January 2012

# Opportunistic Scheduling and Cooperative Relaying in Wireless Networks

Yufeng Wang
*University of South Florida*, ywang2@mail.usf.edu

Opportunistic Scheduling and Cooperative Relaying in Wireless Networks

by

Yufeng Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Electrical Engineering
College of Engineering
University of South Florida

Co-Major Professor: Ravi Sankar, Ph.D.
Co-Major Professor: Salvatore D. Morgera, Ph.D.
Andrew Raij, Ph.D.
Hao Zheng, Ph.D.
Hariharan Srikanth, Ph.D.

Date of Approval:
November 5, 2012

Keywords: throughput-delay tradeoff, adaptive rate transmission, buffer-aware scheduling, system throughput, packet delay, two-hop communications

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ABSTRACT

The demand for ever larger, more efficient, reliable and cost effective communication networks necessitates new network architectures, such as wireless *ad hoc* networks, cognitive radio, relaying networks, and wireless sensor networks. The study of such networks requires a fundamental shift from thinking of a network as a collection of independent communication pipes, to a multi-user channel where users cooperate via conferencing, relaying, and joint source-channel coding.

The traditional centralized networks, such as cellular networks, include a central controller and a fixed infrastructure, in which every node communicates with each other via a centralized based station (BS). However, for a decentralized network, such as wireless *ad hoc* networks and wireless sensor networks, there is no infrastructure support and no central controllers. In such multi-user wireless networks, the scheduling algorithm plays an essential role in efficiently assigning channel resources to different users for better system performance, in terms of system throughput, packet-delay, stability and fairness.

In this dissertation, our main goal is to develop practical scheduling algorithms in wireless *ad hoc* networks to enhance system performance, in terms of throughput, delay and stability. Our dissertation mainly consists of three main parts.

First, we identify major challenges intrinsic to *ad hoc* networks that affect the system performance, in terms of throughput limits, delay and stability condition.

Second, we develop scheduling algorithms for wireless *ad hoc* networks, with various considerations of non-cooperative relays and cooperative relays, fixed-rate

transmission and adaptive-rate transmission, full-buffer traffic model and finite-buffer traffic model. Specifically, we propose an opportunistic scheduling scheme and study the throughput and delay performance, with fixed-rate transmissions in a two-hop wireless *ad hoc* networks. In the proposed scheduling scheme, we prove two key inequalities that capture the various tradeoffs inherent in the broad class of opportunistic relaying protocols, illustrating that no scheduling and routing algorithm can simultaneously yield lower delay and higher throughput. We then develop an adaptive rate transmission scheme with opportunistic scheduling, with the constraints of practical assumptions on channel state information (CSI) and limited feedback, which achieves an optimal system throughput scaling order. Along this work with the consideration of finite-buffer model, we propose a Buffer-Aware Adaptive (BAA) scheduler which considers both channel state and buffer conditions to make scheduling decisions, to reduce average packet delay, while maintaining the queue stability condition of the networks. The proposed algorithm is an improvement over existing algorithms with adaptability and bounded potential throughput reduction.

In the third part, we extend the methods and analyses developed for wireless *ad hoc* networks to a practical Aeronautical Communication Networks (ACN) and present the system performance of such networks. We use our previously proposed scheduling schemes and analytical methods from the second part to investigate the issues about connectivity, throughput and delay in ACN, for both single-hop and two-hop communication models. We conclude that the two-hop model achieves greater throughput than the single-hop model for ACN. Both throughput and delay performances are characterized.

# CHAPTER 1: INTRODUCTION

## 1.1 Wireless Ad Hoc Networks

In the last decade there has been great interest within the research community to improve the performance of wireless *ad hoc* networks. The study of wireless *ad hoc* networks requires a fundamental shift from thinking of a network as a collection of independent communication pipes, to a multi-user channel where users cooperate via conferencing, relaying, and joint source-channel coding. The tradition centralized networks, such as cellular networks, include a central controller and a fixed infrastructure, in which every node communicates with each other via a centralized based station (BS). However, for a decentralized network, such as wireless *ad hoc* networks and wireless sensor networks, there is no infrastructure support and no central controllers. In such multi-user wireless networks, the scheduling algorithm plays an essential role in efficiently assigning channel resources to different users for better system performance, in terms of system throughput, packet-delay, stability and fairness.

For the traditional cellular networks, focusing on throughput performance, the maximum throughput (MT) scheduler is introduced in [38–40], which schedule only users with the best instantaneous channel conditions to transmit in each scheduling interval. MT maximizes sum system throughput at the loss of fairness to cell edge users. Round robin (RR) is the most fair but channel unaware scheduler, in which users' transmissions takes place in a strict numerical order [41]. The MT and RR

schedulers leave room for various schedulers that lie in between them. Proportional fair (PF) scheduler [42–44] weights users' instantaneous transmission rates by their average rates to tradeoff throughput with fairness. PF is the practical scheduling algorithm that currently implemented in most 4G-LTE systems. Although MT, RR and PF algorithms can be directly applied to the centralized networks, such as a cellular network, the implementation of the algorithms require a central controller and perfect channel state information (CSI) knowledge at both transmitters and receivers. For a decentralized network, in which there is no infrastructure support and central controllers, the design of scheduling algorithms and the study of the system performance limits become more challenging and have attracted attention in the research community. In this work, we focus on the design of scheduling algorithms and study of the system performance limits, in terms of system throughput, average packet delay and stability condition for wireless *ad hoc* networks.

## 1.2 Challenges and Constraints in Wireless Ad Hoc Networks

In wireless *ad hoc* networks, due to the lack of infrastructure support, CSI knowledge and central controllers, etc., the design of the scheduling schemes and the system performance have many challenges and constraints. Specifically,

- Throughput limit is unknown in wireless *ad hoc* networks, with the presence of interferences.
- Delay might be large and unbounded in large *ad hoc* networks.
- There is no central coordination among nodes in wireless *ad hoc* networks. The nodes only have access to the channel information, scheduling decision and transmission rates etc., by limited cooperation and feedback.

- In a large system with many nodes, obtaining perfect CSI, especially at the transmitter side, may not be feasible.
- The mobility of *ad hoc* nodes causes dynamic network topology, which may lead to packet losses, network instability, lower throughput and larger delay.
- Broadcast nature of wireless link leads to unavoidable interference and thus causes packet errors.
- Each *ad hoc* node has limited power.
- Network reliability and robustness depends on autonomous nodes' behavior, node density, network load, topology changes, and link disconnections.

Due to the aforementioned challenges, there is no one solution to the above problems. Specific solutions for specific problems are sought by the researchers. The designs of scheduling algorithms have been proposed in [1–3, 7, 8] with focus on the centralized networks, however, the proposed schemes cannot be applied to *ad hoc* networks. The study of throughput limits in wireless *ad hoc* networks has been done by different authors in [8-19], with the considerations of single-hop, two-hope and multi-hop communications. Although these studies have made great strides toward understanding wireless *ad hoc* network capacity, they are not taking the delay performance into consideration. Alongside the body of work on analyzing the throughput performance, and inspired by Grossglauser and Tse in [7], there is a line of work characterizing the delay-throughput trade-off of wireless network in different setups. Among many others, Neely and Modiano [10], El Gamal et al. [12], Toumpis and Goldsmith [13] , Lin and Shroff [14], and Sharma et al. [15] have studied the delay-throughput trade-off of the mobile *ad hoc* networks. These publications

generally follow a similar line in which the authors study the problem by first defining a certain mobility model, and then analyze the delay averaged over the users. Ying et al. [16], Zhang et al. [17] studied the throughput-delay trade-off with network coding.

### 1.3 Research Motivation

Traditional scheduling schemes, such as MT, RR and PF, cannot be implemented in decentralized networks, such as wireless *ad hoc* networks, due to lack of central controller and perfect CSI knowledge. As previously mentioned, the existing studies on wireless *ad hoc* networks focused on throughput limits only. In achieving the throughput limit, the key idea is to schedule at each hop only the subset of nodes that can benefit from multiuser diversity gain. In such schemes, fairness and delay are two concerns that need to be addressed. While the fairness issue is less relevant in the independent and identically distributed (i.i.d.) channel model, since on average every node is afforded the same throughput, the delay consideration is more salient and needs to be quantified. Furthermore, there is no delay guarantee for the transmission of a packet from a sender to a designated user in such schemes. Since both throughput and delay are important figures of merit from an application point of view, it is necessary to design a scheduling algorithm which enhances both throughput and delay performance. It is also important to characterize the throughput and delay performance based on the scheduling algorithms.

## 1.4 Contributions and Organization

This dissertation primarily focuses on the design of the scheduling algorithms, and the corresponding achievable throughput and delay performance in the wireless *ad hoc* networks. The organization of the dissertation is as follows,

In Chapter 2, we propose an opportunistic scheduling scheme and study the delay and throughput trade-off with the help of relays, over channels with random connections, in which the channel connections are independent and identically distributed (i.i.d.). The proposed opportunistic scheduling scheme operates in a completely decentralized fashion, in which there is no infrastructure support or central controller, and only CSI at receivers are available [35]. Our primary contribution is that we show the proposed opportunistic scheduling scheme achieves the optimal throughput in the order[1] of $\log(n)$, with fixed rate transmission in a network with $n$ source-destination pairs and $m$ relays. The scheme provides an upper bound of $O(n)$ delay, including full effects of queuing in the network model. Our second contribution is to use a redundant scheduling scheme to reduce the upper bound of delay scaling to $O(n/\log n)$. Our third contribution is the proof of two key inequalities that capture the various tradeoffs inherent in the broad class of opportunistic relaying protocols, which indicate the tradeoff of $delay/throughput < O(n/\log n)$, illustrating that no scheduling and routing algorithm can simultaneously yield lower delay and higher throughput.

In Chapter 3, based on our previous scheduling for fixed rate transmissions in Chapter 2, we propose an adaptive rate transmission scheme with opportunistic scheduling in a relaying network. We assume only CSI at receivers is available in the

[1]The following notations are used in this Dissertation. For two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ means $|\frac{f(n)}{g(n)}|$ remains bounded as $n \to \infty$; $f(n) = \Theta(g(n))$ denotes $f(n) = O(g(n))$ and $g(n) = O(f(n))$. The function $\log(\cdot)$ indicates the natural logarithm, unless specified otherwise, *e.g.*, $\log_2(\cdot)$.

decentralized network, where $n$ source-to-destination pairs are completely independent and $m$ half-duplex relays cooperate by exchanging their selected source nodes' numbers at the beginning of the two-hop transmission. The primary contribution of this work is that the proposed adaptive rate transmission scheduling scheme achieves a system throughput in the order of $\frac{m}{2}\log(\log n)$ [36]. Furthermore, this is proven to be the same achievable scaling even with perfect CSI assumptions at transmitters and full cooperation among nodes [37]. Our second contribution is to show that the optimal scaling of the number of relays $m$ is $\Theta(\log n)$, under which a linear increase in throughput with m is obtained. Our third contribution is to derive a closed-form expression of average end-to-end packet delay for the proposed scheme.

The scheduling schemes in Chapter 2 and Chapter 3 have focused on improving the system performance with the assumption of full-buffer traffic model. In the full-buffer traffic model, a user has unlimited amount of data to transmit. This model has been extensively adopted in the literature due to its simplicity. However, a more practical traffic model is to assume a user is assigned a finite traffic buffer to transmit, this is called as finite-buffer model, which includes user arrival (birth) and departure (death) process. The finite-buffer model has been less extensively adopted, due to increased complexity.

In Chapter 4, we consider the design of a scheduling algorithm with finite-buffer traffic model for multiuser systems. The primary contribution is that we propose a Buffer-Aware Adaptive (BAA) scheduler which considers both channel state and buffer conditions to make scheduling decisions, to reduce average packet delay, while maintaining the queue stability condition of the networks. The proposed algorithm is an improvement over existing algorithms with adaptability and bounded potential throughput reduction. The second contribution is the establishment of a generalized form of the proposed algorithm, which can be can be implemented to form

a specific scheduling algorithm. Our third contribution is to provide the stability considerations of the proposed algorithm, along with the average throughput lower bound and approximation.

In Chapter 5, we extend the methods and analysis developed for wireless *ad hoc* networks from Chapters 2-4 to a practical Aeronautical Communication Networks (ACN) and present the system performance of such networks in both single [60] and two-hop [61] models. ACN is an emerging concept in which aeronautical stations (AS) are considered as a part of multi-tier network for the future wireless communication system. The goal of ACN is to provide high throughput and cost effective communication network for aeronautical applications, *i.e.*, Air Traffic Control (ATC), Air Traffic Management (ATM) communications and commercial in-flight Internet activities), and terrestrial networks by using aeronautical platforms as a backbone [62]. We use our previously proposed scheduling schemes and analytical methods to investigate the issues about connectivity, throughput and delay in ACN. The primary contribution is to present the topology of ACN as a simple mobile *ad hoc* network and provide the connectivity analysis. Our second contribution is by using information obtained from connectivity analysis to investigate two communication models, *i.e.*, single-hop and two-hop, in which each source AS is communicating with its destination AS with or without the help of intermediate relay AS, respectively. In our throughput analysis, we use the method of finding the maximum number of concurrent successful transmissions to derive ACN throughput upper bounds for the two communication models. We conclude that the two-hop model achieves greater throughput than the single-hop model for ACN. Our third contribution is to characterize the delay performance and derive the closed-form average end-to-end delay for the two-hop model, since delay issue is more salient in two-hop communication.

In Chapter 6, we summarize our contributions from Chapters 2-5 and then propose recommendations for future work.

# CHAPTER 2: FIXED-RATE TRANSMISSIONS

## 2.1 Introduction

In the last decade there has been great interest within the research community to improve the performance of wireless *ad hoc* networks. In this Chapter, we extend the previous work on two-hop opportunistic relaying scheme [8]. The scheme features decentralized operation (in contrast to [1, 2], where various levels of cooperation are needed) and receiver CSI with limited feedback. The system consists of $n$ source-to-destination (S-D) pairs and $m$ relay nodes. It is shown in [8] that under a random channel model, in which the channel connections are independent and identically distributed (i.i.d.), the average system throughput is $m/2$ bits/s/Hz. Moreover, the system throughput in the limit of a large system is given by $\Theta(\log n)$, by which it is understood that the number of relays $m$ can increase (as a function of $n$) as fast as $m = \Theta(\log n)$, while retaining the linearity of throughput in $m$. Linearity breaks down when $m$ increases faster than the order of $\log n$. The work in [8, 34] characterizes the fundamental throughput limits, when node cooperation and full CSI are not available. In achieving the throughput limit, the key idea is to schedule at each hop only the subset of nodes that can benefit from multiuser diversity gain. In schemes of the opportunistic scheduling nature, fairness and delay are two concerns that need to be addressed. While the fairness issue is less relevant in the i.i.d. channel model, since on average every node is afforded the same throughput, the delay consideration is more salient and needs to be quantified. The reason is that,

with opportunistic scheduling, there is no delay guarantee for the transmission of a packet from a sender to a designated user.

## 2.2 Literature Review and Motivation

Focusing on system throughput, numerous schemes have been proposed corresponding to different assumptions on the channel state information (CSI) and levels of cooperations among communicating nodes [1–3, 7, 8]. Alongside the body of work on analyzing the throughput analysis, and inspired by Grossglauser and Tse [7], there is a line of work characterizing the delay-throughput trade-off of wireless network in different setups. Among many others, Neely and Modiano [10], El Gamal et al. [12], Toumpis and Goldsmith [13] , Lin and Shroff [14], and Sharma et al. [15], have studied the delay-throughput trade-off of the mobile *ad hoc* networks proposed in [7]. These publications generally follow a similar line in which the authors study the problem by first defining a certain mobility model, and then analyze the delay averaged over the users. Ying et al. [16], Zhang et al. [17] studied the throughput-delay trade-off with network coding.

Since both throughput and delay are important figures of merit from an application point of view, it is necessary to characterize the delay performance and the delay-throughput relation. The remaining open questions are: What is the average end-to-end packet delay of the two-hop opportunistic relaying scheme? Can we reduce the delay (in the order-of-magnitude sense) by "tweaking" the scheme? What is the relation between delay and throughput? These questions are addressed in this research. To this end, we will make use of delay analysis methodologies developed in [10, 12–15]. However, as will become apparent in Section II, the detailed analysis is quite different (and is more involved) than the previous work based on [7].

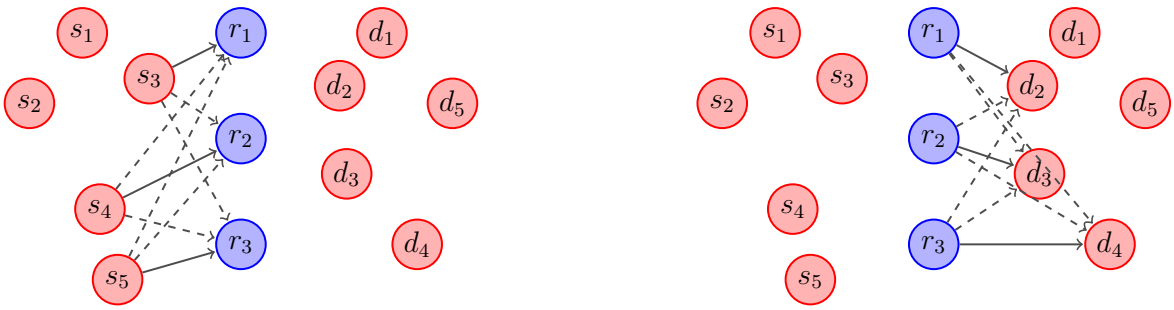


Figure 2.1: A two-hop network model with opportunistic scheduling, note: from [35] ©2011 IEEE

## 2.3 System Model and Assumptions

### 2.3.1 System Model

Consider the wireless network with $n$ S-D pairs and $m$ relay nodes as in Fig. 2.1, in which $n$ source nodes have data traffic to send to their designated destination nodes, while relay nodes have no traffic demand on their own. We consider the two-hop decode-and-forward communication protocol, in which the source nodes communicate with destination nodes only through the half-duplex relays. Specifically, in Phase 1 a subset of sources is scheduled for transmission to relays. The relays then decode and buffer the packets. During Phase 2, the relays forward packets to a subset of destinations (not necessarily the set of destinations associated with the source set in Phase 1). These two phases are interleaved: in the even-indexed time-slots, Phase 1 is run; in the odd-indexed time-slots, Phase 2 is run. The selection process for source/destination sets is of opportunistic nature, that is, in each hop, only a subset of nodes that can benefit from multiuser diversity gain are scheduled for transmission.

### 2.3.2 Channel Model

We consider all the nodes in the network are operating in the same frequency band and in the presence of fading. We assume channel realizations from source nodes to relays and from relays to destination nodes experience i.i.d. Rayleigh fading. Accordingly, the channels are assumed to be constant during the transmission duration $T$ in each phase of the two-hop communication.

### 2.3.3 Opportunistic Scheduling Scheme

The opportunistic scheduling scheme is summarized as follows,

- In the first hop, all $m$ relay nodes operate independently and each selects the source with the strongest channel connection; the selected source nodes then transmit packets to the relay nodes.
- In the second hop, each destination measures the signal to interference and noise ratios (SINR) of all $m$ relays and feedback the index of the relay (if exists) that has $\mathsf{SINR} \geq 1$. Upon receiving the feedback, the relay nodes then forward the packets to the destination nodes.

It has been proven in [8] that $m$ can grow (as a function of $n$) as fast as $\Theta(\log n)$, while still guaranteeing the linear throughput scaling; the linearity breaks down if $m$ grows faster than $\Theta(\log n)$. Hence, when $m$ is less than or equal to $\Theta(\log n)$, the scheduled transmission will be successful despite the SINR. Based on this fact and the above scheduling mechanism, as long as $m$ is less than or equal to $\Theta(\log n)$, the following results can be obtained, which are summarized as:

- The probability for a source node to be scheduled by one of $m$ independent and identical relays is $m/n$, since it is equally likely for each of the $n$ source nodes to have the best channel condition.

- The probability for a given packet from the output of the source node to be transmitted to the first relay node is $1/m$, because each of the $m$ relay nodes are equally likely.

- The probability for a relay node to be scheduled for a packet transmission to the corresponding destination node is $1/n$, since the relay to destination opportunity arises with equal probability for each of the $n$ destination nodes.

Note that the model delineated here is different than [7]. For example, in Phase 1 of [7], the average delay from a source node to a relay node is of the order of $\Theta(1)$ since any *ad hoc* node can serve as a relay. In contrast, in our model there is a limited number $m$ of relays, thus the average delay will be longer than $\Theta(1)$ as in [7]. The combined delay of Phase 1 and 2 is analyzed next.

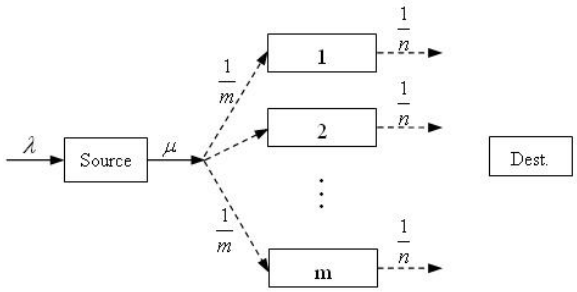


Figure 2.2: Queueing model, note: from [35] ©2011 IEEE

## 2.4 Performance Analysis

### 2.4.1 Average End-to-End Packet Delay

The end-to-end packet delay is defined as the average time that it takes the packet to arrive at the designated destination node from a source node [12], thus, the delay $D$ consists of two parts: $D = D_1 + D_2$, where $D_1$ and $D_2$ denote respectively

the delay in Phase 1 (the time for a packet transmitted from a source node to a relay node) and the delay in Phase 2 (the time for a packet forwarded from the relay node to the corresponding destination node). We introduce the queueing network model illustrated in Fig. 2.2 to analyze the average packet delay. Similar models have been used in [10, 12], which analyzed the delay performance of mobile network [7]. Before proceeding, it is important to note that the opportunistic relaying scheme delineated in Section II is different than [7]. For example, in Phase 1 of [7], $D_1 = \Theta(1)$, since from any particular source's perspective, all other nodes can serve as relays. In contrast, in our model there is a limited number of relays, delay will be longer than $\Theta(1)$. Following the model developed in [10] and [12], each of the $m$ relay nodes keeps a separate queue for each S-D pair. Since the wireless network has random channel connections, all such queues at all relay nodes are identical by symmetry. Thus the average end-to-end packet delay is the average delay at such a queue.

We assume the arrival process is a Poisson process with packet arrival rate $\lambda$. The service has a Bernoulli distribution with an average rate $\mu$ that a transmission opportunity arises. It is noted that a transmission opportunity arises when the source node is scheduled to transmit to a relay, and corresponds to a service opportunity. From the system model, we know service rate $\mu = m/n$. Note that in order to have finite delay, the arrival rate must be strictly smaller than the service rate. To ensure this, let $\lambda = \epsilon\mu$, for some $0 < \epsilon < 1$, so that the queues do not grow to infinity. Now we can represent the source node as a M/G/1 queue with Poisson arrival rate $\lambda$ and Bernoulli service rate $\mu$. With the known results for the M/G/1 queue [20, p. 212],

the average number of packets at a source node can be written as

$$\begin{aligned} \overline{L}_{source} &= \left(\frac{\rho}{1-\rho}\right)\left[1-\frac{\rho}{2}(1-\mu^2\sigma^2)\right] \\ &= \frac{\rho}{1-\rho}+\frac{\rho^2\mu^3}{2(1-\rho)}-\frac{\rho^2}{2(1-\rho)}-\frac{\rho^2\mu^4}{2(1-\rho)}, \end{aligned} \tag{2.1}$$

where $\rho$ is the traffic intensity, $\rho = \lambda/\mu$. Then from Little's Theorem, the delay in Phase 1 can be derived as

$$\begin{aligned} D_1 &= \frac{\overline{L}_{source}}{\lambda} \\ &\leq \frac{n}{m}\frac{2-\epsilon}{2(1-\epsilon)}. \end{aligned} \tag{2.2}$$

From the system model, we know that in every timeslot, a relay independently receives a packet with probability $\widetilde{\lambda} = \lambda/m$, and the relay node is scheduled for a potential packet transmission to the destination with probability $\widetilde{\mu} = 1/n$. Packet arrivals and transmission opportunities can be considered as Bernoulli distributions with mean value of $\widetilde{\lambda}$ and $\widetilde{\mu}$, and variance of $\sigma_a^2$ and $\sigma_s^2$ respectively. It follows a G/G/1 queue. This holds for each relay node, hence using Kingman's upper bound of G/G/1 queues [21, p. 476], the average packet waiting time is bounded as

$$\begin{aligned} \overline{W_i} &\leq \frac{\widetilde{\lambda}^2(\sigma_a^2+\sigma_s^2)}{2(1-\rho^2)}+\frac{1}{\widetilde{\mu}} \\ &= n+\frac{\rho(1+\rho)}{2n^2(1-\rho)}-\frac{\rho(1+\rho^2)}{2n^3(1-\rho)}. \end{aligned} \tag{2.3}$$

Then, the average number of packets in the relay queue is

$$\begin{aligned} \overline{L}_{relay} &= \widetilde{\lambda}\overline{W}_i \\ &\leq \rho+\frac{\rho^2(1+\rho)}{2n^3(1-\rho)}-\frac{\rho^2(1+\rho^2)}{2n^4(1-\rho)}. \end{aligned} \tag{2.4}$$

With $\overline{L}_{relay}$ and the packet arrival rate $\lambda$, the total time spent in Phase 2 can be derived from Little's Theorem as

$$\begin{aligned} D_2 &= m\frac{\overline{L}_{relay}}{\lambda} \\ &\leq n. \end{aligned} \tag{2.5}$$

Thus, the total time spent in the network is

$$\begin{aligned} D &= D_1 + D_2 \\ &\leq n + \frac{n}{m}\frac{2-\epsilon}{2(1-\epsilon)}. \end{aligned} \tag{2.6}$$

This equation shows that it is not possible to overcome the $O(n)$ characteristic of the end-to-end delay by decreasing the input data rate $\lambda$. It follows that the average delay in the order-of-magnitude sense is $D \leq O(n)$.

Furthermore, it is important to point out that the number of relays $m$ must be less than or equal to the order of $\log n$, which is the necessary condition that the scheduled transmission will be successful despite the SINRs.

### 2.4.2 Average Packet Delay Improvement

In this section, we present a fundamental bound on delay performance by adding a redundancy feature to the original opportunistic two-hop relaying scheme. This approach to reducing the delay was also used in [10, 14]. Consider sending a single packet from a source node to the destination node in an empty network, which means that there are no queues for the packet. In Phase 1, we transmit the same packet repeatedly from the source node to many different relay nodes, so that the same packet is then held by more than one relay node. This increases the chance

that the packet be scheduled by its destination node in Phase 2. Let $d_N$ represent the number of duplications by a source node in Phase 1. We write the average end-to-end packet delay expression as $D_r = D_{r(1)} + D_{r(2)}$, where $D_{r(1)}$, $D_{r(2)}$ denote respectively the delays in Phase 1 and 2 for the redundant scheduling algorithm.

#### 2.4.2.1 Delay with Redundant Scheduling in Phase 1

The delay in Phase 1 is the time for the packet to be transmitted to $d_N$ different relay nodes (each relay node receives at most one replica of the packet, since holding more than one replica of the same packet does not help to improve delay). The probability for the packet to be received by one of the $m$ relay nodes is $m/n$; the probability for the second replica is $(m-1)/n$, and so on, until for the last replica the probability is $(m - d_N)/n$. Now it is easy to show the average delay in Phase 1, on the condition that none of the $d_N$ replicas is scheduled by the destined destination node $i$ before all replicas are received, is

$$E[D_{r(1)}|\omega_1] = E\left[\frac{n}{m} + \frac{n}{m-1} + \cdots + \frac{n}{m-d_N}\right], \tag{2.7}$$

where $\omega_1$ represents the event that the none of the $d_N$ replicas is scheduled by the destined destination before all replicas are transmitted. We have

$$\Pr[\omega_1] = \left(1 - \frac{1}{n}\right)^{d_N} \geq e^{-\frac{d_N}{n}}. \tag{2.8}$$

Thus, the average delay in Phase 1 can be written as

$$\begin{aligned} E[D_{r(1)}] &= E[D_{r(1)}|\omega_1] \cdot \Pr[\omega_1] \\ &\leq E\left[\frac{n}{m-d_N} \cdot d_N\right] \cdot \left(1-\frac{1}{n}\right)^{d_N} \\ &\leq \frac{n}{m-d_N} d_N. \end{aligned} \tag{2.9}$$

#### 2.4.2.2 Delay with Redundant Scheduling in Phase 2

Consider now there are $d_N$ relay nodes holding the packet $p_i$. Let $\phi$ represent the probability that a relay node is scheduled by the destined destination node $i$. Based on the system model, we have $\phi = 1/n$. At any given timeslot, the probability for one of these $d_N$ relay nodes to be scheduled by the destined destination node is $\omega_2 = 1-(1-\phi)^{d_N}$. Thus, the average delay in Phase 2 is

$$\begin{aligned} E[D_{r(2)}] &\leq \frac{1}{1-(1-\phi)^{d_N}} \\ &\rightarrow \frac{1}{\phi d_N} = \frac{n}{d_N}. \end{aligned} \tag{2.10}$$

Given $\overline{D}_{r(1)}$ and $\overline{D}_{r(2)}$, we can derive the average end-to-end delay of scheduling with redundancy as

$$\begin{aligned} D_r &= E[D_{r(1)} + D_{r(2)}] \\ &= \frac{n}{m-d_N} \cdot d_N + \frac{n}{d_N}, \end{aligned} \tag{2.11}$$

where $1 < d_N < m$. It is easy to find the optimal value of $d_N = \sqrt{m}$, which minimizes the average end-to-end delay. It follows that

$$D_r \leq O(n/\sqrt{m}). \tag{2.12}$$

Note that one complication arises when implementing the scheduling with redundancy: when a packet has already been delivered to the destination, its leftover duplicated versions must be removed from the network in order to not create excess congestion. This complication can be solved by the *in-cell feedback protocol* described in [10].

### 2.4.3 Throughput Analysis with Redundant Scheduling

The system throughput of the opportunistic relaying scheme is given as

$$T = \min(T_1, T_2), \tag{2.13}$$

where $T_1, T_2$ denote the throughput for Phase 1 and Phase 2, respectively. Since with $m$ relays and as $n \to \infty$, $T_1 = \Theta(m)$ and $T_2 = \Theta(m)$, the system throughput is $T = \Theta(m)$. Now, since all packets are duplicated $\sqrt{m}$-fold by the redundant scheduling algorithm, the throughput is reduced to $T = \Theta(m/\sqrt{m})$. Hence, the system throughput with redundant scheduling becomes $T_r = \Theta(\sqrt{m})$. With the optimal order of $m = \Theta(\log n)$, we have $T_r = \Theta(\sqrt{\log n})$.

## 2.5 Throughput-Delay Tradeoff

In this part, we will provide two key results that capture the delay-throughput tradeoff inherent in wireless *ad hoc* networks with opportunistic scheduling. From the previous analysis, we know that delay or throughput cannot be improved by increasing redundancy or the number of relays. This implies that there are tradeoffs between the average end-to-end delay $D$ and the system throughput $T$, the number of relay nodes $m$, and the number of duplications $d_N$. These tradeoffs are formulated and analyzed next.

### 2.5.1 Tradeoff I: Throughput versus Redundancy

Assume $\mu$ is the throughput of each S-D pair. Thus all destination nodes receive packets at the same rate $\mu$. Let $R_i$ denote the redundancy associated with packets from the S-D pair $i$. That is, $R_i$ is the number of relay nodes holding the packet from a S-D pair $i$. For all destination nodes, the average number of packets that are received per timeslot is given by $\mu \sum_{i=1}^{n} R_i$. Since there is at most 1 packet that can be received by each of the $n$ destination nodes from one of the $m$ relay nodes which holding the desired packet per timeslot, we have

$$\mu \sum_{i=1}^{n} R_i \leq m. \tag{2.14}$$

### 2.5.2 Tradeoff II: Delay versus Redundancy

Suppose the average end-to-end packet delay is $\overline{D}$. In general, the average end-to-end delay of packets from specific S-D pairs could be different, and we denote $\overline{D}_i$ as the resulting average delay of packets from a S-D pair $i$. Then we have:

$$\overline{D} = \frac{1}{n} \sum_{i=1}^{n} \overline{D}_i. \tag{2.15}$$

Let the random variable $D_i$ represent the actual delay for this packet. The end-to-end packet delay $D_i$ has two parts, the delay for Phase 1 $D_{i(1)}$ and the delay for Phase 2 $D_{i(2)}$. Similar to the equation (2.11) obtained in Section 2.4.2, we have

$$\begin{aligned} E[D_i] &= E[D_{i(1)} + D_{i(2)}] \\ &= \frac{n}{m - R_i} \cdot R_i + \frac{n}{R_i}. \end{aligned} \tag{2.16}$$

Summing this equation (2.16) over all S-D pair $i$, we have

$$\begin{aligned} \overline{D} &= \frac{1}{n}\sum_{i=1}^{n}\overline{D}_i \\ &\geq O\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{R_i}\right). \end{aligned} \tag{2.17}$$

From Jensen's inequality, noting that the function $f(R_i) = 1/R_i$ is convex, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{R_i} \geq \frac{1}{\frac{1}{n}\sum_{i=1}^{n}R_i}. \tag{2.18}$$

Thus, we obtain

$$\overline{D} \geq O\left(\frac{n}{\sum_{i=1}^{n}R_i}\right). \tag{2.19}$$

### 2.5.3 Tradeoff III: Delay versus Throughput

From equation (2.14), we have

$$\sum_{i=1}^{n}R_i \leq \frac{m}{\mu}. \tag{2.20}$$

Combining equations (2.19) and (2.20), we obtain

$$\overline{D} \geq O\left(\frac{n}{m/\mu}\right). \tag{2.21}$$

Thus, the *delay/rate* characteristic necessarily satisfies the inequality $\frac{\overline{D}}{\mu} \geq O(\frac{n}{m})$. This is the tradeoff between average end-to-end packet delay and system throughput, the previous analysis of the $O(n)$ and $O(n/\sqrt{m})$ scheduling algorithm also meet this bound. Based on the proof of Tradeoff I and Tradeoff II, it is shown

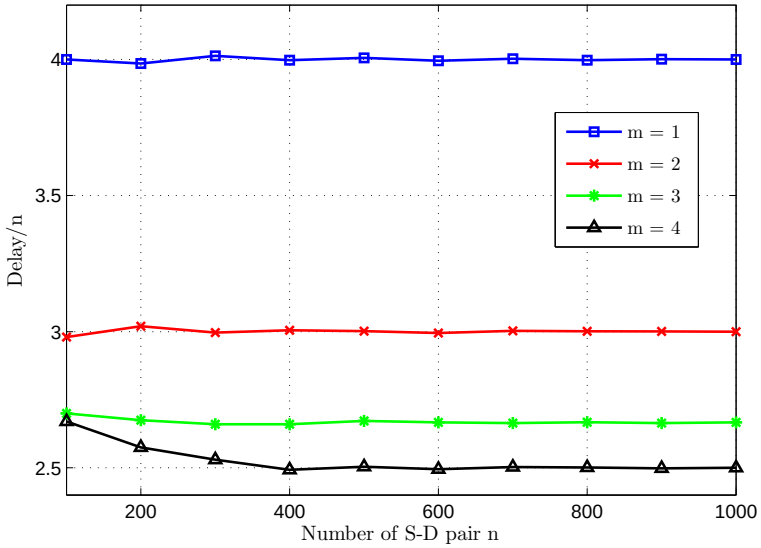


Figure 2.3: Average end-to-end packet delay normalized by the number of S-D pairs $n$ for different relay numbers $m$, note: from [35] ©2011 IEEE

that the relationship between delay and throughput is suited to any opportunistic scheduling protocol that stabilizes the networks with throughput $\mu$ while maintaining bounded average end-to-end delay $\overline{D}$. Furthermore, with the optimal order of $m = \Theta(\log n)$, the delay-throughput trade-off can be expressed as

$$\frac{Delay}{Throughput} \geq O(\frac{n}{\log n}).$$

## 2.6 Numerical Results and Discussions

We now present some numerical examples of the delay in the opportunistic relaying scheme under random channel connections. Consider the ideal condition in Phase 1, in which source nodes always have packets to transmit. Fig. 2.3 plots the average end-to-end packet delay normalized by the number of S-D pairs $n$ for various

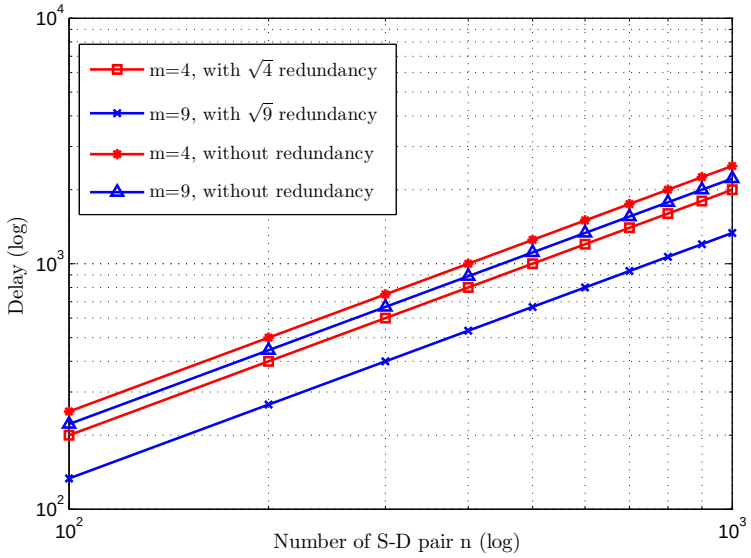


Figure 2.4: Average end-to-end packet delay for $\sqrt{m}$ redundant scheduling and opportunistic scheduling, note: from [35] ©2011 IEEE

cases of the number of relays $m$. It is observed that the normalized end-to-end packet delay is fixed with $n$, and it decreases with the number of relay nodes $m$.

Simulation results for opportunistic scheduling with $\sqrt{m}$ redundancy is shown in Fig. 2.4. Presented are curves with $\sqrt{m}$ redundancy for $m = 4$ and $m = 9$. Another two curves without redundancy used are also presented. From the comparisons between the redundant scheduling and the opportunistic scheduling, clearly that using redundancy can improve the delay performance. The values of the average end-to-end delay as obtained from the simulation results agree closely with the theoretical analysis before.

## 2.7 Concluding Remarks

In this Chapter, we have studied the delay performance and the delay throughput relation with opportunistic scheduling in wireless networks. Our contributions are three-fold. We have derived the upper bound of the average end-to-end packet

delay of the opportunistic relaying scheme which includes full effects of queueing. Our second contribution is to develop $\sqrt{m}$ redundancy scheduling which can improve average delay. Our third contribution is to establish the delay-throughput trade-off as $delay/throughput \geq O(\frac{n}{\log n})$ for the opportunistic relaying scheme. This indicates that no lower delay and higher throughput can be simultaneously achieved with opportunistic scheduling.

# CHAPTER 3: ADAPTIVE RATE TRANSMISSIONS

## 3.1 Introduction and Motivation

In this chapter, we extend the design of scheduling algorithm based on our previous work in Chapter 2. Our previous work in Chapter 2 focuses on fixed-rate transmission only with opportunistic scheduling in wireless networks. For fixed rate transmissions, there is a waste of resources for users with strong channels, as these could have supported higher transmission rates. However, with adaptive rate transmission, both the average system throughput and delay can be improved, if no further requirements or higher complexity are imposed. In this Chapter, we propose an adaptive rate transmission scheme under the same assumptions and with only modest cooperation among relays. Specifically, the proposed scheme operates under the following constraints: available channel state information (CSI) only at receivers and cooperation only among relays in a decentralized network. We have found such constraints to be quite practical; for example, in a network with a large number of nodes, receivers may obtain CSI by measuring pilot signals and relays can be infrastructure nodes that connect to each other through a wired backbone. In such a network, obtaining CSI at transmitters and setting up cooperation among source/destination nodes may not be feasible.

In our proposed adaptive-rate transmission scheme, we use an opportunistic scheduling method, in which only the nodes that benefit from the multiuser diversity are scheduled for transmission and low rate feedback from receivers is employed.

Consider a network with $n$ source-to-destination (S-D) pairs and $m$ relays over independent and identically distributed (i.i.d.) Rayleigh fading. We first show that in the limit of large $n$ and fixed $m$, the system throughput scales as $\frac{m}{2}\log\log n$. We also prove that this throughput scaling result is the same achievable scaling even with perfect CSI assumption at transmitters and full cooperation among nodes, which is a quite interesting result in its own right. To guarantee the linear growth of the system throughput with $m$, we derive the optimal scaling of $m$ as $\Theta(\log n)$. In addition, the closed-form delay expression of the proposed scheme is derived for better understanding of the proposed scheme.

## 3.2 Literature Review

Gupta and Kumar in [1] started on the study of system throughput in wireless *ad hoc* networks. Numerous schemes have since been proposed that apply different assumptions to the channel state information (CSI) and levels of cooperation among communicating nodes [2]-[5]. These studies have contributed to the understanding of system throughput in wireless networks. One common thread among these studies is a focus on the scaling of the system throughput with fixed rate transmissions. For some network architectures, such as *ad hoc*, sensor and CR networks, using adaptive rate transmission may provide better system performance, due to its ability to change transmission parameters. In particular, the transmit power and modulation level can be adjusted according to instantaneous channel conditions using an adaptive resource allocation policy [6]. This approach can provide better system performance and/or extend the lifetime of the *ad hoc* or sensor nodes. Much of the literature on adaptive transmission schemes has dealt with variable coding rates and power allocation schemes in wireless networks, under numerous fading models, power and quality of service (QoS) constraints [7]-[11]. More recently, resource allocation issues

with relay-aided cooperative transmission in distributed MIMO systems are considered in [12], [13]. Our work in this chapter is motivated by the fact that relatively little can be found in the literature on throughput with adaptive rate transmission under limited CSI and relay cooperation.

## 3.3 System Model and Assumptions

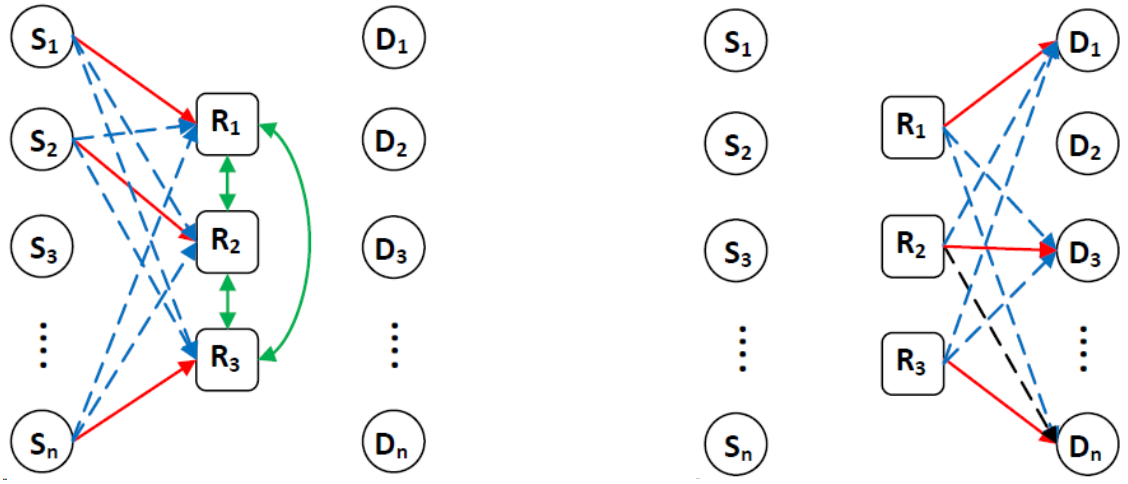


Figure 3.1: A two-hop network with $n$ S-D pairs and 3 relays, note: from [37] ©2012 IEEE

### 3.3.1 Network Model

Consider a wireless network with $n$ source nodes sending data to their designated destination nodes, through the help of $m$ relay nodes as shown in Fig. 3.1. We assume the source nodes are backlogged and relays do not generate their own traffic. The communication protocol is restricted to two-hop decode-and-forward transmissions, in which the source nodes communicate with their destination nodes only through the half-duplex relays. Specifically, in the first hop (Phase 1), a subset of source nodes are scheduled to transmit data to $m$ relays. After decoding and

buffering the received data, relays forward the data to a subset of destination nodes in the second hop (Phase 2).

Due to the fact that there is no cooperation among source or destination nodes, the source nodes encode their packets independently and the destination nodes treat the interference as noise and decode the received packets independently. We assume relays are fixed reliable infrastructure nodes that have no transmission bandwidth, power and processing constraints. Each relay can communicate the selected source ID and its associated SINR to all the other relays. Subsequently, each relay can compare the received information (source IDs and SINRs) from all other relays and decide whether it is the receiver of that source. In addition, we adopt independent encoding and independent decoding at relays in Phase 1 and Phase 2, respectively, in order to avoid complex algorithms that would have to be used in relays for interference cancelation.

### 3.3.2 Channel Model

We assume CSI is available only at the receivers. In many wireless access networks, CSI at receivers can be obtained by measuring pilot signals. The transmitters have no knowledge of CSI, but have access to the feedback information from the receivers. We consider simultaneous transmission in each hop, *i.e.*, all nodes are operating in the same frequency band, with the presence of fading. We assume channel realizations from source nodes to relays and from relays to destination nodes experience i.i.d. Rayleigh fading. Accordingly, the channels are assumed to be constant during the transmission duration $T$ in each phase of the two-hop communication. We denote the channel realizations between source nodes $i$, $1 \leq i \leq n$, and relays $r$, $1 \leq r \leq m$, as $h_{i,r}$ and the channel realizations between relays $r$ and destination nodes $j$, $1 \leq j \leq n$, as $g_{r,j}$. Based on this assumption, the channel gains follow an i.i.d.

exponential distribution, *i.e.*, $|h_{i,r}|^2 \sim Exp(1)$ and $|g_{r,j}|^2 \sim Exp(1)$. This two-hop relaying model is applicable to scenarios for which there is no line-of-sight between transmitter and receiver and relays help to receive and forward signals [30, 31].

### 3.3.3 Opportunistic Scheduling with Cooperative Relaying

We now describe the scheduling method. We use opportunistic scheduling for nodes selections, *i.e.*, only a subset of nodes that benefit from multiuser diversity are scheduled for transmission. We select a specific relay $r$ to show the scheduling method. For Phase 1:

- Relay $r$ measures the channel gains of $|h_{i,r}|^2$, $i = 1, ..., n$ and selects source $i_r$, which has the strongest channel, *i.e.*, $i_r = \arg\max_i |h_{i,r}|^2$.

- Relay $r$ exchanges the selected source ID number with the other $(m-1)$ relays and calculates the corresponding SINR as

$$\mathsf{SINR}_{\mathsf{i_r,r}} = \frac{P|h_{i_r,r}|^2}{N_r + \sum_{k\in\Gamma, k\neq i_r} P|h_{k,r}|^2}, \tag{3.1}$$

  where $P$ is the fixed transmit power, $N_r$ is the Gaussian noise power at relay $r$ and $\Gamma$ denotes the selected source nodes by all the relays. Relay $r$ then feeds back the $\mathsf{SINR}_{\mathsf{i_r,r}}$ value to source $i_r$. In case of multiple relays select the same source, only the relay with the highest SINR will send feedback.

- Upon receiving feedback, source $i_r$ starts transmitting data $x_{i_r}$ at the rate as

$$R_{i_r}^{P1} = \log_2(1 + \mathsf{SINR}_{\mathsf{i_r,r}}). \tag{3.2}$$

- The transmitted data $x_{i_r}$ is received in relay $r$ as

$$y_r = \sqrt{P} h_{i_r,r} x_{i_r} + \sum_{k \in \Gamma, k \neq i_r} \sqrt{P} h_{k,r} x_k + n_r, \tag{3.3}$$

where $x_i$ represents the information symbols from source $i$, having $E[|x_i|^2] = 1$. $n_r$ denotes the additive noise at relay $r$. From an information-theoretic point of view, relay $r$ is able to independently decode the received data at the rate of $R_{i_r}^{P1}$. The decoded data is then buffered in relay $r$.

For Phase 2,

- A specific destination node $j$ is able to calculate SINRs from all the relays, after measuring the channel gains of $|g_{r,j}|^2$, as

$$\mathsf{SINR}_{\mathsf{r,j}} = \frac{P|g_{r,j}|^2}{N_j + \sum_{1 \leq l \leq m, l \neq r} P|g_{l,j}|^2}, \tag{3.4}$$

where $N_j$ is the Gaussian noise power at destination node $j$.

- Among the calculated SINRs, destination node $j$ selects the one which has the highest $\mathsf{SINR}_{\mathsf{r_j,j}}$ and $\mathsf{SINR}_{\mathsf{r_j,j}} \geq \beta$. Then it feeds back to relay $r_j$. If the largest SINR can not satisfy the threshold $\beta$, node $j$ keeps silent. Note that in the limit of large $n$ and fixed $m$, it is necessary to put SINR constraints in Phase 2 for further destination nodes selection to reduce the feedback overhead.

- Upon receiving the feedback information, relay $r_j$ starts transmitting data $y_{r_j}$ at a rate as

$$R_{r_j}^{P2} = \log_2(1 + \mathsf{SINR}_{\mathsf{r_j,j}}). \tag{3.5}$$

- The the transmitted data $y_{r_j}$ is received in destination node $j$ as follows,

$$z_j = \sqrt{P} g_{r_j,j} y_{r_j} + \sum_{1 \le l \le m, l \ne r_j} \sqrt{P} g_{l,j} y_l + n_j, \tag{3.6}$$

where $n_j$ denotes the additive noise at destination node $j$. Since the transmission rate is $\log_2(1+\mathsf{SINR}_{\mathsf{r_j,j}})$, destination node $j$ is able to independently decode the received data.

It is noted that in the steady state of the system, relays have the ability to buffer data received from source nodes, such that a selected relay always has buffered data for the destined destination.

## 3.4 Performance Analysis

### 3.4.1 Sum System Throughput

We now analyze the throughput of the proposed adaptive rate transmission scheme [36]. We first show that the scheme achieves $\Theta(\frac{m}{2}\log\log n)$ throughput in the limit of large $n$ and fixed $m$. We then prove that the achieved throughput is actually the optimal throughput for two-hop relay communications. Furthermore, we derive the optimal scaling of $m$ as $\Theta(\log n)$, which implies that a linear increase in throughput is obtained.

#### 3.4.1.1 Throughput in Phase 1

We select a specific relay r to show the scheduling method without loss of generality, due to the fact that all relays operate independently. According to our proposed scheduling policy in Phase 1, relay $r$ selects a source $i_r$ for transmission. Source node $i_r$ transmits at a rate of $R_{i_r}^{P1} = \log_2(1+\mathsf{SINR}_{\mathsf{i_r,r}})$. Similarly, all the other

$m-1$ relays schedule their corresponding source nodes. We denote the scheduled source nodes constitute a set $\Gamma \subset 1, ..., n$, and $\gamma$ as the elements (scheduled source nodes) in $\Gamma$. Since with $m$ relays, there can be up to $m$ source nodes to be scheduled, hence, $|\Gamma| \leq m$. Specifically, when same source nodes are scheduled by different relays, $\mid \Gamma \mid < m$; when each of the $m$ relays schedules a different source, $|\Gamma| = m$. After the scheduled source nodes receive all the feedback, they start transmitting according to the received SINRs, which leads to the sum-rate throughput in Phase 1 as $R_1 = \sum_{\gamma \in \Gamma} \log(1 + \mathsf{SINR}_\gamma)$, in which $\mathsf{SINR}_\gamma$ denotes the SINRs calculated by each of the scheduled source $\gamma$.

For Rayleigh-fading, it is well studied that the term $\mathsf{SINR}_{\mathsf{i_r,r}} = \frac{P|h_{i_r,r}|^2}{N_r + \sum_{k \in \Gamma, k \neq i_r} P|h_{k,r}|^2}$ scales as $\log n$ due to multiuser diversity gain with a pool of $n$ nodes. Hence, the sum rate throughput can be written as

$$R_1 = \sum_{\gamma \in \Gamma} \log(1 + \log n) \geq m Pr[N_m] \log(\log n), \tag{3.7}$$

in which the term $Pr[N_m]$ is the probability of each of the $m$ relays schedules a different source. The equation is based on the multiuser diversity gain, the inequality is because of the fact that only consider the case of $|\Gamma| = m$.

According to the channel model and opportunistic scheduling, the probability for each source to be scheduled (with the strongest channel) is $\frac{1}{n}$, the probability for $m$ relays schedules a different source, can be written as

$$Pr[N_m] = n(n-1) \cdots (n-m+1)/n^m. \tag{3.8}$$

Hence, $R_1$ can be further written as

$$R_1 \geq m\frac{n(n-1)\cdots(n-m+1)}{n^m}\log\log n. \tag{3.9}$$

Note that for fixed rate scheme in [8], the throughput in Phase 1 can be derived as

$$R_1^{fixed} = m\frac{n(n-1)\cdots(n-m+1)}{n^m}Pr[\mathsf{SINR} \geq 1]. \tag{3.10}$$

It is obvious to find that the adaptive rate transmission achieves a gain of $\log\log n$ $Pr[\mathsf{SINR} \geq 1]$. Note that in Phase 1, there is no threshold requirement on the transmission rate $R_1$, *i.e.*, as long as the source nodes transmit according to the adaptive transmission rate $R_1$, the relays are able to decode the received data.

#### 3.4.1.2 Throughput in Phase 2

The difference between Phases 1 and 2 is that, the receivers in Phase 1 have no *a priori* knowledge of who are the transmitters, while in Phase 2 the transmitters (the relays) are known. We first select a specific destination node $j$ for analysis without loss of generality since all the destination nodes are independent. Based on the scheduling policy, if the highest $\mathsf{SINR}$ satisfies $\mathsf{SINR}_{\mathsf{r_j,j}} \geq \beta$, relay $r_j$ starts transmitting data at the rate of $R_{r_j}^{P2} = \log_2(1+\mathsf{SINR}_{\mathsf{r_j,j}})$. Accordingly, all the other $m-1$ relays receive their corresponding $\mathsf{SINR}$s and start adaptive transmissions. Hence, the sum-rate throughput in Phase 2 can be written as $R_2 = \sum_{r=1}^{m}\log(1+\mathsf{SINR_r})$, in which $\mathsf{SINR_r}$ denotes the $\mathsf{SINR}$s calculated by each of the relays. Similar to Phase 1, we may further calculate $R_2$ as follows,

$$R_2 = \sum_{r=1}^{m}\log(1+\log n) \rightarrow m\log\log n, \mathsf{with}\ n\rightarrow\infty. \tag{3.11}$$

Note that compared with fixed rate transmission, which can be written as

$$R_2^{fixed} = mPr[\text{relay receives a feedback}], \tag{3.12}$$

we see that the adaptive rate transmission scheme achieves a minimum gain of $\log\log n$. Note that in Phase 2, there exists a SINR threshold requirement on the transmission rate. This constraint is imposed to select a subset of best destination nodes and reduce the total feedback overhead. In order to guarantee that the throughput does not decrease in Phase 2, we need to find the optimal $m$. The detailed analysis is provided in the sequel.

#### 3.4.1.3 Throughput of Adaptive Rate Transmission

From the previous analysis, considering the penalty by two-hop transmissions, the overall system throughput can be obtained as,

$$\begin{aligned} R &= \frac{1}{2}\min\{R_1, R_2\} \\ &\rightarrow \frac{m}{2}\log\log n, \text{with } n \rightarrow \infty. \end{aligned} \tag{3.13}$$

We now show that the above achieved throughput (3.13) is equal to the optimal throughput for two-hop opportunistic relaying. It is reasonable to state that for two-hop relaying systems, the optimal throughput can be achieved with full cooperation among the transmitting and receiving nodes and full CSI knowledge at both transmitters and receivers in both hops. This setup is equivalent to the model with MIMO with multiple-access channels (MAC) as Phase 1 and MIMO Broadcast Channel (BC) as Phase 2. The MIMO-MAC and MIMO-BC models have been studied in [32] and [33], in which both are proven to achieve the throughput as

$m \log\log n$, respectively. Taking into the $\frac{1}{2}$ factor due to two-hop transmission, the optimal throughput is given as $\frac{m}{2} \log\log n$.

Note that the proposed scheme employs independent coding at the transmitters and independent decoding at the receivers in both Phases. It means that the scheme operates in a decentralized manner, no joint/cooperative encoding or decoding is needed in this scheme.

#### 3.4.1.4 Optimal Value of the Number of Relays

The system throughput scales as $\frac{m}{2} \log\log n$, in which the term $\log\log n$ is due to the adaptive transmission according to the instantaneous channel conditions, and $m$ relates to the concurrent transmissions. By increasing $m$, we will have more concurrent transmissions; however, it also increases the interference, since all nodes are operating in the same frequency band. A consequence of this is a violation of the SINR threshold constraint in Phase 2. A very interesting research question concerns the tradeoff between increasing the number of concurrent transmissions and also satisfying the SINR requirement and how fast $m$ can grow to guarantee the linear scaling of $m$ in the throughput. In this subsection, we analyze the optimal scaling of $m$.

In Phase 1, each relay node schedules a source for transmission with no SINR threshold requirement, thus, with $n$ source nodes, the growing of relays $m$ can be as fast as $\Theta(n)$, while still retaining the linear throughput growing with $m$ in Phase 1.

In Phase 2, under the constraint of SINR threshold, we have to consider the total amount of interference generated when increasing $m$. We use the well-known genie-aided scheme for our derivations. The concept of concurrent successful transmissions is used to calculate the system throughput in wireless *ad hoc* networks for opportunistic scheduling. With at most $n$ relay-destination (R-D) pairs in the net-

work, the genie scheme can be summarized as follows [34]: First, the scheme selects $c$ $(1 \leq c \leq n)$ active R-D pairs which are scheduled for transmissions; for each selection, if all the $c$ received SINRs are greater than the threshold 1, then the $c$ concurrent transmissions are successful. For each selection, if the $c$ concurrent transmissions are successful, we call the current selection as a valid group. With up to $n$ R-D pairs in Phase 2, we have $\binom{n}{c}$ different possible ways to select $c$ active R-D pairs for transmission. Furthermore, since each relay can be scheduled by any destination nodes, for each $c$ selected active R-D pairs, there are $c!$ different ways to associate the R-D pairs. Thus, there is $\binom{n}{c}c!$ different ways to select the active R-D pairs. We denote $X(c)$ as the total number of valid groups in which all $c$ concurrent R-D pairs are successful. To have at least $c$ concurrent successful transmissions is equivalent to $X(c) \geq 1$. Based on the analysis, we have the following function,

$$\begin{aligned} X(m) &= \sum_{\substack{S \in 1,\ldots,n; \\ |S|=c}} 1(\mathsf{SINR}_{\mathsf{r,j}} \geq \beta, \forall r, j \in S) \\ &= \sum_{\substack{S \in 1,\ldots,n; \\ |S|=c}} 1(\frac{P|g_{r,j}|^2}{\frac{1}{P_R} + \sum_{l \neq r} P|g_{l,j}|^2} \geq \beta, \forall r, j, l \in S) \end{aligned}$$

where $P_R$ is the average SNR of the R-D link. $S$ is the group with selected active nodes,

$$|g_{r,j}| = \max\{|g_{1,j}|^2, \ldots, |g_{n,j}|^2\}, \tag{3.14}$$

and

$$\sum_{l \neq r} |g_{l,j}|^2 = \sum_{\substack{l \in S \\ k \neq r}} |g_{l,j}|^2, \tag{3.15}$$

based on our scheduling policy.

First we upper-bound $Pr[X(c) \geq 1]$ as

$$\begin{aligned} Pr[X(c) \geq 1] &\leq \mathbb{E}[X(c)] \\ &= \binom{n}{c} c!(Pr[\frac{P|g_{r,j}|^2}{\frac{1}{P_R} + P\sum_{l \neq r} |g_{l,j}|^2} \geq \beta])^c, \end{aligned} \tag{3.16}$$

where (5.5) is because of the Markov's inequality, (5.6) is due to the linear property of expectation and the SINRs of the nodes' are i.i.d.

Next we further upper bound the term $(Pr[\frac{P|g_{r,j}|^2}{N_j + P\sum_{l \neq r} |g_{l,j}|^2} \geq \beta])^c$. For simplicity, we denote $M = P|g_{r,j}|^2 = P\max\{|g_{1,j}|^2, \ldots, |g_{n,j}|^2\}$ and the interference from all the other scheduled $c-1$ concurrent transmissions as $Z = P\sum_{\substack{l \in S \\ l \neq r}} |g_{l,j}|^2 = P(\sum_{l \in S} |g_{l,j}|^2 - M)$. The probability term can be written and further upper-bounded as

$$\begin{aligned} Pr[\frac{M}{\frac{1}{P_R} + Z} \geq \beta] &= \int_0^\infty Pr[\frac{M}{\frac{1}{P_R} + Z} \geq 1 | Z = z] f_Z(z) dz \\ &= \int_0^\infty Pr[M \geq (\frac{1}{P_R} + z) | Z = z] f_Z(z) dz \\ &= \int_0^\infty Pr[M - \mu \geq (\frac{1}{P_R} + z) - \mu] f_Z(z) dz \\ &\leq \int_0^\infty \frac{\sigma^2}{\sigma^2 + ((\frac{1}{P_R} + z) - \mu)^2} f_Z(z) dz, \end{aligned} \tag{3.17}$$

where (5.7) is based on one-sided Chebyshev inequality $Pr[X - \mu \geq \omega] \leq \frac{\sigma^2}{\sigma^2 + \omega^2}$. For the distribution of interference which is termed as $f_Z(z)$ in (5.7), we adapt the same approximation as in [8], we can further upper-bound (5.7) as

$$Pr[\frac{M}{\frac{1}{P_R} + Z} \geq \beta] = \frac{e^{-1/P_R}}{2^{c-1}}. \tag{3.18}$$

Combining (5.6) and (5.8), we have

$$\begin{aligned} Pr[X(c) \geq 1] &\leq \binom{n}{c} c!(Pr[\frac{M}{\frac{1}{P_R} + Z} \geq \beta]^c) \\ &= \frac{n!}{(n-c)!}(\frac{e^{-1/P_R}}{2^{c-1}})^c \leq (\frac{2ne^{-1/P_R}}{2^c})^c \\ &= e^{c(\log(2ne^{-1/P_R}) - c\log 2)}. \end{aligned} \tag{3.19}$$

According to the genie scheme, we set the value of $c$ as

$$(\log n + \log 2 + \log e^{-1/P_R})/\log 2, \tag{3.20}$$

so that the term $Pr[X(c) \geq 1] \to 0$, which means that there is no valid groups with the value of $c$ concurrent successful transmissions as

$$c = (\log n + \log 2 + \log e^{-1/P_R})/\log 2. \tag{3.21}$$

It is equivalent that $m$ can grow at most as fast as $\Theta(\log n)$.

Note that from the analysis above, it is noted that the bottle neck of the system throughput occurs in Phase 2. This is reasonable in the sense of CSI availability that in Phase 1, the relays obtain the global CSI by cooperation and each schedule a corresponding source node for transmission with a success-guaranteed transmission rate. While in Phase 2, without global CSI, the destination nodes can only use the SINR threshold constraint to schedule successful transmissions, which limited the throughput in Phase 2.

### 3.4.2 Average Packet Delay

Generally, in opportunistic schemes, fairness and delay are two concerns that need to be addressed. The fairness issue is less relevant in the i.i.d. channel model, since on average every node is scheduled with same probability, and the delay consideration assumes a higher priority [35]. We define average end-to-end packet delay as the average time that it takes the packet to arrive at the designated destination node from a source [12]. For two-hop communications, the average delay $D = D_1 + D_2$, where $D_1$ and $D_2$ denote the delay in Phase 1 and Phase 2, respectively. We introduce the queueing network model illustrated in Fig. 3.2 to derive the average delay.

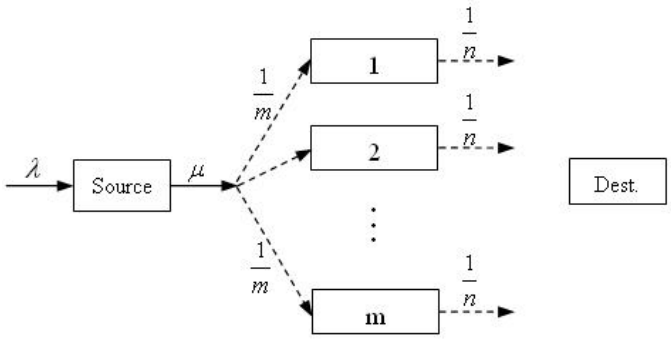


Figure 3.2: Queueing model of the two-hop opportunistic relaying scheme, note: from [37] ©2012 IEEE

#### 3.4.2.1 Delay in Phase 1

We assume the packet-arrival process at a source node is a Markov process with arrival rate $\lambda$, the service has a Bernoulli distribution with a departure rate $\mu$ when a transmission opportunity arises. Since the arriving packets have to be buffered in the source before forwarding to relays, there are potential queues in the source nodes' buffer. The probability for a source to be scheduled by one of $m$ relays is $\mu = \frac{m}{n}$, since it is equally likely for each of the $n$ source nodes to have the best

channel. With the known results for this queuing model [20], the average number of packets at a source is given by,

$$\begin{aligned} \overline{L}_{source} &= \frac{\rho(1-\lambda)}{1-\rho} \\ &= \frac{\lambda(1-\lambda)}{\mu-\lambda}, \end{aligned} \tag{3.22}$$

where $\rho = \lambda/\mu$ is the traffic intensity. Note that we restrict $0 < \rho < 1$ to ensure the queues do not grow to infinity. Using Little's Theorem, we may simply derive the delay in Phase 1 as

$$\begin{aligned} D_1 &= \frac{\overline{L}_{source}}{\lambda} \\ &= \frac{1-\lambda}{\mu-\lambda} \\ &= \frac{n}{m(1-\rho)} - \frac{\rho}{1-\rho}. \end{aligned} \tag{3.23}$$

#### 3.4.2.2 Delay in Phase 2

The probability for transmission of a given packet from the output of a source to a designated relay node is $1/m$, because each of the $m$ relay nodes are equally likely. Thus, packet arrival rate is $\widetilde{\lambda} = \lambda/m$. Since the transmission opportunity of relay to destination node arises with equal probability for each of the $n$ destination nodes, the relay is scheduled for a potential packet transmission with probability $\widetilde{\mu} = 1/n$. The packet arrival process and departure opportunities are mutually independent events in the relay, which follows that the discrete time Markov chain for queue occupancy in the relay [10]. Hence, this queuing system can be modeled as the simple birth-death chain, which is similar to M/M/1 model with arrival rate

$\widetilde{\lambda}$ and departure rate $\widetilde{\mu}$. Accordingly, the average number of packets at a relay is

$$\overline{L}_{relay} = \frac{\widetilde{\lambda}}{\widetilde{\mu} - \widetilde{\lambda}}. \tag{3.24}$$

From the analysis in Section 3.4.1, when $m$ does not grow faster than $\Theta(\log n)$, the SINR threshold can be satisfied. Using Little's Theorem, the delay in Phase 2 can be derived as

$$\begin{aligned} D_2 &= \frac{\overline{L}_{relay}}{\widetilde{\lambda}} \\ &= \frac{1}{\widetilde{\mu} - \widetilde{\lambda}} \\ &= \frac{n}{1-\rho}. \end{aligned} \tag{3.25}$$

#### 3.4.2.3 Total Delay

The average end-to-end packet delay $D = D_1 + D_2$ can now be obtained as

$$\begin{aligned} D &= \frac{\lambda(1-\lambda)}{\mu - \lambda} + \frac{n}{1-\rho} \\ &= \frac{n-\rho}{1-\rho} + \frac{n}{m(1-\rho)}. \end{aligned} \tag{3.26}$$

In the order-of-magnitude sense, (3.26) indicates $D$ grows with the scaling of $n$ and this characteristic can not be removed by decreasing the arrival rate. Furthermore, it is noted that $m$ must be less than or equal to the order of $\Theta(\log n)$, which is the necessary condition that the scheduled transmission will be successful despite the SINR threshold requirement in Phase 2.

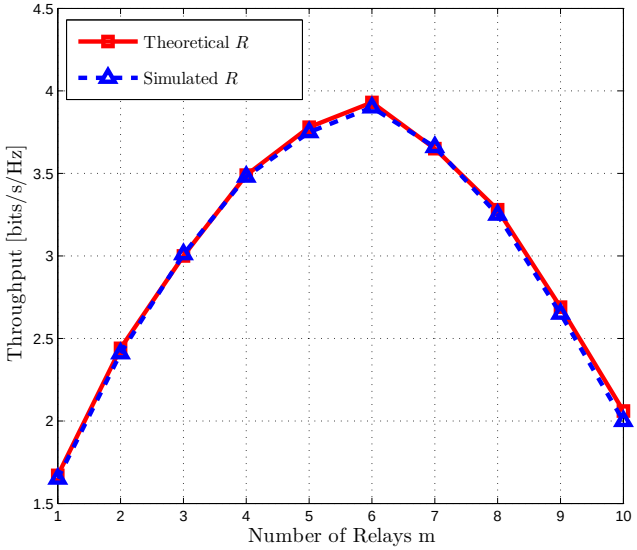


Figure 3.3: Theoretical and simulated throughput, note: from [37] ©2012 IEEE

## 3.5 Numerical Results and Discussions

The proposed adaptive rate transmission scheme is simulated under Rayleigh-fading. Throughout the simulations, we set the SNR in both phases as 10 dB. The data presented is obtained by averaging over 2,000 simulation runs.

Fig. 3.3 and Fig. 3.4 plot the throughput and delay, respectively, illustrating that the simulation results agree closely with our analysis. Noting that $\log 500 = 6$, which is in consistent with our analysis in Section III-D that $m$ can grow at most as fast as $\Theta(\log n)$ to guarantee a linear increase in throughput with $m$.

Simulated throughput comparison between adaptive and fixed rate transmission is shown in Fig. 3.5. Presented are throughput as a function of $m$, the throughput with $n = 500$ is larger than $n = 100$ because of larger multiuser diversity gain. Simulated delay comparison between adaptive and fixed rate transmission schemes is shown in Fig. 3.6, with different values of $\rho$. Presented are average delay as a function of $n$ under $\rho = 0.5$ and $\rho = 0.8$. From the comparisons, clearly the adaptive

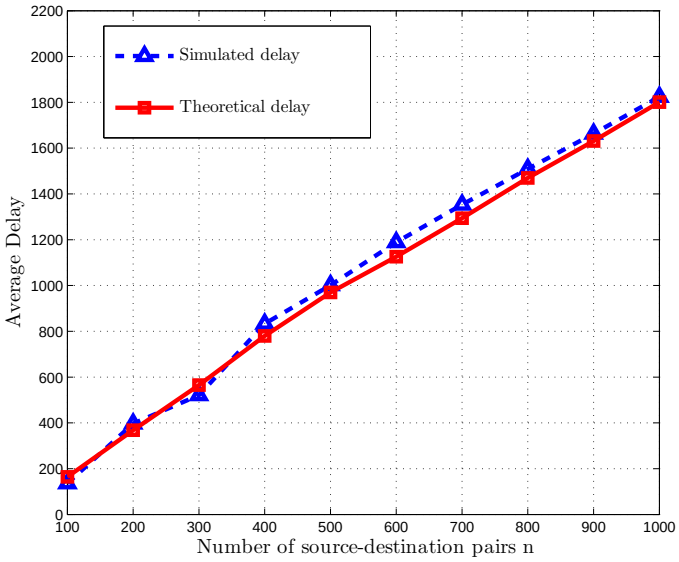


Figure 3.4: Theoretical and simulated delay, note: from [37] ©2012 IEEE

rate transmission provides a higher throughput and lower delay as compared to fixed rate transmission.

## 3.6 Concluding Remarks

In this Chapter, we present an adaptive rate transmission scheme with opportunistic scheduling in a two-hop relay network with CSI only available at receivers. The scheme operates in a decentralized fashion with independent encoding at transmitters and independent decoding at receivers. The proposed scheme achieves a system throughput of $\frac{m}{2} \log \log n$, which is shown to be the optimal throughput for perfect CSI knowledge and full cooperation among nodes. Results are also derived for the optimal scaling of the number of relays in order to guarantee linear growth of the system throughput, as is a closed-form expression for the average packet delay of the proposed scheme. Computer simulations are presented that help to contribute to an understanding of the proposed approach and its advantages vis-*à*-vis other

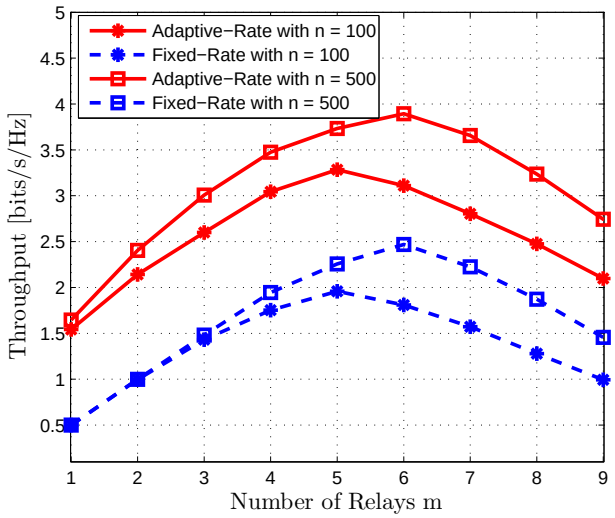


Figure 3.5: Throughput comparison between adaptive and fixed rate transmission schemes, as a function of $m$, note: from [37] ©2012 IEEE

methods. In the future work, we will focus on designing adaptive rate transmission schemes which allow cooperative relays to form a distributed MIMO, under the same constraints of practical CSI assumption (CSIR+feedback), no cooperation among source/destination and no central controller.

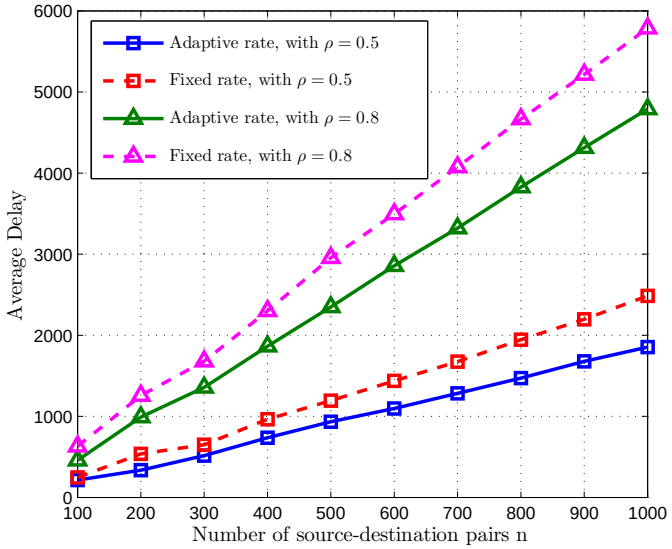


Figure 3.6: Average delay for adaptive and fixed rate transmission schemes under different traffic intensity values $\rho$, note: from [37] ©2012 IEEE

# CHAPTER 4: BUFFER-AWARE PACKET SCHEDULING

## 4.1 Introduction

The 3GPP long-term evolution (LTE) system provides higher date rate transmission with the use of OFDMA-based downlink transmission schemes and MIMO techniques. In LTE, the transmission time is divided into scheduling intervals or subframes. In each subframe, a subset of users are selected as the scheduling candidate set (SCS) based on scheduling metrics in time-domain. Available resource blocks (RBs) are assigned to different users in frequency-domain. The packet scheduler plays the central role in exploiting the users' channel and traffic knowledge for improving system performance.

In this Chapter, we consider the design of a scheduling algorithm focusing on improving the system performance by reducing the packet delay, while maintaining the queue stability condition of the networks. We propose a new buffer-aware scheduling algorithm, termed as buffer-aware adaptive (BAA) scheduler. The proposed scheduler considers both channel and buffer conditions to make scheduling decisions. Specifically, channel conditions are used to select candidate users, *i.e.*, only the users whose transmission rates satisfy a certain condition are considered as candidate users. This guarantees a maximum of throughput deduction. Then the buffer conditions are used to make scheduling decisions among all the candidate users, *i.e.*, the user with largest queue backlog among candidate users will be scheduled, to provide improved delay performance. Furthermore, the generalized form

of the proposed algorithm is compatible with any reasonable metrics, which can be implemented to form a specific scheduling algorithm. Stability considerations of the proposed algorithm are provided, along with the average throughput lower bound and approximation. The numerical results show the consistency with our analysis, and the proposed scheduler outperforms the existing scheduling algorithms in terms of average packet delay.

## 4.2 Literature Review and Motivation

Focusing on throughput performance, the maximum throughput (MT) scheduler is introduced in [38–40], which schedules only users with the best instantaneous channel conditions to transmit in each scheduling interval. MT maximizes sum system throughput at the loss of fairness to cell edge users. Round robin (RR) is the most fair but channel unaware scheduler, in which users' transmissions takes place in a strict numerical order [41]. The MT and RR schedulers leave room for various schedulers that lie in between them. Proportional fair (PF) scheduler [42–44] weights users' instantaneous transmission rates by their average rates to tradeoff throughput with fairness. PF is the practical scheduling algorithm that currently implemented in most LTE systems. Although MT, RR and PF algorithms can be directly applied to the finite-buffer traffic models, the algorithms are actually buffer-unaware. Along this work with the considerations of traffic demands and different quality of service (QoS) requirements, several buffer-aware schedulers are proposed. Sharif and Hassibi in [18] proposed "d-algorithm" which schedules the user with largest buffer size among "d" best candidate users, in terms of their channel condition, to significantly improve the delay performance without sacrificing too much on throughput. Neely in [45] proposed a Longest Connected Queue (LCQ) policy, which schedules the user with the largest product of channel state and queue backlog. Although

these works have made great strides toward reducing average packet delay, implementations of the schemes are less adaptive to channel conditions with unbounded throughput deduction [18] or require more restrictive throughput region and stability condition [45].

## 4.3 System Model and Assumptions

### 4.3.1 Network and Channel Model

We consider a downlink multiuser system with one base station serving $n$ users. We consider that only one channel resource is available to be scheduled for one user at time slot $t$. We assume channel gains from base station to users are independent and the small scale channel fading is independent and identically distributed (i.i.d.) Rayleigh fading. Accordingly, the channels are assumed to be constant during the transmission duration $t$, *i.e.*, the channel coherence time $T \geq t$. We consider both single antenna, *i.e.*, single-input single-output (SISO) and multi-antenna, *i.e.*, multiple-input multiple-output (MIMO) systems.

For SISO systems, We denote the channel gain between base station and user $i$ as $h_i(t)$, which is i.i.d. complex Gaussian, *i.e.*, $h_i \sim \mathcal{N}_C(0,1)$. At time slot $t$, the received signal at user $i$ can be written as

$$y_i(t) = \sqrt{P_i} h_i(t) x_i(t) + \nu_i(t), \quad i = 1, \ldots, n, \tag{4.1}$$

where $P_i$ denotes the transmit power, $\nu_i(t)$ denotes additive white noise with variance $N_i$, $x_i(t)$ denotes the transmitted signal. The instantaneous signal-to-noise ratio (SNR) is then given by $\gamma_i(t) = \frac{P_i |h_i(t)|^2}{N_i}$ and, consequently, the instantaneous

transmission rate of user $i$ is

$$R_i(t) = \log_2(1 + \gamma_i(t)). \tag{4.2}$$

We consider a MIMO system with $M_T$ transmit antennas at the base station, $M_R$ receive antennas at each user. Assuming user $i$ is scheduled at time slot $t$, the received signal for user $i$ with linear precoding is given by

$$\boldsymbol{y}_i(t) = \sqrt{\frac{P_i}{M_T}}\boldsymbol{H}_i(t)\boldsymbol{W}_i(t)\boldsymbol{x}_i(t) + \boldsymbol{\nu}_i(t), \tag{4.3}$$

where the $M_R \times M_T$ complex Gaussian matrix $\boldsymbol{H}_i(t)$ is the channel gain matrix for user $i$ at time slot $t$, $\boldsymbol{W}_i(t) = [\boldsymbol{w}_{i,1}(t), \cdots, \boldsymbol{w}_{i,r}(t)]$ is the $M_T \times r$ linear precoding matrix with the transmit rank $r$ (or $r$ layers), the $r \times 1$ vector $\boldsymbol{x}_i(t)$ denotes the transmitted QAM symbol vector, and $\boldsymbol{\nu}_i(t)$ is the white noise vector. Note that $P_i$ is now the total transmit power for all transmit antennas. Assuming linear minimum mean square error (MMSE) filtering at the receiver, the SNR for the $l$th layer of user $i$ is given by

$$\gamma_{i,l}(t) = \frac{\beta_{i,l}(t)}{1 - \beta_{i,l}(t)}, \tag{4.4}$$

where

$$\beta_{i,l}(t) = \rho_i \boldsymbol{w}_i(t)^\dagger \boldsymbol{H}_i(t)^\dagger (\boldsymbol{I} + \rho_i \boldsymbol{H}_i(t)\boldsymbol{W}_i(t)\boldsymbol{W}_i(t)^\dagger \boldsymbol{H}_i(t)^\dagger)^{-1} \boldsymbol{H}_i(t)\boldsymbol{w}_i(t), \tag{4.5}$$

in which $\rho_i = \sqrt{P_i/M_T}$ is per antenna transmit power, and $^\dagger$ denotes the matrix Hermitian. The total instantaneous rate of user $i$ in the MIMO system is then given by

$$R_i(t) = \sum_{l=1}^{r} \log_2(1 + \gamma_{i,l}(t)). \tag{4.6}$$

### 4.3.2 Traffic Model

Generally, two types of simplified traffic models are considered for system evaluations: 1) full-buffer traffic model: a user has unlimited packets to transmit; 2) finite-buffer traffic model: a user is assigned with a finite traffic to transmit, which includes a packet arrival and departure process. In this report, we consider the finite-buffer traffic model.

We assume packets arrive to the network according to Poisson distribution with a fixed packet size and constant arrival rate $\lambda_i$ at user $i$. The arrived packets are buffered in a separate queue for each user until being scheduled for transmission. Denote $Q_i(t)$ as the buffer backlog waiting for transmission of user $i$ at time slot $t$. The buffer-aware scheduler observes the buffer backlogs before making a scheduling decision. As we can see, the described finite-buffer model is a queueing network with $n$ queues and one server. In order to ensure the stability of the queuing system, the total arrival rate $\lambda_{tot}$ must be less than or equal to the average service rate [10], *i.e.*,

$$\lambda_{tot} = \sum_{i=1}^{n} \lambda_i \leq \mu, \tag{4.7}$$

where $\mu$ is the average throughput of the system. The maximum total arrival rate $\lambda_{tot}$, satisfying this inequality condition, is considered as the maximum throughput of the network that can be stably supported [10].

## 4.4 Proposed Buffer Aware Adaptive Scheduling Scheme

In this Section, we focus on the design of a scheduling algorithm for the multiuser system with finite-buffer traffic model to improve the system performance by reducing the packet delay, while maintaining the queue stability condition of the networks.

### 4.4.1 Existing Scheduling Algorithms

Full-buffer model have been extensively adopted in the literature due to its simplicity. Commonly used MT and PF scheduling algorithms, which are buffer-unaware, can be implemented in this model. We summarize these two scheduling algorithms as follows. Here we assume perfect channel information at the transmitter. Accordingly, the base station computes the instantaneous rate $R_i(t)$ for each user in either single-input single-output (SISO) systems or precoded MIMO systems.

- *MT scheduler*: At time slot $t$, the base station selects the user with the largest instantaneous rate to transmit, *i.e.*, the user to be scheduled is obtained by $k = \arg\max_{i=1,\cdots,n} R_i(t)$.

- *PF scheduler*: Denote $\mu_i$ as the average transmission rate for user $i$. At time slot $t$, the base station selects the user with the largest weighted rate $\frac{R_i(t)}{\mu_i}$ to transmit, *i.e.*, the scheduled user $k = \arg\max_{i=1,\cdots,n} \frac{R_i(t)}{\mu_i}$.

Although MT and PF algorithms can be directly applied to the finite-buffer model, the delay performance might not be good, as the buffer status is not considered in the scheduling algorithms. Several buffer-aware scheduling algorithms have been proposed to reduce the packet delay. One simple solution is LCQ [45], which is described as follows.

- *LCQ scheduler*: At each time slot $t$, given the instantaneous rate $R_i(t)$ and queue backlogs $Q_i(t)$, for $i = 1, \ldots, n$, for all users, the base station schedules the user with the largest product of instantaneous rate and queue backlog, *i.e.*, $k = \arg\max_i \{R_i(t) \cdot Q_i(t)\}$.

Another buffer-aware scheduling algorithm is proposed in [18], which is termed as "$d$-algorithm".

- *"d-algorithm"*: At each time slot $t$, given the instantaneous rate $R_i(t)$ and queue backlogs $Q_i(t)$, $i = 1, \ldots, n$, for all users, with a predetermined value of $d$, the base station first finds $d$ candidate users with the largest rates and forms the set $\mathcal{U}_d(t)$. Then it schedules the user that has the largest queue backlog, *i.e.*, $k = \arg\max_{i \in \mathcal{U}_d(t)} \{Q_i(t)\}$.

It has been shown that compared with buffer-unaware schedulers, the buffer-aware scheduling algorithms are able to reduce the average packet delay.

### 4.4.2 Proposed Buffer-Aware Scheduling Algorithms

We propose a new buffer-aware scheduling algorithm, termed as buffer-aware adaptive (BAA) scheduler. Similar to the existing buffer-aware schedulers, BAA scheduler considers both channel and buffer conditions to make scheduling decisions. Channel conditions are used to select candidate users, *i.e.*, only the users whose transmission rates satisfy a certain condition are considered as candidate users. Then the buffer conditions are used to make a scheduling decision among all the candidate users, *i.e.*, the user with largest queue backlog among candidate users will be scheduled.

The proposed BAA scheduling algorithm is described as follows.

At each time slot $t$:

- Given the instantaneous rate $R_i(t)$, the base station first finds the largest rate $R^\star = \max_i R_i(t)$. Then forms the candidate user set $\mathcal{U}_\alpha(t)$, which consists of users with the instantaneous rates no less than $\alpha R^\star$, *i.e.*, $\mathcal{U}_\alpha(t) = \{i | R_i(t) \geq R^\star, i = 1, \cdots, n\}$, where $\alpha$ is a predefined value satisfying $0 < \alpha \leq 1$.

- Among the considered candidate users in $\mathcal{U}_\alpha(\sqcup)$, the scheduler selects the user $k$ which has the largest queue backlog for transmission, *i.e.*, $k = \arg\max_{i\in\mathcal{U}_\alpha} Q_i(t)$.

An example of the BAA algorithm is depicted in Fig. 4.1. We can see that the proposed algorithm is an improvement over $d$-algorithm but with a certain flexibility and adaptability. With appropriate setting of $\alpha$, the potential throughput reduction over MT algorithm is bounded. For instance, with $\alpha = 0.9$, we have a maximum of 10% throughput reduction, which in turn allows for more flexibility on choosing candidate users. Moreover, when $\alpha = 1$, BAA behaves the same as a MT scheduler, *i.e.*, the user with best channel condition is always scheduled for transmission.

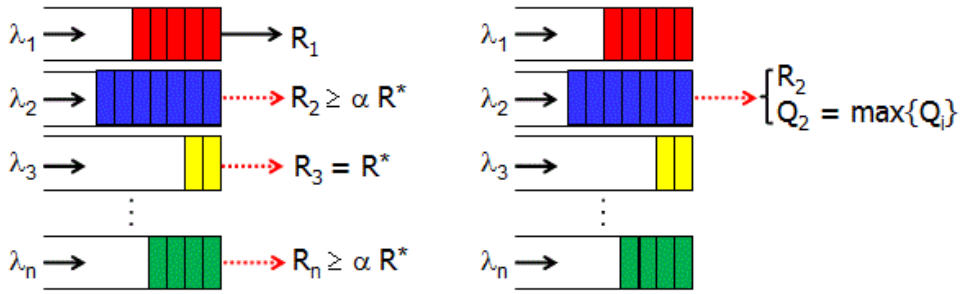


Figure 4.1: A multiuser system with $n$ users

Since in the first step of the proposed algorithm, the candidate user set $\mathcal{U}_\alpha(t)$ is formed based on the maximum rate, we call it MT-BAA algorithm. It is noted that the BAA scheduling algorithm is also compatible with the PF metric $\{\frac{R_i(t)}{\mu_i}\}$. Then the candidate user selection criteria on transmission rate $R_i(t)$ in MT-BAA algorithm is replaced with the weighted transmission rate of $\frac{R_i(t)}{\mu_i}$, *i.e.*,

$$\mathcal{U}_\alpha(t) = \left\{ i \middle| \frac{R_i(t)}{\mu_i} \geq \alpha\kappa,\ i = 1, \cdots, n \right\}, \tag{4.8}$$

where

$$\kappa = \max_i \frac{R_i(t)}{\mu_i}. \tag{4.9}$$

The proposed BAA algorithm can be applied to the wideband, *e.g.*, orthogonal frequency-division multiplexing (OFDM), systems. For the multiuser systems, we consider OFDMA, *i.e.*, each resource block or subchannel in the frequency domain is assigned to one user exclusively. We assume that there are $B$ resource blocks or subchannels. Denote $\mathcal{B}_i$ as the set of subchannels allocated to user $i$. The BAA scheduling in the OFDMA or MIMO-OFDMA can be implemented at each time slot $t$ as follows.

- Initializations: $\mathcal{B}_i = \emptyset$.
- For the subchannel $b = 1, \cdots, B$,
  - Given the rate for subchannel $b$, $R_{i,b}(t)$, $i = 1, \cdots, n,$, $b = 1, \cdots, B$, and queue backlog $Q_i(t)$, apply the BAA (MT-BAA or PF-BAA) scheduling algorithm to determine the user to be scheduled on subchannel $n$. Assuming that the scheduled user is $k$, update $\mathcal{B}_k \leftarrow \mathcal{B}_k \bigcup \{b\}$.
- For user $i = 1, \cdots, n$, transmit data for user $i$ along its allocated subchannels in $\mathcal{B}_i$.

### 4.4.3 Generalized BAA Scheduling Scheme

It is observed that the BAA scheduling algorithm consists of two steps. In the first step, a candidate user set is formed based on either user rate (MT metric) or weighted rate (PF metric) supported by its channel condition. In the second set, a user with the largest queue backlog is selected from the candidate user set. This

scheduling method can be generalized with any two reasonable metrics for these two steps, *i.e.*, forming the candidate user set with one metric and scheduling a user from the candidate user set using another metric. Denote the $\mathcal{M}_1$ and $\mathcal{M}_2$ as the metric in the first and second steps, respectively, which can be a function of instantaneous rate, average rate, queue backlog, etc. The generalized BAA scheduling scheme is then summarized as follows.

- Form the candidate user set $\mathcal{U}_\alpha$ for the users with its metric in the region $\mathcal{R}$, which is defined by the function $\mathcal{F}(\mathcal{M}_1)$ of the first matrix $\mathcal{M}_1$ and some other possible system configured parameters, *i.e.*, $\mathcal{U}_\alpha = \{i|\mathcal{M}_1(i) \in \mathcal{R}_{\mathcal{F}(\mathcal{M}_1)}\}$, where $\mathcal{M}_1(i)$ denotes the metric of the user $i$.

- Select a user from the candidate user set to transmit, which maximizes the second metric $\mathcal{M}_2$, *i.e.*, $k = \arg\max_{i\in\mathcal{U}_\alpha} \mathcal{M}_2(i)$.

The previous proposed BAA algorithms as well as the $d$-algorithm are included in this generalized description. For example, for MT-BAA algorithm, we have $\mathcal{M}_1 = \{R_i(t)\}$, $\mathcal{M}_2 = \{Q_i(t)\}$, $\mathcal{F}(\mathcal{M}_1) = \alpha \cdot \max_i R_i(t)$, and $\mathcal{R}_{\mathcal{F}(\mathcal{M}_1)} = \{x|\alpha \cdot \max_i R_i(t) \leq x \leq \max_i R_i(t)\}$. For PF-BAA algorithm, we have $\mathcal{M}_1 = \{R_i(t)/\mu_i\}$, $\mathcal{M}_2 = \{Q_i(t)\}$, $\mathcal{F}(\mathcal{M}_1) = \alpha \cdot \max_i R_i(t)/\mu_i$, and $\mathcal{R}_{\mathcal{F}(\mathcal{M}_1)} = \{x|\alpha \cdot \max_i R_i(t)/\mu_i \leq x \leq \max_i R_i(t)/\mu_i\}$. For $d$-algorithm, $\mathcal{M}_1$ and $\mathcal{M}_2$ are the same as MT-BAA algorithm, but $\mathcal{F}(\mathcal{M}_1) = R'_d(t)$ where $R'_i(t)$ is the descend ordered $R_i(t)$, and $\mathcal{R} = [R'_d(t)R'_1(t)]$.

Any reasonable metrics can be implemented in this generalized scheme to form one specific scheduling algorithm. Some examples are given as follows.

- The first metric $\mathcal{M}_1$ can be $R_i(t)$ as in MT-BAA or $R_i(t)/\mu_i$ as in PF-BAA algorithm. The second metric $\mathcal{M}_2$ can be the production of rate and queue backlog, *i.e.*, $R_i(t)Q_i(t)$. Thus, in the second step, instead of

selecting a user from the candidate user set with the largest queue backlog size, the scheduler selects the one with the largest connection queue value.

- The first metric $\mathcal{M}_1$ can be $R_i(t)Q_i(t)$. Thus in the first step, the candidate user set is formed with the users having $R_i(t)Q_i(t) \geq \alpha\chi$, $\chi = \max_i(R_i(t)Q_i(t))$.

- Similarly, the first metric $\mathcal{M}_1$ can be $R_i(t)$ as in MT-BAA or $R_i(t)/\mu_i$ as in PF-BAA algorithm. The second metric can be the number of packets in the user buffer.

- The second metric can be the delay of the earliest arrived packet to be scheduled in the user buffer, *i.e.*, the scheduler selects the user from the candidate user set (formed in the first step) with the largest delay of the earliest arrived packet in its buffer.

### 4.5 Performance Analysis

In this Section, we analyze the system performance of the proposed algorithm using the SISO model for simplicity. However, it is straightforward to extend the following approach to MIMO systems, as long as the channel statistics is available. We first derive the probabilities of selecting different number of candidate users based on the values of $\alpha$. Then we obtain the average throughput based on the probability of candidate users. Since the statistics of queue backlogs in different users are not known, we assume the user with the lowest transmission rate among candidate users has the largest queue backlog, to obtain average throughput lower bound. We also derive a throughput approximation, by assuming the scheduled user with the largest queue backlog is uniformly distributed among candidate users. Correspondingly, the stability conditions are given according to average throughput lower bound and

approximation. Note that the stability condition based on throughput approximation may not necessarily lead to a stable system. However, it provides a guidance on approximately how large the total arrival rate can be.

### 4.5.1 Candidate Users Selection

We assume the average $\mathsf{SNR}_i = \frac{P_i}{N_i} = \frac{1}{A}, \forall\ i$. Thus, the instantaneous $\mathsf{SNR}$s $\gamma_1, \ldots, \gamma_n$ are i.i.d. exponential distributed random variables with cumulative distribution function (CDF) as $F(x) = 1 - e^{-Ax}$ and probability density function (pdf) as $f(x) = Ae^{-Ax}$. Let $X_1, \ldots, X_n$ represent the values of $\gamma_i$ in an ascending order.

From the order statistics, the joint distributions for any users $1 \leq p < q \leq n$, $X_p$ and $X_q$ is,

$$\begin{aligned} f_{X_p,X_q}(u,v) =& \frac{n!}{(p-1)!(q-p-1)!(n-q)!} f(u)f(v)[F(u)]^{p-1} \\ & [F(v) - F(u)]^{q-p-1}[1 - F(v)]^{n-q}, \end{aligned} \tag{4.10}$$

in which $-\infty < u < v < \infty$.

Suppose there are $d$ users whose throughput fall in $\alpha * R^\star$ region. Based on the proposed scheduling algorithm, the probability of $d \leq n - j$ is given as

$$\begin{aligned} & Pr[d \leq n-j] = Pr[X_j < \alpha X_n] \\ & = \int_0^\infty \int_0^{(1+v)^\alpha - 1} f_{X_j,X_n}(u,v)dudv \\ & = \frac{n!}{(j-1)!(n-j-1)!} \int_0^\infty Ae^{-Av} \underbrace{\int_0^{(1+v)^\alpha-1} Ae^{-Au}[1-e^{-Au}]^{j-1}[e^{-Au} - e^{-Av}]^{n-j-1}du}_{\triangle} dv, \end{aligned} \tag{4.11}$$

in which $1 \leq j < n-1$.

Since values of $d$ are in the range of $[1, n]$, when $j = n$, $Pr[d \leq 0] = 0$; when $j = 0$, $Pr[d \leq n] = 1$. For $j = n-1$, the index $n - j - 1 = 0$, thus, we have to calculate the probability term $Pr[d \leq 1]$ separately as

$$\begin{aligned} Pr[d \leq 1] &= Pr[X_{n-1} < \alpha X_n] \\ &= \int_0^{\infty}\int_0^{(1+v)^{\alpha}-1} f_{X_{n-1},X_n}(u,v)dudv \\ &= \int_0^{\infty}\int_0^{(1+v)^{\alpha}-1} \frac{n!}{(n-2)!} f(u)f(v)[F(u)]^{n-2}dudv \\ &= \int_0^{\infty}\int_0^{(1+v)^{\alpha}-1} n(n-1)Ae^{-Au}Ae^{-Av}[1-e^{-Au}]^{n-2}dudv \\ &= n\int_0^{\infty} Ae^{-Av}(1-e^{-A(1+v)^{\alpha}+A})^{n-1}dv. \end{aligned} \tag{4.12}$$

For $1 \leq j < n-1$, we derive the inside integral ($\triangle$) in equation (4.11) as follows,

$$\begin{aligned} \triangle &= \int_0^{(1+v)^{\alpha}-1} Ae^{-Au}[1-e^{-Au}]^{j-1}[e^{-Au}-e^{-Av}]^{n-j-1}du \\ &= -\frac{1}{j}\sum_{i=0}^{j-1}(\prod_{k=0}^{i}\frac{j-k}{n-j+k})(1-e^{-Au})^{j-i-1}(e^{-Au}-e^{-Av})^{n-j+i}|_0^{(1+v)^{\alpha}-1} \\ &= -\frac{1}{j}\sum_{i=0}^{j-1}(\prod_{k=0}^{i}\frac{j-k}{n-j+k})[(1-e^{-A(1+v)^{\alpha}+A})^{j-i-1}(e^{-A(1+v)^{\alpha}+A}-e^{-Av})^{n-j+i} \\ &\quad -(1-e^{-Av})^{n-1}]. \end{aligned} \tag{4.13}$$

By inserting ($\triangle$) from equation (4.13) into equation (4.11), we have

$$\begin{aligned}
&Pr[d \le n-j] \\
&= \frac{n!}{(j-1)!(n-j-1)!} \int_0^\infty A e^{-Av} \cdot (\triangle) dv \\
&= -\frac{n!}{(j-1)!(n-j-1)!} \sum_{i=0}^{j-1} (\prod_{k=0}^{i} \frac{j-k}{n-j+k}) G(i,j) \\
&\quad + \underbrace{\frac{n!}{(j-1)!(n-j-1)!} (\prod_{k=0}^{i} \frac{j-k}{n-j+k}) \int_0^\infty A e^{-Av} (1-e^{-Av})^{n-1} dv}_{\triangle'}
\end{aligned} \tag{4.14}$$

in which $G(i,j)$ is defined as

$$G(i,j) \triangleq \int_0^\infty A e^{-Av} [(1-e^{-A(1+v)^\alpha + A})^{j-i-1} (e^{-A(1+v)^\alpha + A} - e^{-Av})^{n-j+i} dv, \tag{4.15}$$

and ($\triangle'$) can be calculated as follows,

$$\begin{aligned}
\triangle' &= \frac{n!}{j!(n-j-1)!} (\prod_{k=0}^{i} \frac{j-k}{n-j+k}) \int_0^\infty A e^{-Av} (1-e^{-Av})^{n-1} dv \\
&= \frac{n!}{j!(n-j-1)!} \cdot \frac{j!}{(n-1)\cdots(n-j)} \int_0^\infty (1-e^{-Av})^{n-1} d - e^{-Av} \\
&= n \cdot \frac{1-e^{-Av}}{n} |_0^\infty = 1.
\end{aligned} \tag{4.16}$$

Insert ($\triangle'$) from equation (4.16) into equation (4.14), for $1 \le j < n-1$, we have

$$Pr[d \le n-j] = 1 - \frac{n!}{j!(n-j-1)!} \sum_{i=0}^{j-1} (\prod_{k=0}^{i} \frac{j-k}{n-j+k}) \cdot G(i,j). \tag{4.17}$$

Similarly, for $1 \leq j < n-1$, the probability of $d \leq n-j-1$ can be obtained as,

$$\begin{aligned} Pr[d \leq n-j-1] &= Pr[X_{j+1} < \alpha X_n] \\ &= 1 - \frac{n!}{(j+1)!(n-j-2)!} \sum_{i=0}^{j} (\prod_{k=0}^{i} \frac{j-k+1}{n-j+k-1}) \\ &\quad \cdot G(i, j+1). \end{aligned} \tag{4.18}$$

From the definition of $G(i,j)$ given in 4.15, we have $G(i,j+1) = G(i-1,j)$, we then rewrite $Pr[d \leq n-j-1]$ as

$$\begin{aligned} Pr[d \leq n-j-1] &= 1 - \frac{j+1}{n-j-1} \cdot \frac{n!}{(j+1)!(n-j-2)!} \cdot G(0, j+1) \\ &\quad - \frac{n!}{(j+1)!(n-j-2)!} \sum_{i=1}^{j} (\prod_{k=0}^{i} \frac{j-k+1}{n-j+k-1}) \cdot G(i, j+1) \\ &= 1 - \frac{n!}{j!(n-j-1)!} \cdot G(0, j+1) - \frac{n!}{j!(n-j-1)!} \sum_{i'=0}^{j-1} (\prod_{k=0}^{i'} \frac{j-k}{n-j+k}) \cdot G(i', j). \end{aligned} \tag{4.19}$$

We then obtain the probability of $d = n-j$, for $1 \leq j < n-1$, as

$$\begin{aligned} Pr[d = n-j] &= Pr[d \leq n-j] - Pr[d \leq n-j-1] \\ &= \frac{n!}{j!(n-j-1)!} \cdot G(0, j+1) + \frac{n!}{j!(n-j-1)!} \sum_{i'=0}^{j-1} (\prod_{k=0}^{i'} \frac{j-k}{n-j+k}) \\ &\quad \cdot G(i', j) - \frac{n!}{j!(n-j-1)!} \sum_{i=0}^{j-1} (\prod_{k=0}^{i} \frac{j-k}{n-j+k}) \cdot G(i, j) \\ &= \frac{n!}{j!(n-j-1)!} \cdot G(0, j+1) \\ &= \frac{n!}{(n-d)!(d-1)!} \cdot G(0, n-d+1). \end{aligned} \tag{4.20}$$

Note that equation (4.20) is for $1 \leq j < n-1$. However, we show that the expression of equation (4.20) can represent the probability terms for $j = 0$ and $j = n-1$, *i.e.*, $Pr[d = n]$ and $Pr[d = 1]$. Using similar procedure for calculating $Pr[d = n - j]$ and previous results, we can obtain $Pr[d = n]$ as

$$\begin{aligned} Pr[d = n] &= Pr[d \leq n] - Pr[d \leq n-1] \\ &= 1 - (1 - \frac{n!}{(n-2)!} \cdot \frac{1}{n-1} \cdot G(0,1)) \\ &= n \cdot G(0,1), \end{aligned} \tag{4.21}$$

Similarly, for $d = 1$,

$$\begin{aligned} Pr[d = 1] &= Pr[d \leq 1] - Pr[d \leq 0] \\ &= n \int_0^\infty A e^{-Av} (1 - e^{-A(1+v)^\alpha + A})^{n-1} dv - 0 \\ &= n \cdot G(0,n). \end{aligned} \tag{4.22}$$

### 4.5.2 Average Throughput

#### 4.5.2.1 Lower Bound

We now proceed to calculate the average throughput lower bound $\mu_{LB}$ of the proposed scheduling algorithm, by assuming the user with the lowest transmission rate always has the largest queue backlog, we have

$$\begin{aligned} \mu_{LB} &\triangleq \sum_{d=1}^{n} Pr[d] \cdot E\{\log_2(1 + X_{n-d+1})\} \\ &= \sum_{d=1}^{n} H(n-d+1) \leq \mu, \end{aligned} \tag{4.23}$$

where the expectation is taken with the condition of $d$ candidate nodes being selected and $H(\cdot)$ is derived as follows.

As described previously, the condition of $d$ candidate nodes being selected can be represented as the joint event of $\log_2(1+X_{n-d+1}) \geq \alpha \log_2(1+X_n)$ and $\log_2(1+X_{n-d}) < \alpha \log_2(1+X_n)$, denoted as $\mathcal{C}_{n-d+1,n}$ and $\mathcal{C}_{n-d,n}$, respectively. We then have

$$\begin{aligned} E\{\log_2(1+X_{n-d+1})\} &\triangleq E_{X_{n-d+1}|X_{n-d},X_n}\{\log_2(1+X_{n-d+1})|\mathcal{C}_{n-d+1,n},\mathcal{C}_{n-d,n}\} \\ &= \frac{E_{X_{n-d+1},X_{n-d},X_n|\mathcal{C}_{n-d+1,n},\mathcal{C}_{n-d,n}}\{\log_2(1+X_{n-d+1})\}}{Pr(\mathcal{C}_{n-d+1,n},\mathcal{C}_{n-d,n})} \end{aligned} \quad (4.24)$$

where the numerator is the expectation over the joint pdf of $X_{n-d+1}, X_{n-d}$, and $X_n$ in the region bounded by $\mathcal{C}_{n-d+1,n}$ and $\mathcal{C}_{n-d,n}$, the denominator is the probability of joint event of $\mathcal{C}_{n-d+1,n}, \mathcal{C}_{n-d,n}$, which is $Pr(d)$. Therefore, we have

$$\mu_{LB} = \sum_{d=1}^{n} E_{X_{n-d+1},X_{n-d},X_n|\mathcal{C}_{n-d+1,n},\mathcal{C}_{n-d,n}}\{\log_2(1+X_{n-d+1})\}. \quad (4.25)$$

Based on the order statistics, the joint pdf $f_{X_{n-d+1},X_{n-d},X_n}(x_{n-d+1}, x_{n-d}, x_n)$ is given by

$$\begin{aligned} f_{X_{n-d+1},X_{n-d},X_n}(x_{n-d+1}, x_{n-d}, x_n) &= \frac{n!f(x_{n-d})f(x_{n-d+1})f(x_n)}{(n-d-1)!(d-2)!}F(x_{n-d})(F(x_n) \\ &\quad -F(x_{n-d+1}))^{d-2}. \end{aligned} \quad (4.26)$$

We then obtain

$$H(n-d+1) \triangleq E_{X_{n-d+1},X_{n-d},X_n|\mathcal{C}_{n-d+1,n},\mathcal{C}_{n-d,n}}\{\log_2(1+X_{n-d+1})\},$$

for $d = 2, \cdots, n-1$ as

$$\begin{aligned} H(n-d+1) &= \frac{n!}{(n-d-1)!(d-1)!} \int_0^\infty \log_2(1+u) A e^{-Au} \int_0^u A e^{-Av} \\ &\quad (1-e^{-Av})^{n-d-1} \Big( (e^{-Au} - e^{-A(1+u)^{\frac{1}{\alpha}}+A})^{d-1} - \\ &\quad (e^{-Au} - e^{-A\max\{(1+v)^{\frac{1}{\alpha}}-1,u\}})^{d-1} \Big) dv du \end{aligned} \tag{4.27}$$

where we let $u = x_{n-d+1}$ and $w = x_n$ for notation simplicity. For $d = 1$ and $d = n$, we obtain $H(n-d+1)$, respectively, given by

$$H(n) = \int_0^\infty n \log_2(1+u) A e^{-Au} (1 - e^{-A(1+u)^\alpha + A})^{n-1} du, \quad d = 1 \tag{4.28}$$

$$H(1) = \int_0^\infty n \log_2(1+u) A e^{-Au} (1 - e^{-A(1+u)^{\frac{1}{\alpha}} + A})^{n-1} du, \quad d = n. \tag{4.29}$$

#### 4.5.2.2 Approximation

We can approximate the throughput of our proposed scheduling algorithm using random selection among the candidate users, instead of selecting the one with largest queue backlog. Assuming i.i.d. traffic arrival model in all $n$ users, by random selection, the probability of being scheduled from the $d$ candidate users is $\frac{1}{d}$. By averaging through all possible values of $d$, we can approximate the mean throughput of our proposed scheduling algorithm as,

$$\mu \approx \sum_{d=1}^{n} Pr[d] \sum_{j=1}^{d} \frac{1}{d} \underbrace{E\{\log_2(1+X_{n-j+1})\}}_{H'(n-j+1)} . \tag{4.30}$$

Similar to the procedure for deriving the lower bound, the conditional expectation can be transformed to the integration over joint pdf of four random variables, *i.e.*, $X_{n-d}$, $X_{n-d+1}$, $X_{n-j+1}$, and $X_n$, which is too complex to be integrated. We approximate it

with the average throughput of $X_{n-j+1}$ under the only condition of $\log_2(1+X_{n-j+1}) \geq \alpha \log_2(1+X_n)$.

For $1 \leq m < n-1$,

$$
\begin{aligned}
H'(m) \approx & \frac{\int_0^\infty \int_u^{(1+u)^{\frac{1}{\alpha}}-1} \log_2(1+u) f_{X_m,X_n}(u,v) dvdu}{Pr[\log_2(1+X_m) \geq \alpha \log_2(1+X_n)]} \\
= & \frac{\int_0^\infty \log_2(1+u) \int_u^{(1+u)^{\frac{1}{\alpha}}-1} \frac{n!}{(m-1)!(n-m-1)!} f(u) f(v) [F(u)]^{m-1} [F(v)-F(u)]^{n-m-1} dvdu}{\int_0^\infty \int_u^{(1+u)^{\frac{1}{\alpha}}-1} \frac{n!}{(m-1)!(n-m-1)!} f(u) f(v) [F(u)]^{m-1} [F(v)-F(u)]^{n-m-1} dvdu} \\
= & \frac{\int_0^\infty \log_2(1+u) A e^{-Au} (1-e^{-Au})^{m-1} \int_u^{(1+u)^{\frac{1}{\alpha}}-1} A e^{-Av} (e^{-Au}-e^{-Av})^{n-m-1} dvdu}{\int_0^\infty A e^{-Au} (1-e^{-Au})^{m-1} \int_u^{(1+u)^{\frac{1}{\alpha}}-1} A e^{-Av} (e^{-Au}-e^{-Av})^{n-m-1} dvdu} \\
= & \frac{\int_0^\infty \log_2(1+u) A e^{-Au} (1-e^{-Au})^{m-1} (e^{-Au}-e^{-A(1+u)^{\frac{1}{\alpha}}+A})^{n-m} du}{\int_0^\infty A e^{-Au} (1-e^{-Au})^{m-1} (e^{-Au}-e^{-A(1+u)^{\frac{1}{\alpha}}+A})^{n-m} du}.
\end{aligned} \tag{4.31}
$$

When $m = n-1$, we have

$$
\begin{aligned}
H'(n-1) \approx & \frac{\int_0^\infty \int_u^{(1+u)^{\frac{1}{\alpha}}-1} \log_2(1+u) f_{X_{n-1},X_n}(u,v) dvdu}{Pr[X_{n-1} \geq \alpha X_n]} \\
= & \frac{\int_0^\infty \log_2(1+u) \int_u^{(1+u)^{\frac{1}{\alpha}}-1} n(n-1) f(u) f(v) [F(u)]^{n-2} dvdu}{\int_0^\infty \int_u^{(1+u)^{\frac{1}{\alpha}}-1} n(n-1) f(u) f(v) [F(u)]^{n-2} dvdu} \\
= & \frac{\int_0^\infty \log_2(1+u) A e^{-Au} (1-e^{-Au})^{n-2} \int_u^{(1+u)^{\frac{1}{\alpha}}-1} A e^{-Av} dvdu}{\int_0^\infty A e^{-Au} (1-e^{-Au})^{n-2} \int_u^{(1+u)^{\frac{1}{\alpha}}-1} A e^{-Av} dvdu} \\
= & \frac{\int_0^\infty \log_2(1+u) e^{-Au} (1-e^{-Au})^{n-2} (e^{-Au}-e^{-A(1+u)^{\frac{1}{\alpha}}+A}) du}{\int_0^\infty e^{-Au} (1-e^{-Au})^{n-2} (e^{-Au}-e^{-A(1+u)^{\frac{1}{\alpha}}+A}) du}.
\end{aligned} \tag{4.32}
$$

Since the probability term $Pr[X_n \geq \alpha X_n] = 1$, for $m = n$,

$$\begin{aligned} H'(n) \approx & \frac{\int_0^\infty \log_2(1+u) n f_{X_n}(u) du}{Pr[X_n \geq \alpha X_n]} \\ = & \int_0^\infty \log_2(1+u) n [F(u)]^{n-1} f(u) du \\ = & \int_0^\infty n \log_2(1+u) A e^{-Au} (1-e^{-Au})^{n-1} du. \end{aligned} \tag{4.33}$$

Combing equations (4.31), (4.32) and (4.33), we have

$$H'(n-j+1) \approx \begin{cases} \frac{\int_0^\infty \log_2(1+u) A e^{-Au} (1-e^{-Au})^{n-j} (e^{-Au} - e^{-A(1+u)^{\frac{1}{\alpha}} + A})^{j-1} du}{\int_0^\infty A e^{-Au} (1-e^{-Au})^{n-j} (e^{-Au} - e^{-A(1+u)^{\frac{1}{\alpha}} + A})^{j-1} du}, & \text{if } 2 \leq j \leq n, \\ \int_0^\infty n \log_2(1+u) A e^{-Au} (1-e^{-Au})^{n-1} du, & \text{if } j = 1. \end{cases}$$

Then by inserting $(H')$ into equation (4.30), we have the average throughput approximation as

$$\mu \approx \sum_{d=1}^{n} Pr[d] \sum_{j=1}^{d} \frac{1}{d} \cdot H'(n-j+1). \tag{4.34}$$

### 4.5.3 Stability Condition

A scheduling algorithm is stable when the total arrival rate $\lambda_{tot}$ is less than or equal to the service rate $\mu$. Based on equation (4.23), we have

$$\lambda_{tot} \leq \mu_{LB} = \sum_{d=1}^{n} H'(n-d+1), \tag{4.35}$$

as the stably supported rate. Note that when users have i.i.d. arrival rate $\lambda_i$, the stability condition for each user is,

$$\lambda_i \leq \frac{1}{n} \sum_{d=1}^{n} H'(n-d+1). \tag{4.36}$$

Furthermore, based on equation (4.30), we can provide a tight stability condition of the network as

$$\lambda_{tot} \leq \sum_{d=1}^{n} Pr[d] \sum_{j=1}^{d} \frac{1}{d} \cdot H(n-j+1). \tag{4.37}$$

Note that this stability condition may not necessarily lead to a stable system, however, it provides a guidance on approximately how large the total arrival rate can be.

Note that when user's channels are i.i.d., the stability condition is more strict for BAA than that for MT scheduler. However, for independent non-identically distributed (i.n.i.d) channels, the BAA algorithm can schedule users that have lower SNR with higher probability than MT scheduler. Thus, the stability condition is easier to be satisfied in BAA than MT algorithm.

## 4.6 Numerical Results and Discussions

In this section, we present the numerical results of the proposed BAA algorithm. We first verify our analysis findings in a simple SISO system. Then we present the simulation results from a system-level simulator and compare with the existing scheduling algorithms. In SISO model, throughout the simulations, we set $n = 10$, BW $= 1$, SNR as 0 dB, packets arrive at system according to Poisson distribution with fixed packet size of 1 bit. We also evaluate the proposed BAA algorithm via the system-level simulation for a single-user (SU)-MIMO-OFDM system which comply the 3GPP LTE standard. The parameters and settings are summarized in Table 4.1.

### 4.6.1 SISO Simulation

We first examine MT-BAA in the SISO system. Fig. 4.2 illustrates the probabilities of candidate user numbers $d$, from both analytical calculations with equa-

Table 4.1: Simulation Parameters for Homogenous Networks

| **Parameter** | **Assumption** |
|---|---|
| Deployment scenario | IMT Urban Micro (UMi) |
| Duplex method, bandwidth | FDD: 10MHz for downlink |
| Cell layout | Hex grid 19 sites, 3 cells/site |
| Transmission power at BS | 46 dBm |
| Number of users per sector | 10 |
| Network synchronization | Synchronized |
| Antenna configuration (BS) | 4 Tx cross-polarized ant., 0.5-$\lambda$ spacing |
| Antenna configuration (user) | 2 Rx cross-polarized ant., 0.5-$\lambda$ spacing |
| Downlink transmission | Dynamic SU-MIMO scheduling |
| Codebook | Rel. 8 codebook [48] |
| Downlink scheduler | MT or PF in time and frequency |
| Scheduling granularity: | 5 RBs |
| Feedback assumptions | 5ms periodicity & 4ms delay;<br>Sub-band CQI and PMI<br>feedback without errors. |
| Sub-band granularity: | 5 RBs |
| Downlink HARQ scheme | Chase Combining |
| Downlink receiver type | LMMSE |
| Channel estimation error | NA |
| Feedback channel error | NA |
| Control channel & reference<br>signal overhead | 3 OFDM symbols for control<br>Use TBS table in TS36.213 [49] |
| Packet arrival rate<br>Packet size | 1.2 per ms<br>1500 bytes |

tion (4.20) and Monte-Carlo simulations, for various values of $\alpha$. It is seen that the probability of larger $d$ increases as $\alpha$ decreases. Fig. 4.3 plots average packet delay for various per user packet arrival rates $\lambda$. It is observed that the packet delay decreases first with $\alpha$, however, after a certain value, keep increasing $\alpha$ will result in worse delay. This indicates that an optimal value of $\alpha$ exists.

We examine in Fig. 4.4 the average throughput lower bound $\mu_{LB}$ and approximation $\mu_{approx}$ of the proposed algorithm, with average throughput $\mu_{MT}$ of the MT scheduler and trivial lower bound $\alpha\mu_{MT}$. The curves of $\mu_{LB}$ and $\mu_{approx}$ are obtained from computations with equations (4.23) and (4.30), respectively. It is observed that

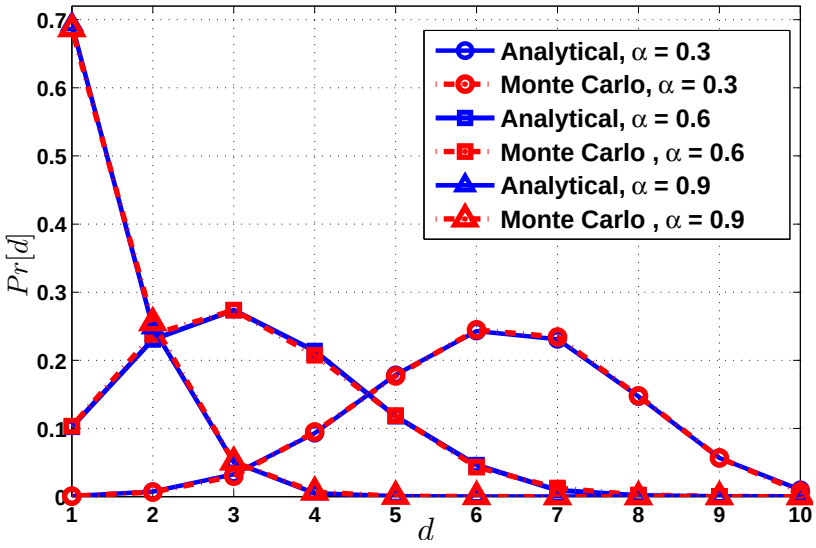


Figure 4.2: Probability of $d$

$\mu_{LB}$ and $\mu_{approx}$ increase with $\alpha$ and are greater than $\alpha\mu_{MT}$ for all presented $\alpha$. Furthermore, when $\alpha$ is large, both $\mu_{LB}$ and $\mu_{approx}$ are close to maximum throughput $\mu_{MT}$. Particulary, for $\alpha = 0.9$, the throughput lower bound $\mu_{LB} = 1.8693$ is very close to $\mu_{MT} = 1.9062$, indicating a throughput loss of 1.9%, which is much less than 10% throughput loss bound.

### 4.6.2 LTE System-Level Simulation

We now evaluate the performance of BAA scheduler, as well as existing schedulers via LTE system level simulation using FTP traffic model with various parameter settings. Specifically we consider $M_T = 4$ and $M_R = 2$ MIMO system with 50 RBs in frequency domain. Each RB is assigned to one user exclusively. In FTP traffic model, we assume packets arrive according to Poisson distribution with fixed packet size of 1500 bytes and various per cell arrival rates $\lambda$. It is noted that due to the hardware limitations and long simulation running time, we adopt a fixed packet size of 1500 bytes, instead of 500,000 bytes as specified in 3GPP TR 36.814 [46]. Correspondingly, we increase the per cell packet arrival rate $\lambda$, to increase the traffic in the

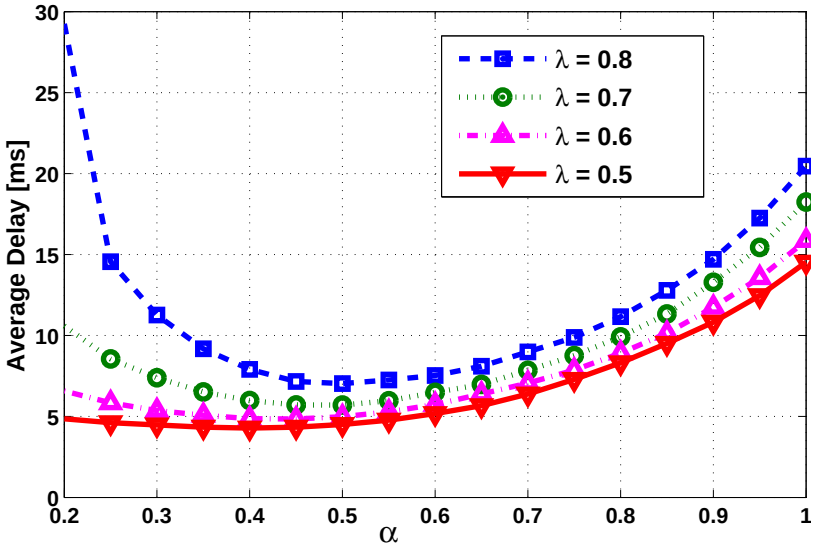


Figure 4.3: Delay as a function of $\alpha$ in a SISO system

system. We first evaluate the system performance with fixed per cell user number as $n = 10$ and various per cell packet arrival rates, *i.e.*, $\lambda = 0.4\ ms$, $0.8\ ms$, $1.2\ ms$ and $1.6\ ms$, corresponding to 25%, 50%, 75%, 90% average system traffic intensities, respectively. In this work, we consider traffic intensity as the ratio of the average generated traffic per cell to the maximum cell throughput in each time slot, *e.g.*, when $\lambda = 0.8$, the average generated traffic per cell is 9.6 ($= 0.8 \cdot 1500 \cdot 8$) Mbps, and the measured maximum cell throughput (assuming full-buffer) is 19.2 Mbps, which leads to a 50% traffic intensity per cell. Then we fix the the traffic intensity in each cell, *i.e.*, $\lambda = 0.8\ ms$, but with different per cell user numbers, *i.e.*, $n = 7$, $n = 10$, $n = 15$, $n = 20$. The following schedulers are evaluated and compared, in terms of average throughput and average packet delay: MT-Based: MT, d-algorithm, LCQ, BAA; PF-Based: PF, d-algorithm, LCQ, BAA.

In addition to the performance metrics of cell throughput and packet delay, we also evaluate the system performance in terms of user packet delay fairness index and cell worst case delay, which are defined as follows.

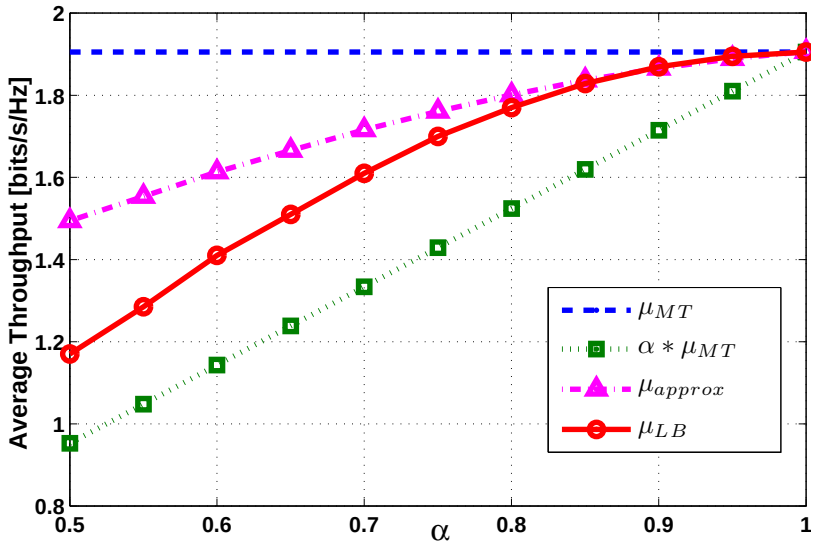


Figure 4.4: Average throughput comparisons in a SISO system

The delay fairness index is defined as the average of the delay fairness indices of all the users, which determines how fair a scheduler algorithm is with respect to the delay in each user. The delay fairness index is calculated according to Jain's fairness index as

$$\mathcal{J} = \frac{(\sum_{i=1}^{N} w_i)^2}{N \sum_{i=1}^{N} w_i^2}. \tag{4.38}$$

$\mathcal{J}$ rates the fairness of a set of values in which $N$ is the total number of users in the network, *e.g.*, when per cell user number $n = 10$, the total number of users $N = 570$ as the network contains 57 cells. $w_i$ is the average packet delay of user $i$. The result ranges from $\frac{1}{N}$ (worst case) to 1 (best case).

The worst case delay is defined as the time it takes for all the users in each cell to receive at least one packet from the base station [18]. This delay metric is of practical interest, since it is related to network efficiency.

Table 4.2: Performance Comparisons With $n = 10$, $\lambda = 0.4$

| Algorithm | | Throughput | Delay | Fairness |
|---|---|---|---|---|
| MT - | MT | 0.4679 | 2.04 | 21.5 |
| | $d$-algorithm, d=3 | 0.4679 | 2.09 | 26.4 |
| | LCQ | 0.4679 | 2.17 | 26.1 |
| | BAA, $\alpha = 90\%$ | 0.4679 | 2.02 | 21.6 |
| PF - | PF | 0.4679 | 2.02 | 22.9 |
| | $d$-algorithm, d=3 | 0.4679 | 2.06 | 27.2 |
| | LCQ | 0.4679 | 2.14 | 30.1 |
| | BAA, $\alpha = 90\%$ | 0.4679 | 2.02 | 23.7 |

#### 4.6.2.1 Fixed n=10, 25% Traffic Intensity

The system performance is summarized in Table 4.2 for both MT and PF based schedulers. With MT based schedulers, when the traffic intensity is low, the system performance is similar for all the considered schedulers, *i.e.*, cell throughput and worst case delay, since most of the time the system buffer is empty. However, with $\alpha = 90\%$, the MT-BAA still provides a slight improvement w.r.t. packet delay. MT-LCQ and MT- $d$-algorithm achieve higher delay fairness index but do not improve the average packet delay. Similar performance is observed with PF based schedulers. Generally, the delay fairness index is higher in PF based scheduler than the corresponding MT based scheduler, due to the weighted rate considered in PF based schedulers.

Table 4.3: Performance Comparisons With $n = 10$, $\lambda = 0.8$

| Algorithm | | Throughput | Delay | Fairness |
|---|---|---|---|---|
| MT - | MT | 0.9146 | 6.64 | 7.6 |
| | $d$-algorithm, d=3 | 0.9177 | 5.74 | 13 |
| | LCQ | 0.9183 | 5.77 | 17.5 |
| | BAA, $\alpha = 30\%$ | 0.9178 | 5.46 | 11.5 |
| PF - | PF | 0.9166 | 5.47 | 9.2 |
| | $d$-algorithm, d=2 | 0.9174 | 5.38 | 10.5 |
| | LCQ | 0.9188 | 5.43 | 18.3 |
| | BAA, $\alpha = 65\%$ | 0.9177 | 4.86 | 10.9 |

#### 4.6.2.2 Fixed n=10, 50% Traffic Intensity

The system performance is summarized in Table 4.3 for both MT and PF based schedulers. With MT based schedulers, when the traffic intensity is moderate, all the three MT based schedulers provide better system performance than MT scheduler, in terms of cell throughput, packet delay, delay fairness and worst case delay. With $\alpha = 70\%$, the MT-BAA provides the best system performance in terms of packet delay. MT-LCQ achieves the best delay fairness index and worst case delay. Note that due to more balanced channel resource allocation and finite traffic buffer, although the three schedulers may not schedule the user with best channel quality, they provide better cell throughput than MT. Similar performance is observed with PF based schedulers. Note that the overall fairness index is higher in PF based scheduler than the corresponding MT based scheduler when under the same condition. When condition is not the same, MT based scheduler may have higher delay fairness index than PF based scheduler, *e.g.*, when $d = 3$,MT-$d$-algorithm achieves higher fairness index than PF-$d$-algorithm with $d = 2$. The MT-BAA achieves an average of 18%, 5%, 5% on delay improvement compared to MT, MT-$d$-algorithm and MT-LCQ, respectively. The PF-BAA achieves an average of 11%, 10%, 11% on delay improvement compared to PF, PF-$d$-algorithm and PF-LCQ, respectively.

Table 4.4: Performance Comparisons With $n = 10$, $\lambda = 1.2$

| **Algorithm** | | **Throughput** | **Delay** | **Fairness** |
|---|---|---|---|---|
| MT - | MT | 1.214 | 29.1 | 10 |
| | $d$-algorithm, d=5 | 1.293 | 20.9 | 14 |
| | LCQ | 1.321 | 19.2 | 23 |
| | BAA,$\alpha$=35% | 1.314 | 18.1 | 16 |
| PF - | PF | 1.295 | 24.8 | 15 |
| | $d$-algorithm, d=4 | 1.312 | 19.3 | 16 |
| | LCQ | 1.322 | 18.1 | 26 |
| | BAA,$\alpha$=45% | 1.321 | 16.7 | 16 |

#### 4.6.2.3 Fixed n=10, 75% Traffic Intensity

When the traffic intensity is increased to 75%, the system performance is summarized in Table 4.4 for both MT and PF based schedulers. It is noted that n/a represents that in this simulation setting, the worst case delay cannot be obtained, since within the limited simulation duration, not all the users in a cell can receive at least one packet. We can see that the three MT based schedulers provide better system performance than MT scheduler, in terms of cell throughput, packet delay, delay fairness and worst case delay. With $\alpha = 45\%$, the MT-BAA provides the best system performance in terms of packet delay. MT-LCQ achieves the best cell throughput, delay fairness index and worst case delay. Similar performance is observed with PF based schedulers. The MT-BAA scheduler achieves an average of 38%, 14%, 6% on delay improvement compared to MT, MT-*d*-algorithm and MT-LCQ, respectively. The PF-BAA achieves 33%, 13% and 8% on delay reduction than PF, PF-*d*-algorithm and PF-LCQ, respectively, which are significant in term of system performance.

Under 75% traffic intensity, Fig. 4.5 illustrates the delay comparisons between PF-BAA and existing scheduling algorithms. It is observed that with $0.35 < \alpha < 0.65$, PF-BAA outperforms all the other algorithms and $\alpha = 0.45$ gives the smallest packet delay. Fig. 4.6 plots the cdf comparisons of packet delay for different PF based scheduling algorithms.

#### 4.6.2.4 Fixed n=10, 100% Traffic Intensity

When the traffic intensity is increased to 100%, the system performance is summarized in Table 4.5 for both MT and PF based schedulers. With the MT based schedulers, MT-LCQ provides the best fairness index and worst case delay; MT-

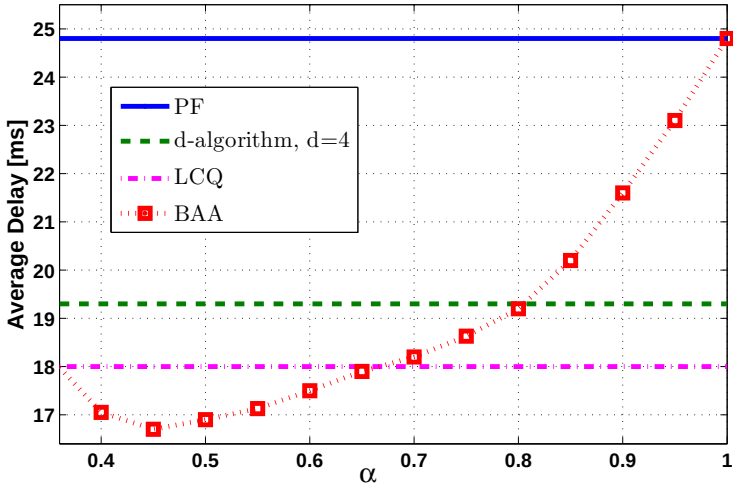


Figure 4.5: Delay as a function of $\alpha$ in a LTE system simulator with 75% traffic intensity

Table 4.5: Performance Comparisons With $n = 10$, $\lambda = 1.6$

| **Algorithm** | | **Throughput** | **Delay** | **Fairness** |
|---|---|---|---|---|
| MT - | MT | 1.376 | 41.4 | 12.2 |
| | *d*-algorithm, d=5 | 1.49 | 38.5 | 15 |
| | LCQ | 1.60 | 50.3 | 39.1 |
| | BAA, $\alpha$=35% | 1.60 | 36.6 | 16.2 |
| PF - | PF | 1.56 | 51.3 | 24.7 |
| | *d*-algorithm, d=5 | 1.59 | 47.3 | 26.3 |
| | LCQ | 1.59 | 52.7 | 42.1 |
| | BAA, $\alpha$=60% | 1.60 | 44.3 | 25.6 |

BAA with $\alpha = 35\%$ achieves the lowest packet delay, both MT-LCQ and MT-BAA achieve the highest cell throughput, but with MT-LCQ the average delay is even worse than the MT scheduler. MT-*d*-algorithm improves packet delay than MT, but the fairness and cell throughput are worse than MT-BAA. Similar performance is observed with PF based schedulers. The MT-BAA scheduler achieves an average of 12%, 5%, 27% on delay improvement compared to MT, *d*-algorithm and LCQ, respectively. The PF-BAA scheduler achieves an average of 14%, 6%, 16% on delay improvement compared to PF, *d*-algorithm and LCQ, respectively.

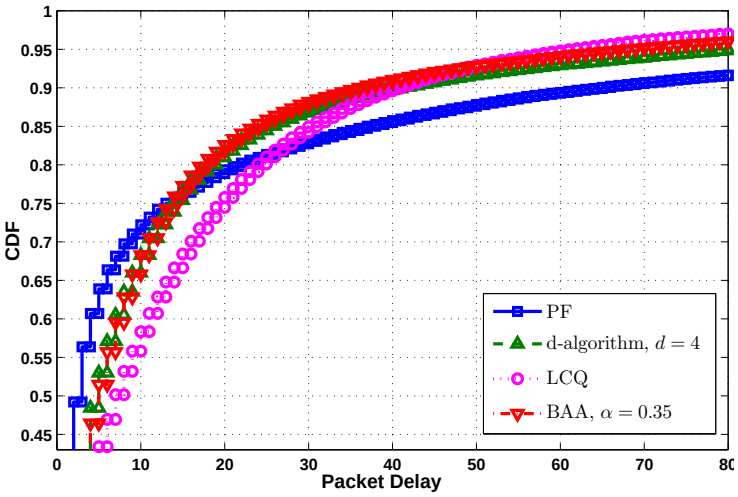


Figure 4.6: CDF comparison of packet delay with 75% traffic intensity

Table 4.6: Performance Comparisons With $n = 7$, $\lambda = 0.8$

| Algorithm | | Throughput | Delay | Fairness |
|---|---|---|---|---|
| MT - | MT | 0.8954 | 5.61 | 6.9 |
| | $d$-algorithm, $d = 3$ | 0.8985 | 4.83 | 9.4 |
| | LCQ | 0.8987 | 4.86 | 17.1 |
| | BAA, $\alpha = 50\%$ | 0.8981 | 4.62 | 9.1 |
| PF - | PF | 0.8982 | 4.61 | 8.8 |
| | $d$-algorithm, $d = 3$ | 0.8993 | 4.43 | 13.1 |
| | LCQ | 0.8996 | 4.39 | 18.4 |
| | BAA, $\alpha = 60\%$ | 0.8994 | 4.08 | 11.8 |

#### 4.6.2.5 Fixed n=7, 50% Traffic Intensity

The system performance is summarized in Table 4.6 for both MT and PF based schedulers. With the MT based schedulers, when the number of users per cell decreases to $n = 7$, similar to the case of $n = 10$, all the three MT based schedulers provide better system performance than MT scheduler, in terms of cell throughput, packet delay, delay fairness and worst case delay. MT-LCQ achieves the best cell throughput, delay fairness index and worst case delay. With $\alpha = 50\%$, the MT-BAA provides the best system performance in terms of packet delay. Similar performance is observed with PF based schedulers. MT-BAA achieves an average of 18%, 4%,

5% on delay improvement compared to MT, $d$-algorithm and LCQ, respectively. PF-BAA achieves an average of 12%, 8%, 7% on delay improvement compared to PF, $d$-algorithm and LCQ, respectively.

Table 4.7: Performance Comparisons With $n = 15$, $\lambda = 0.8$

| **Algorithm** | | **Throughput** | **Delay** | **Fairness** |
|---|---|---|---|---|
| MT - | MT | 0.9264 | 6.29 | 8.8 |
| | $d$-algorithm, $d = 5$ | 0.9290 | 5.78 | 17.6 |
| | LCQ | 0.9292 | 5.72 | 19.3 |
| | BAA, $\alpha = 65\%$ | 0.9287 | 5.21 | 16.4 |
| PF - | PF | 0.9281 | 5.28 | 12.6 |
| | $d$-algorithm, $d = 5$ | 0.9290 | 5.41 | 23.2 |
| | LCQ | 0.9292 | 5.38 | 25.8 |
| | BAA, $\alpha = 70\%$ | 0.9287 | 4.80 | 17.1 |

#### 4.6.2.6 Fixed n=15, 50% Traffic Intensity

With $n = 15$, the average packet delay increases, the system performance is summarized in Table 4.7 for both MT and PF based schedulers. Similar to previous case, with MT based schedulers, under $\alpha = 65\%$, the MT-BAA provides the best system performance in terms of packet delay. MT-LCQ achieves the best delay fairness index and worst case delay. Similar performance is observed with PF based schedulers. The MT-BAA scheduler achieves an average of 17%, 10%, 9% on delay improvement compared to MT, $d$-algorithm and LCQ, respectively. The PF-BAA scheduler achieves an average of 9%, 11%, 11% on delay improvement compared to PF, $d$-algorithm and LCQ, respectively.

#### 4.6.2.7 Fixed n=20, 50% Traffic Intensity

The system performance is summarized in Table 4.8. Similar to previous case of $n = 15$, with MT based schedulers, under $\alpha = 50\%$, the MT-BAA provides the

Table 4.8: Performance Comparisons With $n = 20$, $\lambda = 0.8$

| Algorithm | | Throughput | Delay | Fairness |
|---|---|---|---|---|
| MT - | MT | 0.927 | 6.58 | 7.3 |
| | $d$-algorithm, $d = 3$ | 0.929 | 6.55 | 11.2 |
| | LCQ | 0.929 | 6.38 | 16.3 |
| | BAA, $\alpha = 50\%$ | 0.928 | 5.87 | 10.1 |
| PF - | PF | 0.931 | 5.51 | 10 |
| | $d$-algorithm, $d = 3$ | 0.929 | 6.08 | 14.2 |
| | LCQ | 0.932 | 5.77 | 19.5 |
| | BAA, $\alpha = 60\%$ | 0.931 | 5.24 | 11.3 |

best system performance in terms of packet delay. MT-LCQ achieves the best delay fairness index and worst case delay. Similar performance is observed with PF based schedulers. The MT-BAA scheduler achieves an average of 11%, 10%, 8% on delay improvement compared to MT, $d$-algorithm and LCQ, respectively. The PF-BAA scheduler achieves an average of 5%, 14%, 9% on delay improvement compared to PF, $d$-algorithm and LCQ, respectively.

## 4.7 Concluding Remarks

In this Chapter, we have proposed a new buffer-aware adaptive (BAA) scheduling algorithm, to improve the system performance by reducing average packet delay, while maintaining the queue stability condition of the networks. The proposed scheduler considers only the users whose transmission rates satisfy a certain condition as candidate users. Then the buffer conditions are used to make a scheduling decision among all the candidate users, *i.e.*, the user with largest queue backlog among candidate users will be scheduled. Stability analysis is provided and average throughput lower bound and approximation are derived. Simulation experiments are conducted via both SISO and LTE system level simulators, which indicate consistency with

our analytical findings and the advantages of the proposed algorithm over existing methods.

# CHAPTER 5: AERONAUTICAL COMMUNICATION NETWORKS

## 5.1 Introduction and Motivation

Advances in signal processing, rapid prototyping and an increasing consumer demand for wireless connectivity is opening a new paradigm of data service, "Aeronautical Communication Networks (ACN)". National Aeronautics and Space Administration (NASA) [50], Federal Aviation Administration (FAA), EUROCONTROL and Networking the Sky for Civil Aeronautical Communications (NEWSKY) [51] are all including an aeronautical platform as part of their network infrastructure. The objective is to provide a cost effective data network for aeronautical stations (AS), as well as use it as a relay for ground and airborne nodes. An aeronautical station (AS) could be a commercial plane, helicopter, or any other low orbit station, *i.e.*, Unmanned Air Vehicle, High Altitude Platform. ACN can provide service for ground networks [52], in-flight Internet [53], public safety, and military communications [54].

Two projects within US and European Union (EU) have started the evaluation of a potential ACN based system. NASA's Advanced CNS Architectures and System Technologies (ACAST) are contributing through Technology Assessment and Network Architectures [50] and EU based research is started shaping within the project NEWSKY [51]. The main objective of these studies is to define future wireless communication architecture for air traffic control and management. It will provide a high speed commercial communication service. From networking point of view, there are a couple of studies, in which in-flight Internet with aeronautical *ad hoc* networking

are discussed in [53,54]. There are several patents related to the use of AS as a relay for ACN architecture to provide in-flight services [52,55].

In this Chapter, the system performance of such a communication network is discussed in terms of system throughput and average delay. We consider two communications models with single-hop and two-hop in a mobile *ad hoc* network (MANET) as the models for ACN. The objective is to introduce the concepts and methodologies developed from MANET into ACN and present the system performance of such networks. We derive the ACN throughput upper-bound for the two models, with or without the help of intermediate relay AS. We show that the two-hop model achieves larger throughput than the single-hop model. Since the delay issue is more salient in two-hop communications, in which the data from a source AS has to be buffered in the relay AS until transmitted to the destination AS, we derive the closed-form end-to-end average delay expression analytically. Considering the large communication distances between ASs, it is obvious that ACN as MANET should be formed by establishing wireless multihop paths to reach distant ASs, and also ground stations (GS). For example, to provide connection in oceanic flights, multihop communication seems to be a practical solution, since it is cost effective and delay sensitive, compared to the satellite communications [56, 57]. Application of wireless *ad hoc* strategies to the ACN is investigated in [53,58] by considering the routing protocols.

In the current literature, there is not much progress on understanding the fundamental throughput and delay performances in ACN. The contribution of this work is two-fold: first, we introduce the idea of single-hop and two-hop MANET to ACN and derive the general system performance, respectively, and second, we provide the simulation results for two possible scenarios as ACN architecture. To this end, we will make use of the throughput and delay analysis methods developed for MANETs [7, 10, 15, 34, 35]. These works consider the throughput performance of

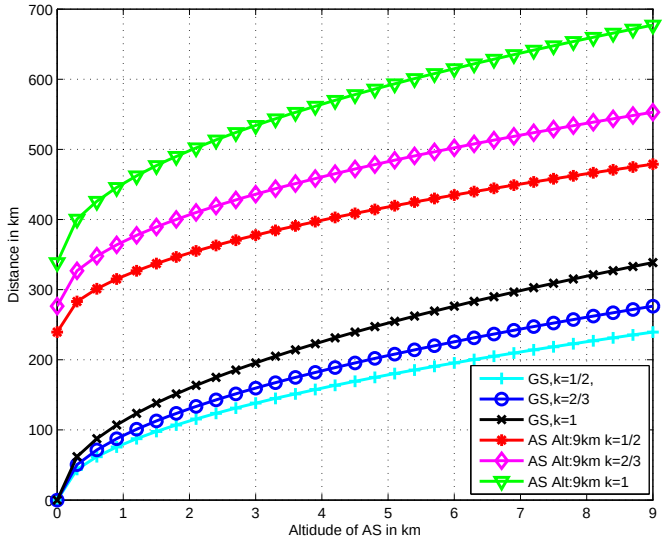


Figure 5.1: Communication zone of an aeronautical station, note: from [60] ©2011 IEEE

MANET with various network structure and mobility models, then derive the system throughput performance; [10, 15, 35] also analyze the delay performance, by defining the delay as the average time for a packet from generating at a source node until arriving at the destination. Furthermore, the throughput and delay relation is also derived, by trade-off throughput to improve the delay. However, in this work, we introduce the idea of both single-hop and two-hop MANET to ACN and provide the analytical and simulation results for the two scenarios.

## 5.2 System Model and Assumptions

### 5.2.1 Aeronautical Geometry and Connectivity

The general aeronautical geometry and connectivity analysis are provided for the investigation of ACN as a MANET. First, we consider evaluation of the geometric relationship between two ASs to their altitudes $(h_1, h_2)$. The altitudes are assumed to be always referred as with respect to sea level. The line-of-sight (LOS)

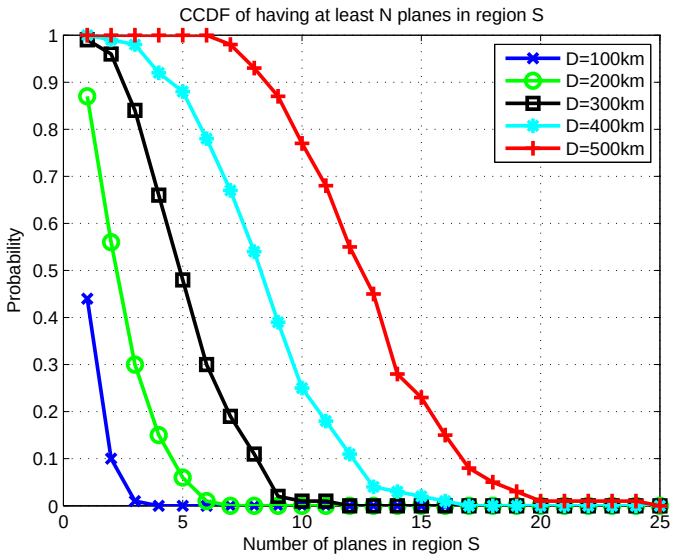


Figure 5.2: CCDF of having N planes in region S, note: from [60] ©2011 IEEE

communication distance (without considering Fresnel and other parameters) between two ASs can be calculated using Pythagoras theorem as follows,

$$D = (h_1 \times [2R + h_2])^{0.5}\,, \tag{5.1}$$

where $R$ is the radius of the Earth which differentiates between [6336 km, 6399 km], but generally assumed 6370 km.

The above formula is only valid, where GS ($h_1 = 0$) and AS ($h_2$) is assumed to be at sea level. Configurations, where either of station is above the seal level is also needs to be accommodated. In this case, the above formula needs additional steps for calculating the communication distance. To simulate various conditions of heights, International Telecommunication Union (ITU) introduced a statistical factor '$k$' to provide more accurate distance measurement as follows:

$$D = (h_1 \times [2Rk + h_2])^{0.5}\,. \tag{5.2}$$

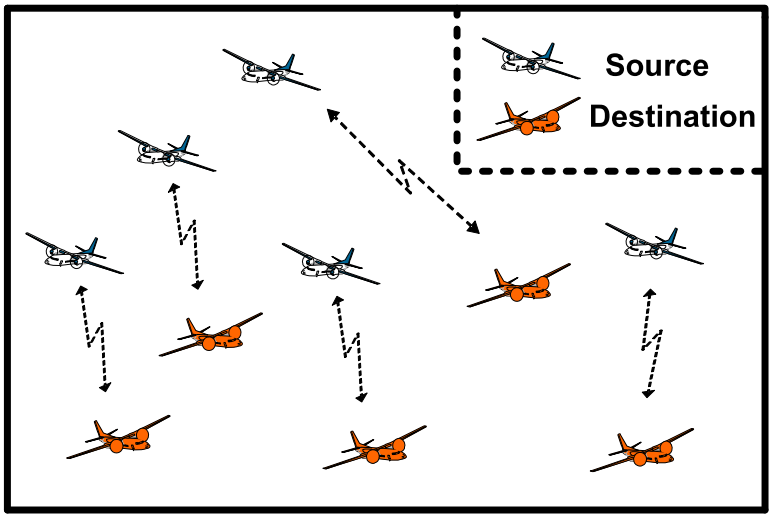


Figure 5.3: Single-hop model, note: from [61] ©2012 IEEE

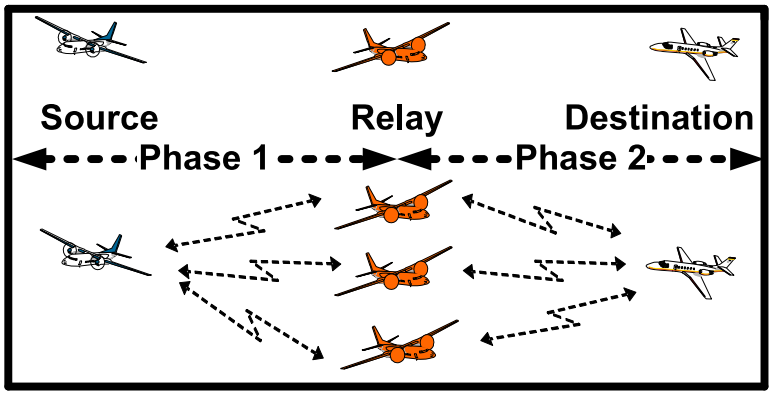


Figure 5.4: Two-hop model, note: from [61] ©2012 IEEE

In order to design a reliable link between transmitter and receiver, the $k$ factor should be less than 1.

Fig. 5.1 shows the maximum communication distances that can be achieved between a GS and an AS, and between two ASs, while an ASs altitude is changing between 0-9 km (9 km is a typical value altitude of a commercial airplane). The communication zone (with radius $D$ given in y-axis) for low altitudes is also very large, meaning even AS in very low altitudes can also communicate very long distances. The jump in the first 2 km altitudes for GS communications can be considered a very low orbit AS which can reach a communication zone of $D = 120$ km. Many

commercial planes flying at the altitude of 9 km can potentially create communication zones about $D = 250$ km with a very conservative approach ($k = 0.5$). The communication distance between two ASs, it can be inferred that it could reach up to $D = 480$ km with $k = 1/2$. Fig. 5.1 prove that ASs could be used as a backhaul or relay for wireless infrastructures, since they have the capability of communicating long distances as compared to wireless ground backhauls. Due to a lower altitude, as compared to a satellite, it is obvious that ACN will have a substantial lower round trip delay. This will allow a low delay telephone and voice over IP service.

According to the proposed ACN structure, providing service to an AS is always robust when an AS is registered to a GS. In the case when it is not registered and forms an ad-hoc network where GS is not available [53, 54]. For commercial aircraft application, a connectivity analysis for AS is presented here as an example. For the purpose of this analysis actual flight data is used for analysis which is obtained from Sivil Havacilik Genel Mudurlugu (SHGM) [59], the governmental agency responsible for civil aeronautics in Turkey. According to the arrival/departure rates and obtained data for each airport in Turkey, it is calculated that there are at least 20 planes in the sky in a given time instant. Turkey as a 1600 km to 800 km region using a conservative approach and assuming the planes have a Poisson distribution within the region, a probabilistic approach for different communication distances (D) is given in Fig. 5.2. S region is defined as the communication zone which is $\pi * D^2$. Fig. 5.2 shows that when D is 100 km, which can be the scenario of GS that can see AS at the altitude of 2 km (see Fig. 5.1), the probability of having a connection between them is 0.35. When $D$ is 500 km, which can be the scenario of two ASs at the altitude of 9 km (see Fig. 5.1), the probability of having connection is up to five AS is 1. The growth in the air traffic increases the number of nodes [54],and the

effect of the number of nodes on throughput and delay is the in the scope of this study and will be investigated in Section 5.5.

### 5.2.2 Problem Statement: Throughput of an ACN

As the connectivity analysis presents in Section 5.2, we can model the ACN as a mobile *ad hoc* network as it is illustrated in Fig. 5.3 for single-hop model and Fig. 5.4 for two-hop model. We assume that ASs are the *ad hoc* nodes, which can transmit data (as sources) or receive data (as destinations), instead of only communicating with the ground base stations [53]. For single-hop model, we consider source ASs communicate with their destination ASs directly, without the help of any relay ASs. For two-hop model, we consider a decode-and-forward communication protocol, in which source ASs communicate with their destination ASs through the help of the intermediate relay ASs. The scheduling method for both models is based on the nearest neighbor transmission. For single-hop, source ASs can communicate with their destination ASs only when they are the nearest neighbors to each other. For two-hop, sources can communicate with any relays and relays can communicate with destinations only when they are the nearest neighbors to each other. We assume every timeslot, nodes move to a new location independent and identically over the network. Since ASs are moving with a determined pattern, the topology of ACN allows clustering. Therefore the capacity calculations in this study can be assumed as the worst case with the assumption of i.i.d. distribution. The clustering topology and its effects on throughput and delay performances is not the scope of this study and left as a future study. We consider a fixed transmission rate in a communication zone and single-user decoding scheme. Accordingly, the interference from multiple concurrent transmissions and noise are incorporated in the signal-to-interference-plus-noise ratio (SINR) model. Thus, for a fixed transmission rate, the transmission

from node $i$ to node $j$ is successful if the received SINR is larger than a threshold $\beta_{th}$, *i.e.*,

$$\mathsf{SINR}_{i,j} = \frac{P|\gamma_{i,j}|^2}{N_j + P\sum_{k\neq i}|\gamma_{k,j}|^2} \geq \beta_{th}, \tag{5.3}$$

where $|\gamma_{i,j}|^2$ denotes the channel gain between the transmitter $i$ and receiver $j$ and $|\gamma_{k,j}|^2$ denotes the channel gain between the transmitter $k$ and receiver $j$, $P$ is the fixed transmission power and $N_j$ is the noise power at receiver $j$. The channel gain is given as $|\gamma_{i,j}|^2 = \frac{1}{|X_i - X_j|^\eta}$, where $X_i$ denotes the location of node $i$ at current time slot and $\eta$ is the path loss exponent. We further assume $\eta = 2$ in ACN, since a 2-ray model with a propagation model is close to free-space in ACN [50]. The SINR threshold $\beta_{th}$ is determined by the transmission rate $R$ according to Shannon's equation $R \leq log(1 + \beta_{th})$. Thus, to find the maximum ACN system throughput is equivalent to find the maximum number of concurrent successful transmissions over the ACN.

## 5.3 Throughput Analysis

In this Section, we derive the throughput upper bounds of ACN under two communication models, *i.e.*, single-hop [60] and two-hop [61], under the assumption of channel attenuation is mainly affected by the large-scale fading, due to distance between the transmitters and receivers [63].

### 5.3.1 Single-Hop Communication

We first present the analytical derivations of the throughput upper bound for the one-hop ACN model. We use the well-known genie-aided scheme for our derivations, which is assumed to have the capability to find the maximum number of concurrent successful transmissions. The concept of concurrent successful trans-

missions is used to calculate the system throughput in wireless *ad hoc* networks for opportunistic scheduling. The genie scheme basically follows two steps: first, the scheme selects $p$ $(1 \leq p \leq n)$ source-destination active pairs which are scheduled for transmissions; for each selection, the scheme tests if all the $p$ received SINRs are greater than the threshold $\beta_{th}$, then the $p$ concurrent transmissions are successful. For each selection, if the $p$ concurrent transmissions are successful, we call the current selection as a valid group. In a network with $n$ source-destination pairs, we have $\binom{n}{p}$ different possible ways to select $p$ active source-destination pairs. We denote $X(p)$ as the total number of valid groups in which all $p$ concurrent source-destination pairs are successful. Thus, to have at least $p$ concurrent successful transmissions is equivalent to $X(p) \geq 1$. Based on the analysis, we have the following function,

$$\begin{aligned} X(p) &= \sum_{\substack{S_1 \in 1,\ldots,n; \\ |S_1|=p}} 1(\mathsf{SINR}_{i,j} \geq \beta_{th}, \forall i \in S_1) \\ &= \sum_{\substack{S_1 \in 1,\ldots,n; \\ |S_1|=p}} 1(\frac{\frac{P}{|X_i - X_j|^\eta}}{N_j + \frac{P}{\sum_{k \neq i} |X_k - X_j|^\eta}} \geq \beta_{th}, \forall i \in S_1) \end{aligned} \tag{5.4}$$

where $S_1$ is the group with selected active nodes, based on our scheduling policy.

First we upper-bound $Pr[X(p)] \geq 1$ as [34]

$$Pr[X(p) \geq 1] \leq \mathbb{E}[X(p)] \tag{5.5}$$

$$= \binom{n}{p} (Pr[\frac{\frac{P}{|X_i - X_j|^\eta}}{N_j + \frac{P}{\sum_{\substack{k \in S_1 \\ k \neq i}} |X_k - X_j|^\eta}} \geq \beta_{th}])^p, \tag{5.6}$$

in which (5.5) is because of the Markov's inequality, (5.6) is due to the linear property of expectation and the SINRs of the nodes' are i.i.d.

Next we further upper bound the probability term of $(Pr[\frac{\frac{P}{|X_i-X_j|^\eta}}{N_j+\frac{P}{\sum_{\substack{k\in S_1\\k\neq i}}|X_k-X_j|^\eta}} \geq \beta_{th}])^p$.

For simplicity, we write $\frac{1}{|X_i-X_j|^\eta} = |\gamma_{i,j}|^2 = \max\{|\gamma_{1,j}|^2, \ldots, |\gamma_{n,j}|^2\}$ and denote the interference from all the other scheduled $(p-1)$ concurrent transmissions as $U = \frac{1}{\sum_{\substack{k\in S_1\\k\neq i}}|X_k-X_j|^\eta} = \frac{1}{\sum_{k\in S_1}|X_k-X_j|^\eta - |X_i-X_j|^\eta}$.

The probability term can be written and further upper-bounded as,

$$\begin{aligned}
&Pr[\frac{P|\gamma_{i,j}|^2}{N_j+PU} \geq \beta_{th}] \\
&= \int_0^\infty Pr[\frac{|\gamma_{i,j}|^2}{N_j/P+U} \geq \beta_{th} | U \leq u] f_U(u) du \\
&= \int_0^\infty Pr[|\gamma_{i,j}|^2 \geq \beta_{th}(N_j/P+u) | U = u] f_U(u) du \\
&= \int_0^\infty Pr[|\gamma_{i,j}|^2 - \mu_1 \geq \beta_{th}(N_j/P+u) - \mu_1] f_U(u) du \\
&\leq \int_0^\infty \frac{\sigma_1^2}{\sigma_1^2 + (\beta_{th}(N_j/P+u) - mu_1)^2} f_U(u) du, \qquad (5.7)
\end{aligned}$$

where $\mu_1$ and $\sigma_1^2$ denote the finite mean and variance of $|\gamma_{k,j}|^2$. (5.7) is based on one-sided Chebyshev inequality $Pr[Z - o \geq \omega] \leq \frac{\sigma^2}{\sigma^2+\omega^2}$.

For the distribution of interference which is termed as $f_U(u)$ in (5.7), we adapt the approximation methods as in [34]. As far as scaling is concerned, we can upper bound (5.7) as

$$Pr[\frac{P|\gamma_{i,j}|^2}{N_j+PU} \geq \beta_{th}] \leq \frac{C_1}{p^2}. \qquad (5.8)$$

where $C_1$ is a constant, which is determined by the values of $\beta_{th}$, $\mu_1$ and $\sigma_1$. Combing (5.8) and (5.6) we have

$$\begin{aligned} Pr[X(p) \geq 1] &\leq \binom{n}{p} (Pr[\frac{|\gamma_{i,j}|^2}{N_j/P + U} \geq \beta_{th}])^p \\ &\leq (\frac{C_1 n}{p^3})^p \\ &= e^{p(\log(C_1 n) - 3\log p)}. \end{aligned} \tag{5.9}$$

According to the genie scheme, we set the value of $p$ as $(1+\varepsilon_1)C_1^{1/3}n^{1/3}$ with $\varepsilon_1 > 0$ , so that the term $Pr[X(p) \geq 1] \to 0$, which means that there is no valid groups with the $p$ concurrent successful transmissions. It is equivalent that $p = (1+\varepsilon_1)C_1^{1/3}n^{1/3}$ is the upper bound of the one-hop ACN system throughput.

### 5.3.2 Two-Hop Communication

We now proceed to derive the system throughput upper bound for two-hop ACN architecture. Similar to the derivations in one-hop model, we use the genie-aided scheme to find the two-hop throughput upper bound.

For a network with $n$ source-destination pairs and $m$ relays, in the first hop, we have $\binom{n}{q}$ different possible ways to select $q$ $(1 \leq q \leq m)$ active source-relay (S-R) pairs. Furthermore, since each source can be scheduled by any relays, for each $q$ selected active S-R pairs, there are $q!$ different ways to associate the S-R pairs. Thus, for Phase 1, there is $\binom{n}{q}q!$ different ways to select the active S-R pairs. We denote $Y(m)$ as the total number of valid groups in which all $q$ concurrent S-R pairs are successful. Thus, to have at least $q$ concurrent successful transmissions is equivalent

to $Y(q) \geq 1$. Based on the analysis, we have the following function,

$$\begin{aligned} Y(q) &= \sum_{\substack{S_2 \in 1,\ldots,n; \\ |S_2|=q}} 1(\mathsf{SINR}_{i,j} \geq \beta_{th}, \forall i,j \in S_2) \\ &= \sum_{\substack{S_2 \in 1,\ldots,n; \\ |S_2|=q}} 1(\frac{P|g_{i,j}|^2}{N_j + P\sum_{k\neq i}|g_{k,j}|^2} \geq \beta_{th}, \forall i,j,k \in S_2) \end{aligned}$$

where $|g_{i,j}|^2 = \max\{|g_{1,j}|^2, \ldots, |g_{n,j}|^2\}$ and $\sum_{k\neq i}|g_{k,j}|^2 = \sum_{\substack{k\in S_2 \\ k\neq i}}|g_{k,j}|^2$, based on our scheduling policy; $S_2$ is the group with selected active nodes.

First we upper-bound $Pr[Y(q) \geq 1]$ as [34]

$$Pr[Y(q) \geq 1] \leq \mathbb{E}[Y(q)] \tag{5.10}$$

$$= \binom{n}{q} q!(Pr[\frac{P|g_{i,j}|^2}{N_j + P\sum_{k\neq i}|g_{k,j}|^2} \geq \beta_{th}])^q, \tag{5.11}$$

where (5.10) is because of the Markov's inequality, (5.11) is due to the linear property of expectation and the $\mathsf{SINRs}$ of the nodes' are i.i.d.

Next we further upper bound the term $(Pr[\frac{P|g_{i,j}|^2}{N_j+P\sum_{k\neq i}|g_{k,j}|^2} \geq \beta_{th}])^q$. For simplicity, we denote $N = |g_{i,j}|^2 = \max\{|g_{1,j}|^2, \ldots, |g_{n,j}|^2\}$ and the interference from all the other scheduled $q-1$ concurrent transmissions as $V = \sum_{\substack{k\in S_2 \\ k\neq i}}|g_{k,j}|^2 = \sum_{k\in S_2}|g_{k,j}|^2 - N$.

The probability term can be written and further upper-bounded as,

$$Pr[\frac{PN}{N_j + PV} \geq \beta_{th}] \leq \int_0^\infty \frac{\sigma_2^2}{\sigma_2^2 + (\beta_{th}(N_j/P + v) - \mu_2)^2} f_V(v)dv, \tag{5.12}$$

where $\mu_2$ and $\sigma_2^2$ denote the finite mean and variance of $|g_{k,j}|^2$, (5.12) is based on one-sided Chebyshev inequality $Pr[Z - o \geq \omega] \leq \frac{\sigma^2}{\sigma^2+\omega^2}$. For the distribution of interference which is termed as $f_V(v)$ in(5.12), we adapt the approximation methods

as in [34]. As far as scaling is concerned, we can upper-bound (5.12) as

$$Pr[\frac{PN}{N_j + PV} \geq \beta_{th}] \leq \frac{C_2}{q^2}. \tag{5.13}$$

where $C_2$ is a constant, which is determined by the values of $\beta_{th}$, $\mu_2$ and $\sigma_2$. Combining (5.13) and (5.11) we have

$$\begin{aligned} & Pr[Y(q) \geq 1] \\ \leq & \binom{n}{q} q!(Pr[\frac{N}{N_j + V} \geq \beta_{th}]^q) \\ \leq & \frac{n!}{(n-q)!}(\frac{C_2}{q^2})^q \\ \leq & (\frac{C_2 n}{q^2})^q = e^{q(\log n - 2\log q + \log C_2)}. \end{aligned} \tag{5.14}$$

According to the genie scheme, we set the value of $q$ as $(1+\varepsilon_2)C_2^{1/2}n^{1/2}$ with $\varepsilon_2 > 0$ , so that the term $Pr[Y(q) \geq 1] \to 0$, which means that there is no valid groups with the $q = (1+\varepsilon_2)C_2^{1/2}n^{1/2}$ concurrent successful transmissions. It is equivalent that in Phase 1, $q = (1+\varepsilon_2)C_2^{1/2}n^{1/2}$ is the throughput upper bound of the two-hop ACN system throughput.

The system throughput in the second hop is similar to that in the first hop. The only difference is that, instead of transmitting from $n$ source nodes to $m$ relays, now the $m$ relays are transmitting to $n$ destinations, which is the reverse side of each other. However, the throughput derivations are mathematically equivalent for the two hops and the throughput upper bound is the same.

With the throughput upper bound derivations for the first and second hops, it is easy to conclude that the two-hop ACN throughput upper bound can be given as $(1+\varepsilon_2)C_2^{1/2}n^{1/2}$.

Note that this throughput result is better than the single-hop communication, where only the communication between source and destination is allowed, since more freedom is allowed in the two-hop communication.

### 5.4 Delay Analysis

We show that the two-hop model achieves better throughput than single-hop model. However, in two-hop model since the packet transmitted from a source node has to be buffered in the relay until arriving at the destination node, the delay issue is more salient. It is necessary to characterize the delay performance for the two-hop model. We define the end-to-end average delay is the time for a packet from generating at the source AS until arriving at the designated destination AS. For the two-hop ACN model, the delay can be written in two parts as $Delay = D_1 + D_2$, where $D_1$ and $D_2$ denote respectively the delay in the first hop (the time for a packet transmitted from a source AS to a relay AS) and the delay in the second hop (the time for a packet forwarded from the relay AS to the corresponding destination AS). Since the source AS can transmit a packet to any relay AS, which is its nearest neighbor in the first hop, while the packet has to be buffered in the relay before it can be forwarded to the corresponding destination AS in the second hop, $D_1$ is relatively small compared to $D_2$ which is actually the major delay component in the two-hop ACN model.

Since packets have to be buffered in the relay before it can be forwarded to the destination AS, there is potential queues in the relays' buffer. For this reason, we introduce the queueing system to obtain the delay performance. We assume the packets arrival process at a source AS is a Markov process with arrival rate $\lambda$, the service has a Bernoulli distribution with departure rate $\mu$. Since there are $(n-2)$ potential nodes to be served as relay AS, the probability for a given packet from the

output of the source AS to be transmitted to the first relay AS is $\mu = 1/(n-2)$. With the known results for such queuing model [20], we have the average number of packets at a source AS as

$$\overline{L}_{source} = \frac{\rho(1-\lambda)}{1-\rho} = \frac{\lambda(1-\lambda)}{1-\mu}, \tag{5.15}$$

where $\rho = \lambda/\mu$ is the traffic intensity. Note that in order to ensure a stable queueing system, the arrival rate must be strictly smaller than the service rate, so the equation $\lambda = \rho\mu$ is on the condition that $0 < \rho < 1$. From Little's Theorem, the delay in the first hop can be derived as

$$D_1 = \frac{\overline{L}_{source}}{\lambda} = \frac{1-\lambda}{\mu-\lambda} = \frac{n(1-\rho)+2\rho}{n(1-\rho)-2(1-\rho)}. \tag{5.16}$$

In the second hop, a packet arrives at the relay AS with the rate value of $\widetilde{\lambda} = \lambda/(n-2)$, since for $(n-2)$ potential relay AS, a given packet from the output of the source node to be transmitted to the first relay node is $1/(n-2)$. The relay AS is scheduled for a potential packet transmission to the destination AS with probability of $\widetilde{\mu} = 1/n$, since the packet from the relay AS has to be forwarded to the designated destination AS out of the $n$ destination AS. For the second hop, the packet arrival process and the departure opportunities are mutually independent events in the relay AS, which follows that the discrete time Markov chain for queue occupancy in the relay AS [10]. Thus, this queuing system can be modeled as the simple birth-death chain, which is similar as the M/M/1 model with arrival rate of $\widetilde{\lambda}$ and departure rate of $\widetilde{\mu}$. Accordingly, with the well-known results, we have the average number of

packets at a relay AS as

$$\overline{L}_{relay} = \frac{\widetilde{\lambda}}{\widetilde{\mu}-\widetilde{\lambda}}. \tag{5.17}$$

Similarly, the delay in the second hop, based on Little's Theorem can be derived as

$$D_2 = \frac{\overline{L}_{relay}}{\widetilde{\lambda}} = \frac{1}{\widetilde{\mu}-\widetilde{\lambda}} = \frac{n}{1-\rho}. \tag{5.18}$$

Thus, the total delay in the two-hop ACN model is

$$\begin{aligned} Delay &= D_1 + D_2 \\ &= \frac{n(1-\rho)+2\rho}{n(1-\rho)-2(1-\rho)} + \frac{n}{1-\rho} \\ &= \frac{n}{1-\rho} + \frac{n}{n-2} + \frac{2\rho}{n(1-\rho)-2(1-\rho)}. \end{aligned} \tag{5.19}$$

This closed-form equation shows that the average end-to-end delay increases with the number of AS nodes $n$. In the order-of-magnitude sense, it indicates *Delay* grows with the scaling of $n$ and this characteristic can not be removed by decreasing the arrival rate.

## 5.5 Numerical Results and Discussions

In this section, we present some numerical examples of the system throughput in ACN under large-scale fading. All the simulation curves were obtained by averaging over 2,000 channel realizations. Note that the system throughput is defined as the number of successful concurrent transmissions in the context of this paper.

Fig. 5.5 plots the single-hop system throughput versus the number of concurrent transmissions for various number of nodes $n$. It is observed that the number of

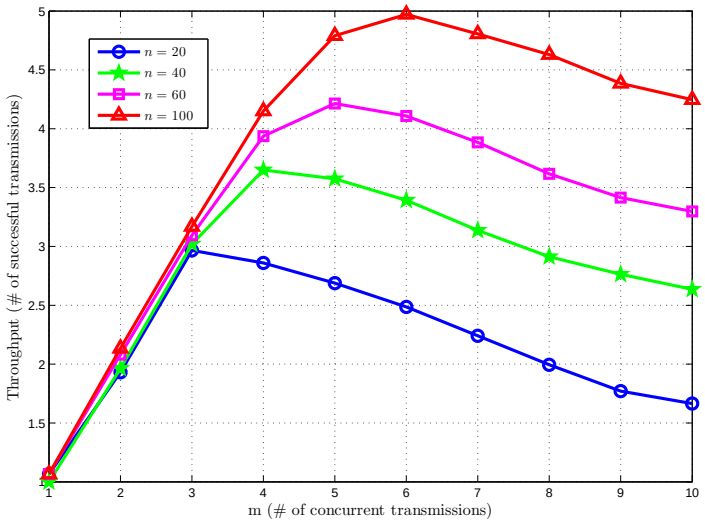


Figure 5.5: Single-hop system throughput versus the number of concurrent transmissions, note: from [61] ©2012 IEEE

successful concurrent transmissions increases with the number of concurrent transmissions at the beginning, however, after a certain number, it starts to decrease. This is because before a certain number of concurrent transmissions, the SINRs of the transmissions are still larger than $\beta_{th}$, which give the increasing number of concurrent successful transmissions. However, increasing the concurrent transmissions also increases the number of interferences, so that the successful transmission starts decreasing after exceeding the threshold, after which the interferences become dominant. Fig. 5.6 presents the single-hop system throughput and its theoretical upper-bound versus the number of nodes $n$. For different number of nodes $n$, the system throughput is obtained with the corresponding optimal number of concurrent transmissions; theoretical upper-bound is derived from the previous analysis. Curves are plotted on a log-log scale, so that with a slope of $\frac{1}{3}$, we can conclude the relationship between throughput and $n$ behavior is $T = C_1 * n^{1/3}$, where $C_1$ is a constant value. Note that $\log T = \frac{1}{3} \log n$ gives the relationship between throughput and the number of nodes $T = C_1 * n^{1/3}$. We can also conclude that the upper-bound is tight, with

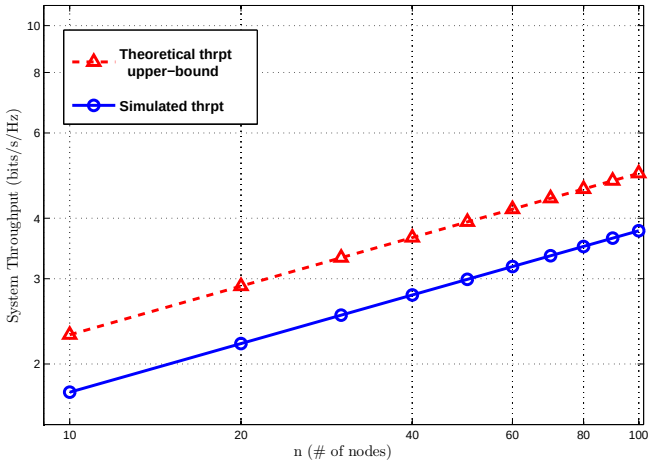


Figure 5.6: Single-hop system throughput upper-bound versus the number of nodes, note: from [61] ©2012 IEEE

the scaling of $n^{1/3}$ and constant value of coefficient $C_1$ under the condition that the aeronautical stations are i.i.d. over the ACN.

Fig. 5.7 plots the two-hop system throughput versus the number of concurrent transmissions for various number of ASs $n$. It is observed that the number of successful concurrent transmissions increases with the number of concurrent transmissions at the beginning. However, after a certain number of concurrent transmissions, it starts to decrease. This is because before a certain number of concurrent transmissions, the SINRs of the transmissions are still larger than $\beta_{th}$, which give the increasing number of concurrent successful transmissions. However, increasing the concurrent transmissions also increases the number of interferences, so that the successful transmission starts decreasing after exceeding the threshold, after which the interferences become dominant. Fig. 5.8 presents the two-hop system throughput and its theoretical upper-bound versus the number of ASs $n$. For different number of nodes $n$, the system throughput is obtained with the corresponding optimal number of concurrent transmissions; theoretical upper-bound is derived from the analysis.

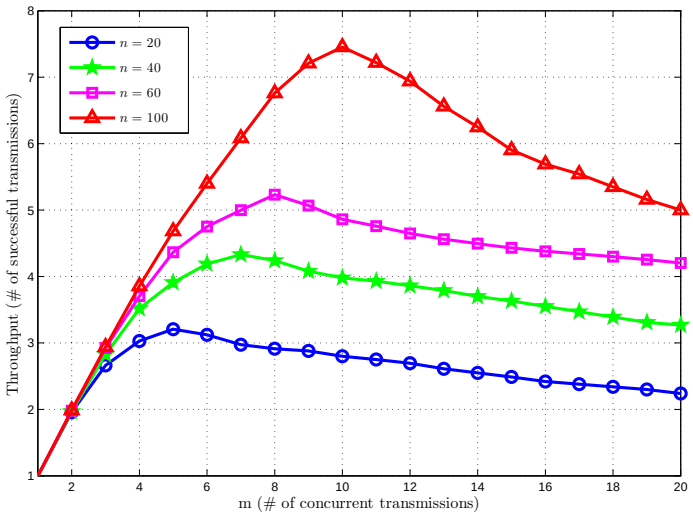


Figure 5.7: Two-hop system throughput versus the number of concurrent transmissions, note: from [61] ©2012 IEEE

Curves are plotted on a log-log scale, so that with a slope of 1/2, we can conclude that the relationship between system throughput $T$ and number of nodes $n$ behavior is $T = C_2 * n^{1/2}$, where $C_2$ is a constant value. Note that $\log T = \frac{1}{2} \log n$ gives the relationship between throughput and the number of nodes as $T = C_2 * n^{1/2}$. Thus, the simulation results confirm our analysis on the throughput upper-bound.

Fig. 5.9 plots the average end-to-end delay with the number of ASs $n$ for two-hop model with different traffic intensity values. It is observed that the average end-to-end delay increases with the number of nodes $n$. Simulation results with different traffic intensity $\rho$ is also shown in Fig. 5.9. Presented are curves with traffic intensity $\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$. From the comparisons between the different traffic intensity rates, clearly that larger traffic intensity will increase the delay. The values of the average end-to-end delay as obtained from the simulation results agree closely with the theoretical analysis.

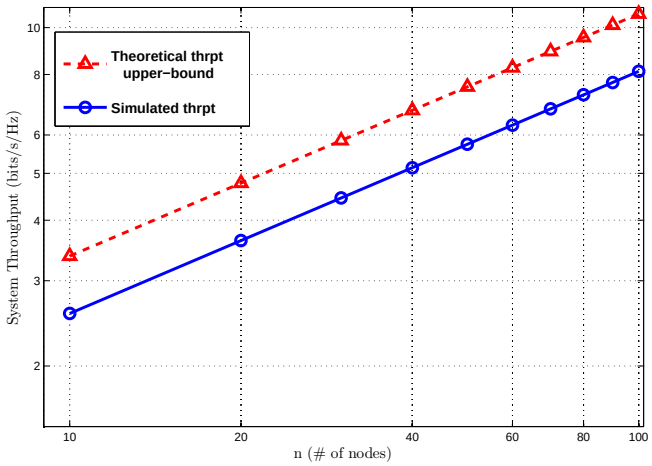


Figure 5.8: Two-hop system throughput upper-bound versus the number of nodes, note: from [61] ©2012 IEEE

## 5.6 Concluding Remarks

In this Chapter, we have investigated a simple ACN system under two communication models, *i.e.*, single-hop and two-hop. The connectivity analysis shows that the number of ASs is feasible to model an ACN as a MANET. We derive the upper bound of the system throughput for the two models, and the closed-form expression of the average delay for the two-hop communication model. We present the corresponding throughput and delay performances of ACN and provide comparisons between single-hop and two-hop models. It is shown that ACN system is worthwhile and feasible in terms of throughput and delay performances, by considering that the number of flights/ASs will keep increasing in the long run. Computer simulations show that our analytical findings on single-hop and two-hop scenarios are aligned with the simulation results for both throughput and delay analysis. The future direction of this study is on more practical mobility and channel models of ASs, considering the effects of correlations among ASs for clustering.

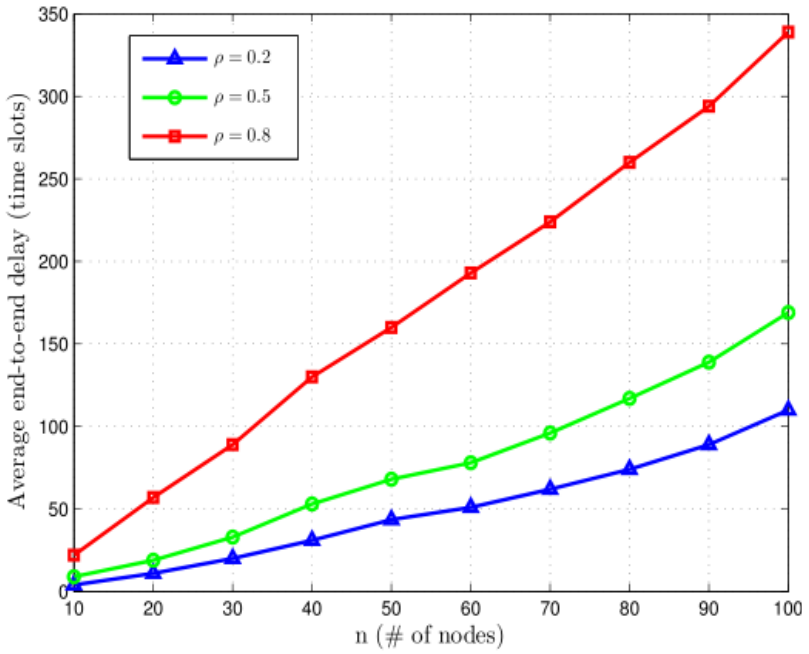


Figure 5.9: Average end-to-end packet delay versus number of nodes with different packet arrival rate, note: from [61] ©2012 IEEE

# CHAPTER 6: CONCLUSION AND FUTURE DIRECTIONS

## 6.1 Main Contributions

In this dissertation, our main goal is to develop practical scheduling algorithms for wireless *ad hoc* networks to enhance system performance, in terms of throughput, delay and stability. This dissertation mainly consists of three main contributions.

Our first contribution is to identify major challenges intrinsic to *ad hoc* networks that affect the system performance, in terms of throughput limits, delay and stability condition.

Our second contribution is that we develop scheduling algorithms for wireless *ad hoc* networks, with various considerations of non-cooperative relays and cooperative relays, fixed-rate transmission and adaptive-rate transmission, full-buffer traffic model and finite-buffer traffic model. Specifically, we propose an opportunistic scheduling scheme and study the throughput and delay performance, with fixed-rate transmissions in a two-hop wireless *ad hoc* networks. In the proposed scheduling scheme, we prove two key inequalities that capture the various tradeoffs inherent in the broad class of opportunistic relaying protocols, illustrating that no scheduling and routing algorithm can simultaneously yield lower delay and higher throughput. We then develop an adaptive rate transmission scheme with opportunistic scheduling, with the constraints of practical assumptions on channel state information (CSI) and limited feedback, which achieves an optimal system throughput scaling order. Along this work with the consideration of finite-buffer model, we propose a Buffer-Aware

Adaptive (BAA) scheduler which considers both channel state and buffer conditions to make scheduling decisions, to reduce average packet delay, while maintaining the queue stability condition of the networks. The proposed algorithm is an improvement over existing algorithms with adaptability and bounded potential throughput reduction.

The third contribution of this dissertation is to extend the methods and analysis developed for wireless *ad hoc* networks to a practical Aeronautical Communication Networks (ACN) and present the system performance of such networks. We use our previously proposed scheduling schemes and analytical methods from the second part to investigate the issues about connectivity, throughput and delay in ACN, for both single-hop and two-hop communication models. We conclude that the two-hop model achieves greater throughput than the single-hop model for ACN. Both throughput and delay performance are characterized.

## 6.2 Future Directions

There are topics that remain unexplored and are closely related to our work. We envision the following extensions to the studies done in this dissertation:

Implementing the proposed vertical layer based cross layer framework under different applications scenarios would be an interesting future task.

- The investigation on the effect of cooperation among relays on throughput and delay tradeoffs with opportunistic scheduling in wireless *ad hoc* networks.

- It would be interesting to study the throughput and delay tradeoffs under different mobility models and optimal resource allocation.

- The study of delay performance with our proposed BAA scheduling algorithm is a remaining work.
- It would be interesting to further improve the system performance in wireless networks using network coding techniques under practical CSI assumption.
- Combine the technique of interference management with the opportunistic scheduling is another interesting research topic.
- A Hybrid Wireless Networks (HWN) is a new kind of wireless networks, combing both wireless *ad hoc* networks, *e.g.*, mobile *ad hoc* networks, sensor networks, and vehicular *ad hoc* networks, and infrastructure wireless networks, *e.g.*, cellular networks, WLANs, and WiMAX networks. It is an interesting research direction to design algorithms for HWN to achieve the advantages of both infrastructure networks and *ad hoc* networks.

## REFERENCES


[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[2] R. Gowaikar, B. M. Hochwald, and B. Hassibi, "Communication over a wireless network with random connections," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2857-2871, Jul. 2006.

[3] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad hoc wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2890-2914, Jul. 2006.

[4] V. I. Morgenshtern and H. Bölcskei, "Crystallization in large wireless networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3319-3349, Oct. 2007.

[5] A. Özgür, O. Lévêque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549-3572, Oct. 2007.

[6] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277-1294, Jun. 2002.

[7] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.

[8] S. Cui, A. M. Haimovich, O. Somekh, and H. V. Poor, "Opportunistic relaying in wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5121-5137, Nov. 2009.

[9] S. Cui, A. M. Haimovich, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Throughput scaling of wireless networks with random connections," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3793-3806, Jul. 2010.

[10] M.J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1917-1937, Jun. 2005.

[11] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks - Part I: the fluid model," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2568-2592, Jun. 2006.

[12] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," *IEEE INFOCOM*, 2004.

[13] S. Toumpis and A. Goldsmith, "Large wireless networks under fading, mobility, and delay constraints," *IEEE INFOCOM*, 2004.

[14] X. Lin and N. B. Shroff, "The fundamental capacity-delay tradeoff in large mobile ad hoc networks," *Third Annual Mediterranean Ad Hoc Networking Workshop*, 2004.

[15] G. Sharma, R. Mazumdar, and N. Shroff, "Delay and capacity trade-offs in mobile ad hoc networks: a global perspective," *IEEE Trans. on Networking*, vol. 15, no. 5, pp. 981-992, Oct. 2007.

[16] L. Ying, S. Yang and R. Srikant, "Optimal Delay-Throughput Tradeoffs in Mobile Ad Hoc Networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4119-4143, Sep. 2008.

[17] C. Zhang, Y. Fang, and X. Zhu, "Throughput-delay tradeoffs in large-scale manets with network coding," *IEEE INFOCOM* 2009.

[18] M. Sharif and B. Hassibi, "Delay considerations for opportunistic scheduling in broadcast fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3353-3363, Sep. 2007.

[19] M. Sharif and B. Hassibi, "On the capacity of MIMO BC channel with partial side information," *IEEE Trans. Wireless Commun.*, no. 2, pp. 506-523, Feb. 2005.

[20] Donald Gross and Carl M. Harris, "Fundamentals of Queueing Theory", 3rd ed. Wiley, John & Sons, 1998.

[21] R. W. Wolff, "Stochastic Modeling and the Theory of Queues", Prentice-Hall, 1989.

[22] V. Asghari and S. Aissa, "Adaptive rate and power transmission in spectrum-sharing systems," *IEEE Trans. Wireless. Comm.*, vol. 9, no. 10, pp 3272-3280, Oct. 2010.

[23] M. S. Alouini and A. J. Goldsmith, "Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Tech.*, vol. 48, no. 4, pp. 1165-1181, Jul. 1999.

[24] T. Keller and L. Hanzo, "Adaptive modulation techniques for duplex OFDM transmission," *IEEE Trans. Veh. Tech.*, vol. 49, no. 5, pp. 1893 - 1906, Sep. 2000.

[25] S. L. Gong, B. G. Kim, and J. W. Lee, "Opportunistic scheduling and adaptive modulation in wireless networks with network coding," *in Proc. IEEE Vehi. Tech. Conf. (VTC'09)*, pp. 1-5, Apr. 2009.

[26] T. Nechiporenko, K. T. Phan, C. Tellambura, and H. H. Nguyen, "On the capacity of Rayleigh fading cooperative systems under adaptive transmission," *IEEE Trans. Wireless. Comm.*, vol. 8, no. 4, pp. 1626-1631, Apr. 2009.

[27] S. Y. Baek, Y. Hong, and D. K. Sung, "Adaptive Transmission Scheme for Mixed Multicast and Unicast Traffic in Cellular Systems," *IEEE Trans. Veh. Tech.*, vol. 58, no. 6, pp. 2899 - 2908, Jul. 2009.

[28] T. Wang and L. Vandendorpe, "Sum rate maximized resource allocation in multiple DF relays aided OFDM transmission", *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1559-1571, 2011.

[29] T. Wang and L. Vandendorpe, "WSR maximized resource allocation in multiple DF relays aided OFDMA downlink transmission", *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3964-3976, 2011.

[30] M. O. Hasna and M. S. Alouin, "End-to-end performance of transmission systems with relays over Rayleigh-fading channels," *IEEE Trans. Wireless. Comm.*, vol. 2, no. 6, pp. 1126-1131, Nov. 2003.

[31] B. Sklar, "Rayleigh fading channels in mobile digital communication systems Part I: Characterization," *IEEE Communications Magazine*, pp. 90-100, Jul. 1997.

[32] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel", *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936-3964, Sep. 2006.

[33] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels", *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768-783, May 2004.

[34] S. Cui, A. M. Haimovich, O. Somekh, H. V. Poor, and S. Shamai, "Throughput Scaling of Wireless Networks With Random Connections," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3793-3806, Aug. 2010.

[35] Y. Wang, S. Cui, R. Sankar, and S. Morgera, "Delay-Throughput Trade-off With Opportunistic Relaying in Wireless Networks," *in Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, Dec. 2011.

[36] Y. Wang, I. Butun, R. Sankar, and S. Morgera, "Adaptive Rate Transmission With Opportunistic Scheduling in Wireless Networks", *in Proc. IEEE Consumer Communications and Networking Conference (CCNC) Workshop*, Las Vegas, NV, Jan. 2012.

[37] Y. Wang, R. Sankar, and S. Morgera, "Adaptive Rate Transmission With Opportunistic Relaying in Wireless Networks", *to appear in IEEE Trans. Veh. Tech.*, Oct. 2012.

[38] R. Knopp and P. Humlet, "Information capacity and power control in single cell multiuser communications," *in Proc. IEEE ICC*, pp. 331-225, 1995.

[39] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing qulaity of service over a shared wireless link," *IEEE Commu. Magazine*, vol. 39, no. 2, pp. 150-154, Feb. 2001

[40] X. Liu, E. K. P. Chong, and N. B. Shroff, "Optimistic transmission scheduling with resource-sharing constaints in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2053-2064, Oct. 2001.

[41] E. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 7, pp. 1024-1039 , Sep. 1991.

[42] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 210-212, Mar. 2005.

[43] Z. Han, Z. Ji, and K. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1366-1376, Aug. 2005.

[44] C. Wengerter, J. Ohlhorst, and A. Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA," *in Proc. IEE Vehicular Tech. Conf. (VTC)*, May. 2005.

[45] M. Neely, "Order Optimal Delay for Opportunistic Scheduling in Multi-User Wireless Uplinks and Downlinks," *IEEE/ACM Trans. Networking*, vol. 16, no. 5, pp. 1188-1199, Oct. 2008.

[46] 3GPP TR 36.814, "Further advancements for E-UTRA physical layer aspects", Rel. 9, v. 9.0.0, Mar. 2010.

[47] IEEE 802.16m, "Evaluation Methodology Document (EMD)", Jan. 2009.

[48] 3GPP TS 36.211, "Physical Channels and Modu- lation," Rel. 8, v. 8.1.0, Nov. 2007.

[49] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA) Physical Layer Procedures," Rel. 8, v. 8.5.0, Dec. 2008.

[50] NASA-ACAST, "Future Aeronautical Communication Infrastructure Technology Investigation," http://acast.grc.nasa.gov/main/projects/, 2008, 215144.

[51] M. Schnell and S. Scalise, "NEWSKY - concept for NEtWorking the SKY for civil aeronautical communications," *IEEE Aero. and Elect. Sys. Mag.*, vol. 22, no. 5, May 2007.

[52] J. Lai, "Broadband wireless communication systems provided by commercial airplanes," *US. Patents 6 285 878*, Sept. 4, 2001.

[53] E. Sakhaee and A. Jamalipour, "The global in-flight internet," *IEEE J. on Sel. Areas in Commun.*, vol. 24, pp. 1748-1757, Sept. 2006.

[54] D. Medina, F. Hoffman, S. Ayaz, and C. H. Rokitansky, "Feasibility of an aeronautical mobile ad hoc network over the north atlantic corridor," *in Proc. IEEE Sensor, Mesh and Ad Hoc Commun. and Net. (SECON)*, June 2008, pp. 109-116.

[55] "Wireless telecommunications system having airborne base station," *Lucent Technologies Inc.*, 1996.

[56] H. Tu and S. Shimamoto, "A proposal for high air-traffic oceanic flight routes employing ad-hoc networks," *in Proc. IEEE Wireless Commun. and Net. Conf. (WCNC)*, Apr. 2009.

[57] D. Medina, F. Hoffmann, S. Ayaz, and C.-H. Rokitansky, "Topology characterization of high density airspace aeronautical ad hoc networks," *in Proc. IEEE Int. Conf. on Mobile Ad Hoc and Sens. Sys. (MASS)*, Oct. 2008, pp. 295-304.

[58] M. Iordanakis, D. Yannis, K. Karras, G. Bogdos, G. Dilintas, M. Amirfeiz, G. Colangelo, and S. Baiotti, "Ad-hoc routing protocol for aeronautical mobile ad-hoc networks," *in Proc. Int.Symp. on Commun. Sys., Net. and Digital Sig. Process. (CSNDSP)*, vol. 6, Jul. 2006.

[59] "Flight statistics," *Sivil Havacilik Genel Mudurlugu*, 2009. [Online]. Available: http://www.shgm.gov.tr

[60] Y. Wang, M. Erturk, H. Arslan, R. Sankar, and S. Morgera, "Throughput analysis in aeronautical data networks," *in Proc. IEEE Wireless and Microwave Technology Conf. (WAMICON)*, Clearwater, FL, Apr. 2011.

[61] Y. Wang, M. Erturk, H. Arslan, R. Sankar, I. Ra, and S. Morgera, "Throughput and Delay Analysis in Aeronautical Data Networks," *in Porc. International Conference on Computing, Networking and Communications (ICNC)*, Maui, HI, Jan. 2012.

[62] Y. Wang, M. Erturk, I. Ra, R. Sankar, and S. Morgera, "Throughput and Delay Analysis for Aeronautical Communication Networks," *to appear in International Journal of Research and Reviews in Wireless Sensor Networks (IJRRWSN)*, Aug. 2012.

[63] E. Haas, "Aeronautical channel modeling," *IEEE Trans. on Veh. Tech.*, vol. 51, no. 2, pp. 254-264, March 2002.

# APPENDICES

# Appendix A: Copyrights Permissions

## A.1 IEEE Copyright Permission for Use of Figures