Proc. IEEE GLOBECOM 2010*, Miami, USA, DEC. 2010, pp.1–5.

[31] R. Yao, Y. Liu, J. Liu, P. Zhao, and S. Ci "Perceptual experience oriented transmission scheduling for scalable video streaming over cognitive radio networks," *Proc. IEEE GLOBECOM 2013*, Atlanta, USA, DEC. 2013, pp.1681–1686.

[32] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol.54, no.5, pp.2053–2071, May 2008.

[33] X. Zhang and H. Su "CREAM-MAC: Cognitive radio-enabled multi-channel MAC protocol over dynamic spectrum access networks," *IEEE J. Sel. Topics Signal Process.*, vol.17, no.9, pp.110–123, Feb. 2011.

[34] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol.26, no.1, pp.118–129, Jan. 2008.

[35] H.A.B Salameh, M. Krunz, and D. Manzi, "Spectrum bonding and aggregation with guard-band awareness in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol.13, no.3, pp.569–581, 2014.

[36] M. Wien, H. Schwarz, and T. Oelbaum "Performance Analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol.5, no.1, pp.1194–1203, Sept. 2007.

[37] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali *Linear Programming and Network Flows*, 4th edition, *Wiley*, New York, NY, 2009.

[38] Y.C. Liang, Y. Zheng, E. Peh, and A.T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol.7, no.4, pp.1326–1337, Apr. 2008.

[39] W. Wang, B. Kasiri, J. Cai, and A.S. Alfa. "Channel assignment of cooperative spectrum sensing in multi-channel cognitive radio networks," *Proc. IEEE ICC'11*, kyoto, Japan, June. 2011, pp.1–5.

[40] A. W. Min, Kang G. Shin, "Joint optimal sensor selection and scheduling in dynamic spectrum access networks," *IEEE Trans. Mobile Comput.*, vol.12, no.8, pp.1532–1545, Aug. 2013.

[41] Y. Chen, Y. Wu, B. Wang, and K. J. R. Liu, "Spectrum auction games for multimedia streaming over cognitive radio networks," *IEEE Trans. Commun.*, vol.58, no.8, pp.2381–2390, Aug. 2010.

[42] S. Ali and F. Yu, "Cross-layer QoS provisioning for multimedia transmissions in cognitive radio networks," *Proc. IEEE WCNC'09*, Budapest, Hungary, Apr. 2009, pp.1–5.

[43] N. Tang, S. Mao, and S. Kompella, "Power control in full duplex underlay cognitive radio networks: A control theoretic approach," *Proc. IEEE MILCOM'14*, Baltimore, MA, Oct. 2014, pp.949–954.

[44] K. Xiao, S. Mao, and J. Tugnait, "Congestion control for infrastructure-based CRNs: A multiple model predictive control approach," *Proc. IEEE GLOBECOM'16*, Washington DC, Dec. 2016.

[45] Z. Jiang and S. Mao, "Opportunistic spectrum sharing in LTE-unlicensed with Lyapunov optimization based auction," *IEEE Transactions on Vehicular Technology*, DOI:10.1109/TVT.2016.2613444.

[46] M. Feng, S. Mao, and T. Jiang, "Duplex mode selection and channel allocation for full-duplex cognitive femtocell networks," *Proc. IEEE WCNC 2015*, New Orleans, LA, Mar. 2015, pp.1900–1905.

[47] Y. Wang and S. Mao, "Distributed power control in full duplex wireless networks," *Proc. IEEE WCNC 2015*, New Orleans, LA, Mar. 2015, pp.1165–1170.

[48] X. Wang, S. Mao, and M.X. Gong, "A survey of LTE Wi-Fi coexistence in unlicensed bands," *ACM Mobile Computing and Communications Review (MC2R)*

[49] Y. Wang, S. Mao, and R.M. Nelms, "Optimal hierarchical power scheduling for cooperative microgrids," *Proc. IEEE MASS 2014*, Philadelphia, PA, Oct. 2014, pp.497–498.

[50] H. Zhou, D. Hu, S. Mao, and P. Agrawal, "Joint relay selection and power allocation in cooperative FSO networks," *Proc. IEEE GLOBECOM 2013*, Atlanta, GA, Dec. 2013, pp.2418–2423.

[51] Y. Xu and S. Mao, "Distributed interference alignment in cognitive radio networks," *Proc. IEEE ICCCN 2013*, Nassau, Bahamas, July/Aug. 2013, pp1–7.

[52] Z. Jiang and S. Mao, "Online channel assignment, transmission scheduling, and transmission mode selection in multi-channel full-duplex wireless LANs," *Proc. The 10th International Conference on Wireless Algorithms, Systems, and Applications*, LNCS 9204, K. Xu and J. Zhu (Eds.), Qufu, P.R. China, Aug. 2015, pp.243–252.

[53] Y. Wang, S. Mao, and R. M. Nelms, "A distributed online algorithm for optimal real-time energy distribution in smart grid," *Proc. IEEE GLOBECOM 2013*, Atlanta, GA, Dec. 2013, pp.1644–1649.

[54] G.B. Dantzig and P. Wolfe, "The decomposition principle for linear programs," *Operations Research*, vol.8, no.1, pp.101–111, Feb. 1960.

[55] J.J. Lyu, H. Luh, and M.C. Lee "Solving linear programming problems on the parallel virtual machine environment," *American J. Appl. Sci.*, vol.1, no.2, pp.90–94, 2004.

[56] R. Garfinkel and G.L. Nemhauser *Integer Programming*, *Wiley*, New York, NY, 1972.

[57] I. Griva, S.G. Nash, and A. Sofer *Linear and Nonlinear Optimization*, 2nd edition, *Society for Industrial and Applied Mathematics*, Philadelphia, PA, 2009.

[58] T. Jiang, H. Wang, and A. Vasilakos, "QoE-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol.30, no.7, pp.1215–1224, Aug. 2012.

[59] A. Khan, L. Sun, and E. Ifeachor. "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol.14, no.2, pp.431–442, Apr. 2012.

[60] Z. He, S. Mao, and S. Kompella, "Quality of Experience driven multi-user video streaming in cellular cognitive radio networks with single channel access," in *IEEE Transactions on Multimedia*, DOI: 10.1109/TMM.2016.2564104.

[61] Z. He, S. Mao, and S. Kompella, "QoE driven video streaming in cognitive radio networks: Case of single channel access," in *Proc. IEEE GLOBECOM 2014*, Austin, TX, Dec. 2014.

[62] Cisco, "Visual Networking Index (VNI)," Feb. 2014. [Online]. Available: http://www.cisco.com/.

[63] D. Hu, and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol.11, no.9, pp.3501–3511, Nov. 2011.

[64] K. Yamagishi and T. Hayashi, "Opinion model using psychological factors for interactive multimodal services," *IEICE Trans. Communication.*, E89-B(2):281–288, Feb. 2006.

[65] J. You, U. Reiter, M. Hannuksela, M, Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio-visual services: A images," *Signal Processing: Image Communication.*, vol.25, no.7, pp.482–501, Aug. 2010.

[66] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," in *Proc. IEEE ICC'09.*, Dresden, Germany, June 2009, pp.1–5.

[67] K. Yamagishi, T. Tominaga, T. Hayashi, and A. Takashi, "Objective quality evaluation model for videophone services," *NTT Technical Review*, vol.5, no.6, pp.1–5, June 2007.

[68] A. Eden, "No-reference image quality analysis for compressed video sequences," *IEEE Trans. Broadcast.*, vol.54, no.3, pp.691–697, Sept. 2008.

[69] Q. Huynh-Thu and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," *IEEE Trans. Broadcast.*, vol.54, no.3, pp.641–651, Sept. 2008.

[70] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali *Linear Programming and Network Flows*, 4th edition, *Wiley*, New York, NY, 2009.

[71] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol.14, no.2, pp.431–442, Apr. 2012.

[72] Y. Zhao, S. Mao, J. Neel, and Jeffrey H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodologies," *Proceedings of the IEEE*, vol.97, no.4, pp.642-659, Apr. 2009.

[73] Y.C. Liang, Y. Zheng, E. Peh, and A.T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol.7, no.4, pp.1326–1337, Apr. 2008.

[74] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol.54, no.5, pp.2053–2071, May 2008.

[75] W. Wang, B. Kasiri, J. Cai, and A.S. Alfa. "Channel Assignment of Cooperative Spectrum Sensing in Multi-Channel Cognitive Radio Networks," *Proc. IEEE ICC'11*, kyoto, Japan, June. 2011, pp.1–5.

[76] A. W. Min, K. G. Shin, "Joint Optimal Sensor Selection and Scheduling in Dynamic Spectrum Access Networks," *IEEE Trans. Mobile Computing*, vol.12, no.8, pp.1532–1545, Aug. 2013.

[77] J. Mitola III, "Cognitive radio for flexible mobile multimedia communications," in *Mobile Netw. Appl. J. (MONET)*, vol.6, no.5, pp.435–441, Sept. 2001.

[78] Y. Chen, Y. Wu, B. Wang, and K. J. R. Liu, "Spectrum auction games for multimedia streaming over cognitive radio networks," *IEEE Trans. Commun.*, vol.58, no.8, pp.2381–2390, Aug. 2010.

[79] T. Jiang, H. Wang, and A. Vasilakos, "QoE-Driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol.30, no.7, pp.1215–1224, Aug. 2012.

[80] W. Wang, B. Kasiri, J. Cai, and A.S. Alfa. "Distributed cooperative multi-channel spectrum sensing based on dynamic coalitional game." *Proc. IEEE GLOBECOM'10*, Miami, FL, Dec. 2010, pp.1–5.

[81] B. Wang, K.J. Ray Liu, and T.Charles. Clancy. "Evolutionary cooperative spectrum sensing game: How to collaborate?" *IEEE Trans. Commun.*, vol.58, no,3, pp.890–900, Mar. 2010.

[82] S. Zhang, T. Wu, and R.K.N. Lau, "A low-overhead energy detection based cooperative sensing protocol for cognitive radio systems," *IEEE Trans. Wireless. Commun.*, vol.11, no.8, pp.5575–5581, Nov. 2009.

[83] K. Kar, L. Xiang, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," *IEEE Trans. Wireless. Commun.*, vol.7, no.7, pp.2619–2629, July 2008.

[84] H. W. Kuhn, "The Hungarian Method for the assignment problem," *Naval Research Logistics Quarterly*, vol.2, no.1/2, pp.83-97, Mar. 1955.

[85] H. Su, and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol.26, no.1, pp.118–129, Jan. 2008.

[86] R. Garfinkel and G.L. Nemhauser *Integer Programming*, *Wiley*, New York, NY, 1972.

[87] I. Griva, S.G. Nash, and A. Sofer *Linear and Nonlinear Optimization*, 2nd edition, *Society for Industrial and Applied Mathematics*, Philadelphia, PA, 2009.

[88] D. Hu and S. Mao, "A sensing error aware MAC protocol for cognitive radio networks," *EAI Endorsed Trans. Mobile Commun. Appl.*, vol.12, no.7/9, pp.e1, Aug. 2012.

[89] L. Zhou, Z. Yang, Y. Wen, H. Wang, and M. Guizani, "Resource allocation with incomplete information for QoE-driven multimedia communications." *IEEE Trans. Wireless. Commun.*, vol.12, no.8, pp.3733–3745, 2013.

[90] A. Khan, I.H. Mkwawa, L. Sun, and E. Ifeachor, "QoE-driven sender bitrate adaptation scheme for video applications over IP multimedia subsystem," *Proc. IEEE ICC'11*, Kyoto, Japan, June. 2011, pp.1–6.

[91] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Content classification-based and QoE-driven video send bitrate adaptation scheme," *Proc. ICST Mobimedia09*, London, UK, Sept., 2009, pp.16–19.

[92] W. Song, and D .Tjondronegoro, "Acceptability-based QoE models for mobile video." *IEEE Trans. Multimedia*, vol.16, no.3, pp.738–750, 2014.

[93] Y. Wu, F. Hu, S. Kumar, and Y. Zhu, "A Learning-Based QoE-Driven Spectrum Handoff Scheme for Multimedia Transmissions over Cognitive Radio Networks," *IEEE J. Sel. Areas Commun.*, vol.32, no.11, pp.2134–2148, Nov. 2014.

[94] R. Yao, Y. Liu, J. Liu, P. Zhoa, and C. Song, "Utility-Based H. 264/SVC Video Streaming Over Multi-Channel Cognitive Radio Networks," *IEEE Trans. Multimedia*, vol.17, no.3, pp.434–449, 2015.

[95] H. Saki, M. Maria, and M. Shikh-BahaeiJ, "Multi-user scalable video transmission over cognitive radio networks," *Proc. IEEE ICC'15*, London, June. 2015, pp.7564–7569.

[96] T. Ojanper, M. Luoto, M. Majanen, P. Mannersalo, and P.T. Savolainen, "Cognitive network management framework and approach for video streaming optimization in heterogeneous networks," *Springer Wireless Personal Communications*, vol.84, pp.1–31, 2015.

[97] I. Son, S. Mao, M.X. Gong, and Y .Li, "On frame-based scheduling for directional mmWave WPANs," *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar., 2012, pp.2149–2157.

[98] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A Tutorial on Video Quality Assessment," *IEEE Commun. Surveys and Tutorials*, no. 99, Oct. 2014, pp. 1–41.

[99] A. Khan, L. Sun, and E. Ifeachor, "QoE Prediction Model and Its Application in Video Quality Adaptation over UMTS Networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, 2012, pp. 431–42.

[100] Y. Wu et al., "A Learning-based QoE-Driven Spectrum Handoff Scheme for Multimedia Transmissions over Cognitive Radio Networks," *IEEE JSAC*, vol. 32, no. 11, 2014, pp. 2134–48.

[101] H. Saki, A. Shojaeifard, and M. Shikh-Bahaei, "Cross-Layer Resource Allocation for Video Streaming over OFDMA Cognitive Radio Networks with Imperfect Cross-Link CSI," *Proc. ICNC 2014*, Honolulu, HI, Feb. 2014, pp. 98–104.

[102] T. Jiang, H. Wang, and A. V. Vasilakos "QoE-Driven Channel Allocation Schemes for Multimedia Transmission of Priority-Based Secondary Users over Cognitive Radio Networks," *IEEE JSAC*, vol. 30, no. 7, Aug. 2012, pp. 1215–24.

[103] Z. He, S. Mao, and S. Kompella, "QoE Driven Video Streaming in Cognitive Radio Networks: The Case of Single Channel Access," *Proc. IEEE GLOBECOM 2014*, Austin, TX, Dec. 2014, pp. 138893.

[104] Y. Ding and L. Xiao, "Video On-Demand Streaming in Cognitive Wireless Mesh Networks," *IEEE Trans. Mobile Computing*, vol. 12, no. 3, Mar. 2013, pp. 412–23.

[105] D. Hu et al., "Scalable Video Multicast in Cognitive Radio Networks," *IEEE JSAC*, vol. 28, no. 3, Apr. 2010, pp. 334–44.

[106] H. Luo, S. Ci, and D. Wu, "A Cross-Layer Design for the Performance Improvement of Real-Time Video Transmission of Secondary Users over Cognitive Radio Networks," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 21, no. 8, Aug. 2011, pp. 1040–48.

[107] H. Kushwaha et al., "Reliable Multimedia Transmission Over Cognitive Radio Networks Using Fountain Codes," *Proc. IEEE*, vol. 96, no. 1, Jan. 2008, pp. 155–65.

[108] R. Yao et al., "Perceptual Experience Oriented Transmission Scheduling for Scalable Video Streaming over Cognitive Radio Networks," *Proc. IEEE GLOBECOM 2013*, Atlanta, GA, Dec. 2013, pp. 1681–86.

[109] S. Li, T. H. Luan, and X. Shen, "Channel Allocation for Smooth Video Delivery Over Cognitive Radio Networks," *Proc. IEEE GLOBECOM 2010*, Miami, FL, Dec. 2010, pp. 1–5.

[110] Z. Guan et al., "On the Effect of Cooperative Relaying on the Performance of Video Streaming Applications in Cognitive Radio Networks," *Proc. IEEE ICC 2011*, Kyoto, Japan, June 2011, pp. 1–6.

[111] Z. Gao et al., "Location Privacy in Database-Driven Cognitive Radio Networks: Attacks and Countermeasures," *Proc. IEEE INFOCOM 2013*, Turin, Italy, Apr. 2013, pp. 2751–59.

Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Netw.*, pp. 1–20, Apr. 2015.

[112] I. K. Son, S. Mao, Y. Li, M. Chen, M. X. Gong, and T. T. S. Rappaport, "Frame-based medium access control for 5G wireless networks," *Mobile Netw. Appl.*, pp. 1–10, Jan. 2015.

[113] K. Han, Z. Hu, J. Luo, and X. Liu, "RUSH: Routing and scheduling for hybrid data center networks," *Proc. IEEE INFOCOM 2015*, HongKong, China, Apr. 2015, pp. 415–423.

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

[115] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.

[116] R. Mudumbai, S. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2871–2875.

[117] S. Singh, R. Mudumbai, and U. Madhow, "Distributed coordination with deaf neighbors: efficient medium access for 60 GHz mesh networks," in *Proc. IEEE INFOCOM 2010*, San Diego, CA, Mar. 2010, pp. 1–9.

[118] H. Chu and P. Xu, "Relay selection with feedback beamforming information for NLoS 60GHz mmWave WLANs/WPANs," in *Proc. IEEE ICC 2014*, Sydney, Australia, June 2014, pp. 5514–5519.

[119] Z. Liu, X. Tao, W. U. Rehman, Z. Xu, and X. Xu, "Opportunistic relay selection and outage performance analysis for 60GHz wireless system," in *Proc. 2013 IEEE Globecom*, Atlanta, GA, Dec. 2013, pp. 328–332.

[120] S. Singh, F. Ziliotto, U. Madhow, E. Belding, and M. Rodwell, "Blockage and directivity in 60 GHz wireless personal area networks: from cross-layer model to multihop mac design," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1400–1413, Oct. 2009.

[121] T. Bai, R. Vaze, and R. W. Heath Jr, "Using random shape theory to model blockage in random cellular networks," in *Proc. SPCOM 2012*, Bangalore, India, July 2012, pp. 1–5.

[122] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 922–928, May 2004.

[123] Z. Genç, U. H. Rizvi, E. Onur, and I. Niemegeers, "Robust 60 GHz indoor connectivity: is it possible with reflections?" in *Proc. IEEE VTC 2010-Spring*, Taipei, Taiwan, May 2010, pp. 1–5.

[124] C. Yiu and S. Singh, "Empirical capacity of mmwave wlans," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1479–1487, Oct. 2009.

[125] X. An, C.-S. Sum, R. V. Prasad, J. Wang, Z. Lan, J. Wang, R. Hekmat, H. Harada, and I. Niemegeers, "Beam switching support to resolve link-blockage problem in 60 GHz WPANs," in *Proc. IEEE PIMRC 2009*, Toyko, Japan, Sept. 2009, pp. 390–394.

[126] K. Song, R. Cai, and D. Liu, “A fast relay selection algorithm over 60GHz mm-wave systems,” in *Proc. IEEE ICCT 2013*, Guilin, China, Nov. 2013, pp. 676–680.

[127] J. Qiao, L. X. Cai, X. Shen, and J. Mark, “Enabling multi-hop concurrent transmissions in 60 GHz wireless personal area networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3824–3833, Nov. 2011.

[128] Z. Lan, C.-S. Sum, J. Wang, T. Baykas, F. Kojima, H. Nakase, and H. Harada, “Relay with deflection routing for effective throughput improvement in Gbps millimeter-wave WPAN systems,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1453–1465, Oct. 2009.

[129] X. Zhang, S. Zhou, X. Wang, Z. Niu, X. Lin, D. Zhu, and M. Lei, “Improving network throughput in 60GHz WLANs via multi-AP diversity,” in *Proc. IEEE ICC 2012*, Ottawa, Canada, June 2012, pp. 4803–4807.

[130] P. Hyunhee, Y. Kim, T. Song, and S. Pack, “Multiband directional neighbor discovery in self-organized mmWave ad hoc networks,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1143–1155, Mar. 2015.

[131] Z. He, S. Mao, and T. T. S. Rappaport, “Minimum time length link scheduling under blockage and interference in 60GHz networks,” in *Proc. IEEE WCNC 2015*, New Orlean, LA, Mar. 2015, pp. 837–842.

[132] L. Li, K. Josiam, and R. Taori, “Feasibility study on full-duplex wireless millimeter-wave systems,” in *Proc. IEEE ICASSP 2014*, Florence, Italy, May 2014, pp. 2769–2773.

[133] G. Zheng, H. Cunqing, R. Zheng, and Q. Wang, “A robust relay placement framework for 60GHz mmwave wireless personal area networks,” in *Proc. IEEE GLOBECOM’13*, Atlanta, GA, Dec. 2013, pp. 4816–4822.

[134] W.K. Ching, X. Huang, M.K. Ng, T.K. Siu, “Markov Chains: Models, Algorithms and Applications,” Springer, 2006.

[135] M.X. Gong, R. Stacey, D. Akhmetov, and S. Mao, “A Directional CSMA/CA Protocol for mmWave Wireless PANs,” in *Proc. IEEE WCNC 2010*, Sydney, Australia, Apr. 2010, pp. 1–6.

[136] C. Yiu and S. Singh, “Link Selection for Point-to-Point 60GHz Networks,” in *Proc. IEEE ICC 2010*, Cape Town, May. 2010, pp. 1–6.

[137] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, “On the Efficient Beam-Forming Training for 60GHz Wireless Personal Area Networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 504 – 515, Dec. 2013.

[138] G. Kalfas, N. Pleros, K. Tsagkaris, L. Alonso, and C. Verikoukis, “Performance Analysis of a Medium-Transparent MAC Protocol for 60GHz Radio-over-Fiber Networks,” in *Proc. IEEE GLOBECOM 2011*, Houston, TX, Dec. 2011, pp. 1–5.

[139] C. Yiu and S. Singh, "SDMA for 60GHz Gigabit Wireless Networks," in *Proc. IEEE ICC 2009*, Dresden, June 2009, pp. 1–6.

[140] P. Maniotis, G. Kalfas, L. Alonso, C. Verikoukis, and N. Pleros, "Throughput and delay fairness through an agile medium-transparent MAC protocol for 60GHz fiber-wireless LAN networks," in *Proc. IEEE WCNC 2012*, Shanghai, China, Apr. 2012, pp. 3769 – 3773.

[141] IEEE 802.15.3 Working Group, "Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)," in *IEEE Unapproved Draft Std P802.15.3c/DIO*, June 2009.

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

[143] S. Mao and T.S. Rappaport, "Millimeter wave wireless networks: A medium access control perspective," in *Wireless Network Performance Enhancement via Directional Antennas: Models, Protocols, and Systems*, J.D. Matyjas, F. Hu, and S. Kumar (editors). New York, NY: CRC Press, 2015. ISBN: 149870753X.

[144] S.K. Yong, P. Xia, and A.V. Garcia, *mmWave Technology for Gbps, WLAN, and WPAN: From Theory to Practice*, John Wiley & Sons: Hoboken, NJ, 2011.

[145] R.C. Daniels and R.W. Heath, "mmWave wireless communications: Emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol.2, no.3, pp.41–50, Sept. 2007

[146] S. Singh, R. Mudumbai, and U. Madhow. "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol.19, no.5, pp.1513–1527, Oct. 2011.

[147] T.S. Rappaport, E. Ben-Dor, N.N. Murdock, and Y. Qiao, "38 GHz and mmWave angle-dependent propagation for cellular and peer-to-peer wireless communications," in *Proc. IEEE ICC'12*, Ottawa, Canada, June 2012, pp.4568–4573.

[148] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. Rodwell, "Blockage and directivity in mmWave wireless personal area networks: From cross-layer model to multihop MAC design," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 14001413, Oct. 2009.

[149] T.S. Rappaport, R.W. Heath Jr., R.C. Daniels, and J.N. Murdock, *Millimeter Wave Wireless Communications*. Pearson, Prentice-Hall: Upper Saddle River, NJ, 2014.

[150] T,. Bai, R. Vaze, and R.W. Heath, "Using random shape theory to model blockage in random cellular networks," in *Proc. IEEE SPCOM'12*, Bangalore, India, July 2012, pp.1–5.

[151] C.R. Anderson and T.S. Rappaport, "In-building wideband partition loss measurements at 2.5 and mmWave," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp.922–928, Mar. 2004.

[152] S. Sun, T.S. Rappaport, R.W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun.*, vol.52, no.12, pp.110–121, Dec. 2014.

[153] T. Dinc, A. Chakrabarti, and H. Krishnaswamy, "A mmWave same-channel full-duplex CMOS transceiver and link based on reconfigurable polarization-based antenna cancellation," in *Proc. 2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, Phoenix, AZ, May 2015, pp.1–4.

[154] Y. Niu, Y. Li, D. Jin, L. Su, A.V. VasilakosA, "Survey of millimeter Wave (mmWave) communications for 5G: Opportunities and challenges," arXiv:1502.07228, Feb. 2015.

[155] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali, *Linear Programming and Network Flows*. John Wiley & Sons: Hoboken, NJ, 2009.

[156] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Co.: San Francisco, CA, 1979.

[157] S. Kompella, S. Mao, Y. T. Hou, and H. D. Sherali, "On path selection and rate allocation for video in wireless mesh networks," *IEEE/ACM Trans. Netw.*, vol.17, no.1, pp.212–224, Feb. 2009.

[158] S. Kompella, J.E. Wieselthier, A. Ephremides, H.D. Sherali, and G.D. Nguyen, "On optimal SINR-based scheduling in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol.18, no.6, pp.1713–1724, Nov. 2010.

[159] Z. He and S. Mao, "QoS driven multi-user video streaming in cellular CRNs: The case of multiple channel access," in *Proc. IEEE MASS'14*, Philadelphia, PA, Oct. 2014.

[160] P. Bjrklund, P. Vrbrand, and D. Yuan, "A column generation method for spatial TDMA scheduling in ad hoc networks," *Elsevier Ad Hoc Netw.*, vol.2, no.4, pp.405–418, Oct. 2004.

[161] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol.54, no.5, pp.2053–2071, May. 2008.

[162] I.K. Son, S. Mao, M. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp.2149–2157.

[163] I.-K. Son, S. Mao, Y. Li, M. Chen, M.X. Gong, and T.S. Rappaport, "Frame-based medium access control for 5G wireless networks, *Springer MONET Journal*, to appear.

[164] X. An and R. Hekmat, "Directional MAC protocol for millimeter wave based wireless personal area networks," in *Proc. IEEE VTC-Spring 2008*, Singapore, Singapore, May 2008, pp.1636–1640.

[165] M.X. Gong, Stacey, R. Stacey, D. Akhmetov and S. Mao, "A Directional CSMACA protocol for mmWave wireless PANs," in *Proc. IEEE WCNC 2010*, Sydney, Australia, Apr. 2010, pp.1–6.

[166] L.X. Cai, L. Cai, X. Shen, and N.W. Mark, "Rex: A randomized EXclusive region based scheduling scheme for mmWave WPANs with directional antenna," *IEEE Trans. Wireless Commun.*, vol.9, no.1, pp.113–121, Jan. 2010.

[167] Z. He, S. Mao, and T.S. Rappaport "Minimum time length link scheduling under blockage and interference in mmWave networks," in *Proc. IEEE WCNC 2015*, New Orleans, US, Mar. 2015, pp.1–6.

Appendices

Appendix A

Publications

A.1 Conference Publications

1. **Zhifeng He** and Shiwen Mao, Multiple Description Coding for uncompressed video streaming over 60GHz networks, in *Proc. First ACM Workshop on Cognitive Radio Architectures for Broadband (CRAB 2013), in conjunction with ACM MobiCom 2013*, pp.61-68, Miami, FL, October 2013.

2. **Zhifeng He** and Shiwen Mao, QoS driven multi-user video streaming in cellular CRNs: The case of multiple channel access, in *Proc. IEEE MASS 2014*, Philadelphia, PA, Oct. 2014, pp.28-36.

3. **Zhifeng He**, Shiwen Mao, and Sastry Kompella, QoE driven video streaming in cognitive radio networks: Case of single channel access, in *Proc. IEEE GLOBECOM 2014*, Austin, TX, Dec. 2014, pp.1388-1393.

4. **Zhifeng He**, Shiwen Mao, and Theodore (Ted) S. Rappaport, Minimum time length link scheduling under blockage and interference in 60GHz networks, in *Proc. IEEE WCNC 2015*, New Orlean, LA, Mar. 2015, pp.837–842.

5. **Zhifeng He**, Shiwen Mao, Sastry Kompella, and Ananthram Swami, Minimum time length scheduling under blockage and interference in multi-hop mmWave networks, in *Proc. IEEE GLOBECOM 2015*, San Diego, CA, Dec. 2015, pp.1-7.

6. **Zhifeng He** and Shiwen Mao, A decomposition principle for link and relay selection in dual-hop 60 GHz networks, in *Proc. IEEE INFOCOM 2016*, San Francisco, CA, Apr. 2016, pp.1683-1691.

7. **Zhifeng He**, Shiwen Mao, Sastry Kompella, and Ananthram Swami, Link scheduling and channel assignment with a graph spectral clustering approach, in *Proc. IEEE MILCOM 2016*, Baltimore, MD, Nov. 2016.

A.2 Journal Publications

1. **Zhifeng He** and Shiwen Mao, Adaptive multiple description coding and transmission of uncompressed video over 60GHz networks, *ACM Mobile Computing and Communications Review (MC2R)*, vol.18, no.1, pp.14-24, Jan. 2014.

2. **Zhifeng He**, Shiwen Mao, and Theodore (Ted) S. Rappaport, On link scheduling under blockage and interference in 60 GHz ad hoc networks, *IEEE Access Journal, Special Section on Ultra Dense Cellular Networks*, vol.3, pp.1437-1449, Sept. 2015.

3. **Zhifeng He**, Shiwen Mao, and Sastry Kompella, A decomposition approach to quality of service driven multi-user video streaming in cellular cognitive radio networks, *IEEE Transactions on Wireless Communications*, vol.15, no.1, pp.728-739, Jan. 2016.

4. **Zhifeng He**, Shiwen Mao, and Tao Jiang, A survey of QoE driven video streaming over cognitive radio networks, *IEEE Network, Special Issue on Quality-of-Experience (QoE)-Aware Design in Next-Generation Wireless Networks*, vol.29, no.6, pp.20-25, Nov./Dec. 2015.

5. **Zhifeng He**, Shiwen Mao, and Sastry Kompella, Quality of Experience driven multi-user video streaming in cellular cognitive radio networks with single channel access, *IEEE Transactions on Multimedia*, vol.18, no.7, pp.1401-1413, July 2016.