

Topics in Modeling, Analysis and Optimisation of Wireless Networks

A Thesis

Submitted for the Degree of

Doctor of Philosophy

in the Faculty of Engineering

by

Venkatesh Ramaiyan



Electrical Communication Engineering
Indian Institute of Science, Bangalore
Bangalore – 560 012 (INDIA)

January 2009

*to my thatha,
late Mr. Rengarajulu Naidu*

Abstract

The work in this thesis is concerned with two complementary aspects of wireless networks research; performance analysis and resource optimization. The first part of the thesis focusses on the performance analysis of IEEE 802.11(e) wireless local area networks. We study the distributed coordination function (DCF) and the enhanced distributed channel access (EDCA) MAC of the IEEE 802.11(e) standard. We consider n IEEE 802.11(e) DCF (EDCA) nodes operating as a single cell; by single cell, we mean that every packet transmission can be heard by every other node. Packet loss is attributed only to simultaneous transmissions by the nodes (i.e., collisions). Using the well-known decoupling approximation [18], we characterize the collision behaviour and the throughput performance of the WLAN with a set of fixed point equations involving the backoff parameters of the nodes. We observe that the fixed point equations can have multiple solutions, and in such cases, the system exhibits multistability and short-term unfairness of throughput. Also, the fixed point analysis fails to characterize the average system behaviour when the system has multiple solutions. We then obtain sufficient conditions (in terms of the backoff parameters of the nodes) under which the fixed point equations have a unique solution. For such cases, using simulations, we observe that the fixed point analysis predicts the long term time average throughput behaviour accurately. Then, using the fixed point analysis, we study throughput differentiation provided by the different backoff parameters, including minimum contention window (CW_{min}), persistence factor and arbitration interframe space (AIFS) of the IEEE 802.11e standard. Finally, we extend the above results to the case where the receiver supports physical layer capture.

In the second part of the thesis, we study resource allocation and optimization problems for a variety of wireless network scenarios. For a dense wireless network, deployed over a small area and with a network average power constraint, we show that single cell operation (the channel supports only one successful transmission at any time) is throughput efficient in the asymptotic regime (in which the network average power is made large). We show that, for a realistic path loss model and a physical interference model (SINR based), the maximum aggregate bit rate among arbitrary transmitter-receiver pairs scales only as $\Theta(\log(\bar{P}))$, where \bar{P} is the network average power. Spatial reuse is ineffective and direct transmission between source-destination pairs is the throughput optimal strategy. Then, operating the network with only a single successful transmission permitted at a time, and with CSMA being used to select the successful transmitter-receiver pair, we consider the situation in which there is stationary spatio-temporal channel fading. We study the optimal hop length (routing strategy) and power control (for a fading channel) that maximizes the network aggregate throughput for a given network power constraint. For a fixed transmission time scheme, we study the throughput maximizing schedule under homogeneous traffic and MAC assumptions. We also characterize the optimal operating point (hop length and power control) in terms of the network power constraint and the channel fade distribution.

It is now well understood that in a multihop network, performance can be enhanced if, instead of just forwarding packets, the network nodes create output packets by judiciously combining their input packets, a strategy that is called “network coding.” For a two link slotted wireless network employing a network coding strategy and with fading channels, we study the optimal power control and optimal exploitation of network coding opportunities that minimizes the average power required to support a given arrival rate. We also study the optimal power-delay tradeoff for the network and show that the minimum average queueing delay scales as $\Omega\left(\frac{1}{v}\right)$ for an excess average power of $O(v)$.

Finally, we study a vehicular network problem, where vehicles are used as relays to transfer data between a pair of stationary source and destination nodes. The source node has a file to transfer to the destination node and we are interested in the delay minimizing

schedule for the vehicular network. We characterize the average queueing delay (at the source node) and the average transit delay of the packets (at the relay vehicles) in terms of the vehicular speeds and their interarrival times, and study the asymptotically optimal tradeoff achievable between them.

Contents

Abstract	ii
1 Introduction	1
I Performance Analysis of IEEE 802.11(e) WLANs	8
2 Fixed Point Analysis of Single Cell IEEE 802.11 DCF WLANs : Multi-stability and Fairness	9
2.1 Introduction	9
2.1.1 Comments on the Literature	11
2.1.2 Outline of the Chapter	11
2.2 The Generalized Backoff Model	12
2.3 The Fixed Point Equation	13
2.4 Nonunique Fixed Points and Multistability	16
2.4.1 Example 1	16
2.4.2 Example 2	21
2.4.3 Example 3	24
2.4.4 Short-term Fairness in Examples 1, 2 and 3	25
2.5 Uniqueness of the Fixed Point	27
2.6 Summary	29
2.7 Appendix	29
2.7.1 Proofs of Theorems and Lemmas	29

3	Fixed Point Analysis of Single Cell IEEE 802.11e EDCA WLANs	33
3.1	Introduction	33
3.1.1	A Survey of the Literature	34
3.1.2	Outline of the Chapter	35
3.2	The Nonhomogeneous Case	36
3.3	Analysis of the AIFS Mechanism	39
3.3.1	The Fixed Point Equations	39
3.3.2	Uniqueness of the Fixed Point	43
3.3.3	Numerical Study and Discussion	43
3.4	Multiple Access Categories per Node	47
3.4.1	Without AIFS	47
3.4.2	With AIFS	48
3.4.3	Numerical Study and Discussion	50
3.5	Throughput Differentiation: An Analytical Study	51
3.5.1	Case 1: Differentiation by b_0	52
3.5.2	Case 2: Differentiation by p	55
3.5.3	Case 3: Differentiation by AIFS	56
3.5.4	Numerical Study and Discussion	57
3.6	Summary	59
3.7	Appendix	60
3.7.1	Proofs of Theorems and Lemmas	60
4	Fixed Point Analysis of IEEE 802.11(e) WLANs with Capture	77
4.1	Introduction	77
4.1.1	Literature Survey	78
4.2	Fixed Point Framework with Capture	80
4.2.1	Validity of the Model	81
4.3	Uniqueness of the Fixed Point	87
4.3.1	Counterexample	88
4.3.2	A Contraction Mapping	90

4.3.3	Infrastructure Mode with Uplink Traffic	94
4.4	Summary	97
II Topics in Resource Optimization in Wireless Networks		98
5	Spatial Reuse and Cooperative Communication in Dense Wireless Networks	99
5.1	Introduction	99
5.1.1	Outline of the Chapter	101
5.2	Network Model and Assumptions	101
5.2.1	Path Loss Model	102
5.2.2	Objective	102
5.3	Scaling Laws for Large \bar{P}	103
5.3.1	An Upper Bound on the Network Throughput	103
5.3.2	The Optimization Problem	104
5.4	Spatial Reuse and Cooperative Communication	106
5.5	Summary	110
5.6	Appendix	110
5.6.1	Proofs of Theorems and Lemmas	110
6	Power Control and Routing for a Single Cell, Dense Wireless Network	114
6.1	Introduction	114
6.1.1	Preview of Contributions	116
6.1.2	Motivation for Single Cell Operation	117
6.1.3	Outline of the Chapter	117
6.2	Network Model and Assumptions	118
6.2.1	Channel Model: Path Loss, Fading and Transmission Rate	119
6.2.2	Fixed Transmission Time Strategy	120
6.3	Multihop Transport Capacity	121
6.4	Optimising the Transport Capacity	123

6.4.1	Optimization over $\{P(h)\}$ for a fixed d	124
6.4.2	Optimization over d	125
6.4.3	Characterisation of the Optimal d	127
6.5	Summary	131
6.6	Appendix	132
6.6.1	Stationary Points of $d \times \Gamma(\pi(d))$	132
6.6.2	Proof of Theorem 6.4.1	138
6.6.3	Discrete Fading States	143
6.6.4	Fixed Transmission Time vs Fixed Packet Size	147
7	Network Coding and Power Control for a Two Link Wireless Network	153
7.1	Introduction	153
7.1.1	Literature Survey	156
7.1.2	Outline of the Chapter	157
7.2	Network Model	157
7.2.1	Transmission Strategies and Rate Vectors	158
7.3	Capacity Region	161
7.3.1	Minimizing the Average Power Requirement	162
7.3.2	Optimal Power Allocation	163
7.3.3	Power Benefits of Network Coding	164
7.4	Power - Delay Tradeoff	165
7.5	Summary	169
7.6	Appendix	169
7.6.1	Delay is $\Omega\left(\frac{1}{v}\right)$ if Excess Power is $O(v)$	169
8	Delay Optimal Scheduling in a Two Hop Vehicular Relay Network	174
8.1	Introduction	174
8.1.1	A Survey of the Literature	177
8.1.2	Outline of the Chapter	178
8.2	The Network Model	178

8.3	The Optimization Problem	180
8.4	Finite File Size	181
8.4.1	One Shot Problem ($z = 1$)	184
8.4.2	Piecewise Transmission Problem ($z > 1$)	185
8.5	Infinite File Size	189
8.6	Summary	195
8.7	Appendix	195
8.7.1	Proof of Theorems and Lemmas	195
9	Conclusion	199
	Bibliography	203

Chapter 1

Introduction

Information is “key” to success in the modern world, and connectivity is essential to share information. From cellular telephony to satellite television, from the public Internet to ad hoc sensor networks, the nature of information and the network that carries the information varies greatly. The increasing user demands have lead to a variety of network architectures, from the simple point-to-point wired networks to the more complex distributed ad hoc wireless networks, each having evolved with a class of applications in mind.

Communication networks can be broadly classified into wired and wireless systems depending on the medium used as the channel. Recent advances in the areas of optical communication and switching have permitted overprovisioning as a viable solution for wired networks, to meet modern day traffic demands. However, wireless networks continue to be resource limited, since the available RF spectrum needs to be shared among several competing wireless devices. The increasing popularity of wireless networks, due to their ease of deployment and their support for mobility, has only increased the performance expectations from these systems. It has thus become increasingly important to understand the performance limitations of the various wireless network technologies, and to devise efficient resource allocation algorithms for them.

A fundamental understanding of the capabilities and the limitations of the wireless channel and the users is essential here. For example, when the nodes in the network

have an average or max power constraint, then the maximum bit rate (or throughput) achievable is bounded by the physics of the wireless channel. Similarly, when the nodes are not collocated, then we need to seek a distributed scheduling algorithm to operate the network; a centralized scheduler like a base station is not an option. Now, resource optimization involves the need to design such algorithms that achieve the various network performance objectives, given the spectrum and user constraints. Resource optimization helps us know the fundamental limits of the network, and also provides us the way to operate the network efficiently. For example, in [50], the authors study the capacity of large ad hoc wireless networks and show that the end-to-end user throughput scales as $O\left(\frac{1}{\sqrt{n}}\right)$, where n is the number of users in the network. Also, [50] suggests multihopping and spatial reuse to achieve the optimal throughput.

While multihopping is clearly the optimal strategy for large ad hoc wireless networks, it might still be practical to operate the networks in a single hop mode (direct transmission), due to its ease of operation and the economics of operating the network. In this sense, the way a practical network is operated might be very different from the optimal operation suggested by the analysis. Also, most analyses do not account for the overheads associated with operating the network (like interference management, routing overheads), which are specific to the standards and the protocols involved in the system. Performance modeling and analysis helps one understand the deployed networks, to ensure that the network is provisioned appropriately and that the users see the desired quality of service while using minimal resources. Also, system modeling and performance analysis is the first step in resource optimization and in the design of resource allocation algorithms. We see performance analysis and resource optimization as two facets of the same problem and we study them both in this thesis.

In this thesis, we study scheduling and resource allocation problems in a variety of wireless networks, including WiFi networks, ad hoc (and sensor) networks and vehicular networks. We consider a collection of wireless nodes whose channel access is determined either by a centralized scheduler (e.g., an access point or a base station) or by a distributed medium access protocol (e.g., the IEEE 802.11(e) MAC). The nodes are assumed to be

resource constrained, for example, the nodes could be average or max power limited. Our performance metrics of interest are the long term time average throughput and the long term time average queueing delay in the system. Our approach is to model the system performance in terms of the channel, MAC and node parameters, and then we analyse and optimize it to provide the required quality of service.

The work in this thesis can be categorized into two parts. The first part of the thesis focusses on the performance analysis of IEEE 802.11(e) WLANs. We study the distributed coordination function (DCF) and the enhanced distributed channel access (EDCA) MAC of the IEEE 802.11(e) standard. We consider n IEEE 802.11(e) DCF (EDCA) nodes operating as a single cell. By single cell, we mean that every packet transmission can be heard by every other node. Hence, the backoff engine and the transmission attempts of the nodes are synchronized within the cell. We assume that the nodes are saturated, i.e., the nodes always have packets to send, and we are interested in the long term time average throughput of the nodes. Using a decoupling approximation introduced by Bianchi in [18] for IEEE 802.11 DCF nodes and a generalization of the node response formula reported in [6], we characterize the throughput performance of the WLAN with a set of fixed point equations involving the average collision probabilities (γ) and the average attempt rates (β) of the nodes.

In Chapter 2, we are concerned with the saturation throughput analysis of single cell IEEE 802.11 DCF wireless local area networks. We assume a *pure collision* channel (without capture, fading or frame error) in which packets are lost only due to collision of simultaneous transmissions. We study the vector fixed point equations (of the collision probabilities of the nodes) arising out of the analysis of the saturation throughput of a single cell IEEE 802.11 DCF WLAN. We consider both the balanced and unbalanced solutions of the fixed point equations arising in the WLAN; we say that a fixed point is balanced, when all the coordinates of the fixed point are equal. We are concerned, in particular, with (i) whether the fixed point is balanced, and (ii) whether the fixed point is unique. Our simulations show that when multiple unbalanced fixed points exist, then the time behaviour of the system demonstrates severe short-term unfairness (or

multistability). Also, in such cases, the balanced fixed point does not represent the actual system performance accurately. Implications for the use of the fixed point formulation for performance analysis are also discussed. Finally, we provide a condition for the fixed point solution to be balanced, and also a condition for uniqueness. We have shown that the default IEEE 802.11 parameters satisfy these sufficient conditions.

IEEE 802.11e defines enhanced distributed channel access (EDCA) MAC to provide quality of service (QoS) to competing nodes. Throughput differentiation is provided by defining different traffic classes and by allowing different channel access parameters, including different minimum and maximum contention windows (CW_{min}, CW_{max}) and arbitration interframe space (AIFS) values. In Chapter 3, we study the saturation throughput performance of IEEE 802.11e EDCA WLANs with nonhomogeneous backoff parameters (including AIFS and multiple queues per node). Here again, we assume a single cell model with a *pure collision* channel. We obtain the fixed point equations characterizing the system performance by modeling AIFS based differentiation and the concept of virtual collision (when there are multiple traffic classes per node). Like in Chapter 2, we provide a condition for the fixed point equations to have a unique solution. We also show that the default IEEE 802.11e EDCA parameters satisfy the uniqueness conditions. The fixed point analysis is then used to obtain insights into the throughput differentiation provided by different initial backoffs, persistence factor, and AIFS, for finite number of nodes and for differentiation parameter values similar to those in the IEEE 802.11e standard. An asymptotic analysis of the fixed point is also provided for the case in which packets are never abandoned, and the number of nodes goes to infinity. Simulation results have been provided to validate the accuracy of the analysis.

Chapters 2 and 3 assume a *pure collision* channel model with no fading, frame error or capture. In Chapter 4, we extend the saturation throughput analysis of single cell IEEE 802.11 and 802.11e WLANs to channels with frame capture at the receiver. Most analyses of single cell WLANs assume that, when nodes attempt simultaneously in a slot, collision occurs and all the attempts fail. However, in practice, it is possible that the receiver can decode a signal with sufficient strength even in the presence of interfering transmissions,

i.e., the receiver can capture a frame. Hence, we allow frame capture at the receiver and develop a general framework to study saturation throughput of single cell WLANs. We first discuss the network scenarios which can be modeled using our fixed point framework. Using simulations, we then show that capture can introduce multistability in single cell scenarios (even when the sufficient conditions of uniqueness for a *pure collision* channel hold). We observe that the system of equations characterizing such WLANs can have multiple fixed points. Then, we obtain some sufficient conditions to guarantee a unique solution to the fixed point equations. Finally, we prove a general uniqueness result for the infrastructure setup with uplink traffic.

In the second part of the thesis, we study resource allocation and optimization problems for a variety of wireless network scenarios. In Chapter 5, we consider a dense ad hoc wireless network comprising n nodes confined to a given two dimensional region of fixed, finite area A . We assume a realistic interference model (SINR based) and path loss model (the channel gains are bounded above by 1, and are bounded below by a strictly positive number, because of the finite area assumption). For the Gupta-Kumar random traffic model [50], we study the scaling of the aggregate end-to-end throughput with respect to the network average power constraint, \bar{P} , and the number of nodes, n . The network power constraint \bar{P} is related to the per node power constraint, p , as $\bar{P} = np$. For large \bar{P} , we show that the throughput saturates as $\Theta(\log(\bar{P}))$, irrespective of the number of nodes in the network. We observe that single cell operation is optimal for large network power constraints and direct transmission between the source-destination pair is the throughput optimal strategy. For moderate \bar{P} , which can accommodate spatial reuse to improve end-to-end throughput, we observe that the amount of spatial reuse feasible in the network is limited by the diameter of the network. In fact, we observe that the end-to-end path loss in the network and the amount of spatial reuse feasible in the network are inversely proportional. This puts a restriction on the gains achievable using the cooperative communication techniques studied in [60] and [9], as these rely on direct long distance communication over the network.

Motivated by the results in Chapter 5 for large power networks, in Chapter 6, we study

a throughput maximization problem in an average power constrained, dense, ad hoc wireless network operated as a single cell. Data packets are sent between source-destination pairs by multihop relaying. We assume that nodes self-organize into a multihop network such that all hops are of length d meters, where d is a design parameter. There is a contention based multiaccess scheme, and it is assumed that every node always has data to send, either originated from it or a transit packet (saturation assumption). In this scenario, we seek to maximize a measure of the transport capacity of the network (measured in bit-meters per second) over power controls (in a fading environment) and over the hop distance d , subject to a network average power constraint. More specifically, for a fading channel and for a fixed transmission time strategy (akin to the IEEE 802.11 TXOP), we find that there exists an intrinsic aggregate bit rate (Θ_{opt} bits per second, depending on the contention mechanism and the channel fading characteristics) carried by the network, when operating at the optimal hop length and power control. The optimal transport capacity is of the form $d_{opt}(\bar{P}_t) \times \Theta_{opt}$ with d_{opt} scaling as $\bar{P}_t^{\frac{1}{\eta}}$, where \bar{P}_t is the available time average transmit power and η is the path loss exponent. Under certain conditions on the fading distribution, we then provide a simple characterization of the optimal operating point.

In Chapter 7, we study an optimal power allocation and network coding strategy for a two link wireless network. We consider a two link slotted wireless network carrying bidirectional traffic. Packets originate at the terminal nodes, and are routed via the central node to their destinations. All the nodes have sufficient buffer space and the packets are queued until transmission (i.e., there are no packet loss due to buffer overflow). We assume that the central node can network code packets belonging to the two routes (of the bidirectional traffic) and broadcast the coded packet simultaneously to the terminal nodes. In this scenario, we study the optimal power allocation and network coding strategy that stabilizes the queues at the central node. Using the framework developed in [44], we obtain the power allocation policy and the network coding strategy that minimizes the average power required to support a given packet arrival rate. Then, we study the asymptotic tradeoff achievable between the average transmission power and the average queueing

delay for the network.

In Chapter 8, we study a scheduling problem in a wireless network where vehicles are used as store-and-forward relays. A fixed source node wants to transfer a file to a fixed destination node, located beyond its communication range. In the absence of any infrastructure connecting the two nodes, we consider the possibility of communication using vehicles passing by. Vehicles arrive at the source node at renewal instants and are known to travel towards the destination node with speed v sampled from a given probability distribution. The source node communicates packets of the file to the destination node using these vehicles as relays. We assume that the vehicles communicate with the source node and the destination node only, and hence, every packet communication involves two hops. In this setup, we study the source node's sequential decision problem of transferring packets of the file to vehicles as they pass by, with the objective of minimizing the average delay in the network. We study both the finite file size case and the infinite file size case. In the finite file size case, using a Markov decision process framework, we study the expected total delay minimization problem. In the infinite file size case, we study the optimal tradeoff achievable between the average queueing delay at the source node buffer and the average transit delay in the relay vehicle.

Part I

Performance Analysis of IEEE 802.11(e) WLANs

Chapter 2

Fixed Point Analysis of Single Cell IEEE 802.11 DCF WLANs : Multistability and Fairness

2.1 Introduction

This chapter is concerned with the saturation throughput analysis of single cell IEEE 802.11 DCF wireless local area networks. We consider a single cell WLAN; single cell meaning that all nodes are within control channel range of each other and every packet transmission can be heard by every other node. We assume a *pure collision* channel (without capture, fading or frame error) and packets are lost only due to collision of simultaneous transmissions. There can be only one successful transmission in the channel at any time, and the network does not support spatial reuse. IEEE 802.11 standard [1] defines a CSMA/CA based distributed medium access control (MAC) protocol, called the distributed coordination function (DCF). DCF permits nodes to have a single queue each and all have identical backoff parameters (which governs the channel access). We are interested in the expected throughput performance of the DCF WLAN when the nodes always have packet to transmit (i.e., saturation assumption).

Much work has been reported on the performance analysis of the DCF MAC. Most

of the analytical work reported has been based on a decoupling approximation proposed initially by Bianchi in [18]. While keeping this basic decoupling approximation, in [6], Kumar et al. presented a significant simplification and generalization of the analysis of the IEEE 802.11 backoff mechanism. Their analysis led to a certain one dimensional fixed point equation for the collision probability experienced by the nodes.

In this chapter, we study the vector fixed point equations (of the collision probabilities of the nodes) arising out of the analysis of the saturation throughput of a single cell IEEE 802.11 DCF WLAN. We consider both the balanced and the unbalanced solutions of the fixed point equations arising in the WLAN; we say that a fixed point is balanced, when all the coordinates of the fixed point are equal. We are concerned, in particular, with (i) whether the fixed point is balanced, and (ii) whether the fixed point is unique. Our simulations show that when multiple unbalanced fixed points exist, then the time behaviour of the system demonstrates severe short-term unfairness (or *multistability*). Also, in such cases, the balanced fixed point does not represent the actual system performance accurately. Implications for the use of the fixed point formulation for performance analysis are also discussed. Finally, we provide a condition for the fixed point solution to be balanced, and also a condition for uniqueness.

Our approach in this chapter builds upon the one provided in [6]. The main contributions of this chapter are the following:

1. We provide examples in which, even though a unique balanced fixed point exists, there can be multiple unbalanced fixed points, thus suggesting *multistability*. We demonstrate by simulation that, in such cases, significant short-term unfairness can be observed and the unique balanced fixed point fails to capture the system performance.
2. Next, in the case where the backoff increases multiplicatively (as in IEEE 802.11 DCF), we establish a simple sufficient condition for the uniqueness of the solution of the multidimensional fixed point equation.

2.1.1 Comments on the Literature

There has been much research activity on modeling the performance of IEEE 802.11 DCF WLANs. The general approach has been to use the decoupling approximation and the constant collision probability assumption introduced by Bianchi in [18] and characterize the system performance using a Markov chain. The decoupling approximation is to assume that the aggregate attempt process of the competing nodes is independent of the backoff state of the tagged node. By the constant collision probability assumption, we mean that the aggregate behaviour of the competing nodes is characterized by a constant collision probability in every slot. In this chapter, in Section 2.3, we will provide a generalization and simplification of this approach. In the previous literature, it is assumed that the collision rate experienced by any node is constant over time. There appears to have been no attempt to study the phenomenon of short-term unfairness in the fixed point framework. A related work is [40] which identifies short-term unfairness in Ethernet (using experimentation and simulation), and suggests modifications to the protocol to eliminate it. Also, all the existing works assume that the collision probabilities of all the nodes (in an IEEE 802.11 DCF WLAN) are the same. Thus there appears to have been no earlier work on studying the possibility of unbalanced solutions of the fixed point equations. In addition, the possibility of nonuniqueness of the solution of the fixed point equations arising in the analyses seems to have been missed in the earlier literature. In this chapter, we study the fixed point equations of IEEE 802.11 DCF WLANs and take into account all these possibilities.

2.1.2 Outline of the Chapter

In Section 2.2 we review the generalized backoff model that was first presented in [6]. In Section 2.3 we develop the multidimensional fixed point equations for IEEE 802.11 DCF WLANs and obtain the necessary and sufficient conditions satisfied by the solutions to the fixed point equations. We provide examples in Section 2.4 to show that there can exist multiple unbalanced fixed points and show the consequence of this. In Section 2.5, we analyse the fixed point equations and obtain a condition for the existence of only one

fixed point. Section 2.6 summarizes the results in this chapter.

2.2 The Generalized Backoff Model

There are n nodes, indexed by $i, 1 \leq i \leq n$. We adopt the notation in [6], whose authors consider a generalization of the backoff behaviour of the nodes, and define the following backoff parameters (for node i)

$K_i :=$ At the $(K_i + 1)$ th attempt either the packet being attempted by node i succeeds or is discarded

$b_{i,k} :=$ The *mean* backoff (in slots) at the k th attempt for a packet being attempted by node $i, 0 \leq k \leq K_i$

Definition 2.2.1 *A system of n nodes is said to be **homogeneous**, if all the backoff parameters of the nodes, like, $K_i, b_{i,k}, 0 \leq k \leq K_i$ are the same for all $i, 1 \leq i \leq n$. A system of nodes is called **nonhomogeneous** if the backoff parameters of the nodes are not identical.* ■

Remark: IEEE 802.11e permits different backoff parameters to differentiate channel access obtained by the nodes in an attempt to provide QoS. The above definitions capture the possibility of having different CW_{min} and CW_{max} values, different exponential backoff multiplier values and even different number of permitted attempts. In this chapter, we consider IEEE 802.11 DCF WLANs whose nodes have identical backoff parameters and hence, are homogeneous. ■

It has been shown in [6] (and later in [20]) that under the decoupling assumption, introduced by Bianchi in [18], the attempt probability of node i (in a backoff slot, and conditioned on being in backoff) for given collision probability γ_i is given by,

$$G_i(\gamma_i) := \frac{1 + \gamma_i + \cdots + \gamma_i^{K_i}}{b_{i,0} + \gamma_i b_{i,1} + \cdots + \gamma_i^{K_i} b_{i,K_i}} \quad (2.1)$$

Remarks 2.2.1

1. We will assume that $b_{i,\cdot}$ are such that $0 \leq G_i(\gamma_i) \leq 1$ for all $\gamma_i, 0 \leq \gamma_i \leq 1$ and $G_i(\gamma_i) < 1$ whenever $\gamma_i > 0$.
2. When the system is homogeneous (e.g., IEEE 802.11 DCF nodes) then we will drop the subscript i from $G_i(\cdot)$, and write the function simply as $G(\cdot)$.

2.3 The Fixed Point Equation

It is important to note that in the present discussion all rates are conditioned on being in the backoff periods; i.e., we have eliminated all durations other than those in which nodes are counting down their backoff counters, in order to obtain the collision probability γ_i of node i and its attempt probability $\beta_i (= G_i(\gamma_i))$. Later one brings back the channel activity periods in order to compute the throughput in terms of the attempt probabilities (see [6]). Now consider a homogeneous system of n nodes. Let $\boldsymbol{\gamma}$ be the vector of collision probabilities of the nodes. With the slotted model for the backoff process and the decoupling assumption, the natural mapping of the attempt probabilities of other nodes to the collision probability of a node is given by

$$\gamma_i := \Gamma_i(\beta_1, \beta_2, \dots, \beta_n) = 1 - \prod_{j=1, j \neq i}^n (1 - \beta_j)$$

where $\beta_j = G_j(\gamma_j)$. We could now expect that the equilibrium behaviour of the system will be characterized by the solutions of the following system of equations, for $1 \leq i \leq n$,

$$\gamma_i = \Gamma_i(G_1(\gamma_1), \dots, G_n(\gamma_n))$$

We write these n equations compactly in the form of the following multidimensional fixed point equation.

$$\boldsymbol{\gamma} = \mathbf{\Gamma}(\mathbf{G}(\boldsymbol{\gamma})) \tag{2.2}$$

Since $\mathbf{\Gamma}(\mathbf{G}(\boldsymbol{\gamma}))$ is a composition of continuous functions it is continuous. We thus have a continuous mapping from $[0, 1]^n$ to $[0, 1]^n$. Hence by Brouwer's fixed point theorem there exists a fixed point in $[0, 1]^n$ for the equation $\boldsymbol{\gamma} = \mathbf{\Gamma}(\mathbf{G}(\boldsymbol{\gamma}))$.

Consider the i^{th} component of the fixed point equation, i.e.,

$$\gamma_i = 1 - \prod_{1 \leq j \leq n, j \neq i} (1 - G_j(\gamma_j))$$

or equivalently,

$$(1 - \gamma_i) = \prod_{1 \leq j \leq n, j \neq i} (1 - G_j(\gamma_j))$$

Multiplying both sides by $(1 - G_i(\gamma_i))$, we get,

$$(1 - \gamma_i)(1 - G_i(\gamma_i)) = \prod_{1 \leq j \leq n} (1 - G_j(\gamma_j))$$

Thus a *necessary and sufficient condition* for a vector of collision probabilities $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ to be a fixed point solution is that, for all $1 \leq i \leq n$,

$$(1 - \gamma_i)(1 - G_i(\gamma_i)) = \prod_{j=1}^n (1 - G_j(\gamma_j)) \quad (2.3)$$

where the right-hand side is seen to be independent of i .

Define $F_i(\boldsymbol{\gamma}) := (1 - \gamma_i)(1 - G_i(\boldsymbol{\gamma}))$. From (2.3) we see that if $\boldsymbol{\gamma}$ is a solution of (2.2), then for all $i, j, 1 \leq i, j \leq n$,

$$F_i(\boldsymbol{\gamma}_i) = F_j(\boldsymbol{\gamma}_j) \quad (2.4)$$

Notice that this is only a *necessary condition*. For example, in a homogeneous system of nodes (with $G_i(\cdot) = G(\cdot)$ and hence, $F_i(\cdot) = F(\cdot)$), the vector $\boldsymbol{\gamma}$ such that $\gamma_i = \gamma$ for all $1 \leq i \leq n$, satisfies (2.4) for any $\gamma, 0 \leq \gamma \leq 1$, but not all such points are solutions of the fixed point equation (2.2).

Definition 2.3.1 *We say that a fixed point γ (i.e., a solution of $\gamma = \mathbf{\Gamma}(\mathbf{G}(\gamma))$) is a **balanced fixed point** if $\gamma_i = \gamma_j$ for all $1 \leq i, j \leq n$; otherwise, γ is said to be an **unbalanced fixed point**. ■*

Remarks 2.3.1

1. It is clear that if there exists an unbalanced fixed point for a homogeneous system (like IEEE 802.11 DCF WLAN), then every permutation is also a fixed point and hence, in such cases, we do not have a unique fixed point.
2. In the homogeneous case (like IEEE 802.11 DCF WLAN), by symmetry, the average collision probability must be the same for every node. If the collision probabilities correspond to a fixed point (see 3, next), then this fixed point will be of the form $(\gamma, \gamma, \dots, \gamma)$ where γ solves $\gamma = \Gamma(G(\gamma)) = 1 - (1 - G(\gamma))^{n-1}$ (since $\Gamma_i(\cdot) = \Gamma(\cdot)$ and $G_i(\cdot) = G(\cdot)$ for all $1 \leq i \leq n$). Such a fixed point of $\gamma = \Gamma(G(\gamma))$ is guaranteed by Brouwer's Fixed Point. The uniqueness of such a balanced fixed point was studied in [6]. We reproduce this result in Theorem 2.5.1.
3. There is, however, the possibility that even in the homogeneous case, there is an unbalanced solution of $\gamma = \mathbf{\Gamma}(\mathbf{G}(\gamma))$. By simulation examples we observe in Section 2.4 that when there exist unbalanced fixed points, the balanced fixed point of the system does not characterize the average performance, even if there exists only one balanced fixed point. In Section 2.5, we provide a condition for homogeneous IEEE 802.11 type nodes (with exponential backoff) under which there is a unique balanced fixed point and no unbalanced fixed point. In such cases, it is now well established, that the unique balanced fixed point accurately predicts the saturation throughput of the system.
4. For the homogeneous case the backoff process can be exactly modeled by a positive recurrent Markov chain (see [6]). Hence the attempt process and the collision processes will be ergodic and, by symmetry, the nodes will have equal attempt and collision probabilities. In such a situation the existence of multiple unbalanced

fixed points will suggest short-term unfairness or multistability. We will observe this phenomenon in Section 2.4.

5. Consider a system of homogeneous nodes having unbalanced solutions for the fixed point equation $\gamma = \mathbf{\Gamma}(\mathbf{G}(\gamma))$ (i.e., there exists i, j such that $\gamma_i \neq \gamma_j$), then from (2.4), we see that $F(\gamma_i) = F(\gamma_j)$, or the function F is many-to-one. Hence for a homogeneous system of nodes, if the function F is one-to-one then there cannot exist unbalanced fixed points.

2.4 Nonunique Fixed Points and Multistability

In this section, we will show (using simulation examples) that the fixed point equations of a homogeneous system can have multiple solutions. We observe that such systems exhibit significant short-term unfairness in the throughput received and the balanced fixed point does not represent the average system performance.

2.4.1 Example 1

Consider a homogeneous system (let us call it System-I) with $n = 10$ nodes. The function $G(\cdot)$ of the nodes is given by,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \gamma^3 + \dots}{1 + \gamma + \gamma^2 + \gamma^3 + 64(\gamma^4 + \gamma^5 + \dots)}$$

The system corresponds to the case where $K = \infty$, $b_0 = b_1 = b_2 = b_3 = 1$ and $b_4 = b_5 = b_6 = \dots = 64$ (b_i are distributed uniformly over the integers in $[1, CW_i]$ for appropriate CW_i). From the form of function $G(\cdot)$, we can see that a node which is currently at backoff stage 0 is more likely to remain at that stage as it *takes 4 successive collisions* to make the attempt rate of the node < 1 . Likewise, a node that is in the larger backoff stages $b_4 = b_5 = \dots = 64$, will retry continuously with mean inter-attempt slots of 64 until it succeeds. Observe that only one node can be at backoff stage 0 at any time. This leads to the apparent multistability of the system.

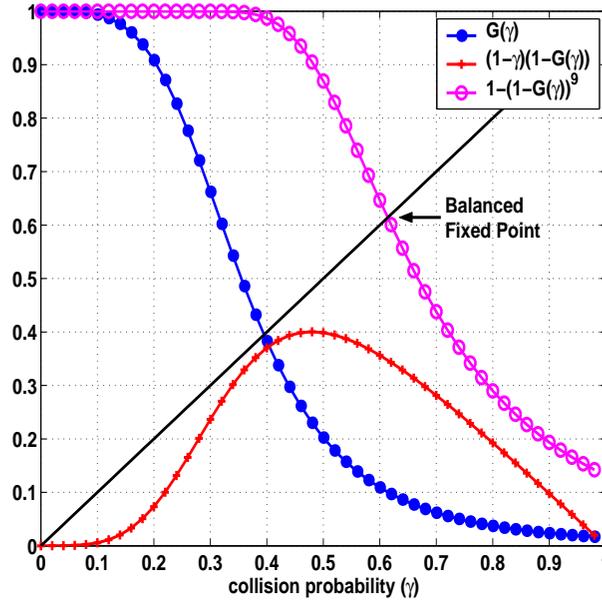


Figure 2.1: Example System-I: The balanced fixed point. Plots of $G(\gamma)$, $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$ and $1 - (1 - G(\gamma))^9$ vs. the collision probability γ ; we also show the “ $y=x$ ” line.

Figure 2.1 plots $G(\gamma)$, the corresponding $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$ and shows the balanced fixed point of the system for $n = 10$ nodes. The balanced fixed point of the system shown in the figure is obtained using the fixed point equation $\gamma = 1 - (1 - G(\gamma))^9$. Observe that the function $F(\cdot)$ is not one-to-one (the function $F(\cdot)$ not being one-to-one does not imply that there exist multiple fixed point solutions; see Remarks 2.3.1, 5).

Figure 2.2 shows the existence of unbalanced fixed points for System-I. These fixed points are obtained as follows. Assume that we are interested in fixed points such that $\gamma_1 \neq \gamma_2 = \dots = \gamma_n$. Given $\gamma_2 = \dots = \gamma_n$, the attempt probability of the nodes $2, \dots, n$ is given by $G(\gamma_2)$. Hence, the collision probability of node 1 is given by $\gamma_1 = 1 - (1 - G(\gamma_2))^{n-1}$. The attempt probability of node 1 would then be $G(\gamma_1)$. Using the decoupling assumption, the collision probability of any of the other $n - 1$ nodes would then be, $1 - (1 - G(\gamma_2))^{n-2}(1 - G(\gamma_1)) = \gamma_2$. Thus we obtain a fixed point equation for γ_2 (and hence for all the other $\gamma_j, 3 \leq j \leq n$). In Figure 2.2 we plot $1 - (1 - G(\gamma))^8(1 - G(1 - (1 - G(\gamma))^9))$ (plotted as the line marked with dots), the intersection of which with the “ $y=x$ ” line shows the solutions for $\gamma_2 (= \dots = \gamma_n)$. In the same way, we obtain the fixed point equation for γ_1 by eliminating $\gamma_2, \dots, \gamma_n$ from the multidimensional system of equations. This function

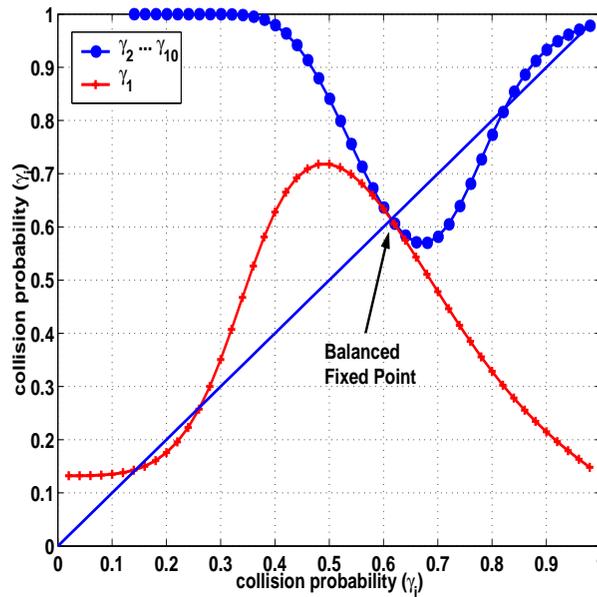


Figure 2.2: Example System-I: Demonstration of unbalanced fixed points. Plots of $\gamma_2 = 1 - (1 - G(\gamma))^8(1 - G(1 - (1 - G(\gamma))^9))$ (the curve drawn with dots and lines) and the expression for the fixed point equation for γ_1 (see text) using pluses and lines.

is plotted in Figure 2.2 using pluses and lines and the intersection of this curve with the “ $y=x$ ” line shows the corresponding solutions for γ_1 . We see that there are three solutions in each case. The smallest values of γ_1 (approx. 0.14) pairs up with the largest value of $\gamma_2 = \dots = \gamma_n$ (approx. 0.97). Notice that the balanced fixed point of the system is also a fixed point in the plot (compare with Figure 2.1). Then there is one remaining unbalanced fixed point whose values can be read off the plot. We note that there could exist many other unbalanced fixed points for this system of equations, as we have considered only a particular variety of fixed points that have the property that $\gamma_1 \neq \gamma_2 = \dots = \gamma_n$.

In order to examine the consequences of multiple unbalanced fixed points we simulated the backoff process with the backoff parameters of System-I. The following remarks summarise our simulation approach.

Remarks 2.4.1 (On the Simulation Approach used)

1. We have developed an event-driven simulator written in the “C” language based on the coupled multidimensional backoff process of the various nodes, to compare with the analytical results. In this simulator, we do not simulate the detailed wireless

LAN system (as is done in an ns-2 simulator), but only the backoff slots. We will refer to this as the CMP (Coupled Markov Process) simulator. The main aim of the CMP simulator is to understand the backoff behaviour of the nodes and its dependence on the different backoff parameters. From the point of view of performance analysis, it may also be noted that once the backoff behaviour is correctly modelled the channel activity can easily be added analytically, and thus throughput results can be obtained (see [18] and [6]). In addition, for some cases, ns-2 simulations have also been provided in comparison with the CMP simulator and the analytical results.

2. Our simulation is programmed as follows. The system evolves over backoff slots. All the nodes are assumed to be in perfect slot synchronization. The actual coupled evolution of the backoff process is modeled. The backoff values are chosen from the uniform distribution and the backoff stage of the node and the residual backoff counter value is the state for each node. At every slot, depending on the state of the backoff process, there are three possibilities: the slot is idle, there is a successful transmission, or there is a collision. This causes further evolution of the backoff process.
3. Our simulation approach, which we primarily use to study the backoff behaviour of the nodes, takes few seconds to complete a simulation run, in comparison with the ns-2 simulations which takes any time between few minutes to an hour depending on the number of nodes in the system. The coupled backoff evolution approach we use captures all the essential features of a single cell system where there is perfect synchronization among the nodes. The simulation provides the attempt rates and collision probabilities directly, which can be used with the throughput formula provided in [6] to obtain the throughput of the nodes.
4. In all our simulations, b_i are distributed uniformly over the integers in $[1, CW_i]$ for appropriate CW_i .
5. In Figures 2.3, 2.6 and 2.8, for the purpose of reporting the short-term unfairness

results, the entire duration of simulation is divided into k frames, where the size of each frame is 10,000 slots. The short-term average of the collision probability of each node j , $1 \leq j \leq n$, is calculated as $\frac{C_j(i)}{A_j(i)}$ where $C_j(i)$ and $A_j(i)$ correspond to the number of collisions and attempts in frame i , $1 \leq i \leq k$, for node j . The long-term average is similarly calculated as $\frac{1}{n} \sum_{j=1}^n \frac{\sum_{i=1}^k C_j(i)}{\sum_{i=1}^k A_j(i)}$ where n is the number of nodes. Notice that the long-term average collision rate is a batch biased average of the short-term collision rates. Hence, when looking at the graphs, it will be incorrect to visually average the short-term collision rate plots in an attempt to obtain the long-term average collision rate. This is because when a node is shown to have a low collision probability, it is the one that is attempting every slot (while the other nodes attempt with a mean gap of 64 slots), and hence it sees a low probability of collision. In this case $A_j(\cdot)$ is large and $C_j(\cdot) \ll A_j(\cdot)$. On the other hand, when a node is shown to have a high collision probability it is attempting at an average rate of $\frac{1}{64}$ and almost all its attempts collide with the node that is then attempting in every slot. In this case $A_j(\cdot)$ is small and $C_j(\cdot) \approx 1$. Thus, in obtaining the overall average, it is essential to account for the large variation in $A_j(\cdot)$ between the two cases. ■

In Figure 2.3 we plot a (simulation) snap shot of the short-term time average collision rate of two of the 10 nodes of System-I and the average collision probability of the nodes (The average is calculated over all frames and all nodes. Since the nodes are identical, the average collision probability is the same for all the nodes). Observe that the short-term average has a huge variance around the long term average. It is evident that over 1000's of slots one node or the other monopolises the channel (and the remaining nodes see a collision probability of 1 during those slots). This could be described as multistability. A look into the fairness index (see Figure 2.9) plotted as a function of the frame size used to calculate throughput suggests that System-I exhibits significant unfairness in service even over reasonably large time intervals.

Implication for the use of the balanced fixed point: Notice also that the average collision rate shown in Figure 2.3 is about 0.25, whereas the balanced fixed point shown in

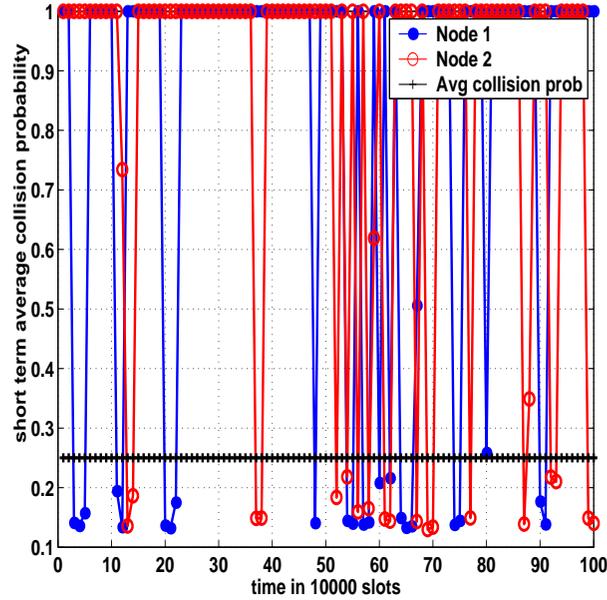


Figure 2.3: Example System-I: Snap-shot of short-term average collision rate of 2 of the 10 nodes. Also plotted is the average collision probability of the nodes (averaged *over all frames and nodes*). The 95% confidence interval for the average collision probability lies within 0.7% of the mean value.

Figure 2.1 shows a collision probability of about 0.62. *Hence we see that in this case, where there are multiple fixed points, the balanced fixed point does not capture the actual system performance.*

2.4.2 Example 2

Let us now consider yet another homogeneous example (let us call it System-II) with $n = 20$ nodes. The function $G(\cdot)$ of the nodes is given by,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \dots + \gamma^7}{1 + 3\gamma + 9\gamma^2 + 27\gamma^3 + \dots + 2187\gamma^7}$$

The system corresponds to the case where $K = 7$, $b_0 = 1$, $p = 3$ and $b_k = p^k b_0$ for all $0 \leq k \leq K$ (b_i are uniformly distributed in $[1, CW_i]$ for appropriate CW_i). We notice that in this example the way the backoff expands is similar to the way it expands in the IEEE 802.11 standard, except that the initial backoff is very small (1 slot) and the multiplier is 3, rather than 2. Figure 2.4 plots $G(\gamma)$, the corresponding $F(\gamma) =$

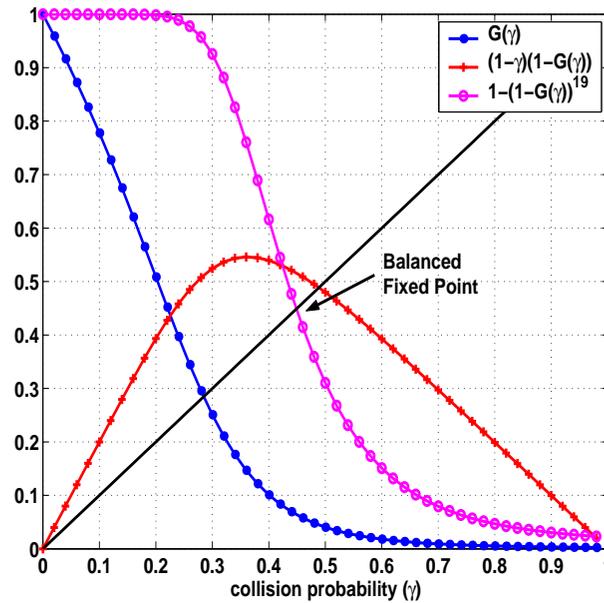


Figure 2.4: Example System-II: The balanced fixed point. Plots of $G(\gamma)$, $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$ and $1 - (1 - G(\gamma))^{19}$ vs. the collision probability γ ; the line “ $y=x$ ” is also shown. Notice that the function F is not one-to-one.

$(1 - \gamma)(1 - G(\gamma))$ and the balanced fixed point of the system for $n = 20$ nodes. The balanced fixed point of the system shown in the figure is obtained using the fixed point equation $\gamma = 1 - (1 - G(\gamma))^{19}$.

As in the case of System-I, Figure 2.5 shows the existence of multiple unbalanced fixed points for System-II. The fixed points we have shown correspond to the case where $\gamma_1 \neq \gamma_2 = \dots = \gamma_n$ and are obtained just as discussed for System-I.

Figure 2.6 plots a snap shot of the short-term average collision probability (from simulation) of two of the 20 nodes and the average collision probability of the nodes (same for all the nodes). Observe that the short-term averages vary a lot as compared to the long term average, suggesting multistability. Again, as in the case of System-I, comparing the average collision probability with the balanced fixed point of the system in Figure 2.4, we see that the fixed point does not capture the actual system performance.

Discussion of Examples 1 and 2: From the simulation examples, we can make the following inferences.

1. When there are multiple unbalanced fixed points in a homogeneous system then the

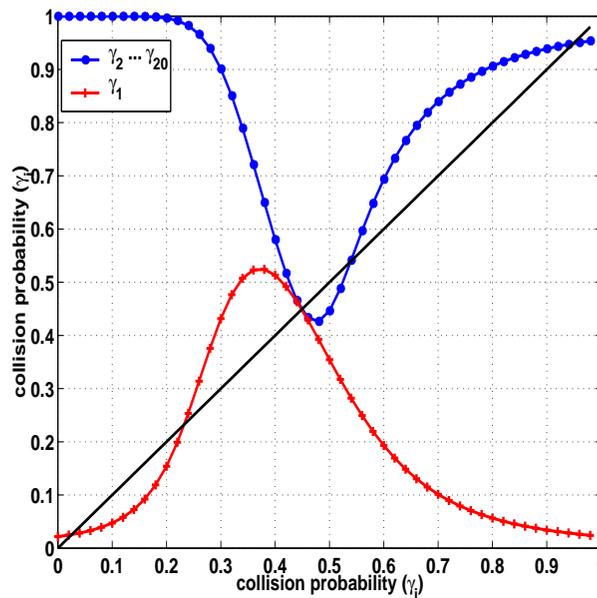


Figure 2.5: Example System-II: Demonstration of unbalanced fixed points. Plots of $\gamma_2 = 1 - (1 - G(\gamma))^{18}(1 - G(1 - (1 - G(\gamma))^{19}))$ (the curve drawn with dots and lines) and the function for the fixed point equation for γ_1 (see text) using pluses and lines.

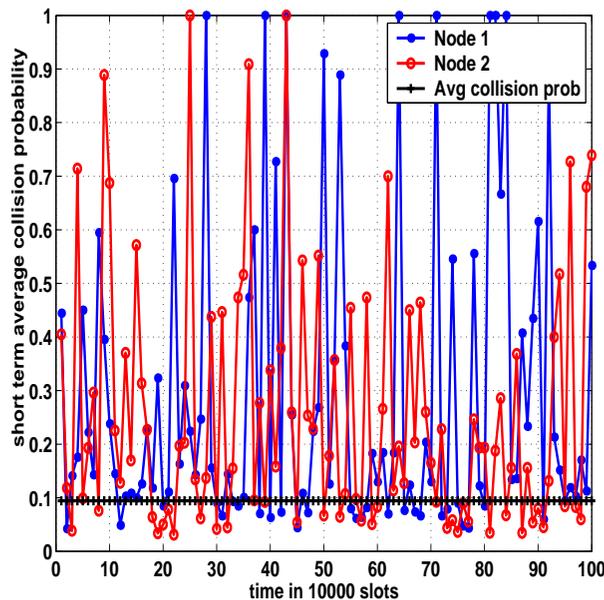


Figure 2.6: Example System-II: Snap-shot of short-term average collision probability of 2 of the 20 nodes. The average collision probability is also plotted in the figure (averaged over all slots and nodes). The 95% confidence interval for the average collision rate lies within 0.7% of the mean value.

system can display multistability, which manifests itself as significant short-term unfairness in channel access.

2. When there are multiple unbalanced fixed points in a homogeneous system then the collision probability obtained from the balanced fixed point may be a poor approximation to the long term average collision probability. ■

It appears that the existence of multiple-fixed points is a consequence of the form of the $G(\cdot)$ function in the above examples, where $G(\cdot)$ is similar to a switching curve; see, for example, Figure 2.1 where there is a very high attempt probability at low collision probabilities and a very low attempt probability at high collision probabilities.

2.4.3 Example 3

Consider a homogeneous system, with $n = 10$ nodes, in which the backoff increases multiplicatively as in IEEE 802.11 DCF (let us call it System-III). The function $G(\cdot)$ is given by,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \dots + \gamma^7}{16 + 32\gamma + 64\gamma^2 + \dots + 2048\gamma^7}$$

The system corresponds to the case where $K = 7$, $p = 2$ and $b_0 = 16$ and $b_k = p^k b_0$ for all $0 \leq k \leq K$ (b_i are uniformly distributed in $[1, CW_i]$ for appropriate CW_i). These parameters are similar to those used in the IEEE 802.11 standard. Figure 2.7 plots $G(\cdot)$, the corresponding $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$ and the unique balanced fixed point of the system. (Notice that F is one-to-one and uniqueness of the fixed point will be proved in Section 2.5.) The balanced fixed point of the system is obtained using the fixed point equation $\gamma = 1 - (1 - G(\gamma))^9$. The balanced fixed point yields a collision probability of approximately 0.29.

Figure 2.8 plots a snap shot of the short-term average collision probability (from simulation) of two of the 10 nodes and the average collision probability of the nodes of the Example System-III. Notice that the short-term average collision rate is close to the average collision rate (the vertical scale in this figure is much finer than in the

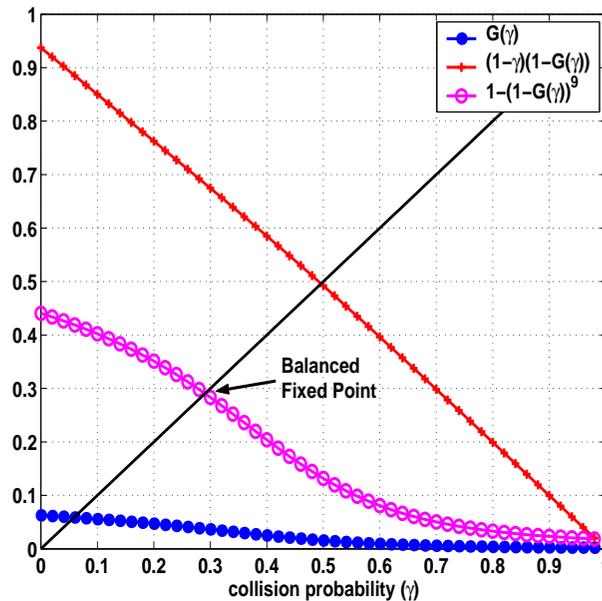


Figure 2.7: Example System-III: Plots of $G(\gamma)$, $F(\gamma) = (1-\gamma)(1-G(\gamma))$ and $1-(1-G(\gamma))^9$ vs. the collision probability γ ; the line “ $y=x$ ” is also shown.

corresponding figures for System-I and System-II). Also, the average collision rate matches well with the balanced fixed point solution obtained in Figure 2.7.

Remark: Thus we see that in a situation in which there is a unique fixed point not only is there a lack of multistability, but also the fixed point solution yields a good approximation to the long run average behaviour. ■

2.4.4 Short-term Fairness in Examples 1, 2 and 3

Figure 2.9 plots the throughput fairness index $\frac{1}{n} \frac{(\sum_{i=1}^n \tau_i)^2}{\sum_{i=1}^n \tau_i^2}$ (where τ_i is the average throughput of node i over the measurement frame, see [53]) against the frame size used to measure throughput. The fairness index is obtained for each frame size and is averaged over the duration of the simulation. Also plotted in the figure is the 95% confidence interval. We note that values of this index will lie in the interval $[0, 1]$, and smaller values of the index correspond to greater unfairness between the nodes. The performance of all the three example systems are compared. Notice that Example System-III (similar to IEEE 802.11 DCF) has the best fairness properties. The system achieves fairness of 0.9 over 1000's of slots. However, for Example System-I and II, similar performance is achieved only

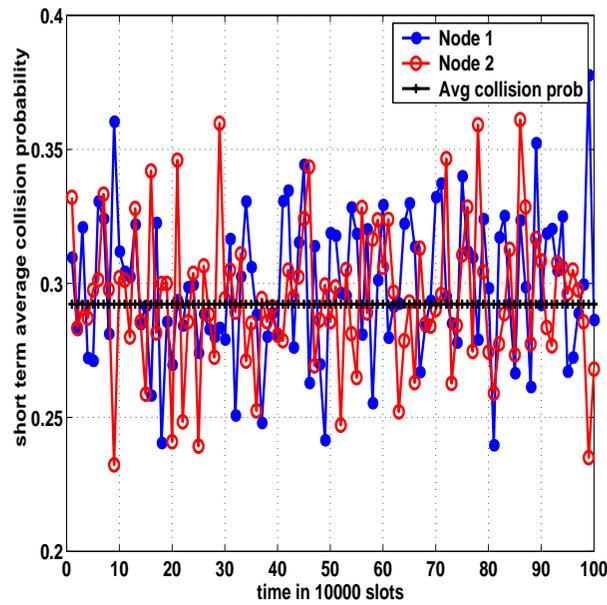


Figure 2.8: Example System-III: Snap-shot of short-term average collision probability of 2 of the 10 nodes. Also plotted is the average collision probability of the nodes. The 95% confidence interval of the average collision probability lies within 0.2% of the mean value.

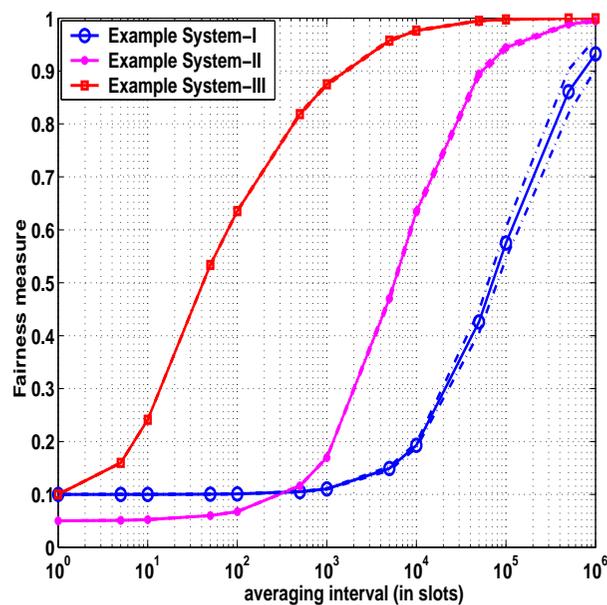


Figure 2.9: Throughput fairness index plotted against the frame size (number of slots used to measure throughput). The dotted lines mark the 95% confidence interval.

over 1,000,000 and 100,000 slots. The unfairness of Example Systems-I and II can be attributed to their apparent multistability.

In Section 2.5 we will establish conditions for uniqueness of the solution of the multi-dimensional fixed point equation.

2.5 Uniqueness of the Fixed Point

The following two results are adopted from [6].

Lemma 2.5.1 *$G(\gamma)$ is nonincreasing in γ if $b_k, k \geq 0$, is a nondecreasing sequence. In that case, unless $b_k = b_0$ for all k , $G(\gamma)$ is strictly decreasing in γ . ■*

Theorem 2.5.1 *Define $\Gamma := 1 - (1 - G(\gamma))^{n-1}$. For a homogeneous system of nodes, $\Gamma(G(\gamma)) : [0, 1] \rightarrow [0, 1]$, has a unique fixed point if $b_k, k \geq 0$, is a nondecreasing sequence. ■*

Remark: The fixed point $(\gamma, \gamma, \dots, \gamma)$ is the unique balanced fixed point for $\boldsymbol{\gamma} = \boldsymbol{\Gamma}(\mathbf{G}(\boldsymbol{\gamma}))$. From (2.4), we see that a *necessary* condition for the existence of unbalanced fixed points in a homogeneous system of nodes is that the function $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$ needs to be many-to-one. In other words, if the function $(1 - \gamma)(1 - G(\gamma))$ is one-to-one and if $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_n)$ is a solution of the system $\boldsymbol{\gamma} = \boldsymbol{\Gamma}(\mathbf{G}(\boldsymbol{\gamma}))$, then $\gamma_i = \gamma_j$ for all i, j . ■

Consider the exponentially increasing backoff case for which $G(\cdot)$ is given by,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \dots + \gamma^K}{b_0(1 + p\gamma + p^2\gamma^2 + \dots + p^K\gamma^K)} \quad (2.5)$$

Clearly, $G(\gamma)$ is a continuously differentiable function and so is $F(\gamma) = (1 - \gamma)(1 - G(\gamma))$.

The following simple lemma is a consequence of the mean value theorem.

Lemma 2.5.2 *$F(\gamma)$ is one-to-one in $0 \leq \gamma \leq 1$ if $F'(\gamma) \neq 0$ for all $0 \leq \gamma \leq 1$. ■*

Remarks 2.5.1

When $F(\cdot)$ is one-to-one in $0 \leq \gamma \leq 1$ and $G(\cdot)$ is such that $0 \leq G(\gamma) \leq 1$ for all $0 \leq \gamma \leq 1$, the following hold

- (i) $F(\gamma) = 0$ iff $\gamma = 1$,
- (ii) $F(0) > 0$, and
- (iii) $F(\gamma)$ is a decreasing function of γ . ■

Now the derivative of F is

$$F'(\gamma) = -1 + G(\gamma) - G'(\gamma)(1 - \gamma)$$

Lemma 2.5.3 For $G(\cdot)$ as in (2.5), if $K \geq 1$ and $p \geq 2$, then $G'(\gamma) < 0$ and $|G'(\gamma)| \leq \frac{2p}{b_0}$ for all $0 \leq \gamma \leq 1$.

Proof: See Appendix 2.7.1. ■

Clearly, $G(\gamma) \leq \frac{1}{b_0}$ and $1 \geq (1 - \gamma) \geq 0$ for all $0 \leq \gamma \leq 1$. Substituting into the expression for $F'(\gamma)$, we get,

$$F'(\gamma) \leq -1 + \frac{1 + 2p}{b_0}$$

Thus, if in addition to the other condition in Lemma 2.5.3, if $b_0 > 1 + 2p$, then $F'(\gamma) < 0$ and the following result holds by virtue of the remark following Theorem 2.5.1.

Theorem 2.5.2 For a function $G(\cdot)$ defined as in (2.5) if $K \geq 1, p \geq 2$ and $b_0 > 2p + 1$, then the system of equations $\gamma = \Gamma(\mathbf{G}(\gamma))$ has a unique fixed point which is balanced. ■

Remark: It can be shown that if Lemma 2.5.3 holds for $G(\cdot)$ as in (2.5) it also holds for any case in which $b_k = p^k b_0$ for $0 \leq k \leq m \leq K$ and $b_k = p^m b_0$ for $m < k \leq K$. The latter situation closely matches the IEEE 802.11 standard (with $b_0 = 16, p = 2, K = 7, m = 5$). Hence a homogeneous IEEE 802.11 WLAN has a unique fixed point which is also balanced. In general, if the function $G(\cdot)$ is arbitrary (as in (2.1)) but monotone decreasing, there exists a unique balanced fixed point for the system as long as the function $(1 - \gamma)(1 - G(\gamma))$ is one-to-one. ■

2.6 Summary

In this chapter we have studied a multidimensional fixed point equation arising from a model of the backoff process of the DCF access mechanism in IEEE 802.11 wireless LANs. Our first concern was the consequences of the nonuniqueness of the fixed point solution and conditions for uniqueness. We demonstrated via examples of homogeneous systems that even when the balanced fixed point is unique, the existence of unbalanced fixed points coexists with the observation of severe short-term unfairness in simulations. Further, in such examples the balanced fixed point solution does not capture the long run average behaviour of the system. With these observations in mind, we concluded that it is desirable to have systems in which there is a unique fixed point. We have provided simple sufficient conditions on the node backoff parameters that guarantee that a unique fixed point exists. We have shown that the default IEEE 802.11 parameters satisfy these sufficient conditions.

The fixed point approach is simply a heuristic that is found to work well in some cases. Our work suggests where it might not work and where it might work. In a recent work [11], the authors have proved that for random backoff algorithms, when the number of sources grow large, the system is indeed decoupled, providing a theoretical justification of decoupling arguments used in the analysis.

2.7 Appendix

2.7.1 Proofs of Theorems and Lemmas

Proof of Lemma 2.5.3

Define $G(\gamma) := \frac{u(\gamma)}{v(\gamma)}$. We have

$$\frac{u(\gamma)}{v(\gamma)} = \frac{1 + \gamma + \gamma^2 \cdots + \gamma^K}{b_0(1 + \gamma p + \cdots + \gamma^k p^k + \cdots + \gamma^K p^K)}$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$$

Since, by Lemma 2.5.1, $G'(\cdot) \leq 0$, $\left(\frac{u}{v}\right)' \leq 0$ for all $0 \leq \gamma \leq 1$. Also, with $K \geq 1$, u, u', v and v' are nonnegative for all $0 \leq \gamma \leq 1$. Hence, for all $0 \leq \gamma \leq 1$

$$\left|\left(\frac{u}{v}\right)'\right| \leq \frac{v'u}{v^2}$$

Differentiating v , we get,

$$v' = b_0(p + 2p^2\gamma + 3p^3\gamma^2 + \dots + Kp^K\gamma^{K-1})$$

Multiplying with u , we have,

$$\begin{aligned} v'u &= b_0(p + 2p^2\gamma + 3p^3\gamma^2 + \dots + Kp^K\gamma^{K-1})(1 + \gamma + \gamma^2 + \dots + \gamma^K) \\ &= b_0p(1 + 2p\gamma + 3p^2\gamma^2 + \dots + Kp^{K-1}\gamma^{K-1})(1 + \gamma + \gamma^2 + \dots + \gamma^K) \\ &= b_0p(1 + \gamma(1 + 2p) + \gamma^2(1 + 2p + 3p^2) + \gamma^3(1 + 2p + 3p^2 + 4p^3) + \dots \\ &\quad + \gamma^{K-1}(1 + 2p + \dots + Kp^{K-1}) + \gamma^K(1 + 2p + \dots + Kp^{K-1}) \\ &\quad + \gamma^{K+1}(2p + \dots + Kp^{K-1}) + \dots + \gamma^{2K-2}((K-1)p^{K-2} + Kp^{K-1}) \\ &\quad + \gamma^{2K-1}(Kp^{K-1})) \end{aligned}$$

We see that,

$$\begin{aligned} v'u &\leq b_0p(1 + \gamma(2 + 2p) + \gamma^2(3 + 3p + 3p^2) + \gamma^3(4 + 4p + 4p^2 + 4p^3) + \dots \\ &\quad + \gamma^{K-1}(K + Kp + \dots + Kp^{K-1}) + \gamma^K(K + Kp + \dots + Kp^{K-1}) \\ &\quad + \gamma^{K+1}(Kp + \dots + Kp^{K-1}) + \dots + \gamma^{2K-1}(Kp^{K-2} + Kp^{K-1}) \\ &\quad + \gamma^{2K-1}(Kp^{K-1})) \end{aligned}$$

For $p \geq 2$,

$$1 + p + p^2 + \dots + p^n < p^{n+1}$$

Hence,

$$\begin{aligned}
 v'u &\leq b_0p((1+1) + \gamma(2p+2p) + \gamma^2(3p^2+3p^2) + \gamma^3(4p^3+4p^3) + \dots \\
 &\quad + \gamma^{K-1}(Kp^{K-1} + Kp^{K-1}) + \gamma^K(Kp^{K-1} + Kp^{K-1}) \\
 &\quad + \gamma^{K+1}(Kp^{K-1} + Kp^{K-1}) + \dots + \gamma^{2K-1}(Kp^{K-1} + Kp^{K-1}) \\
 &\quad + \gamma^{2K-1}(Kp^{K-1} + Kp^{K-1})) \\
 &\leq b_02p(1 + \gamma(2p) + \gamma^2(3p^2) + \gamma^3(4p^3) + \dots + \gamma^{K-1}(Kp^{K-1}) + \gamma^K(Kp^{K-1}) \\
 &\quad + \gamma^{K+1}(Kp^{K-1}) + \dots + \gamma^{2K-1}(Kp^{K-1}) + \gamma^{2K-1}(Kp^{K-1}))
 \end{aligned}$$

But we know that,

$$\begin{aligned}
 v^2 &= b_0^2(1 + p\gamma + p^2\gamma^2 + \dots + p^K\gamma^K)^2 \\
 &= b_0^2(1 + \gamma(2p) + \gamma^2(3p^2) + \gamma^3(4p^3) + \dots + \gamma^{K-1}(Kp^{K-1}) + \gamma^K((K+1)p^K) \\
 &\quad + \gamma^{K+1}(Kp^{K+1}) + \gamma^{K+2}((K-1)p^{K+2}) + \dots + \gamma^{2K-1}(2p^{2K-1}) + \gamma^{2K}(p^{2K}))
 \end{aligned}$$

We see that, for $x \geq 2, y \geq 2, (x-1)(y-1) \geq 1 \Rightarrow xy \geq x+y$. Hence, for $K \geq 2, p \geq 2, K \leq (K-1)p$. Repeating the above argument for $(K-1)$ and p and so on, we get $K \leq (K-n)p^n$ for $0 \leq n \leq K-1$.

Now, comparing $v'u$ and v^2 term by term in powers of γ and using the fact that $K \leq (K-n)p^n$ for $K \geq 2, p \geq 2$ and $0 \leq n \leq K-1$, we see that,

$$\frac{v'u}{v^2} \leq \frac{2p}{b_0}$$

For the case $K = 1$ and $p \geq 2$, we have $v' = b_0p$ and $v'u = b_0p(1 + \gamma)$. Also, $v^2 = b_0^2(1 + 2p\gamma + \gamma^2)$. Hence,

$$\begin{aligned}
 \frac{v'u}{v^2} &= \frac{b_0p(1 + \gamma)}{b_0^2(1 + 2p\gamma + \gamma^2)} \\
 &= \frac{p}{b_0} \frac{(1 + \gamma)}{(1 + 2p\gamma + \gamma^2)}
 \end{aligned}$$

$$\leq \frac{p}{b_0} \leq \frac{2p}{b_0}$$

■

Chapter 3

Fixed Point Analysis of Single Cell IEEE 802.11e EDCA WLANs

3.1 Introduction

A new component of the IEEE 802.11e [2] medium access control is enhanced distributed channel access (EDCA), which provides differentiated channel access to packets by allowing different backoff parameters. Several traffic classes (or access categories) are supported, the classes being distinguished by channel priorities and backoff parameters. Thus, whereas in the legacy DCF all nodes have a single queue, and a single backoff “state machine”, all with the same backoff parameters (we say that the nodes are *homogeneous*), in EDCA the nodes can have multiple queues with separate backoff state machines with different parameters, and hence are permitted to be *nonhomogeneous*.

This chapter is concerned with the saturation throughput analysis of single cell IEEE 802.11e EDCA wireless LANs. We consider a single cell network of IEEE 802.11e nodes with a *pure collision* channel model (no capture, fading or frame error). For ease of understanding, much of our presentation is for the case in which each node has only one EDCA queue (of some access category). The analysis, however, applies to the general case of multiple EDCA queues (of different access categories) per node and we show this in Section 3.4.

In Chapter 2, using the generalization proposed in [6], we obtained the multidimensional fixed point equations for homogeneous IEEE 802.11 DCF nodes. In this chapter we consider *multidimensional fixed point equations* for a nonhomogeneous system of nodes. The nonhomogeneity arises due to different initial backoffs, or different backoff multipliers, or different amounts of time that nodes wait after a transmission before restarting their backoff counters (i.e., the AIFS (Arbitration InterFrame Space) mechanism of IEEE 802.11e), or different number of access categories per node. We consider both the balanced and unbalanced solutions of the fixed point equations and provide a condition for the fixed point solution to be balanced within an access category, and also a condition for uniqueness. Finally the fixed point equations are used to obtain insights into the throughput differentiation provided by different initial backoffs, persistence factors, AIFS and multiple traffic classes, for finite number of nodes and for differentiation parameter values similar to those in the IEEE 802.11e standard. An asymptotic analysis of the fixed point is also provided for the case in which packets are never abandoned, and the number of nodes goes to ∞ . Simulation results (from the CMP simulator and ns-2) validate the accuracy of our analysis.

3.1.1 A Survey of the Literature

Without modeling the AIFS mechanism, the extension is straightforward. Only the initial backoff, and the backoff multiplier (*persistence factor*) are modeled. In [77], [76] and [10], such a scheme is studied by extending Bianchi's Markov model per traffic class.

The AIFS technique is a further enhancement in IEEE 802.11e that provides a sort of priority to nodes that have smaller values of AIFS. After any successful transmission, whereas high priority nodes (with $\text{AIFS} = \text{DIFS}$) wait only for DIFS (DCF Interframe Space) to resume counting down their backoff counters, low priority nodes (with $\text{AIFS} > \text{DIFS}$) defer the initiation of countdown for an additional $\text{AIFS} - \text{DIFS}$ slots. Thus a high priority node decrements its backoff counter earlier than a low priority node and also has fewer collisions.

Among the approaches that have been proposed for modeling the AIFS mechanism

(for example, [75], [78], [79], [23], [80], [34], [24] and [28]) the ones in [34], [24] and [28] come much closer to capturing the service differentiation provided by the AIFS feature. In [34] the authors propose a Markov model to capture both the different backoff window expansion approach and AIFS. AIFS is modeled by expanding the state-space of the Markov chain to include the number of slots elapsed since the previous transmission attempt on the channel. In [28] the authors observe that the system exists in states in which only nodes of certain access categories can attempt. The approach is to model the evolution of these states as a Markov chain. The transition probabilities of this Markov chain are obtained from the assumed, decoupled attempt probabilities. This approach yields a fixed point formulation. This is the approach we will discuss in Section 3.3. [24] uses a Markov chain on the number of slots elapsed from the previous transmission to model AIFS based service differentiation. [80] extends the Bianchi's analysis to multiple traffic classes per node case using the Markov chain approach.

We note that the analyses in [34] and [28] are based on Bianchi's approach to modeling the residual backoff by a Markov chain. In this chapter, we have extended the simplification reported in [6] (which was for a homogeneous system of nodes) to nonhomogeneous nodes with different backoff parameters and AIFS based priority schemes. Also, we model the case of multiple queues (of different access categories) per node (see [80]). Thus, in our work, we have provided a simplified and integrated model to capture all the essential backoff based service differentiation mechanisms of the IEEE 802.11e EDCA.

3.1.2 Outline of the Chapter

In Section 3.2 we will develop the fixed point equations for the nonhomogeneous case with different backoff parameters (and without AIFS) and prove uniqueness results. In Section 3.3, we extend the analysis to include AIFS based differentiation. In Section 3.4 we analyse the case of multiple EDCA queues per node. An analytical study of the service differentiation provided by the various differentiation mechanisms is done in Section 3.5. Section 3.6 summarizes the results in the chapter.

3.2 The Nonhomogeneous Case

In this section, we will extend our results (of Chapter 2 for homogeneous nodes) to systems with nonhomogeneous nodes. AIFS will be introduced in Section 3.3. Nonhomogeneity is introduced by allowing different values of initial backoff (b_0), backoff multiplier (p) and different number of retries (K) per node.

Consider a nonhomogeneous system of n nodes, with $G_i(\cdot)$, the i^{th} node's response formula defined as in Chapter 2, Section 2.2. In Chapter 2, Section 2.5, we showed that, for a homogeneous system of nodes to have a unique solution to the fixed point equations, a sufficient condition is that the function $G(\cdot)$ is monotone decreasing and the function $F(\cdot)$ is one-to-one. Here as well, we will assume that $G_i(\cdot)$ is monotone decreasing and $F_i(\cdot)$ is one-to-one for all $1 \leq i \leq n$.

Let γ be the vector of collision probabilities of the nodes. With the slotted model for the backoff process and the decoupling assumption, the natural mapping of the attempt probabilities of other nodes to the collision probability of a node is given by

$$\gamma_i = \Gamma_i(\beta_1, \beta_2, \dots, \beta_n) = 1 - \prod_{j=1, j \neq i}^n (1 - \beta_j)$$

where $\beta_j = G_j(\gamma_j)$. The equilibrium behaviour of the system will be characterized by the solutions of the following system of equations, for $1 \leq i \leq n$,

$$\gamma_i = \Gamma_i(G_1(\gamma_1), \dots, G_n(\gamma_n))$$

A *necessary and sufficient condition* for a vector of collision probabilities $\gamma = (\gamma_1, \dots, \gamma_n)$ to be a fixed point solution is that, for all $1 \leq i \leq n$,

$$(1 - \gamma_i)(1 - G_i(\gamma_i)) = \prod_{j=1}^n (1 - G_j(\gamma_j)) \quad (3.1)$$

Define $F_i(\gamma) := (1 - \gamma_i)(1 - G_i(\gamma_i))$. From (3.1), we see that a *necessary condition* for a

vector of collision probabilities γ to be a fixed point solution is that, for all $i, j, 1 \leq i, j \leq n$,

$$F_i(\gamma_i) = F_j(\gamma_j) \quad (3.2)$$

Let there be two fixed point solutions $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$ and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ for the above system, and there exists $k, 1 \leq k \leq n$, such that $\gamma_k \neq \lambda_k$. From the necessary condition (equations (3.2)) we require that, for all i , and for some $J_1 > 0$ and $J_2 > 0$ (clearly, $J_1, J_2 \neq 0$),

$$(1 - \gamma_i)(1 - G_i(\gamma_i)) = J_1$$

$$(1 - \lambda_i)(1 - G_i(\lambda_i)) = J_2$$

Since $(1 - \gamma)(1 - G_i(\gamma))$ is one-to-one for all i , applying this to γ_k and λ_k , we require $J_1 \neq J_2$. Without loss of generality, assume $J_1 < J_2$. Hence, $\gamma_i > \lambda_i$ for all i (see Chapter 2, Remarks 2.5.1). Using (3.1) we have,

$$\begin{aligned} \lambda_i &= 1 - \prod_{j \neq i} (1 - G_j(\lambda_j)) \\ &\geq 1 - \prod_{j \neq i} (1 - G_j(\gamma_j)) \\ &= \gamma_i \end{aligned}$$

a contradiction. Hence, it must be that $J_1 = J_2$ or there exists a unique fixed point.

Notice that the arguments above immediately imply the following result.

Theorem 3.2.1 *If $G_i(\gamma)$ is a decreasing function of γ for all i and $(1 - \gamma)(1 - G_i(\gamma))$ is a strictly monotone function on $[0, 1]$, then the system of equations $\beta_i = G_i(\gamma_i)$ and $\gamma_i = \Gamma_i(\beta_1, \dots, \beta_i, \dots, \beta_n)$ has a unique fixed point. ■*

Where nodes use exponentially increasing backoff, the next result then follows.

Theorem 3.2.2 *For a system of nodes $1 \leq i \leq n$, with $G_i(\cdot)$ as in Chapter 2, equation (2.5), that satisfy $K_i \geq 1$, $p_i \geq 2$ and $b_{0_i} > 2p_i + 1$, there exists a unique fixed point for the system of equations, $\gamma_i = 1 - \prod_{j \neq i} (1 - G_j(\gamma_j))$ for $1 \leq i \leq n$.*

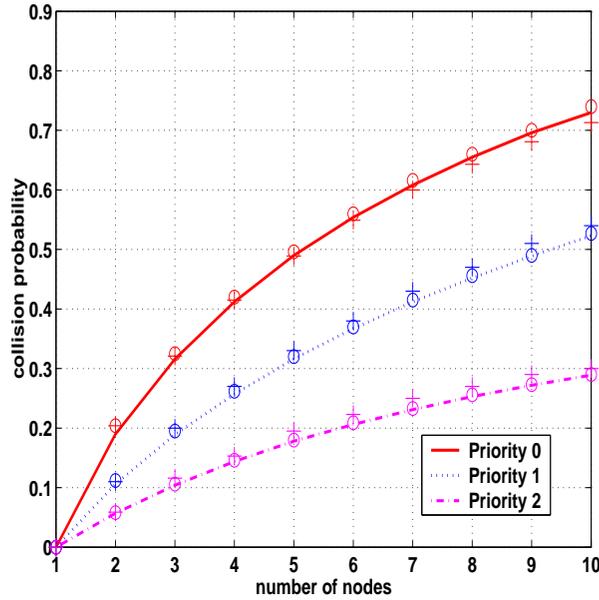


Figure 3.1: Plots of collision probability for a homogeneous system of nodes. Three different cases are considered, Priority 0 (AC_VO), Priority 1 (AC_VI) and Priority 2 (AC_BE). The lines correspond to the fixed point analysis, the “+” correspond to the ns-simulations and “O” correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

Proof: In Chapter 2, Theorem 2.5.2, we showed that when $G_i(\cdot)$ uses an exponentially increasing backoff, that satisfies $K_i \geq 1$, $p_i \geq 2$ and $b_{0_i} > 2p_i + 1$, then $G_i(\cdot)$ is monotone decreasing and $F_i(\cdot)$ is one-to-one. The uniqueness result now follows from Theorem 3.2.1. ■

Remark: The above result has relevance in the context of the IEEE 802.11e standard where the proposal is to use differences in backoff parameters to differentiate the throughputs obtained by the various nodes. While Theorem 3.2.2 only states a sufficient condition, it does point to a caution in choosing the backoff parameters of the nodes.

Figure 3.1 compares the collision probability obtained using the fixed point analysis for a homogeneous system, with ns-2 simulation and the CMP simulator. The plot shows 3 different cases, Priority 0, 1 and 2, corresponding to the IEEE 802.11e EDCA default settings for AC_VO, AC_VI and AC_BE. Observe that the fixed point analysis accurately predicts the system performance (from the CMP simulator and the ns-2 simulations).

3.3 Analysis of the AIFS Mechanism

Our approach for obtaining the fixed point equations when the AIFS mechanism is included is the same as the one developed in [28]. However, we develop the analysis in the more general framework introduced in [6] and extended here in Section 3.2. We will show that under the condition that $F(\cdot)$ is one-to-one there exists a unique fixed point for this problem as well. The analysis is presented here for two different AIFS class case, but can be extended to any number of classes. Also in this section, we consider only the case in which there is one queue (of an AIFS class) in each node. Extension to the case of multiple queues per node is done in Section 3.4.

Let us begin by recalling the basic idea of AIFS based service differentiation (see [63]). In legacy DCF, a node decrements its backoff counter, and then attempts to transmit only after it senses an idle medium for more than a DCF interframe space (DIFS). However, in EDCA (Enhanced Distributed Channel Access), based on the access category of a node (and its AIFS value), a node attempts to transmit only after it senses the medium idle for more than its AIFS. Higher priority nodes have smaller values of AIFS, and hence obtain a lower average collision probability, since these nodes can decrement their backoff counters, and even transmit, in slots in which lower priority nodes (waiting to complete their AIFSs) cannot. Thus, *nodes of higher priority (lower AIFS) not only tend to transmit more often but also have fewer collisions compared to nodes of lower priority (larger AIFS)*. The model we use to analyze the AIFS mechanism is quite general and accommodates the actual nuances of AIFS implementations (see [19] for how AIFS and DIFS differs) when the AIFS parameter values and the sampled backoff values are suitably adjusted.

3.3.1 The Fixed Point Equations

Let us consider two classes of nodes of two different priorities. The priority for a class is supported by using AIFS as well as b_0, p and K . All the nodes of a particular priority have the same values for all these parameters. There are $n^{(1)}$ nodes of Class 1 and $n^{(0)}$ nodes of Class 0. Class 1 corresponds to a higher priority of service. The AIFS for Class 0 exceeds

the AIFS of Class 1 by l slots. Thus, after every transmission activity in the channel, while Class 0 nodes wait to complete their AIFS, Class 1 nodes can attempt to transmit in those l slots. Also, if there is any transmission activity (by Class 1 nodes) during those l slots, then again the Class 0 nodes wait for another additional l slots compared to the Class 1 nodes, and so on.

As in [18] and [6], we need to model only the evolution of the backoff process of a node (i.e., the backoff slots after removing any channel activity such as transmissions or collisions) to obtain the collision probabilities. For convenience, let us call the slots in which only Class 1 nodes can attempt as *excess AIFS* slots, which will correspond to the subscript EA in the notation. In the *remaining* slots (corresponding to the subscript R in the notation) nodes of either class can attempt. Let us view such groups of slots, where different sets of nodes contend for the channel, as different *contention periods*. Let us define

$\beta_i^{(1)} :=$ the attempt probability of a Class 1 node for all $i, 1 \leq i \leq n^{(1)}$, in the slots in which a Class 1 node can attempt (i.e., all the slots)

$\beta_i^{(0)} :=$ the attempt probability of a Class 0 node for all $i, 1 \leq i \leq n^{(0)}$, in the contention periods during which Class 0 nodes can attempt (i.e., slots that are not Excess AIFS slots)

Note that in making these definitions we are modeling the attempt probabilities for Class 1 as being constant over all slots, i.e., the Excess AIFS slots and the remaining slots. This simplification has been shown to yield results that match well with simulations (see [28]). We also provide results using our simulation approach in Section 3.3.3.

Now the collision probabilities experienced by nodes will depend on the contention period (*excess AIFS* or *remaining* slots) that the system is in. The approach is to model the evolution over contention periods as a Markov Chain over the states $(0, 1, 2, \dots, l)$, where the state $s, 0 \leq s \leq (l - 1)$, denotes that an amount of time equal to s slots has elapsed since the end of the AIFS for Class 1. These states correspond to the *excess AIFS* period in which only Class 1 nodes can attempt. In the *remaining* slots, when the state is $s = l$, all nodes can attempt.

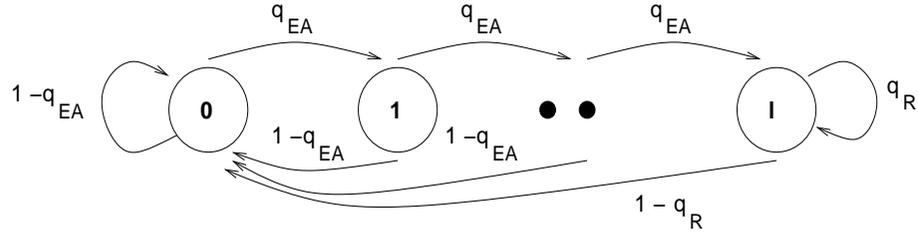


Figure 3.2: AIFS differentiation mechanism: Markov model for remaining number of AIFS slots.

In order to obtain the transition probabilities for this Markov chain we need the probability that a slot is idle. Using the decoupling assumption, the idle probability in any slot during the *excess AIFS* period is obtained as,

$$q_{EA} = \prod_{i=1}^{n^{(1)}} (1 - \beta_i^{(1)}) \quad (3.3)$$

Similarly, the idle probability in any of the remaining slots is obtained as,

$$q_R = \prod_{i=1}^{n^{(1)}} (1 - \beta_i^{(1)}) \prod_{j=1}^{n^{(0)}} (1 - \beta_j^{(0)}) \quad (3.4)$$

The transition structure of the Markov chain is shown in Figure 3.2. As compared to [28], we have used a simplification that the maximum contention window is much larger than l . If this were not the case then some nodes would certainly attempt before reaching l . In practice, l is small (e.g., 1 slot or 5 slots; see [2]) compared to the maximum contention window.

Let $\pi(EA)$ be the stationary probability of the system being in the *excess AIFS* period; i.e., this is the probability that the above Markov chain is in states 0, or 1, or \dots , or $(l-1)$. In addition, let $\pi(R)$ be the steady state probability of the system being in the remaining slots, i.e., state l of the Markov chain. Solving the balance equations for the steady state probabilities, we obtain,

$$\pi(EA) = \frac{1 + q_{EA} + q_{EA}^2 + \dots + q_{EA}^{l-1}}{1 + q_{EA} + q_{EA}^2 + \dots + q_{EA}^{l-1} + \frac{q_{EA}^l}{1 - q_R}}$$

$$\pi(R) = \frac{\frac{q_{EA}^l}{1-q_R}}{1 + q_{EA} + q_{EA}^2 + \cdots + q_{EA}^{l-1} + \frac{q_{EA}^l}{1-q_R}} \quad (3.5)$$

The average collision probability of a node is then obtained by averaging the collision probability experienced by a node over the different contention periods. The average collision probability for Class 1 nodes is given by, for all i , $1 \leq i \leq n^{(1)}$,

$$\begin{aligned} \gamma_i^{(1)} &= \pi(EA) \left(1 - \prod_{j=1, j \neq i}^{n^{(1)}} (1 - \beta_j^{(1)}) \right) \\ &+ \pi(R) \left(1 - \left(\prod_{j=1, j \neq i}^{n^{(1)}} (1 - \beta_j^{(1)}) \prod_{j=1}^{n^{(0)}} (1 - \beta_j^{(0)}) \right) \right) \end{aligned} \quad (3.6)$$

Similarly, the average collision probability of a Class 0 node is given by, for all i , $1 \leq i \leq n^{(0)}$,

$$\gamma_i^{(0)} = 1 - \left(\prod_{j=1}^{n^{(1)}} (1 - \beta_j^{(1)}) \prod_{j=1, j \neq i}^{n^{(0)}} (1 - \beta_j^{(0)}) \right) \quad (3.7)$$

Our analysis in the remaining section now generalizes the analysis of [28] and also establishes uniqueness of the fixed point and the property that the fixed point is balanced over nodes in the same class. Define $G^{(1)}(\cdot)$ and $G^{(0)}(\cdot)$ as in Chapter 2, equation (2.1) (except that the superscripts here denote the class dependent backoff parameters, with nodes within a class having the same parameters). Then the average collision probability obtained from the previous equations can be used to obtain the attempt rates by using the relations

$$\beta_i^{(1)} = G^{(1)}(\gamma_i^{(1)}), \text{ and } \beta_j^{(0)} = G^{(0)}(\gamma_j^{(0)}) \quad (3.8)$$

for all $1 \leq i \leq n^{(1)}$, $1 \leq j \leq n^{(0)}$. We obtain the fixed point equations for the collision probabilities by substituting the attempt probabilities from (3.8) into (3.6) and (3.7) (and also into equations (3.3) and (3.4)). We have a continuous mapping from $[0, 1]^{n^{(1)}+n^{(0)}}$ to $[0, 1]^{n^{(1)}+n^{(0)}}$. It follows from Brouwer's fixed point theorem that there exists a fixed point.

3.3.2 Uniqueness of the Fixed Point

Lemma 3.3.1 *If $F^{(\cdot)}$ is one-to-one, then collision probabilities of all the nodes of the same class are identical; i.e., the fixed points are balanced within each class.*

Proof: See Appendix 3.7.1. ■

Theorem 3.3.1 *The set of equations (3.6), (3.7) and (3.8) together with (3.5), (3.3) and (3.4), representing the fixed point equations for the AIFS model, has a unique solution if the corresponding functions $G^{(1)}$ and $G^{(0)}$ are monotone decreasing and $F^{(1)}$ and $F^{(0)}$ are one-to-one.*

Proof: See Appendix 3.7.1. ■

Remark: It follows from the earlier results in this chapter (see, example, Theorem 3.2.2) that if $G^{(0)}(\cdot)$ and $G^{(1)}(\cdot)$ have exponential backoffs, and if $K^{(i)} \geq 1$, $p^{(i)} \geq 2$, and $b_0^{(i)} > 2p^{(i)} + 1$, for $i = 0, 1$, then the fixed point equations will have a unique solution.

3.3.3 Numerical Study and Discussion

Although the numerical accuracy of the fixed point analysis has been reported before (see [18], [28]), for completeness, in Figures 3.3 and 3.4, we compare the collision probability obtained using the fixed point analysis with ns-2 simulation and the CMP simulator. Figure 3.3 plots the collision probabilities of AC_VO (access category for voice; the high priority) nodes and AC_BE (access category for best-effort traffic, e.g., TCP; the low priority) nodes, with the number of AC_BE nodes fixed to 4. Figure 3.4 plots the collision probabilities of AC_VI (access category for video; the high priority) nodes and AC_BE (the low priority) nodes with the number of AC_BE nodes fixed to 12. AC_VO, AC_VI and AC_BE correspond to the IEEE 802.11e EDCA access categories. As observed in the plots, the AIFS model works very well whenever $l \ll CW_{min}$ of the traffic classes. Additional plots comparing the analysis with the CMP simulator have been provided in Figures 3.5, 3.6 and 3.7.

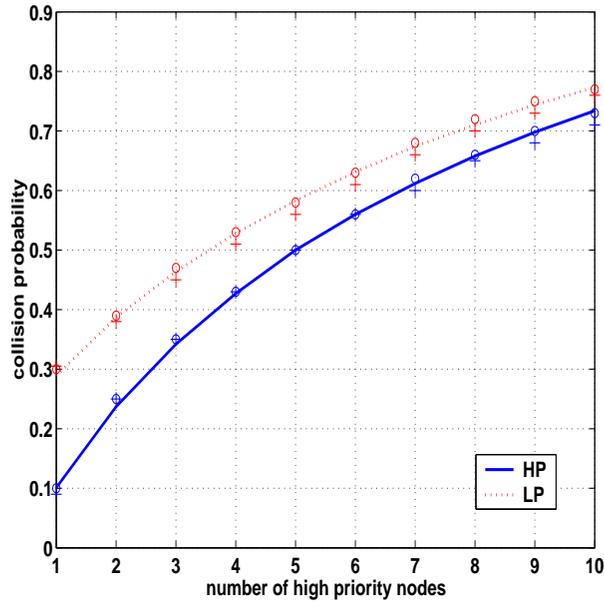


Figure 3.3: Plots of collision probability of HP - Priority 0 (AC_VO) nodes and LP - Priority 2 (AC_BE) nodes with the number of Priority 2 nodes fixed to 4. The lines correspond to the fixed point analysis, the “+” correspond to the ns-simulations and “o” correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

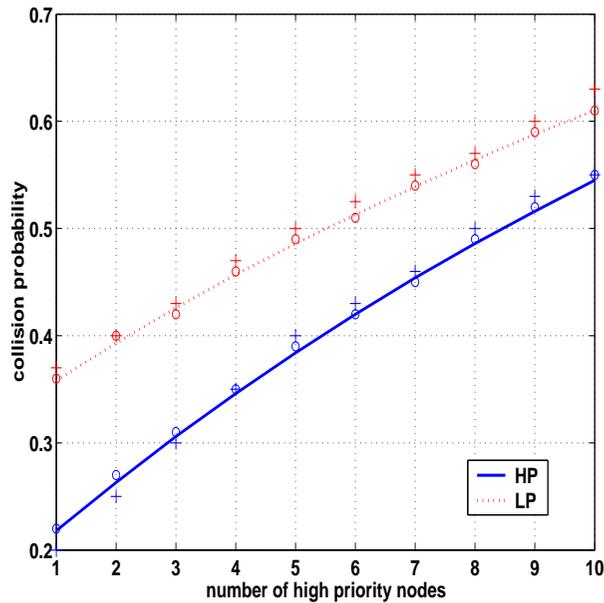


Figure 3.4: Plots of collision probability of HP - Priority 1 (AC_VI) nodes and LP - Priority 2 (AC_BE) nodes with the number of Priority 2 nodes fixed to 12. The lines correspond to the fixed point analysis, the “+” correspond to the ns-simulations and “o” correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

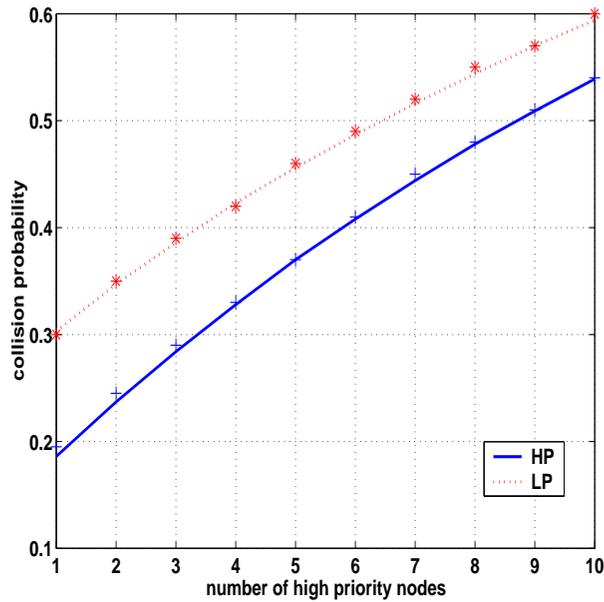


Figure 3.5: Plots of collision probability of HP - Priority 1 (AC_VI) nodes and LP - Priority 2 (AC_BE) nodes with the number of Priority 2 nodes fixed to 8. The lines correspond to the fixed point analysis and the symbols correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

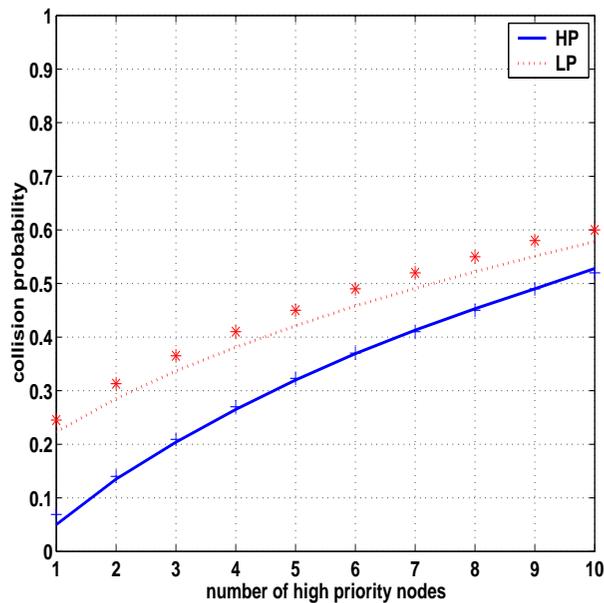


Figure 3.6: Plots of collision probability of HP - Priority 1 (AC_VI) nodes and LP - Priority 3 (AC_BK) nodes with the number of Priority 3 nodes fixed to 4. The lines correspond to the fixed point analysis and the symbols correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

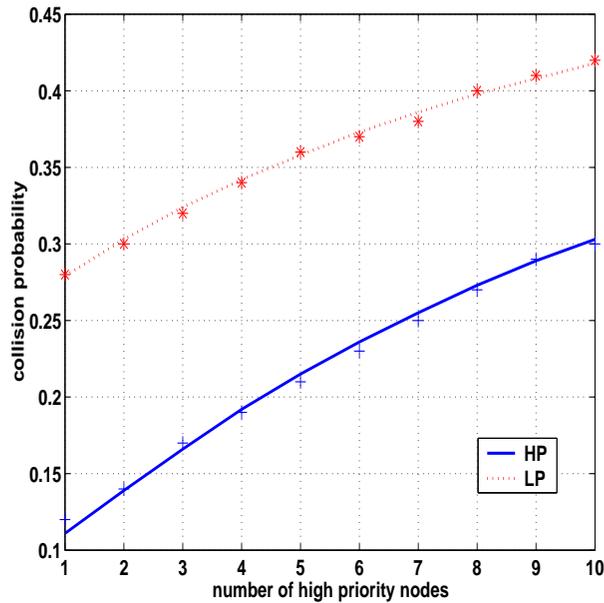


Figure 3.7: Plots of collision probability of HP - Priority 2 (AC_BE) nodes and LP - Priority 3 (AC_BK) nodes with the number of Priority 3 nodes fixed to 8. The lines correspond to the fixed point analysis and the symbols correspond to the CMP simulator. The 95% confidence interval lies within 1% of the simulation estimate.

Remarks 3.3.1 (AIFS Differentiation and Multistability) *It has been observed that (see, e.g., Section 3.5) as the number of nodes in the system increases, AIFS provides non-preemptive service to high priority nodes, starving the low priority nodes. This may lead to long periods of time when high priority nodes get serviced while the low priority nodes wait. We capture this behaviour using the Markov model in Figure 3.2. This cannot be viewed as multistability (as seen in Chapter 2), because AIFS always gives preferential access to the high priority nodes, while starving the low priority nodes, and never the other way. Further, in our analysis on AIFS, the attempt probability $\beta^{(i)}$ of a class i corresponds to only those slots in which class i can attempt (rather than all slots). The variation in attempt rate and collision probability, due to AIFS, is captured using the Markov model shown in Figure 3.2.*

3.4 Multiple Access Categories per Node

In this section we further generalize our fixed point analysis to include the possibility of multiple access categories (or queues) per node. We consider n nodes and c_i access categories (ACs) per node i ; the ACs can be of either AIFS class (for simplicity, we consider only two AIFS classes) and $c_i = c_i^{(1)} + c_i^{(0)}$ (the superscripts referring to the AIFS classes as before). The ACs in a node need not have the same $G(\cdot)$. Since there are multiple ACs per node, each with its own backoff process, it is possible that two or more ACs in a node complete their backoffs at the same slot. This is then called *Virtual Collision*, and is resolved in favour of the queue with the highest *Collision Priority* in the node. We label the ACs from 1 to c_i , with AC 1 corresponding to the highest collision priority in the node and AC c_i corresponding to the least collision priority. Unlike the single access category per node case where a collision is caused whenever any two nodes (equivalently, ACs) attempt in a slot, here, a AC sees a collision in a slot only when a AC of some other node or a higher collision priority AC of the same node attempts in that slot. A lower collision priority AC of a node cannot cause collision to a higher collision priority AC in the same node. In Section 3.4.1 we will study multiple access categories per node without AIFS (i.e., all the ACs wait only for DIFS) and consider AIFS later in Section 3.4.2.

We assume that, in a node (say i), the AIFS of Class 0 ACs (with $c_i^{(0)}$ ACs) exceeds the AIFS of the Class 1 ACs (with $c_i^{(1)}$ ACs) by l slots. Further, we will assume that the class 1 ACs have higher collision priority than the class 0 ACs. This assumption conforms with the way access categories are defined in the IEEE 802.11e standard. Also, when collision priorities are interchanged with AIFS priorities, the actual performance of the system would be hard to characterize.

3.4.1 Without AIFS

Let $\gamma_{i,j}$ be the collision probability of AC j of node i and $\beta_{i,j}$ be the attempt probability of AC j of node i , when the AC can attempt. The fixed point equations for this system

are, for all $i = 1, \dots, n$ (and $j = 1, \dots, c_i$),

$$\beta_{i,j} = G_{i,j}(\gamma_{i,j}) \quad (3.9)$$

$$\gamma_{i,j} = 1 - \prod_{m=1}^{j-1} (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l}) \quad (3.10)$$

where $G_{i,j}(\cdot)$ depend on the backoff parameters of AC j of node i . The term $\prod_{m=1}^{j-1} (1 - \beta_{i,m})$ in the above equation corresponds to the higher collision priority ACs in the same node. Observe that the $G_{i,j}(\cdot)$ definition allows the possibility of different backoff parameters (b_0, p, K) within a node.

Theorem 3.4.1 *The fixed point equations for γ , obtained by substituting (3.9) in (3.10) have a unique solution when $G_{i,j}$ is monotone decreasing and $F_{i,j}(\gamma) := (1 - \gamma)(1 - G_{i,j}(\gamma))$ is one-to-one for all $i = 1, \dots, n$ and $j = 1, \dots, c_i$.*

Proof: See Appendix 3.7.1. ■

3.4.2 With AIFS

In this section, we analyse the system where nodes have ACs of either AIFS class (the case where there are only Class 1 ACs can be modeled using the approach in Section 3.4.1). Define for $1 \leq i \leq n$, $1 \leq j \leq c_i$, $C_{i,j} \in \{0, 1\}$ to be the AIFS class of AC j in node i . Writing the fixed point equations for i, j s.t. $C_{i,j} = 1$, we have,

$$\begin{aligned} \gamma_{i,j} = & 1 - \left(\pi(EA) \prod_{m=1}^{j-1} (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{\{1 \leq l \leq c_k : C_{k,l}=1\}} (1 - \beta_{k,l}) \right. \\ & \left. + \pi(R) \prod_{m=1}^{j-1} (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l}) \right) \end{aligned} \quad (3.11)$$

and for i, j s.t. $C_{i,j} = 0$, we have,

$$\gamma_{i,j} = 1 - \prod_{m=1}^{j-1} (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l}) \quad (3.12)$$

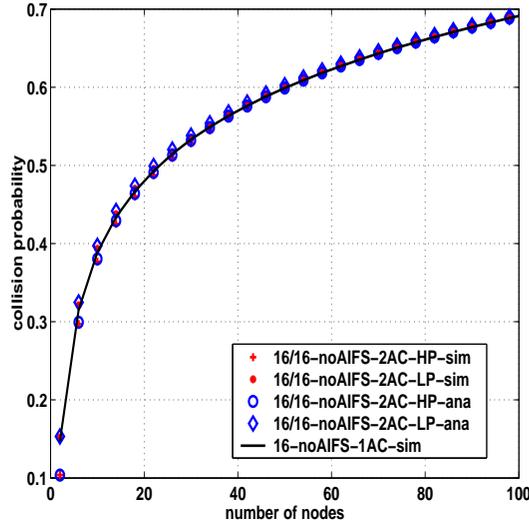


Figure 3.8: Collision probability of high priority AC (HP) and low priority AC (LP) in a system of nodes with two ACs. Both simulation (sim) and analysis (ana) are plotted. The backoff parameters of both the ACs (in all the nodes) are identical with $b_0 = 16$ and AIFS = DIFS. Also plotted is the collision probability (obtained from simulation) for single AC per node case with same backoff parameters and twice the number of nodes. In all the cases $p = 2$ and $K = 7$. For the simulation results, the 95% confidence interval lies within 1% of the mean value.

and $\beta_{i,j} = G_{i,j}(\gamma_{i,j})$. $\pi(EA)$ and $\pi(R)$ are defined as before (see equation (3.5)), with q_{EA} and q_R defined as

$$\begin{aligned}
 q_{EA} &= \prod_{k=1}^n \prod_{\{1 \leq l \leq c_k : C_{k,l}=1\}} (1 - \beta_{k,l}) \\
 q_R &= \prod_{k=1}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l})
 \end{aligned} \tag{3.13}$$

Theorem 3.4.2 *The fixed point equations (3.11) and (3.12) have a unique solution when $G_{i,j}$ are monotone decreasing and $F_{i,j}(\cdot)$ are one-to-one for all $i = 1, \dots, n$ and for each $i, j = 1, \dots, c_i$.*

Proof: See Appendix 3.7.1. ■

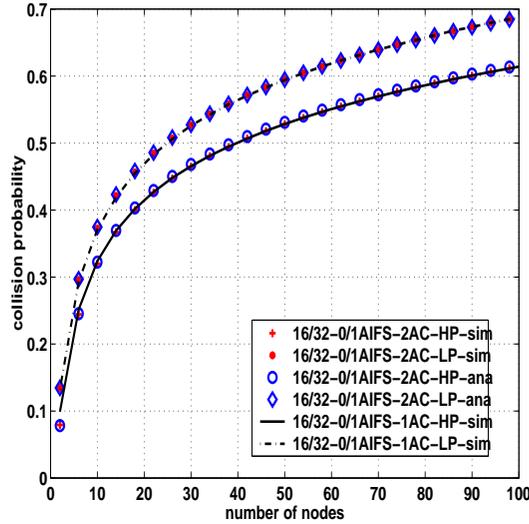


Figure 3.9: Collision probability of high priority AC (HP) and low priority AC (LP) in a system of nodes with two ACs. Both simulation (sim) and analysis (ana) are plotted. For the high priority AC, $b_0 = 16$ and AIFS = DIFS, while for the low priority AC we have $b_0 = 32$ and AIFS = DIFS + 1 slot. Also plotted is the collision probability (from simulation) for the case when the two ACs of a node are considered as independent ACs in separate nodes. In all the cases $p = 2$ and $K = 7$. For the simulation results, the 95% confidence interval lies within 1% of the mean value.

3.4.3 Numerical Study and Discussion

Figures 3.8 and 3.9 plot performance results for the multiple ACs per node case. In Figure 3.8, we consider a set of homogeneous nodes each with two access categories. The backoff parameters for either AC are the same ($b_0 = 16, p = 2, K = 7$ and AIFS = DIFS). The figure plots the collision probability of the higher priority (HP) AC and the low priority (LP) AC in simulation as well as the analysis. Also plotted in comparison is the collision probability (from simulation) for the single AC per node case with twice the number of nodes. Notice that, except for small n , the performance of the high priority AC and the low priority AC are almost identical (the backoff parameters are identical), and close to the performance of the single AC per node case (see Remark 3.4.1 below).

In Figure 3.9, we again consider a set of nodes each with two access categories. The higher priority AC has $b_0 = 16$ and AIFS = DIFS, while the low priority AC has $b_0 = 32$ and AIFS = DIFS + 1 slot. $p = 2$ and $K = 7$ for either case. Figure 3.9 plots the collision probability of the high priority AC and the low priority AC from simulation as

well as the analysis. Also plotted is the collision probability (from simulation) obtained by modeling the two ACs in a node as independent ACs in separate nodes. Notice again that except for small n , the performance of the multiple queue per node case is close to the performance of the single queue case.

Remarks 3.4.1 *The above observations from Figures 3.8 and 3.9 can be understood as follows. From the fixed point equations in Section 3.4, we see that for the high priority AC in any node, only one term corresponding to the low priority AC of the same node is missing (for the systems in Figures 3.8 and 3.9 with two ACs), in comparison to the case in which all the ACs are in $2n$ separate nodes. Hence, as n increases, the effect of this single AC in the same node diminishes, and the performance of the multiple queue per node case coincides with the performance of the single queue per node case each with one of the original ACs.*

3.5 Throughput Differentiation: An Analytical Study

It should be noted that all the results in this section are for the fixed point solution. Hence, when we use the term “collision probability” and “attempt rate” it is only in so far as a good match between the fixed point analysis and simulation has already been reported in earlier literature (see Section 3.1).

We will consider two alternatives for K , the maximum retransmission attempts allowed for a packet, namely $K = \infty$ and K finite. In this section, for the finite K case, the form of the function $G(\gamma)$, for all γ , $0 \leq \gamma \leq 1$ is,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \dots + \gamma^K}{b_0(1 + p\gamma + p^2\gamma^2 + \dots + p^K\gamma^K)} \quad (3.14)$$

It is clear that for finite K the attempt rate of a node is lower bounded, and hence as the number of nodes increases to infinity the collision probability of any node goes to 1. Hence, for this case, we will obtain insights regarding performance differentiation only for a finitely large number of nodes. For the infinite K case, however, we will study (as in [6]) the asymptotics of performance differentiation as the number of nodes tends to ∞ .

In the $K = \infty$ case, the function $G(\gamma)$ simplifies to,

$$G_{\infty}(\gamma) = \begin{cases} \frac{(1-\gamma p)}{b_0(1-\gamma)} & 0 \leq \gamma < \frac{1}{p} \\ 0 & \gamma \geq \frac{1}{p} \end{cases} \quad (3.15)$$

In the nonhomogeneous case we will write $G_{\infty}^{(1)}(\gamma)$ and $G_{\infty}^{(0)}(\gamma)$. For the homogeneous case with $K = \infty$, the (balanced fixed point) asymptotic analysis as $n \rightarrow \infty$ was performed in [6].

Consider a set of nodes, divided into two classes, Class 1 and Class 0, with Class 1 corresponding to a higher priority of service. *For simplicity, we assume that $n^{(1)}$ and $n^{(0)}$, the number of nodes of Class 1 and Class 0 respectively, are related as, $n^{(1)} = \alpha n$, $n^{(0)} = (1 - \alpha)n$ for some n and α , $0 < \alpha < 1$.* Let $\gamma^{(1)}(K, n)$ and $\beta^{(1)}(K, n)$ be the fixed point solutions for the collision probability and attempt rate of a Class 1 node for a given K and total number of nodes n . Similarly, let $\gamma^{(0)}(K, n)$ and $\beta^{(0)}(K, n)$ be the corresponding values for a Class 0 node.

We will study three cases:

Case 1: $b_0^{(1)} < b_0^{(0)}$, $p^{(1)} = p^{(0)} = p$, $AIFS^{(1)} = AIFS^{(0)} = DIFS$

Case 2: $b_0^{(1)} = b_0^{(0)} = b_0$, $p^{(1)} < p^{(0)}$, $AIFS^{(1)} = AIFS^{(0)} = DIFS$

Case 3: $b_0^{(1)} = b_0^{(0)} = b_0$, $p^{(1)} = p^{(0)} = p$, $AIFS^{(1)} < AIFS^{(0)}$

Note that in the analysis in earlier sections, we used the Binomial model for the number of attempts in a slot. With $n \rightarrow \infty$, in this section, we will use the Poisson batch model for the number of attempts in a slot (as in [6]).

3.5.1 Case 1: Differentiation by b_0

$K = \infty$, Asymptotic Analysis as $n \rightarrow \infty$

With the random number of attempts of each class in a backoff slot being modeled as Poisson distributed, the collision probabilities $\gamma^{(\cdot)}(\infty, n)$ and the attempt rates $\beta^{(\cdot)}(\infty, n)$

are related by

$$\begin{aligned}\gamma^{(1)}(\infty, n) &= 1 - e^{-((n^{(1)}-1)\beta^{(1)}(\infty, n) + n^{(0)}\beta^{(0)}(\infty, n))} \\ \gamma^{(0)}(\infty, n) &= 1 - e^{-(n^{(1)}\beta^{(1)}(\infty, n) + (n^{(0)}-1)\beta^{(0)}(\infty, n))}\end{aligned}\quad (3.16)$$

Substituting $\beta^{(\cdot)}(\infty, n) = G_{\infty}^{(\cdot)}(\gamma^{(\cdot)}(\infty, n))$ in the above equations gives the desired fixed point equations governing the system. Trivially, we see that,

$$(1 - \gamma^{(1)}(\infty, n))e^{-\beta^{(1)}(\infty, n)} = (1 - \gamma^{(0)}(\infty, n))e^{-\beta^{(0)}(\infty, n)} \quad (3.17)$$

Lemma 3.5.1 For $i \in \{0, 1\}$, $F_{\infty}^{(i)}(\gamma) := (1 - \gamma)e^{-G_{\infty}^{(i)}(\gamma)}$ is one-to-one for all $\gamma, 0 \leq \gamma \leq 1$ if $b_0^i \geq 2p + 1$.

Proof: See Appendix 3.7.1. ■

Theorem 3.5.1 In Case 1, with $K = \infty$, when $F_{\infty}^{(i)}$ is one-to-one for $i \in \{0, 1\}$,

1. $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n
2. $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) \uparrow \frac{1}{p}$, $\lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) \uparrow \frac{1}{p}$
3. $\lim_{n \rightarrow \infty} (n^{(1)}\beta^{(1)}(\infty, n) + n^{(0)}\beta^{(0)}(\infty, n)) \uparrow \ln(\frac{p}{p-1})$

Proof: See Appendix 3.7.1. ■

Theorem 3.5.2 In Case 1, with $K = \infty$, the ratio of the throughputs of Class 1 and Class 2 converges to $\frac{b_0^{(0)} - p}{b_0^{(1)} - p}$ as $n \rightarrow \infty$.

Proof: See Appendix 3.7.1. ■

Remark: Thus, for example, if $b_0^{(1)} = 16$, $b_0^{(0)} = 32$, and $p = 2$ then the ratio of the Class 1 to Class 0 node throughput will be approximately 30/14 for large n .

Finite K , Approximate Analysis for Large n

With finite K , as the number of nodes increases, the collision probability of either class increases to 1 (since the attempt rate is lower bounded) and $G^{(\cdot)}$ is small (since it decreases like $\frac{1}{b_0 p^{K+1}}$, see equation (3.14)). Then the difference between the collision probabilities is given by (we drop the arguments K and n in the following),

$$\gamma^{(1)} - \gamma^{(0)} = (G^{(0)}(\gamma^{(0)}) - G^{(1)}(\gamma^{(1)}))(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}-1}(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}-1}$$

also becomes insignificant. Hence, we can assume that $\gamma^{(1)} \approx \gamma^{(0)}$. For equal packet length transmission, the ratio of the throughputs of a Class 1 node to a Class 0 node corresponds to the ratio of their success probabilities, hence the throughput ratio is given by,

$$\frac{G^{(1)}(\gamma^{(1)})(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}-1}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}}{G^{(0)}(\gamma^{(0)})(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}-1}} = \frac{\frac{G^{(1)}(\gamma^{(1)})}{(1 - G^{(1)}(\gamma^{(1)}))}}{\frac{G^{(0)}(\gamma^{(0)})}{(1 - G^{(0)}(\gamma^{(0)}))}} \quad (3.18)$$

Using $\gamma^{(1)} \approx \gamma^{(0)}$ and writing this as γ , and using the fact that $G^{(\cdot)}(\gamma) \approx 0$ for large n , we have,

$$(3.18) \approx \frac{\frac{G^{(1)}(\gamma)}{(1 - G^{(1)}(\gamma))}}{\frac{G^{(0)}(\gamma)}{(1 - G^{(0)}(\gamma))}} \approx \frac{G^{(1)}(\gamma)}{G^{(0)}(\gamma)} = \frac{b_0^{(0)}}{b_0^{(1)}}$$

It follows that when service differentiation is provided by the backoff window, for a large number of nodes, the throughput ratio roughly corresponds to $\frac{b_0^{(0)}}{b_0^{(1)}}$, which, for large values of $b_0^{(0)}$ and $b_0^{(1)}$ is almost that same as that obtained for the asymptotic analysis with $K = \infty$ in Theorem 3.5.2

Remark: For finite K case, this observation (throughput ratio is approximately equal to $\frac{b_0^{(0)}}{b_0^{(1)}}$) is well known. This result has been shown analytically (using similar approximations) and also has been observed in simulations (see [10], [34] and [30]). It has been observed in [6] that for a given number of nodes, n , there will exist a $K(n)$ such that the system performance will not vary much for all $K > K(n)$. Hence, an asymptotic analysis would suffice for such cases.

3.5.2 Case 2: Differentiation by p

It may be noted that in the current version of IEEE 802.11e standard this mechanism no longer exists [2].

$K = \infty$, **Asymptotic Analysis as $n \rightarrow \infty$**

The fixed point equations governing the collision probability and the attempt rate is the same as (3.16). The following theorem summarizes the main results for Case 2.

Theorem 3.5.3 *In Case 2, with $K = \infty$, when $F_\infty^{(i)}$ is one-to-one for $i \in \{0, 1\}$, the following hold:*

1. $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n
2. $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) \uparrow \frac{1}{p^{(1)}}, \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) \uparrow \frac{1}{p^{(1)}}$
3. $\lim_{n \rightarrow \infty} n^{(1)} \beta^{(1)}(\infty, n) \uparrow \ln\left(\frac{p^{(1)}}{p^{(1)}-1}\right)$
4. $\lim_{n \rightarrow \infty} n^{(0)} \beta^{(0)}(\infty, n) = 0$

Proof: See Appendix 3.7.1. ■

Remark: Thus we see that, with $K = \infty$ and a large number of nodes, unlike initial backoff based differentiation, the persistence factor based differentiation completely suppresses the class with the larger value of p . ■

Finite K , Approximate Analysis for Large n

For finite K , with the approximation $\gamma^{(1)} \approx \gamma^{(0)}$ and the fact that $G^{(\cdot)}(\gamma^{(\cdot)}) \approx 0$, the throughput ratio approximates to $\frac{(1+p^{(0)}\gamma+p^{(0)^2}\gamma^2+\dots+p^{(0)^K}\gamma^K)}{(1+p^{(1)}\gamma+p^{(1)^2}\gamma^2+\dots+p^{(1)^K}\gamma^K)}$ (see equation (3.18)). Hence, as the collision probability of the system increases with load, the ratio of the throughputs of Class 1 to Class 0 also increases (depending on $p^{(1)}$, $p^{(0)}$ and the value of K). We note that as $n \rightarrow \infty$, the throughput ratio for the finite K case is finite, unlike the asymptotic case ($K = \infty$). However, the ratio tends to infinity when we consider $K \rightarrow \infty$.

3.5.3 Case 3: Differentiation by AIFS

$K = \infty$, Asymptotic Analysis for $n \rightarrow \infty$

In this case service differentiation is provided only by AIFS and we let $G_\infty^{(1)} = G_\infty^{(0)} = G_\infty$ (i.e., the backoff parameters b_0 and p are the same). With the assumption that the number of attempts in each slot is Poisson distributed, the fixed point equations for the AIFS model are (see equations (3.6) and (3.7))

$$\begin{aligned}\gamma^{(1)}(\infty, n) &= \pi(EA)(1 - e^{-(n^{(1)}-1)\beta^{(1)}(\infty, n)}) + \\ &\quad \pi(R)(1 - e^{-(n^{(1)}-1)\beta^{(1)}(\infty, n) - n^{(0)}\beta^{(0)}(\infty, n)}) \\ \gamma^{(0)}(\infty, n) &= (1 - e^{-n^{(1)}\beta^{(1)}(\infty, n) - (n^{(0)}-1)\beta^{(0)}(\infty, n)})\end{aligned}$$

Theorem 3.5.4 *In Case 3, with $K = \infty$, when $F_\infty^{(i)}$ is one-to-one for $i \in \{0, 1\}$,*

1. $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n
2. $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) \uparrow \frac{1}{p}$, $\lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) \uparrow \frac{1}{p}$
3. $\lim_{n \rightarrow \infty} n^{(1)}\beta^{(1)}(\infty, n) \uparrow \ln\left(\frac{p}{p-1}\right)$
4. $\lim_{n \rightarrow \infty} n^{(0)}\beta^{(0)}(\infty, n) = 0$

Proof: See Appendix 3.7.1. ■

Remark: Again we see that using AIFS for differentiation (for $K = \infty$ and n large) completely suppresses the class with the larger value of AIFS. Observe that Parts 3 and 4 of Theorem 3.5.4 imply that the individual node attempt ratio $\frac{\beta^{(1)}(\infty, n)}{\beta^{(0)}(\infty, n)}$ goes to ∞ as $n \rightarrow \infty$. Some insight into this result will be obtained from the analysis in the following sections.

Finite K , Approximate Analysis

Lemma 3.5.2 *In Case 3 for finite K , with $l = 1$, if the fixed point collision probabilities are $\gamma^{(1)}$ and $\gamma^{(0)}$, then the ratio of the throughputs of Class 1 to Class 0 is given by*

$$\frac{\frac{G^{(1)}(\gamma^{(1)})}{(1-G^{(1)}(\gamma^{(1)}))} \frac{1}{q_R}}{\frac{G^{(0)}(\gamma^{(0)})}{(1-G^{(0)}(\gamma^{(0)}))} \frac{1}{q_R}}$$

Proof: See Appendix 3.7.1. ■

Using this result and approximating $(1 - G^{(i)}(\gamma^{(i)})) \approx 1$ as before, the ratio of throughput equals

$$\frac{\frac{G^{(1)}(\gamma^{(1)})}{(1-G^{(1)}(\gamma^{(1)}))} \frac{1}{q_R}}{\frac{G^{(0)}(\gamma^{(0)})}{(1-G^{(0)}(\gamma^{(0)}))} \frac{1}{q_R}} \approx \frac{G^{(1)}(\gamma^{(1)})}{G^{(0)}(\gamma^{(0)})} \frac{1}{q_R} \quad (3.19)$$

For general l , we can expect a factor like $\frac{1}{q_R}$ in the previous expression. For low loads, when q_R is not close to 0, the dominating term in the previous expression is $\frac{G^{(1)}(\gamma^{(1)})}{G^{(0)}(\gamma^{(0)})}$. At high loads, both the terms contribute to throughput differentiation depending on the values of $n^{(1)}$ and $n^{(0)}$.

3.5.4 Numerical Study and Discussion

In Figure 3.10 we plot throughput ratios obtained from the CMP simulator for two classes of nodes differentiated by either b_0, p or AIFS. We note that this is the throughput ratio if the packet sizes of the two classes are equal. If the packet sizes are unequal then we only need to multiply the throughput ratio plotted here by the ratio of the packet lengths of the two classes. Also plotted is the analytical results obtained from our fixed point approach. The following remarks help in interpreting the results in Figure 3.10.

Remarks 3.5.1

1. Consider AIFS based differentiation. For finite K the attempt rates are bounded below, and the term $\frac{G^{(1)}(\gamma^{(1)})}{G^{(0)}(\gamma^{(0)})}$ is bounded, but as $(n^{(1)} + n^{(0)}) \rightarrow \infty$ the idle probability

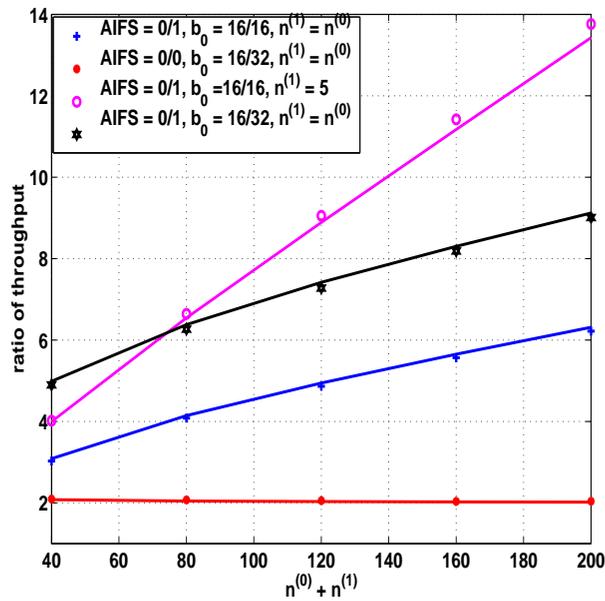


Figure 3.10: Ratio of the throughput of a Class 1 (higher priority) node to the throughput of a Class 0 node (lower priority). Analysis results (solid lines) and simulation results (symbols). Four cases are considered: +: differentiation only by AIFS with equal number of nodes, $n^{(1)} = n^{(0)}$; *: differentiation by AIFS and by b_0 with equal number of nodes, $n^{(1)} = n^{(0)}$; •: differentiation only by b_0 with equal number of nodes, $n^{(1)} = n^{(0)}$; o: differentiation only by AIFS with, $5 = n^{(1)} \ll n^{(0)}$. In all cases $p = 2$ and $K = 7$ for either class. For the simulation results, the 95% confidence interval lies within 1% of the average value.

$q_R \rightarrow 0$ ensuring (see equation (3.19)) that the individual node throughput ratio goes to ∞ for finite K as well (similar to the asymptotic results in Theorem 3.5.4). In addition, when $n^{(1)}$ increases, $\pi(EA)$ increases to 1. Hence, the lower priority nodes (with larger AIFS) rarely get a chance to attempt and the throughput ratio goes to infinity; this is demonstrated by the simulation results in Figure 3.10, plots with + and \star . When $n^{(1)}$ is kept constant and $n^{(0)}$ is increased (which is more typical), the collision probability of Class 0 nodes increases to 1 and their success probability tends to 0. However, the collision probability of Class 1 nodes remains much less than 1 depending on the value of $n^{(1)}$ and hence again the throughput ratio tends to ∞ (see Figure 3.10, plots with \circ). Figure 3.10 also shows the throughput ratio when only b_0 is used for differentiation (plots with \bullet); notice that, as shown earlier, the throughput ratio is just the reciprocal of the ratios of the initial backoff durations, and does not change with n .

2. For Case 3, in general, $\gamma^{(1)}$ and $\gamma^{(0)}$ are different, unlike in Cases 1 and 2. This is captured by the first term in the expression $\frac{G^{(1)}(\gamma^{(1)})}{G^{(0)}(\gamma^{(0)})} \frac{1}{q_R}$.
3. Notice that similar results for AIFS hold even when the functions $G^{(1)}$ and $G^{(0)}$ are not identical (see Figure 3.10, plot with \star). A comparison between the plots with + and \star in Figure 3.10 shows the effect of using both b_0 and AIFS for throughput differentiation. Adding b_0 based differentiation causes the entire curve to shift up (in favour of the higher priority class), and AIFS still causes the ratio to increase with increasing n . ■

3.6 Summary

In this chapter we have studied a multidimensional fixed point equation arising from a model of the backoff process of the EDCA access mechanism in IEEE 802.11e wireless LANs. We have provided simple sufficient conditions on the node backoff parameters that guarantee that a unique fixed point exists. We have shown that the default IEEE 802.11e parameters satisfy these sufficient conditions. The IEEE 802.11e standard motivated

us to consider the nonhomogeneous case, and our results suggest certain *safe* ranges of parameters that guarantee the uniqueness of the fixed point.

Using the fixed point analysis, we were also able to obtain insights into how the different backoff parameters provide throughput differentiation between the nodes in a nonhomogeneous system. We observed that using initial backoff window, in general, a fixed throughput ratio can be achieved. On the other hand, using p and *AIFS* the service can be significantly biased towards the high priority class, with the differentiation increasing in favour of the high priority class as the load in the system increases. We also observed that the effect of collision priority, where there are multiple access categories per node, decreases when the number of nodes increases.

This chapter is concerned with the saturation throughput analysis of an IEEE 802.11e single cell WLAN without fading and capture. In Chapter 4, we have developed a general framework to analyse single cell systems with capture. Performance analysis of IEEE 802.11 networks comprising interfering co-channel cells was studied in [39], using a similar fixed point approach discussed in this chapter.

3.7 Appendix

3.7.1 Proofs of Theorems and Lemmas

Proof of Lemma 3.3.1

Rewriting (3.6), for all i , $1 \leq i \leq n^{(1)}$, we get,

$$(1 - \gamma_i^{(1)}) = \prod_{j=1, j \neq i}^{n^{(1)}} (1 - \beta_j^{(1)}) \left[\pi(EA) + \pi(R) \prod_{k=1}^{n^{(0)}} (1 - \beta_k^{(0)}) \right]$$

Multiplying by $(1 - \beta_i^{(1)})$ and using the fact that $\beta_i^{(1)} = G^{(1)}(\gamma_i^{(1)})$, we have,

$$(1 - \gamma_i^{(1)})(1 - G^{(1)}(\gamma_i^{(1)})) = \pi(EA)q_{EA} + \pi(R)q_R \quad (3.20)$$

Observing (3.20), we see that the right hand side is independent of i . Hence, if the left hand side function, $F^{(1)}(\gamma) := (1 - \gamma)(1 - G^{(1)}(\gamma))$, is one to one, then $\gamma_i^{(1)} = \gamma_j^{(1)}$ for all $1 \leq i, j, \leq n^{(1)}$. Similarly, we can see from (3.7) that, for all i , $1 \leq i \leq n^{(0)}$,

$$(1 - \gamma_i^{(0)})(1 - G^{(0)}(\gamma_i^{(0)})) = q_R \quad (3.21)$$

Hence again, $\gamma_i^{(0)} = \gamma_j^{(0)}$ for all $1 \leq i, j, \leq n^{(0)}$, if $F^{(0)}$ is one to one.

Proof of Theorem 3.3.1

From Lemma 3.3.1, we already know that the fixed point is balanced within a class. Now, assume that there exist two vector fixed point solutions, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$, with the first $n^{(1)}$ elements of $\boldsymbol{\gamma}$ are $\gamma^{(1)}$ and the remaining $n^{(0)}$ elements are $\gamma^{(0)}$. Similarly, the first $n^{(1)}$ elements of $\boldsymbol{\lambda}$ are $\lambda^{(1)}$ and the next $n^{(0)}$ elements are $\lambda^{(0)}$.

Let us, in this proof, denote the value of q_R (see equation (3.4)) for the fixed point $\boldsymbol{\gamma}$ as $q_R(\boldsymbol{\gamma})$ and for the fixed point $\boldsymbol{\lambda}$ as $q_R(\boldsymbol{\lambda})$; similarly, we do for q_{EA} and for other variables.

Lemma 3.7.1 *Let $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ be two fixed point solutions and let $F^{(0)}$ be one-to-one. If $\gamma^{(1)} < \lambda^{(1)}$, then $\gamma^{(0)} > \lambda^{(0)}$. Also, $\gamma^{(1)} = \lambda^{(1)}$ iff $\gamma^{(0)} = \lambda^{(0)}$.*

Proof: Without loss of generality, let $\gamma^{(1)} < \lambda^{(1)}$. Then $G^{(1)}(\gamma^{(1)}) > G^{(1)}(\lambda^{(1)})$ (see Chapter 2, Lemma 2.5.1). Hence,

$$(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}} < (1 - G^{(1)}(\lambda^{(1)}))^{n^{(1)}}$$

If we assume $\gamma^{(0)} < \lambda^{(0)}$, then $q_R(\gamma^{(0)}) > q_R(\lambda^{(0)})$ (see equation (3.21)). Hence,

$$(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}} > (1 - G^{(1)}(\lambda^{(1)}))^{n^{(1)}}(1 - G^{(0)}(\lambda^{(0)}))^{n^{(0)}}$$

Since, $(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}} < (1 - G^{(1)}(\lambda^{(1)}))^{n^{(1)}}$, we now have,

$$(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}} > (1 - G^{(0)}(\lambda^{(0)}))^{n^{(0)}}$$

which implies $\gamma^{(0)} > \lambda^{(0)}$, which is a contradiction.

If $\gamma^{(0)} = \lambda^{(0)}$, then $q_R(\gamma^{(0)}) = q_R(\lambda^{(0)})$. Hence, $(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}} = (1 - G^{(1)}(\lambda^{(1)}))^{n^{(1)}}$, Or, $\gamma^{(1)} = \lambda^{(1)}$. Hence, if $\gamma^{(1)} < \lambda^{(1)}$, then $\gamma^{(0)} > \lambda^{(0)}$. Let $\gamma^{(0)} \neq \lambda^{(0)}$, then $q_R(\gamma^{(0)}) \neq q_R(\lambda^{(0)})$. Hence, $(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}} \neq (1 - G^{(1)}(\lambda^{(1)}))^{n^{(1)}}$, Or, $\gamma^{(1)} \neq \lambda^{(1)}$. ■

Now, using (3.5), write the right hand side of (3.20) as

$$J(q_{EA}, q_R, l) := \frac{q_{EA}(1 + q_{EA} + \cdots + q_{EA}^{l-1}) + q_R \frac{q_{EA}^l}{1 - q_R}}{1 + q_{EA} + q_{EA}^2 + \cdots + q_{EA}^{l-1} + \frac{q_{EA}^l}{1 - q_R}}$$

Lemma 3.7.2 *If $\gamma^{(1)} < \lambda^{(1)}$, then $J(q_{EA}(\gamma), q_R(\gamma), l) < J(q_{EA}(\lambda), q_R(\lambda), l)$.*

Proof: Consider $J(q_{EA}, q_R, l)$.

$$J(q_{EA}, q_R, l) = \frac{q_{EA}(1 + q_{EA} + \cdots + q_{EA}^{l-1}) + q_R \frac{q_{EA}^l}{1 - q_R}}{1 + q_{EA} + \cdots + q_{EA}^{l-1} + \frac{q_{EA}^l}{1 - q_R}}$$

Expanding and rewriting the above equation, we get,

$$= \frac{q_{EA} + q_{EA}(q_{EA} - q_R) + \cdots + q_{EA}^{l-1}(q_{EA} - q_R)}{q_{EA} + q_{EA}(q_{EA} - q_R) + \cdots + q_{EA}^{l-1}(q_{EA} - q_R) + (1 - q_R)}$$

which is of the form $\frac{f_1}{f_1 + f_2}$. When $\gamma^{(1)} < \lambda^{(1)}$, then $\gamma^{(0)} > \lambda^{(0)}$ (from the previous lemma).

Hence,

$$\begin{aligned} q_{EA}(\gamma) - q_R(\gamma) &= \prod_{i=1}^{n^{(1)}} (1 - G^{(1)}(\gamma^{(1)})) (1 - \prod_{i=1}^{n^{(0)}} (1 - G^{(0)}(\gamma^{(0)}))) \\ &< \prod_{i=1}^{n^{(1)}} (1 - G^{(1)}(\lambda^{(1)})) (1 - \prod_{i=1}^{n^{(0)}} (1 - G^{(0)}(\lambda^{(0)}))) \\ &= q_{EA}(\lambda) - q_R(\lambda) \end{aligned}$$

Also, we can see that,

$$q_{EA}(\gamma) < q_{EA}(\lambda)$$

$$q_R(\gamma) < q_R(\lambda)$$

Using the above three inequalities, we can see that,

$$J(q_{EA}(\boldsymbol{\gamma}), q_R(\boldsymbol{\gamma}), l) < J(q_{EA}(\boldsymbol{\lambda}), q_R(\boldsymbol{\lambda}), l)$$

■

If $\gamma^{(1)} < \lambda^{(1)}$, then $(1 - \gamma^{(1)})(1 - G^{(1)}(\gamma^{(1)})) > (1 - \lambda^{(1)})(1 - G^{(1)}(\lambda^{(1)}))$. However, from the above lemma and the right hand side of (3.20), we see that we have a contradiction. Hence, the fixed point equations have a unique solution when $G^{(1)}$ and $G^{(0)}$ are monotone decreasing and $F^{(1)}$ and $F^{(0)}$ are one-to-one.

Proof of Theorem 3.4.1

The fixed point equations are, for all $i = 1, \dots, n$ (and $j = 1, \dots, c_i$),

$$\gamma_{i,j} = 1 - \prod_{m=1}^{j-1} (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l})$$

where $\beta_{i,j} = G_{i,j}(\gamma_{i,j})$. Clearly, by Brouwer's fixed point theorem, there exists a fixed point solution for the above system of equations. Rewriting the above equation, we get,

$$(1 - \gamma_{i,j})(1 - \beta_{i,j}) = \prod_{m=1}^j (1 - \beta_{i,m}) \prod_{\{k=1, k \neq i\}}^n \prod_{l=1}^{c_k} (1 - \beta_{k,l})$$

Notice that, for $2 \leq j \leq c_i$,

$$(1 - \gamma_{i,j})(1 - \beta_{i,j}) = (1 - \gamma_{i,j-1})(1 - \beta_{i,j-1})(1 - \beta_{i,j})$$

or,

$$(1 - \gamma_{i,j}) = (1 - \gamma_{i,j-1})(1 - \beta_{i,j-1}) \tag{3.22}$$

when $(1 - \beta_{i,j}) > 0$.

Let us assume that there exists two fixed point solutions ($\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$) for the system. Without loss of generality, assume that for some node i and its AC j , $\gamma_{i,j} < \lambda_{i,j}$. Then,

the following lemma shows that $\gamma_{k,l} < \lambda_{k,l}$ for all $k = 1, \dots, n$ and $l = 1, \dots, c_k$.

Lemma 3.7.3 *Whenever γ and λ are the fixed point solutions, and if $\gamma_{i,j} < \lambda_{i,j}$ for some $i = 1, \dots, n$ and $j \in \{1, \dots, c_i\}$, then $\gamma_{k,l} < \lambda_{k,l}$ for all $k = 1, \dots, n$ and all $l = 1, \dots, c_k$.*

Proof: Let $\gamma_{i,j} < \lambda_{i,j}$ for some $i \in 1, \dots, n$ and $j \in \{1, \dots, c_i\}$. Then, using the fact the $F_{i,j}$ are strictly monotone decreasing, we have

$$(1 - \gamma_{i,j})(1 - G_{i,j}(\gamma_{i,j})) > (1 - \lambda_{i,j})(1 - G_{i,j}(\lambda_{i,j}))$$

Using (3.22), we see that,

$$(1 - \gamma_{i,j+1}) > (1 - \lambda_{i,j+1})$$

i.e., $\gamma_{i,j+1} < \lambda_{i,j+1}$ when ever $j + 1 \in \{1, \dots, c_i\}$ and, again using (3.22), we have

$$(1 - \gamma_{i,j-1})(1 - G_{i,j-1}(\gamma_{i,j-1})) > (1 - \lambda_{i,j-1})(1 - G_{i,j-1}(\lambda_{i,j-1}))$$

Or, $\gamma_{i,j-1} < \lambda_{i,j-1}$ when ever $j - 1 \in \{1, \dots, c_i\}$. Arguing as above, we see that $\gamma_{i,l} < \lambda_{i,l}$ for all $l = 1, \dots, c_i$.

From the fixed point equations, we observe that for all $k = 1, \dots, n$,

$$\begin{aligned} (1 - \gamma_{k,c_k})(1 - G_{k,c_k}(\gamma_{k,c_k})) &= \prod_{l=1}^n \prod_{m=1}^{c_l} (1 - G_{l,m}(\gamma_{l,m})) \\ (1 - \lambda_{k,c_k})(1 - G_{k,c_k}(\lambda_{k,c_k})) &= \prod_{l=1}^n \prod_{m=1}^{c_l} (1 - G_{l,m}(\lambda_{l,m})) \end{aligned}$$

But we know that

$$(1 - \gamma_{i,c_i})(1 - G_{i,c_i}(\gamma_{i,c_i})) > (1 - \lambda_{i,c_i})(1 - G_{i,c_i}(\lambda_{i,c_i}))$$

since $\gamma_{i,c_i} < \lambda_{i,c_i}$. Hence, we have,

$$(1 - \gamma_{k,c_k})(1 - G_{k,c_k}(\gamma_{k,c_k})) > (1 - \lambda_{k,c_k})(1 - G_{k,c_k}(\lambda_{k,c_k}))$$

Or, $\gamma_{k,c_k} < \lambda_{k,c_k}$ for all $1 \leq k \leq n$. Arguing as before for node i , we thus have $\gamma_{k,l} < \lambda_{k,l}$ for all $k = 1, \dots, n$ and $l = 1, \dots, n$. ■

Hence, if $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are two fixed point solutions for the system of equations, we see that $\gamma_{i,k} < \lambda_{i,k}$ for all $i = 1, \dots, n$ (and $k = 1, \dots, c_i$), which is clearly a contradiction (the proof is similar to that in Section 3.2 and is not provided). Hence, the system of equations for the multiple access categories per node case (without AIFS) has a unique fixed point.

Proof of Theorem 3.4.2

Consider c_i access categories per node i with $c_i^{(1)}$ ACs $(1, \dots, c_i^{(1)})$ with $AIFS^{(1)}$, and the remaining $c_i^{(0)}$ ACs $(c_i^{(1)} + 1, \dots, c_i)$ with $AIFS = AIFS^{(1)} + l$ slots. The fixed point equations for the system are given in (3.11) and (3.12).

As before, by Brouwer's fixed point theorem, there exists a fixed point for the system of equations. Assume that there exist two fixed point solutions for the above system of equations, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ with $\gamma_{i,j}$ and $\lambda_{i,j}$ as elements.

Let us, in this proof, denote the value of q_R (see equation (3.13)) for the fixed point $\boldsymbol{\gamma}$ as $q_R(\boldsymbol{\gamma})$ and for the fixed point $\boldsymbol{\lambda}$ as $q_R(\boldsymbol{\lambda})$; similarly, we do for q_{EA} and for other variables.

In a node i , consider two ACs of the same AIFS class, i.e., j and $j-1$ s.t. $C_{i,j} = C_{i,j-1}$. As in the proof of Theorem 3.4.1, it can be shown from (3.11) or (3.12), that

$$(1 - \gamma_{i,j}) = (1 - \gamma_{i,j-1})(1 - G_{i,j-1}(\gamma_{i,j-1}))$$

or,

$$(1 - \gamma_{i,j}) = F_{i,j-1}(\gamma_{i,j-1})$$

Hence, using the one-to-one property of $F_{i,j}(\cdot)$ if $\gamma_{i,j} < \lambda_{i,j}$, then $\gamma_{i,k} < \lambda_{i,k}$ for all k such that $C_{i,j} = C_{i,k}$,

Now consider all those nodes with $C_{i,c_i} = 0$, i.e., the least collision priority AC in a

node is of AIFS class 0. We then have, using (3.12) and (3.13),

$$\begin{aligned}(1 - \gamma_{i,c_i})(1 - G_{i,c_i}(\gamma_{i,c_i})) &= q_R(\boldsymbol{\gamma}) \\ (1 - \lambda_{i,c_i})(1 - G_{i,c_i}(\lambda_{i,c_i})) &= q_R(\boldsymbol{\lambda})\end{aligned}$$

i.e., $F_{i,c_i}(\gamma_{i,c_i}) = q_R(\boldsymbol{\gamma})$ and $F_{i,c_i}(\lambda_{i,c_i}) = q_R(\boldsymbol{\lambda})$. If $q_R(\boldsymbol{\gamma}) > q_R(\boldsymbol{\lambda})$, then $\gamma_{i,c_i} < \lambda_{i,c_i}$ for all i s.t. $C_{i,c_i} = 0$. If $q_R(\boldsymbol{\gamma}) = q_R(\boldsymbol{\lambda})$, then $\gamma_{i,c_i} = \lambda_{i,c_i}$ for all i s.t. $C_{i,c_i} = 0$. Combining the above two results, we see that for all i, j s.t. $C_{i,j} = 0$, either $\gamma_{i,j} > \lambda_{i,j}$ or $\gamma_{i,j} = \lambda_{i,j}$ or $\gamma_{i,j} < \lambda_{i,j}$.

Without loss of generality, assume that the collision probability of Class 0 ACs is more in $\boldsymbol{\gamma}$ than in $\boldsymbol{\lambda}$ ($\boldsymbol{\gamma}^{(0)} > \boldsymbol{\lambda}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$ are the vector of collision probabilities corresponding to AIFS class 0 in the vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ respectively). Hence, $q_R(\boldsymbol{\gamma}) < q_R(\boldsymbol{\lambda})$. Also, $q_{EA}(\boldsymbol{\gamma}) < q_{EA}(\boldsymbol{\lambda})$ (the proof is similar to that provided for AIFS with single AC per node and is not provided), which implies $\boldsymbol{\gamma}^{(1)} < \boldsymbol{\lambda}^{(1)}$.

Now consider the expression $F(\cdot)$ for the least collision priority Class 1 AC, say j , of any node i ,

$$\begin{aligned}(1 - \gamma_{i,j})(1 - G_{i,j}(\gamma_{i,j})) &= \pi(EA, \boldsymbol{\gamma})q_{EA}(\boldsymbol{\gamma}) + \pi(R, \boldsymbol{\gamma})q_{R(i,j)}(\boldsymbol{\gamma}) \\ (1 - \lambda_{i,j})(1 - G_{i,j}(\lambda_{i,j})) &= \pi(EA, \boldsymbol{\lambda})q_{EA}(\boldsymbol{\lambda}) + \pi(R, \boldsymbol{\lambda})q_{R(i,j)}(\boldsymbol{\lambda})\end{aligned}$$

where $q_{R(i,j)} = \prod_{m=1}^j (1 - \beta_{i,m}) \prod_{\{1 \leq k \leq n, k \neq i\}} \prod_{l=1}^{c_k} (1 - \beta_{k,l})$. Notice that $q_{R(i,j)}^{(i,j)}$ is similar to q_R except for terms corresponding to the Class 0 (with lower collision priority) ACs in node i . Hence, if $\boldsymbol{\gamma}^{(0)} > \boldsymbol{\lambda}^{(0)}$, then not only is $q_{EA}(\boldsymbol{\gamma}) < q_{EA}(\boldsymbol{\lambda})$ and $q_R(\boldsymbol{\gamma}) < q_R(\boldsymbol{\lambda})$, but also, $q_{R(i,j)}(\boldsymbol{\gamma}) < q_{R(i,j)}(\boldsymbol{\lambda})$. Expanding $(1 - \cdot_{i,j})(1 - G_{i,j}(\cdot_{i,j}))$, we get,

$$\begin{aligned}(1 - \cdot_{i,j})(1 - G_{i,j}(\cdot_{i,j})) &= \frac{(1 + q_{EA} + q_{EA}^2 + \cdots + q_{EA}^{l-1})q_{EA} + \frac{q_{EA}^l}{1 - q_R} q_{R(i,j)}}{1 + q_{EA} + q_{EA}^2 + \cdots + q_{EA}^{l-1} + \frac{q_{EA}^l}{1 - q_R}} \\ &= \frac{q_{EA} + q_{EA}(q_{EA} - q_R) + \cdots + q_{EA}^{l-1}(q_{EA} - q_R) + q_{EA}^l(q_{R(i,j)} - q_R)}{q_{EA} + q_{EA}(q_{EA} - q_R) + \cdots + q_{EA}^{l-1}(q_{EA} - q_R) + (1 - q_R)}\end{aligned}$$

where $q_{EA} - q_R = q_{EA}(1 - \prod_{k=1}^N \prod_{\{l=1, C_l^k=0\}}^{n_k} (1 - \beta_{k,l}))$ and $q_{R(i,j)}^{(i,j)} - q_R = q_{R(i,j)}^{(i,j)}(1 - \prod_{\{l=1, C_l^i=0\}}^{n_i} (1 -$

$\beta_{i,l}$). Clearly, if $\gamma^{(0)} > \lambda^{(0)}$, then $q_{EA}(\gamma) - q_R(\gamma) < q_{EA}(\lambda) - q_R(\lambda)$ and $q_{R_{i,j}}(\gamma) - q_R(\gamma) < q_{R_{i,j}}(\lambda) - q_R(\lambda)$. Also, we know that $1 - q_R(\gamma) > 1 - q_R(\lambda)$. From the above observations, we see that, $(1 - \gamma_{i,j})(1 - G_{i,j}(\gamma_{i,j})) < (1 - \lambda_{i,j})(1 - G_{i,j}(\lambda_{i,j}))$, which clearly implies that $\gamma_{i,j} > \lambda_{i,j}$. Hence we have $\gamma^{(1)} > \lambda^{(1)}$ which is a contradiction.

Also, we can see that $\gamma^{(1)} = \lambda^{(1)}$ iff $\gamma^{(0)} = \lambda^{(0)}$ (the proof is similar to that in Theorem 3.3.1 and is not provided here).

Proof of Lemma 3.5.1

Consider γ such that $0 \leq \gamma \leq \frac{1}{p}$. Then, $G_\infty(\gamma) = \frac{(1-\gamma p)}{b_0(1-\gamma)}$. Differentiating $(1 - \gamma)e^{-G_\infty(\gamma)}$, we have,

$$= e^{-G_\infty(\gamma)}(-1) + (1 - \gamma)e^{-G_\infty(\gamma)}(-G'_\infty(\gamma))$$

But $G'_\infty(\gamma) = \frac{1}{b_0} \frac{(1-p)}{(1-\gamma)^2}$. Substituting it in the previous equation, we get,

$$\begin{aligned} &= e^{-G_\infty(\gamma)} \left(-1 - (1 - \gamma) \frac{1}{b_0} \frac{(1-p)}{(1-\gamma)^2} \right) \\ &= e^{-G_\infty(\gamma)} \left(-1 - \frac{1}{b_0} \frac{(1-p)}{(1-\gamma)} \right) \end{aligned}$$

$e^{-G_\infty(\gamma)}$ is always positive (since $G_\infty(\gamma) < 1$). For $0 \leq \gamma \leq \frac{1}{p}$, the absolute value of $\frac{1}{b_0} \frac{(1-p)}{(1-\gamma)}$ is maximum when $\gamma = \frac{1}{p}$, at which the value equals, $\frac{1}{b_0} \frac{(1-p)}{(1-\frac{1}{p})} = -\frac{p}{b_0}$. Hence, the second term is always less than $(-1 + \frac{p}{b_0})$. But, if $b_0 \geq 2p + 1$, clearly, the second term is negative. Hence, the derivative is always negative and never equal to zero for all $0 \leq \gamma \leq \frac{1}{p}$. Hence, the function $(1 - \gamma)e^{-G_\infty(\gamma)}$ is one-to-one in the range $0 \leq \gamma \leq \frac{1}{p}$. For $\frac{1}{p} \leq \gamma \leq 1$, $G_\infty(\gamma) = 0$. Hence, $(1 - \gamma)e^{-G_\infty(\gamma)}$ is one-to-one for all γ , $\frac{1}{p} \leq \gamma \leq 1$. Also, the function is decreasing in both the intervals $0 \leq \gamma \leq \frac{1}{p}$ and $\frac{1}{p} \leq \gamma \leq 1$. Hence, $(1 - \gamma)e^{-G_\infty(\gamma)}$ is one-to-one for all $0 \leq \gamma \leq 1$.

Proof of Theorem 3.5.1

We shall prove Theorem 3.5.1 by first proving Lemmas 3.7.4 to 3.7.8.

Lemma 3.7.4 *In Case 1, with $K = \infty$, $\gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n)$ for all n .*

Remark: Thus, as expected, the collision probability for the higher priority class is smaller, for each n .

Proof: Since $b_0^{(0)} > b_0^{(1)}$, we see from (3.15) that, for every $\gamma \in [0, 1]$, $G_\infty^{(1)}(\gamma) \geq G_\infty^{(0)}(\gamma)$. Hence, $e^{-G_\infty^{(1)}(\gamma)} \leq e^{-G_\infty^{(0)}(\gamma)}$. Hence, $(1 - \gamma)e^{-G_\infty^{(1)}(\gamma)} \leq (1 - \gamma)e^{-G_\infty^{(0)}(\gamma)}$. Since the fixed point satisfies (3.17), it is necessary that $\gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n)$ holds (since $(1 - \gamma)e^{-G_\infty^{(1)}(\gamma)}$ is one-to-one decreasing). ■

Lemma 3.7.5 *In Case 1, with $K = \infty$, $\gamma^{(1)}(\infty, n)$ and $\gamma^{(0)}(\infty, n)$ are strictly increasing functions of n .*

Proof: Consider $n1 < n2$. We know that

$$\begin{aligned} (1 - \gamma^{(1)}(\infty, n1))e^{-\beta^{(1)}(\infty, n1)} &= (1 - \gamma^{(0)}(\infty, n1))e^{-\beta^{(0)}(\infty, n1)} \\ (1 - \gamma^{(1)}(\infty, n2))e^{-\beta^{(1)}(\infty, n2)} &= (1 - \gamma^{(0)}(\infty, n2))e^{-\beta^{(0)}(\infty, n2)} \end{aligned}$$

If $\gamma^{(1)}(\infty, n1) = \gamma^{(1)}(\infty, n2)$, then using Lemma 3.5.1 we see that $\gamma^{(0)}(\infty, n1) = \gamma^{(0)}(\infty, n2)$. Hence $\beta^{(0)}(\infty, n1) = \beta^{(0)}(\infty, n2)$ and $\beta^{(1)}(\infty, n1) = \beta^{(1)}(\infty, n2)$. Since both $\beta^{(0)}(\infty, \cdot)$ and $\beta^{(1)}(\infty, \cdot)$ cannot be zero, and as $n1 < n2$, substituting in (3.16), we get a contradiction.

Assume that $\gamma^{(1)}(\infty, n1) > \gamma^{(1)}(\infty, n2)$. Then, from (3.17) and Lemma 3.5.1, it follows that $\gamma^{(0)}(\infty, n1) > \gamma^{(0)}(\infty, n2)$. Also if collision probabilities decrease with n , it would imply that the attempt rates increase with n , i.e., $\beta^{(1)}(\infty, n1) \leq \beta^{(1)}(\infty, n2)$ and $\beta^{(0)}(\infty, n1) \leq \beta^{(0)}(\infty, n2)$. But from (3.16), we see that,

$$\begin{aligned} \gamma^{(1)}(\infty, n1) &= 1 - e^{-((n1^{(1)}-1)\beta^{(1)}(\infty, n1) + n1^{(0)}\beta^{(0)}(\infty, n1))} \\ &\leq 1 - e^{-((n2^{(1)}-1)\beta^{(1)}(\infty, n2) + n2^{(0)}\beta^{(0)}(\infty, n2))} \\ &= \gamma^{(1)}(\infty, n2) \end{aligned}$$

Thus we have a contradiction and the result is proved. ■

Lemma 3.7.6 *In Case 1, with $K = \infty$, the attempt rates $\beta^{(1)}(\infty, n)$ and $\beta^{(0)}(\infty, n)$ tend to zero as $n \rightarrow \infty$.*

Proof: If not, the exponent in the collision probability equation (3.16) tends to $-\infty$ taking the collision probabilities to 1. However, we know that the attempt rate is zero for all $\gamma \geq \frac{1}{p}$, leading to a contradiction (since we are interested only in the case $p > 1$). ■

Lemma 3.7.7 *In Case 1, with $K = \infty$, $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) = \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n)$*

Proof: We have

$$\begin{aligned} 0 < \gamma^{(1)}(\infty, n) - \gamma^{(0)}(\infty, n) &= e^{-((n^{(1)}-1)\beta^{(1)}(\infty, n) + (n^{(0)}-1)\beta^{(0)}(\infty, n))} (e^{-\beta^{(1)}(\infty, n)} - e^{-\beta^{(0)}(\infty, n)}) \\ &\leq (e^{-\beta^{(1)}(\infty, n)} - e^{-\beta^{(0)}(\infty, n)}) \rightarrow_{n \rightarrow \infty} 0 \end{aligned}$$

which proves the required result. ■

Lemma 3.7.8 *In Case 1, with $K = \infty$, $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n) < \frac{1}{p}$ for all n .*

Proof: We first observe that $\gamma^{(1)}(\infty, n) < \frac{1}{p}$ for all n . Otherwise, by Lemma 3.7.4 and Lemma 3.7.5, $\frac{1}{p} \leq \gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n)$ for all $n > N$ for some N . Hence, $\beta^{(1)}(\infty, n) = \beta^{(0)}(\infty, n) = 0$ for all $n > N$. However, substituting in (3.16) gives a contradiction.

Now assume that $\gamma^{(0)}(\infty, n) \geq \frac{1}{p}$ for all $n \geq N$ for some N . Since, $\gamma^{(\cdot)}(\infty, n)$ is a strictly increasing function of n , we can, without loss of generality, assume that $\gamma^{(0)}(\infty, n) > \frac{1}{p}$ for all $n \geq N$ for some N . Hence, $\beta^{(0)}(\infty, n)$ is zero for all $n \geq N$. But we know that the collision probability of Class 1 also increases with n and the limit of the collision probability of Class 1 and Class 0 are equal. Hence, $\gamma^{(1)}(\infty, n)$ exceeds $\frac{1}{p}$ for all $n \geq N'$ for some N' , which is a contradiction.

Since $\gamma^{(1)}(\infty, n)$ and $\gamma^{(0)}(\infty, n)$ are less than $\frac{1}{p}$, the inequality in Lemma 3.7.4 becomes strict, i.e., $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n (when $b_0^{(0)} > b_0^{(1)}$, $G_\infty^{(0)}(\gamma) < G_\infty^{(1)}(\gamma)$ for all $0 \leq \gamma \leq \frac{1}{p}$). ■

Combining the above Lemmas, we see that $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n (From Lemma 3.7.8). Using the fact that $\beta^{(\cdot)}(\infty, n) \rightarrow 0$ as $n \rightarrow \infty$ and $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n) < \frac{1}{p}$ for all n , we get $\lim_{n \rightarrow \infty} \gamma^{(\cdot)}(\infty, n) = \frac{1}{p}$ as $n \rightarrow \infty$ (from (3.15)). Substituting $\gamma^{(\cdot)}(\infty, n) \rightarrow \frac{1}{p}$ as $n \rightarrow \infty$ in (3.16), we see that $\lim_{n \rightarrow \infty} (n^{(1)}\beta^{(1)}(\infty, n) + n^{(0)}\beta^{(0)}(\infty, n)) \uparrow \ln(\frac{p}{p-1})$, thus completing the proof of Theorem 3.5.1.

Proof of Theorem 3.5.2

In the following, for notational simplicity, we drop the argument (∞, n) . Consider the necessary condition, that a fixed point solution satisfies.

$$(1 - \gamma^{(1)})(1 - G_{\infty}^{(1)}(\gamma^{(1)})) = (1 - \gamma^{(0)})(1 - G_{\infty}^{(0)}(\gamma^{(0)}))$$

Since we are interested only in the range $0 \leq \gamma \leq \frac{1}{p}$, we can substitute for $G_{\infty}(\gamma) = \frac{(1-\gamma p)}{b_0(1-\gamma)}$, and further simplify the equation to,

$$(1 - \gamma^{(1)}) - \frac{1}{b_0^{(1)}}(1 - \gamma^{(1)}p) = (1 - \gamma^{(0)}) - \frac{1}{b_0^{(0)}}(1 - \gamma^{(0)}p)$$

Rearranging the terms, we have,

$$(\gamma^{(0)} - \gamma^{(1)}) = \frac{1}{b_0^{(1)}}(1 - \gamma^{(1)}p) - \frac{1}{b_0^{(0)}}(1 - \gamma^{(0)}p)$$

Further

$$\frac{b_0^{(1)}}{(1 - \gamma^{(1)}p)}(\gamma^{(0)} - \gamma^{(1)}) = 1 - \frac{b_0^{(1)}}{b_0^{(0)}} \frac{(1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)}$$

Let us rewrite the left hand side of this equation as follows

$$\begin{aligned} \frac{b_0^{(1)}}{(1 - \gamma^{(1)}p)}(\gamma^{(0)} - \gamma^{(1)}) &= \frac{b_0^{(1)}}{p} \frac{(\gamma^{(0)}p - \gamma^{(1)}p)}{(1 - \gamma^{(1)}p)} \\ &= \frac{b_0^{(1)}}{p} \frac{(1 - \gamma^{(1)}p) - (1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)} \\ &= \frac{b_0^{(1)}}{p} \left(1 - \frac{(1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)} \right) \end{aligned}$$

Substituting back this expression for the left hand side into the original equation, we have

$$\frac{b_0^{(1)}}{p} \left(1 - \frac{(1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)} \right) = 1 - \frac{b_0^{(1)}}{b_0^{(0)}} \frac{(1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)}$$

Rearranging terms, we obtain

$$\frac{b_0^{(1)}}{p} - 1 = \frac{(1 - \gamma^{(0)}p)}{(1 - \gamma^{(1)}p)} \left(\frac{b_0^{(1)}}{p} - \frac{b_0^{(1)}}{b_0^{(0)}} \right)$$

Finally the calculation yields

$$\frac{(1 - \gamma^{(1)}p)/b_0^{(0)}}{(1 - \gamma^{(0)}p)/b_0^{(1)}} = \frac{(b_0^{(0)} - p)}{(b_0^{(1)} - p)} \quad (3.23)$$

For equal packet length case, the ratio of the throughput of the nodes equals the ratio of their success probabilities in a slot (see, for example, [18] and [6]) which, upon simplification, yields (we reintroduce the dependence on n in the notation)

$$\frac{\beta^{(1)}(\infty, n)}{(1 - \beta^{(1)}(\infty, n))} \frac{(1 - \beta^{(0)}(\infty, n))}{\beta^{(0)}(\infty, n)}$$

As $n \rightarrow \infty$, $\beta^{(1)}(\infty, n)$ and $\beta^{(0)}(\infty, n)$ tends to 0. Also we know that $\beta^{(\cdot)}(\infty, n)$ is of the form $\frac{(1 - \gamma^{(\cdot)}(\infty, n)p)}{b_0^{(\cdot)}(1 - \gamma^{(\cdot)}(\infty, n))}$ and $\gamma^{(\cdot)}(\infty, n) < \frac{1}{p}$ for all n . Since

$$\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) = \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n)$$

we have $\lim_{n \rightarrow \infty} (1 - \gamma^{(1)}(\infty, n)) = \lim_{n \rightarrow \infty} (1 - \gamma^{(0)}(\infty, n))$. Hence, using (3.23), the ratio of the throughputs, as $n \rightarrow \infty$, can be seen to converge to $\frac{(b_0^{(0)} - p)}{(b_0^{(1)} - p)}$ as $n \rightarrow \infty$.

Proof of Theorem 3.5.3

Lemma 3.7.9 *For Case 2, with $K = \infty$, and $F_\infty^{(i)}$ one-to-one*

1. $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n
2. $\gamma^{(1)}(\infty, n)$ and $\gamma^{(0)}(\infty, n)$ strictly increase with n
3. $\beta^{(1)}(\infty, n)$ and $\beta^{(0)}(\infty, n)$ tend to 0 as $n \rightarrow \infty$
4. $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n) < \frac{1}{p^{(1)}}$, $\forall n$

$$5. \lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) = \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) = \frac{1}{p^{(1)}}$$

■

The proof follows in similar lines as in Lemmas 3.7.4 - 3.7.8 and hence is not provided here.

Lemma 3.7.10 *In Case 2, with $K = \infty$, $n^{(0)}\beta^{(0)}(\infty, n) \rightarrow 0$ as $n \rightarrow \infty$.*

Proof: Since

$$\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) = \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) = \frac{1}{p^{(1)}} > \frac{1}{p^{(0)}}$$

we have, $\beta^{(0)}(\infty, n) = 0$ for all $n > N$ for some N . Hence, $\lim_{n \rightarrow \infty} n^{(0)}\beta^{(0)}(\infty, n) = 0$. ■

Remark: Thus, the aggregate attempt rate of the Class 0 goes to zero, while the aggregate attempt rate of the Class 1 governs the system performance.

From Lemma 3.7.9, we see that $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n and $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) \uparrow \frac{1}{p^{(1)}}$, $\lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n) \uparrow \frac{1}{p^{(1)}}$. Lemma 3.7.10 shows that $\lim_{n \rightarrow \infty} n^{(0)}\beta^{(0)}(\infty, n) = 0$. Hence, substituting in the fixed point equations for Case 2, we get $\lim_{n \rightarrow \infty} n^{(1)}\beta^{(1)}(\infty, n) \uparrow \ln\left(\frac{p^{(1)}}{p^{(1)}-1}\right)$ completing the proof of Theorem 3.5.3.

Proof of Theorem 3.5.4

Lemma 3.7.11 *In Case 3, $\gamma^{(1)}(\infty, n)$ and $\gamma^{(0)}(\infty, n)$ are strictly increasing functions of n .*

Proof: Rewriting the fixed point equations for AIFS, we have,

$$\begin{aligned} (1 - \gamma^{(1)}(\infty, n))e^{-\beta^{(1)}(\infty, n)} &= e^{-n^{(1)}\beta^{(1)}(\infty, n)}(\pi(EA) + \pi(R)e^{-n^{(0)}\beta^{(0)}(\infty, n)}) \\ (1 - \gamma^{(0)}(\infty, n))e^{-\beta^{(0)}(\infty, n)} &= e^{-n^{(1)}\beta^{(1)}(\infty, n) - n^{(0)}\beta^{(0)}(\infty, n)} \end{aligned} \quad (3.24)$$

Consider $n_1 < n_2$.

Assume that $\gamma^{(0)}(\infty, n_1) > \gamma^{(0)}(\infty, n_2)$ (hence, $\beta^{(0)}(\infty, n_1) \leq \beta^{(0)}(\infty, n_2)$). As $\gamma^{(0)}(\infty, n)$ decreases with n , $(1 - \gamma)e^{-G(\gamma)}$ increases. Hence, $q_R = e^{-n^{(1)}\beta^{(1)}(\infty, n) - n^{(0)}\beta^{(0)}(\infty, n)}$ increases

with n . Since q_R increases with n and $\beta^{(0)}(\infty, n)$ is non-decreasing with n , we require $q_{EA} = e^{-n^{(1)}\beta^{(1)}(\infty, n)}$ strictly increase with n . Hence, $\beta^{(1)}(\infty, n)$ strictly decreases with n (or $\gamma^{(1)}(\infty, n)$ strictly increase with n). From Lemma 3.7.2, we see that as q_{EA} and q_R both increase with n , the R.H.S. of the first expression of (3.24) also increases with n . From the monotonicity of $(1 - \gamma)e^{-G(\gamma)}$, we have $\gamma^{(1)}(\infty, n)$ decreasing with n which yields a contradiction.

Assume that $\gamma^{(1)}(\infty, n1) > \gamma^{(1)}(\infty, n2)$ (hence $\beta^{(1)}(\infty, n)$ increases with n). Hence, $q_{EA} = e^{-n^{(1)}\beta^{(1)}(\infty, n)}$ decreases with n . From the second expression of (3.24), we see that if q_{EA} decreases, then $\gamma^{(0)}(\infty, n)$ must strictly increase with n (otherwise, the R.H.S. will decrease with n , and from the monotonicity of the L.H.S., we get a contradiction). Since $\gamma^{(0)}(\infty, n)$ increases with n , q_R decreases with n . Using the fact that q_{EA} and q_R decreases with n and from Lemma 3.7.2, we see that the R.H.S. of the first equation also decreases with n , which implies that $\gamma^{(1)}(\infty, n)$ increases with n , which is a contradiction.

Assume that $\gamma^{(1)}(\infty, n1) = \gamma^{(1)}(\infty, n2)$ (clearly, $\beta^{(1)}(\infty, n) > 0$ for all n). Then q_{EA} decreases with n . So from the R.H.S. of the second expression of (3.24), we need that $\gamma^{(0)}(\infty, n)$ strictly increase with n . So, q_R also decreases with n . Hence, the R.H.S. of the first equation decreases from the Lemma 3.7.2 and hence we obtain a contradiction.

Similarly, if $\gamma^{(0)}(\infty, n1) = \gamma^{(0)}(\infty, n2)$, q_R is constant. Since $e^{-n^{(0)}\beta^{(0)}(\infty, n)}$ is non-increasing, we require that q_{EA} be non-decreasing ($q_R = e^{-n^{(0)}\beta^{(0)}(\infty, n)}q_{EA}$). Hence, we require $\beta^{(1)}(\infty, n)$ strictly decreasing with n . Hence, $\gamma^{(1)}(\infty, n)$ strictly increases with n . Hence, the L.H.S. of the first equation decreases with n . However, since q_R is a constant and q_{EA} is non-decreasing, we have the R.H.S. of the first equation non-decreasing, which is a contradiction.

Hence, $\gamma^{(\cdot)}(\infty, n)$ strictly increases with n . ■

From the above lemma, we can see that $\beta^{(\cdot)}(\infty, n)$ goes to zero as $n \rightarrow \infty$.

Lemma 3.7.12 *In Case 3, with $K = \infty$, $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n) < \frac{1}{p}$ for all n .*

Proof: From (3.24) we can easily see that $(1 - \gamma^{(1)}(\infty, n))e^{-\beta^{(1)}(\infty, n)} \geq (1 - \gamma^{(0)}(\infty, n))e^{-\beta^{(0)}(\infty, n)}$. Since we assumed that the function $G_\infty(\cdot)$ is the same for both

the classes, and since we know that $(1 - \gamma)e^{-G_\infty(\gamma)}$ is a strict monotone decreasing function, we have, $\gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n)$ for all n .

As in Lemma 3.7.8, it can be seen that $\gamma^{(1)}(\infty, n) < \frac{1}{p}$ for all n . Now suppose that, that $\gamma^{(0)}(\infty, n) > \frac{1}{p}$ for all $n \geq N$ for some N ($\gamma^{(\cdot)}(\infty, n)$ are strictly increasing functions of n). With $\beta^{(0)}(\infty, n) = 0$ for all $n \geq N$, the only factor that governs the collision probability of Class 1 and 0 is $(n^{(1)} - 1)\beta^{(1)}$ and $n^{(1)}\beta^{(1)}$. However, we know that $\beta^{(1)}(\infty, n)$ goes to zero, or $\gamma^{(1)}(\infty, n) \rightarrow \gamma^{(0)}(\infty, n)$, which requires $\gamma^{(1)}(\infty, n) > \frac{1}{p}$ for some $n > N'$, leading to a contradiction. Hence, $\gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n) < \frac{1}{p}$ for all n . Also, when $\gamma^{(\cdot)} < \frac{1}{p}$, the inequality between the collision probabilities becomes strict, i.e., $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ (We already know that $\gamma^{(1)}(\infty, n) \leq \gamma^{(0)}(\infty, n)$). The result follows from the (3.24) and the fact that $G_\infty(\gamma)$ is a strictly decreasing function of γ when $0 \leq \gamma \leq \frac{1}{p}$. ■

Lemma 3.7.13 *In Case 3, with $K = \infty$, $n^{(0)}\beta^{(0)}(\infty, n) \rightarrow 0$ as $n \rightarrow \infty$.*

Proof: Since $\beta^{(0)}(\infty, n) \geq 0$

$$1 - e^{-(n^{(1)}-1)\beta^{(1)}(\infty, n)} \leq 1 - e^{-(n^{(1)}-1)\beta^{(1)}(\infty, n) - n^{(0)}\beta^{(0)}(\infty, n)}$$

If $n^{(0)}\beta^{(0)}(\infty, n)$ converges to a positive value, then this inequality becomes strict in the limit. Hence, $\lim_{n \rightarrow \infty} \gamma^{(1)}(\infty, n) < \lim_{n \rightarrow \infty} \gamma^{(0)}(\infty, n)$, which is a contradiction, since both $\gamma^{(1)}(\infty, n)$ and $\gamma^{(0)}(\infty, n)$ tend to $\frac{1}{p}$ as $n \rightarrow \infty$ (this follows since $\gamma^{(\cdot)}(\infty, n) < \frac{1}{p}$ for all n and $\beta^{(\cdot)}(\infty, n)$ tend to 0 as $n \rightarrow \infty$). Hence, $n^{(0)}\beta^{(0)}(\infty, n)$ goes to zero. ■

Using Lemmas 3.7.11 and 3.7.12, we see that $\gamma^{(1)}(\infty, n) < \gamma^{(0)}(\infty, n)$ for all n and $\lim_{n \rightarrow \infty} \gamma^{(\cdot)}(\infty, n) \uparrow \frac{1}{p}$. From the previous Lemma, we see that $n^{(0)}\beta^{(0)}(\infty, n) \rightarrow 0$. From the fixed point equations for AIFS and Lemma 3.7.13, we obtain $\lim_{n \rightarrow \infty} n^{(1)}\beta^{(1)}(\infty, n) \uparrow \ln\left(\frac{p}{p-1}\right)$, completing the proof of Theorem 3.5.4.

Proof of Lemma 3.5.2

Consider the case of finite K with $n^{(1)}$ Class 1 nodes and $n^{(0)}$ Class 0 nodes. The success probability for a Class 1 node is given by (we drop K and n in the notation)

$$G^{(1)}(\gamma^{(1)})\pi(EA)(1 - G^{(1)}(\gamma^{(1)}))^{(n^{(1)}-1)} + \pi(R)(1 - G^{(1)}(\gamma^{(1)}))^{(n^{(1)}-1)}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}$$

and the success probability for a Class 0 node equals

$$G^{(0)}(\gamma^{(0)})\pi(R)(1 - G^{(1)}(\gamma^{(1)}))^{n^{(1)}}(1 - G^{(0)}(\gamma^{(0)}))^{(n^{(0)}-1)}$$

The ratio of throughput of a Class 1 node to a Class 0 node is then given by,

$$\begin{aligned} & \frac{G^{(1)}(\gamma^{(1)})\pi(EA) + \pi(R)(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}}{G^{(0)}(\gamma^{(0)})\pi(R)(1 - G^{(1)}(\gamma^{(1)}))(1 - G^{(0)}(\gamma^{(0)}))^{(n^{(0)}-1)}} \\ &= \frac{\frac{G^{(1)}(\gamma^{(1)})}{(1 - G^{(1)}(\gamma^{(1)}))}(\pi(EA) + \pi(R)(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}})}{\frac{G^{(0)}(\gamma^{(0)})}{(1 - G^{(0)}(\gamma^{(0)}))}\pi(R)(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}}} \\ &= \frac{\frac{G^{(1)}(\gamma^{(1)})}{(1 - G^{(1)}(\gamma^{(1)}))}}{\frac{G^{(0)}(\gamma^{(0)})}{(1 - G^{(0)}(\gamma^{(0)}))}} \left(\frac{\pi(EA)}{\pi(R)(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}} + 1 \right) \end{aligned}$$

Consider the term inside the bracket,

$$\frac{\pi(EA)}{\pi(R)(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}} + 1$$

Let $l = 1$. From (3.5), we see that $\frac{\pi(EA)}{\pi(R)} = \frac{1 - q_R}{q_{EA}}$. Substituting back, we have,

$$\frac{1 - q_R}{q_{EA}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}}} + 1$$

We know that $q_{EA}(1 - G^{(0)}(\gamma^{(0)}))^{n^{(0)}} = q_R$. Hence, the above expression simplifies to

$$\frac{1 - q_R}{q_R} + 1 = \frac{1}{q_R}$$

Hence, the throughput ratio simplifies to

$$\frac{\frac{G^{(1)}(\gamma^{(1)})}{(1-G^{(1)}(\gamma^{(1)}))}}{\frac{G^{(0)}(\gamma^{(0)})}{(1-G^{(0)}(\gamma^{(0)}))}} \left(\frac{\pi(EA)}{\pi(R)(1-G^{(0)}(\gamma^{(0)}))^{n^{(0)}}} + 1 \right) = \frac{\frac{G^{(1)}(\gamma^{(1)})}{(1-G^{(1)}(\gamma^{(1)}))}}{\frac{G^{(0)}(\gamma^{(0)})}{(1-G^{(0)}(\gamma^{(0)}))}} \frac{1}{q_R}$$

Chapter 4

Fixed Point Analysis of IEEE 802.11(e) WLANs with Capture

4.1 Introduction

We consider a single cell IEEE 802.11(e) WLAN with DCF (EDCA) as the medium access protocol. By “single cell”, we imply that the nodes are synchronized and every transmission in the channel is perceived by every receiver as a busy channel. In a *pure collision* channel, an attempt succeeds when there are no simultaneous transmissions in the channel. In Chapters 2 and 3, we studied the saturation throughput performance of IEEE 802.11(e) WLANs for a *pure collision* channel. The *pure collision* channel model is an approximation aimed at simplification. In practice, the power levels of different transmissions as heard at the receiver(s) are different, due to path loss, shadowing and multipath fading. Then, in case of simultaneous transmissions, it is possible that one of the signals has sufficient strength to be decoded by the receiver, i.e., one of the transmissions is able to capture the receiver. This is particularly true when a spread spectrum modulation is used, as is the case in the IEEE 802.11b physical layer. Also, it has been observed in the literature that capture increases system throughput as well as induces unfairness among nodes. Hence, it is important to model capture to accurately represent the physical reality.

In this chapter, we extend the analysis presented in Chapters 2 and 3 to include the

possibility of frame capture at the receiver. We consider a single cell WLAN with IEEE 802.11(e) type nodes. We model backoff differentiation due to b_0, p and K , but we do not model AIFS based service differentiation. We assume that every transmission is heard by every other node and there are no hidden nodes in the system. We consider both the *infrastructure* WLAN model as well as the *ad hoc* WLAN model. Also, we assume that the nodes always have a packet to transmit (i.e., we do saturation throughput analysis).

We believe that, an exact analysis using a complex Markovian model is prohibitive. Hence, our aim in this work is to propose a tractable but accurate model for capture. In Section 4.2, we propose a general framework to model capture in single cell WLANs and obtain the fixed point equations in terms of the collision probabilities and the attempt rates. We then discuss the network scenarios which can be satisfactorily studied using the proposed framework. An important concern is the uniqueness of the solution of the fixed point equations. In Section 4.3, we first show by an example that capture introduces multiple fixed points even when the sufficient conditions for uniqueness (for a *pure collision* channel) from Chapters 2 and 3 hold. In such cases, we observe that the system exhibits multistability and that the framework fails to track the average behaviour. Then, we obtain some sufficient conditions to guarantee a unique solution to the fixed point equations (with capture). Finally, we prove a general uniqueness result for the infrastructure setup with uplink traffic. Section 4.4 summarizes the results in this chapter.

4.1.1 Literature Survey

The idea of capture is that a packet with sufficient strength can be received even in the presence of other overlapping transmissions. Successful capture depends on a variety of factors such as received power, the type of modulation and receiver synchronization techniques. Many models have been developed in the literature to study frame capture probability. The models focus both on physical layer as well as MAC layer techniques. Physical layer models assume that a packet can be successfully captured provided that the power of the desired signal exceeds the power of the interfering signals, or, the cumulative power of the interfering signals, by at least a capture ratio. MAC models assume that the

receiver (randomly) locks on to one of the signals and successful packet reception occurs if the packet is decoded without bit error. In [12], [7], [13], [33] and [22], a variety of capture models have been studied under different channel conditions (Rayleigh, Nakagami fading) and spatial distribution of nodes. Throughput and Delay performance of the wireless systems is obtained by incorporating the capture probabilities into the system equations.

Early works on capture studied the performance of slotted ALOHA systems for different channel and traffic scenarios. In [29], the authors study the capacity of a slotted ALOHA system, with Poisson traffic and Rayleigh fading channel, for coherent and incoherent addition of signals. The throughput performance of slotted ALOHA protocol in the presence of Rayleigh fading and log-normal shadowing was studied in [46]. A similar performance analysis for Nakagami fading was done in [59]. Initial works on 802.11 assumed simple models for the backoff behaviour of the nodes. In [25], the authors study the throughput and delay of CSMA/CA MAC in the presence of Rayleigh fading and shadowing. They assume an infrastructure setup and study the performance of CSMA/CA under different handshake protocols.

Based on the Markov chain model proposed by Bianchi in [18] and the general capture analysis for wireless systems, in [81] and [82], the authors study the performance of single cell IEEE 802.11 WLANs with capture. In [81], the authors obtain the balanced fixed point equations for the infrastructure setup; however, they did not solve the equations but use an approximation (from simulations) to estimate the system performance. Similar analysis for the infrastructure setup under homogeneous assumptions has been done in [69], [72], [71], [22] and [17]. Using an approximation for the attempt probabilities of the nodes, in [22], the authors develop an iterative algorithm to solve the balanced fixed point equations. In our work, we propose to develop the fixed point equations using the generalization proposed in [6] (and developed in Chapters 2 and 3), which makes our work more rigorous and much simpler.

Most works on single cell scenarios have focussed on the infrastructure setup with uplink traffic. Models on ad hoc network have been restrictive and the focus has been more on multihop and hidden node scenarios in such cases. Also, all previous literature

have assumed a homogeneous model for the backoff parameters as well as the channel distribution. The fixed point equations have been analyzed only for the balanced solution without studying the possibility of multiple fixed points. In our work, we generalize the framework and obtain a vector fixed point equation for the system. We study the possibility of multiple solutions to the fixed point equations and provide conditions under which the system has a unique solution.

4.2 Fixed Point Framework with Capture

Consider a single cell WLAN with n IEEE 802.11e nodes using EDCA as the medium access protocol. We assume that the nodes are saturated and denote by \mathcal{R}_i the unique receiver of transmitter i . Let $\boldsymbol{\gamma} := (\gamma_1, \dots, \gamma_n)$ be the vector of the average collision probabilities of the nodes, and $\boldsymbol{\beta} := (\beta_1, \dots, \beta_n)$ be the vector of their average attempt probabilities. Here again, we assume that all rates are conditioned on being in the backoff periods; i.e., we have eliminated all durations other than those in which nodes are counting down their backoff counters. As explained in [6] this suffices to obtain the collision probabilities for a single cell scenario. Define \mathcal{N} as the set of all subsets of $\{1, \dots, n\}$. \mathcal{N} corresponds to the set of all simultaneous transmitters in a slot. For $\alpha \in \mathcal{N}$ (a set of transmitters), define p_α as the probability of the event α in a given slot. We note here that p_α is a function of the attempt probabilities ($\boldsymbol{\beta}$) of the nodes. Define c_α^i as the probability of capture for node i at its receiver \mathcal{R}_i conditioned on the event α (for $i \in \alpha, \alpha \in \mathcal{N}$).

We make the following assumptions in our model as in [18], decoupling of the attempt process of the nodes and the nodes perceive an average collision probability independent of their backoff state. Further, we assume stationary capture probabilities with mean c_α^i . The decoupling approximation is to assume that the aggregate attempt process of the competing nodes is independent of the backoff process of the tagged node. By the constant collision probability assumption, we mean that the aggregate behaviour of the competing nodes is characterized by a constant collision probability in every slot. The

collision probability of a node, γ_i , is now given by, for all $i, 1 \leq i \leq n$,

$$\gamma_i = \sum_{\alpha \in \mathcal{N}_i} p_\alpha \times (1 - c_{i,\alpha}^i) \quad (4.1)$$

where \mathcal{N}_i is the set of all non-empty subsets of $\{1, \dots, i-1, i+1, \dots, n\}$, $c_{i,\alpha}^i$ corresponds to the capture probability of node i in the event $\{i\} \cup \alpha$ and p_α is given by $p_\alpha := \prod_{j \in \alpha} \beta_j \prod_{\{j \notin \alpha, j \neq i\}} (1 - \beta_j)$. We can now obtain the attempt probabilities of the nodes, β_i , from $\beta_i = G_i(\gamma_i)$ (where $G_i(\cdot)$ is the node response formula discussed in Chapter 2, Section 2.2). Equations (4.1) with $\beta_i = G_i(\gamma_i)$ for all $1 \leq i \leq n$ constitute the fixed point equations of the system.

Remark: Similar models have been proposed to study WLANs with capture in [81], [69], [72], [71], [22] and [17]. The major improvement in our model is the use of the analytical expression $G(\cdot)$ for the attempt process of the nodes. Our model is very general as it allows non-homogeneity in the attempt process, $G_i(\cdot)$ for each i , as well as in the channel perceived by the nodes, c_α^i for each i and every α . The non-homogeneity permits us to model IEEE 802.11e EDCA as well as different network scenarios such as the infrastructure and the ad hoc modes of traffic. For example, in an infrastructure setup, only one transmission can be successful in a slot. Hence, we require $\sum_{i:i \in \alpha} c_\alpha^i \leq 1$ for all $\alpha \in \mathcal{N}$. However, for an ad hoc model, we can have simultaneous successful transmissions in a slot, and hence, we allow $\sum_{i:i \in \alpha} c_\alpha^i > 1$. Our formulation allows us to model a variety of such cases.

4.2.1 Validity of the Model

In this section, we elaborate on the network scenarios which can be modeled using our framework. We first discuss the applicability of the framework to the infrastructure and the ad hoc network models. We then note that the decoupling assumption and the constant collision probability assumption puts a restriction on the channel model, mobility and the spatial reuse in the network. The restrictions discussed hold not only for our work but also for the literature described in Section 4.1.1. Simulation results have

been included, when necessary, to explain our viewpoint.

Infrastructure Mode

IEEE 802.11(e) permits two different network models: infrastructure and ad hoc. In the infrastructure mode, all traffic in the network is through the access point (AP). Hence, at any point of time, even with capture, the system supports only one transmission in the network. In terms of c_α^i , this implies that $\sum_{i:i \in \alpha} c_\alpha^i \leq 1$ for all $\alpha \in \mathcal{N}$.

Consider an infrastructure setup with 8 IEEE 802.11e AC_BE stations (access category best-effort as defined in the standard 802.11e) indexed from 1 to 8. We assume that the stations are saturated and there is only uplink traffic (AP does not contend with the stations). We will assume the following capture model in the system. In case of simultaneous transmissions, the node with the least index (assumed to have the best channel) among the simultaneous transmitters captures the receiver with certainty, i.e.,

$$\begin{aligned} c_\alpha^i &= 1 && \text{if } i = \arg \min\{\alpha\} \\ &= 0 && \text{otherwise} \end{aligned}$$

Substituting the capture probabilities in (4.1) and simplifying the equations, we have, for all $1 \leq i \leq n$,

$$\gamma_i = 1 - \prod_{j=1}^{i-1} (1 - \beta_j)$$

Figure 4.1 plots the collision probability and the attempt rate of the 8 nodes from the fixed point analysis and simulation. Notice that the fixed point analysis provides an excellent match to the observed system performance. As expected, the collision probabilities and the attempt rates of the nodes are ordered, with node 1 seeing no collision and node 8 seeing the largest collision in the channel.

In the previous example, $c_\alpha^i \in \{0, 1\}$ for all $\alpha \in \mathcal{N}$ and $i \in \alpha$. However, for a Rayleigh fading channel between the stations and the AP, we do not restrict ourselves to discrete values for c_α^i and let $c_\alpha^i \in [0, 1]$. Consider an infrastructure setup with 8 IEEE 802.11e AC_BE stations. Suppose that the transmitters see a homogeneous Rayleigh

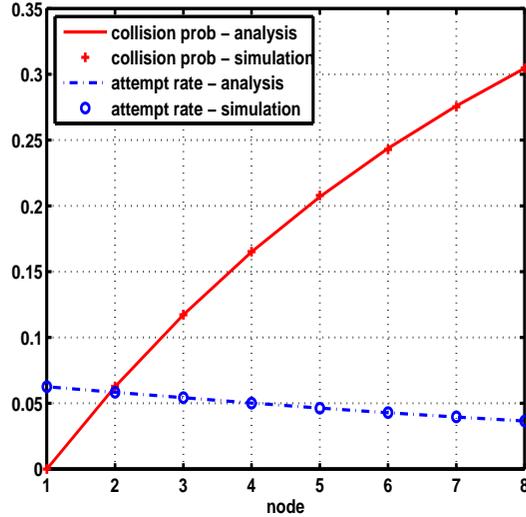


Figure 4.1: Collision probability and attempt rate of 8 IEEE 802.11e AC_BE nodes in an infrastructure setup with uplink traffic. The capture probabilities of the system are given by $c_\alpha^i = 1$ when $i = \arg \min\{\alpha\}$ and $c_\alpha^i = 0$ otherwise.

fading channel towards the AP and our capture model is such that the node with the highest SINR succeeds in packet transmission. Then, $c_\alpha^i = \frac{1}{|\alpha|}$ for all $\alpha \in \mathcal{N}$ and $i \in \alpha$. Figure 4.2 plots the collision probabilities and attempt rates for the scenario from the analysis and simulation. Observe that the collision probabilities (and the attempt rates) are the same for all nodes (as expected for a homogeneous system of nodes).

From the plots, we observe that the general framework proposed in (4.1) models the infrastructure setup very well, and has been reported in [81], [69], [72], [71] and [22] as well. Later, in Section 4.3.3, we study the fixed point equations for the infrastructure setup and show that the fixed point is unique.

Ad hoc Mode

The ad hoc network model permits nodes to communicate directly among themselves. With capture, this implies that the system can support multiple transmissions simultaneously in the network. We restrict ourselves to a single cell model, i.e., whenever there is a transmission in the channel, every other node can perceive it and freezes its backoff. Capture permits simultaneous transmissions in the system only when the nodes attempt in the same slot.

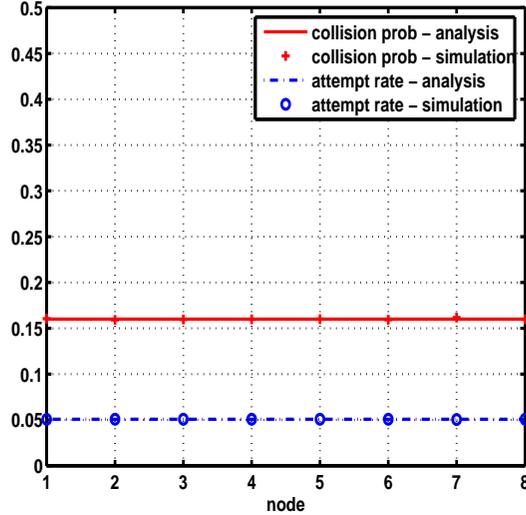


Figure 4.2: Collision probability and the attempt rates of 8 IEEE 802.11e AC_BE nodes in an infrastructure setup with uplink traffic. The capture probabilities of the nodes are given by $c_{\alpha}^i = \frac{1}{|\alpha|}$.

Most works in the literature assume that the proposed framework models the ad hoc network very well. In Section 4.3, we show using an example that in an ad hoc setup, capture can lead to multistability even in single cell systems (similar to those studied in [67]) and the framework fails to characterize the average system performance. The counter-example suggests that careful modeling is required while studying ad hoc networks with capture.

Stationary Capture Probabilities

In equation (4.1), we have assumed a constant collision probability γ_i , for a node i , in every slot. This assumption requires that the capture probabilities be stationary, because, variation in capture probabilities can lead to unacceptable levels of variation in the collision probabilities perceived by the nodes, which may not be tracked by our framework. Changes in capture probabilities of the nodes may be attributed to slow node mobility or slow fade changes in the links. Our framework assumes time homogeneity and does not model such scenarios.

Consider a single cell WLAN with 8 IEEE 802.11e “type” nodes in an infrastructure

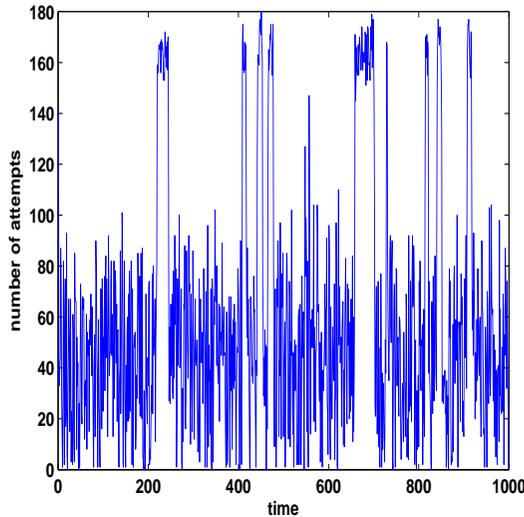


Figure 4.3: Attempt process of a mobile node in real time (over 1000 slot window).

setup with uplink traffic. We consider an exponentially increasing backoff $G(\cdot)$ (like in Chapter 2, equation 2.5) with $b_k = p \times b_{k-1}$, $b_0 = 4$, $p = 2$, $K = 7$. We note that the backoff parameters are very similar to the backoff values defined in the IEEE 802.11e standard. Also, for a pure collision channel and for the backoff response function $G(\cdot)$ defined above, we note that the fixed point equations have a unique solution. We will assume the following node mobility model and channel capture model. At the end of every backoff slot, nodes change their positions with probability 0.0001, and the node locations are chosen uniformly within the cell. The transmission attempt of a node succeeds only when there are no simultaneous transmissions in the channel, or, when the node is the closest to the access point (AP). Figure 4.3 plots the attempt process of a tagged node in real time for the above network. Observe that the attempt rate of a node significantly varies over different time intervals. When the node is the closest to the AP, most of its attempts succeed and the average attempt rate is large (> 100 in the figure). When the tagged node is not the closest to the AP, the probability of collision increases and hence, it decreases its attempt rate (≈ 50 in the figure). The average collision probability and the attempt rate of the nodes from simulations was found to be 0.31 and 0.09 respectively. Substituting 0.31 in the backoff equation, we get $G(0.31) = 0.14$, which is much larger than the average attempt rate obtained from the simulations. The error between the simulation

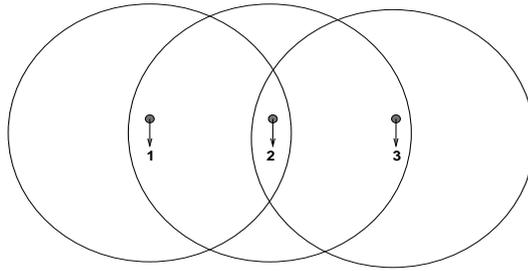


Figure 4.4: A 3 link wireless network. The circles represent the sensing and the interference range of the communication link located at the center of the circle. Links 1 and 3 are hidden from each other, while link 2 interferes with both the links 1 and 3.

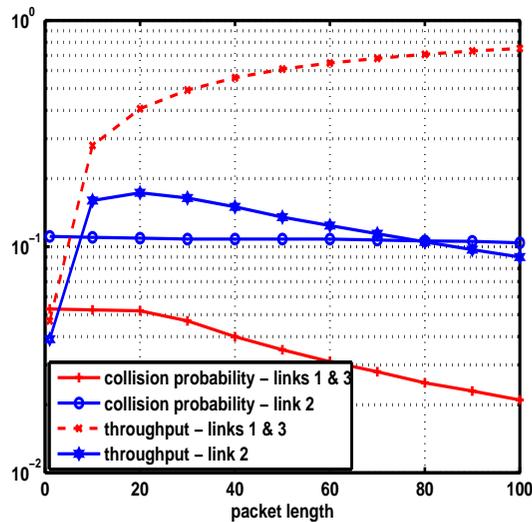


Figure 4.5: Collision probability and long term average throughput from simulation for the example network shown in Figure 4.4.

and the analysis clearly shows that the framework does not support time correlation in the channel.

Multihop and Hidden Nodes

One of the key assumptions in the proposed framework is the decoupling assumption. This assumption works when the nodes are synchronized, i.e., the channel evolution needs to be identical for all the nodes in the network. We observe that in multihop and hidden node scenarios, the channel evolution is not identical, and hence, our framework need not capture the system performance. Consider a simple three link example shown in Figure 4.4. Links 1 and 3 are hidden from each other and they do not interfere. Link 2

interferes with both the links 1 and 3. The fixed point equations of the system are

$$\gamma_1 = \gamma_3 = \beta_2, \gamma_2 = 1 - (1 - \beta_1)(1 - \beta_3)$$

Observe that the fixed point equations are independent of the packet size (packet transmission duration). Figure 4.5 plots the average collision probability and the long term average throughput obtained from simulations for increasing packet size (as a ratio of the backoff slot). Observe that for large packet sizes, the throughput of links 1 and 3 increases, while the throughput of link 2 decreases. Also, the collision probability of links 1 and 3 decreases with increasing packet size. Transmissions of links 1 and 3 are not synchronized. For large packet sizes, one of the two links (1 or 3) freezes the backoff engine of the link 2, thereby decreasing the attempt rate (and throughput) of link 2 and the collision probability of the links 1 and 3. The decoupling approximation, assumed in the framework, does not hold in this scenario.

Our observations show that to model correlations in channel, mobility and scenarios like multihop and hidden nodes, one needs to enhance the framework by incorporating additional features specific to the scenario. We do not study such enhancements in this work, and is beyond the scope of this chapter.

4.3 Uniqueness of the Fixed Point

In this section, we will study the uniqueness of the fixed point for the set of equations (4.1). We will first show, using a counterexample, that capture can lead to multistability (and multiple fixed points) in single cell WLANs, even when the uniqueness conditions for a *pure collision* channel hold. Then, using a contraction mapping, we will obtain sufficient conditions to guarantee a unique solution to the fixed point equations (4.1). Finally, we will prove a general uniqueness result for the infrastructure setup with uplink traffic.

In the previous chapters, we observed the importance of uniqueness of the fixed point for single cell WLANs (for a *pure collision* channel). Multiple fixed points suggested multistability, and further, the fixed point analysis did not capture the average system

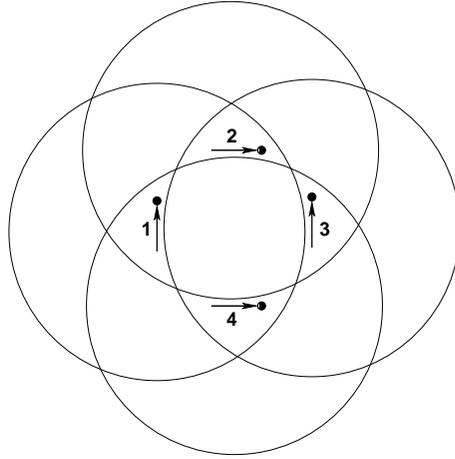


Figure 4.6: A four link wireless network operating as a single cell WLAN. The circles represent the interference range of the communication link located at the center of the circle.

performance as well. The backoff behaviour of the nodes was the reason. We also observed that when the backoff function $G(\gamma)$ is such that $(1 - \gamma)(1 - G(\gamma))$ is not one-to-one, the system may have multiple fixed points. Trivially, for such a $G(\cdot)$, it is easy to construct examples for systems with capture, where the fixed point equations have multiple solutions. We are not interested in such examples, and our focus is to study the impact of capture on multistability, and the interaction between the backoff process and the capture probabilities. Hence, we will always assume that $G(\cdot)$ is monotone and $F(\cdot)$ is one-to-one and decreasing.

4.3.1 Counterexample

Consider an 8 node system comprising four links 1, 2, 3 and 4 as shown in Figure 4.6. All the nodes can hear each other and there are no hidden nodes in the system (a single cell scenario). The links are located such that links 1 and 3 do not interfere with each other, in case of simultaneous transmissions. Similarly, links 2 and 4 do not interfere with each other in case of simultaneous transmissions. The capture probabilities of this system are given by,

$$c_1^1 = 1, c_2^2 = 1, c_3^3 = 1, c_4^4 = 1$$

$$\begin{aligned}
c_{1,2}^1 &= 0, c_{1,2}^2 = 0, c_{1,3}^1 = 1, c_{1,3}^3 = 1, c_{1,4}^1 = 0, c_{1,4}^4 = 0 \\
c_{2,3}^2 &= 0, c_{2,3}^3 = 0, c_{2,4}^2 = 1, c_{2,4}^4 = 1 \\
c_{3,4}^3 &= 0, c_{3,4}^4 = 0 \\
c_{1,2,3} &= 0, c_{1,3,4} = 0, c_{2,3,4} = 0 \\
c_{1,2,3,4} &= 0
\end{aligned}$$

Substituting the above capture probabilities in the fixed point equations (4.1) and simplifying the equations, we get,

$$\begin{aligned}
\gamma_1 &= 1 - (1 - G_2(\gamma_2))(1 - G_4(\gamma_4)) = \gamma_3 \\
\gamma_2 &= 1 - (1 - G_1(\gamma_1))(1 - G_3(\gamma_3)) = \gamma_4
\end{aligned}$$

Suppose that all the nodes use the following backoff parameters, with G modeling the attempt process of the nodes given by,

$$G(\gamma) = \frac{1 + \gamma + \gamma^2 + \gamma^3 + \gamma^4 + \gamma^5 + \gamma^6 + \gamma^7}{2(1 + (2\gamma) + (2\gamma)^2 + (2\gamma)^3 + (2\gamma)^4 + (2\gamma)^5 + (2\gamma)^6 + (2\gamma)^7)}$$

We note here that $G(\gamma)$ is monotone decreasing and $(1 - \gamma)(1 - G(\gamma))$ is one-to-one. Hence, without capture, the system of fixed point equations will have a unique solution. Also, the backoff parameters are similar to the IEEE 802.11e setup, albeit with a smaller CW_{min} . Substituting $G(\cdot)$ in the above equations, we get,

$$\begin{aligned}
\gamma_1 &= 1 - (1 - G(\gamma_2))^2 \\
\gamma_2 &= 1 - (1 - G(\gamma_1))^2
\end{aligned}$$

Figure 4.7 plots $1 - (1 - G(1 - (1 - G(\gamma))^2))^2$ (line with plus) with respect to γ , the intersection of which with the "y=x" line shows the fixed point solutions. Also, plotted in the figure is $(1 - \gamma)(1 - G(\gamma))$ (lines with dots). Observe that even though $(1 - \gamma)(1 - G(\gamma))$ is one-to-one and decreasing, the system allows multiple fixed points. The three fixed points read from the graph for $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are $(0.09, 0.69, 0.09, 0.69)$, $(0.69, 0.09, 0.69, 0.09)$ and

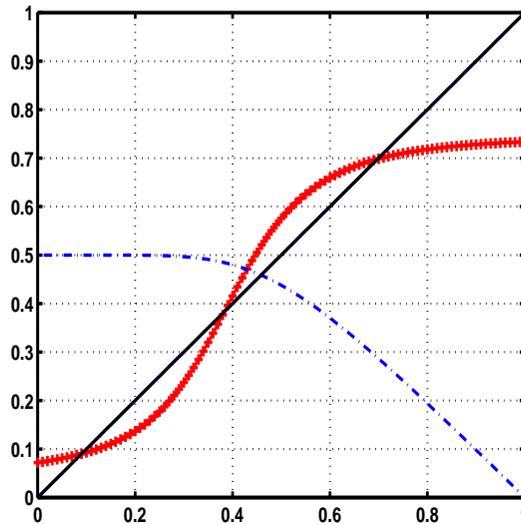


Figure 4.7: Figure plots $1 - (1 - G(1 - (1 - G(\gamma))^2))^2$ (line with plus) and $(1 - \gamma)(1 - G(\gamma))$ (lines with dots) with respect to γ . Observe that the function $(1 - \gamma)(1 - G(\gamma))$ is one-to-one. Yet, the system permits multiple fixed points.

(0.38, 0.38, 0.38, 0.38). Clearly, the fixed point solutions suggest a multistability between the pair of links 1, 3 and 2, 4. Simulating the backoff process of the nodes, we observed multistability in channel access and short-term unfairness in the throughput received. The average collision probability of the nodes (from simulation) was found to be 0.117, whereas, the balanced fixed point from the analysis (Figure 4.7) yields 0.38. Thus, we see that capture induces multistability in single cell scenarios. The impact was observed as multiple fixed points. And, the balanced fixed point did not match the system performance.

The example clearly shows that modeling capture in a single cell setup requires more care, and the framework is not appropriate for every single cell WLAN.

4.3.2 A Contraction Mapping

In the previous section, we observed that the sufficient conditions to guarantee a unique fixed point for a *pure collision* channel model, were not sufficient in the presence of capture. In this section, using contraction mapping, we will obtain sufficient conditions to guarantee uniqueness of the fixed point for the equations (4.1).

Lemma 4.3.1 *Let $\left|\frac{\partial \Gamma_i}{\partial \beta_j}\right| \leq 1$ for all $1 \leq i, j \leq n$. Then the system of equations (4.1) has a unique fixed point, whenever*

$$n \left(\sup_{\{i:1 \leq i \leq n, \gamma:0 \leq \gamma \leq 1\}} G'_i(\gamma) \right) < 1$$

Proof: Let us rewrite the fixed point equations (4.1) in a compact form, in terms of the vector of the attempt probabilities $\boldsymbol{\beta}$ as, $\boldsymbol{\beta} = \mathbf{G}(\boldsymbol{\Gamma}(\boldsymbol{\beta}))$, where $\beta_i = G_i(\Gamma_i(\boldsymbol{\beta})) = G_i(\gamma_i)$. For any two vectors $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$, we have from the mean value theorem,

$$|\Gamma_i(\boldsymbol{\beta}^1) - \Gamma_i(\boldsymbol{\beta}^2)| \leq n \left(\sup_{\{j:1 \leq j \leq n\}} \left(|\beta_j^1 - \beta_j^2| \sup_{\{\beta_j:0 \leq \beta_j \leq 1\}} \left| \frac{\partial \Gamma_i}{\partial \beta_j} \right| \right) \right)$$

Using the hypothesis $\left|\frac{\partial \Gamma_i}{\partial \beta_j}\right| \leq 1$, we get,

$$|\Gamma_i(\boldsymbol{\beta}^1) - \Gamma_i(\boldsymbol{\beta}^2)| \leq n \left(\sup_{\{j:1 \leq j \leq n\}} |\beta_j^1 - \beta_j^2| \right) \quad (4.2)$$

Suppose that $\boldsymbol{\beta}^1$ and $\boldsymbol{\beta}^2$ are distinct fixed point solutions of $\boldsymbol{\beta} = \mathbf{G}(\boldsymbol{\Gamma}(\boldsymbol{\beta}))$. Then,

$$\begin{aligned} |\beta_i^1 - \beta_i^2| &= |G_i(\Gamma_i(\boldsymbol{\beta}^1)) - G_i(\Gamma_i(\boldsymbol{\beta}^2))| \\ &\leq \left(\sup_{\{\gamma:0 \leq \gamma \leq 1\}} |G'_i(\gamma)| \right) |\Gamma_i(\boldsymbol{\beta}^1) - \Gamma_i(\boldsymbol{\beta}^2)| \end{aligned}$$

where the last inequality follows from the mean value theorem. Substituting from (4.2) in the previous inequality, we have,

$$|\beta_i^1 - \beta_i^2| \leq n \left(\sup_{\{j:1 \leq j \leq n\}} |\beta_j^1 - \beta_j^2| \right) \left(\sup_{\{\gamma:0 \leq \gamma \leq 1\}} |G'_i(\gamma)| \right)$$

Taking supremum over i on both the sides, we get,

$$\sup_{\{i:1 \leq i \leq n\}} |\beta_i^1 - \beta_i^2| \leq n \left(\sup_{\{j:1 \leq j \leq n\}} |\beta_j^1 - \beta_j^2| \right) \left(\sup_{\{i:1 \leq i \leq n, \gamma:0 \leq \gamma \leq 1\}} |G'_i(\gamma)| \right)$$

Thus, when

$$n \left(\sup_{\{i:1 \leq i \leq n, \gamma:0 \leq \gamma \leq 1\}} |G'_i(\gamma)| \right) < 1$$

we conclude that $\beta^1 = \beta^2$, or, the set of equations (4.1) has a unique fixed point solution. \blacksquare

We will now show that the inequality $|\frac{\partial \Gamma_i}{\partial \beta_j}| \leq 1$ holds true for our setup. Without AIFS, all the nodes attempt with a constant attempt probability $\beta_i (= G_i(\gamma_i))$ in every slot. Then, p_α is given by,

$$p_\alpha := \left(\prod_{k \in \alpha} \beta_k \right) \left(\prod_{k \notin \alpha} (1 - \beta_k) \right)$$

Differentiating $\gamma_i = \Gamma_i(\beta)$ w.r.t. β_j , we have from (4.1)

$$\begin{aligned} \frac{\partial \Gamma_i}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left(\sum_{\alpha \in \mathcal{N}_i} p_\alpha (1 - c_{i,\alpha}^i) \right) \\ &= \frac{\partial}{\partial \beta_j} \left(\sum_{\alpha \in \mathcal{N}_i, j \in \alpha} p_\alpha (1 - c_{i,\alpha}^i) + \sum_{\alpha \in \mathcal{N}_i, j \notin \alpha} p_\alpha (1 - c_{i,\alpha}^i) \right) \\ &= \frac{\partial}{\partial \beta_j} \left(\beta_j \sum_{\alpha \in \mathcal{N}_i, j \in \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i) + (1 - \beta_j) \sum_{\alpha \in \mathcal{N}_i, j \notin \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i) \right) \\ &= \left(\sum_{\alpha \in \mathcal{N}_i, j \in \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i) \right) - \left(\sum_{\alpha \in \mathcal{N}_i, j \notin \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i) \right) \end{aligned}$$

where $p_{\alpha - \{j\}}$ is the probability of the event $\alpha - \{j\}$ and is not a function of β_j . Since $p_{\alpha - \{j\}} \geq 0$ for all $\alpha \in \mathcal{N}_i$ and the capture probabilities are non-negative, we have,

$$\left| \frac{\partial \Gamma_i}{\partial \beta_j} \right| \leq \left(\sum_{\alpha \in \mathcal{N}_i, j \in \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i), \sum_{\alpha \in \mathcal{N}_i, j \notin \alpha} p_{\alpha - \{j\}} (1 - c_{i,\alpha}^i) \right)$$

But $(1 - c_{i,\alpha}^i) \leq 1$ and $\sum_{\alpha \in \mathcal{N}_i, j \in \alpha} p_{\alpha - \{j\}} \leq 1$ for all i and α . Hence, we have,

$$\left| \frac{\partial \Gamma_i}{\partial \beta_j} \right| \leq 1$$

Theorem 4.3.1 summarizes the above discussions.

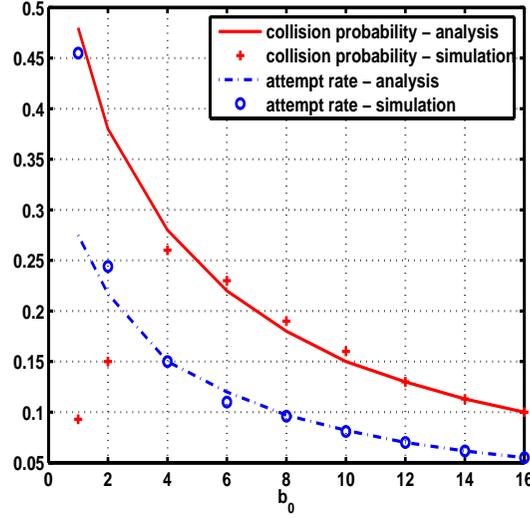


Figure 4.8: Collision probability and attempt rate, from analysis (balanced fixed point) and simulations for the example network shown in Figure 4.6. $b_k = p \times b_{k-1}$, $p = 2$ and $K = 7$.

Theorem 4.3.1 For n IEEE 802.11e type nodes, with G_i defined as in Chapter 2, equation (2.1), $b_{k,i} = b_{0,i}p_i^k$, $K_i \geq 1$, $p_i \geq 2$, and $n < \max_i \left\{ \frac{b_{0,i}}{2p_i} \right\}$, the system of equations (4.1) has a unique fixed point.

Proof: For $G_i(\gamma)$ defined as in equation (2.1) with $b_{k,i} = b_{0,i}p_i^k$, $K_i \geq 1$ and $p_i \geq 2$, lemma 2.5.3 from Chapter 2 showed that

$$\sup_{\{\gamma: 0 \leq \gamma \leq 1\}} |G'_i(\gamma)| \leq \frac{2p_i}{b_{0,i}}$$

Hence, from Lemma 4.3.1, we see that the set of fixed point equations (4.1) has a unique fixed point whenever $n < \max_i \left\{ \frac{b_{0,i}}{2p_i} \right\}$. ■

Remarks 4.3.1

1. Figure 4.8 plots the collision probability and the attempt rate, from analysis (balanced fixed point) and simulations, for the example network shown in Figure 4.6 ($n = 4$). We have assumed the same capture model (and probabilities) discussed in the example. $G(\cdot)$ is defined as in (2.1), and we have assumed $b_k = p \times b_{k-1}$, $p = 2$

and $K = 7$ in the example. As expected, for large b_0 , we see an excellent match between the analysis and the simulations.

2. In [26], the authors study the fixed point equations of a single cell IEEE 802.11 WLAN for a *pure collision* channel, using contraction mapping techniques. They use a game-theoretic model to study the backoff behaviour of the nodes. In our work, we have used a renewal argument from [6] and have extended the results to WLANs with capture.

4.3.3 Infrastructure Mode with Uplink Traffic

In Section 4.3.2, we obtained a sufficient condition to guarantee uniqueness of the fixed point of (4.1). In this section, we will focus on the infrastructure setup with uplink traffic and homogeneous channel conditions. We will show that, for the infrastructure setup, the sufficient conditions for a *pure collision* channel guarantee a unique solution for the fixed point equations (4.1) *even with capture*.

Consider n homogeneous nodes in an infrastructure setup with $G_i(\gamma) = G(\gamma)$ for all $1 \leq i \leq n$. We assume that the function $G(\gamma)$ is monotone decreasing and the function $F(\gamma) := (1 - \gamma)(1 - G(\gamma))$ is one-to-one and decreasing. Also, we assume that the nodes are saturated and there is only uplink traffic. We will make the following assumption on the channel.

Assumption 1 (Homogeneous Capture Model) *There is a sequence of numbers $c_k, k \geq 0, 0 \leq c_k \leq 1$, such that for each $\alpha \in \mathcal{N}_i, c_{i,\alpha}^i = c_{|\alpha|}$. $c_0^i := 1$.*

The assumption states that the capture probabilities are independent of the competing nodes (the event α) and depend only on the number of competing nodes ($|\alpha|$). The above assumption could model the situation in which the nodes are all the same distance from the AP and have identical Rayleigh fading channel.

Using the framework developed in Section 4.2 and the homogeneous capture assumption, the collision probabilities of the STAs are now given by, for all $1 \leq i \leq n$,

$$\gamma_i := 1 - \left(\prod_{j \neq i} (1 - \beta_j) + c_1 \sum_{j \neq i} \beta_j \prod_{k \notin \{i,j\}} (1 - \beta_k) + c_2 \sum_{j \neq i, k \neq i, k > j} \beta_j \beta_k \prod_{m \notin \{i,j,k\}} (1 - \beta_m) + \dots \right) \quad (4.3)$$

where $\beta_i = G(\gamma_i)$.

Lemma 4.3.2 *If $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ is a fixed point solution to (4.3), then $\gamma_i = \gamma_{i'}$ for all $1 \leq i, i' \leq n$.*

Proof: Rearranging (4.3) and multiplying $(1 - \beta_i)$ on either side of the equation, we get, for $1 \leq i \leq n$,

$$(1 - \gamma_i)(1 - \beta_i) = \prod_j (1 - \beta_j) \left(1 + c_1 \sum_{k \neq i} \frac{\beta_k}{(1 - \beta_k)} + c_2 \sum_{k \neq i, l \neq i, l > k} \frac{\beta_k}{(1 - \beta_k)} \frac{\beta_l}{(1 - \beta_l)} + \dots \right)$$

For any two nodes i and i' , $1 \leq i, i' \leq n$,

$$(1 - \gamma_i)(1 - \beta_i) - (1 - \gamma_{i'})(1 - \beta_{i'}) = \prod_j (1 - \beta_j) \left(c_1 \left(\frac{\beta_{i'}}{(1 - \beta_{i'})} - \frac{\beta_i}{(1 - \beta_i)} \right) + c_2 \left(\frac{\beta_{i'}}{(1 - \beta_{i'})} - \frac{\beta_i}{(1 - \beta_i)} \right) \sum_{k \notin \{i, i'\}} \frac{\beta_j}{(1 - \beta_j)} + \dots \right)$$

Suppose that $\gamma_i < \gamma_{i'}$. Then $F(\gamma_i) = (1 - \gamma_i)(1 - G(\gamma_i)) > (1 - \gamma_{i'})(1 - G(\gamma_{i'})) = F(\gamma_{i'})$ (since $F(\cdot)$ is one-to-one and also decreasing). Hence, the left hand side of the above equation is strictly greater than 0. Consider the R.H.S., we have $\prod_j (1 - \beta_j) \geq 0$ and $c_1, c_2, \dots \geq 0$. When $\gamma_i < \gamma_{i'}$, then $\beta_i = G(\gamma_i) \geq G(\gamma_{i'}) = \beta_{i'}$. Hence, $\frac{\beta_i}{(1 - \beta_i)} \geq \frac{\beta_{i'}}{(1 - \beta_{i'})}$, or the R.H.S. is non-positive, which yields a contradiction. Since i and i' were chosen arbitrarily, we see that $\gamma_i = \gamma_{i'}$ or the fixed point solution for the system is balanced. ■

Let us denote this common value by γ . With the observation that the fixed point solution for the collision probabilities is balanced, it follows as well that the attempt probabilities β_i , are all equal; denote the common value by $\beta (= G(\gamma))$. The collision

probability equation (4.3) now simplifies to

$$\gamma = \sum_{k=1}^{n-1} \frac{(n-1)!}{(n-1-k)!k!} \beta^k (1-\beta)^{n-1-k} (1-c_k) \quad (4.4)$$

Thus the multidimensional fixed point equation becomes a one-dimensional equation $\gamma = \Gamma(G(\gamma))$. To establish uniqueness, and knowing that $G(\cdot)$ is nonincreasing, it suffices to show that (4.4) is nondecreasing in its argument. We will show that if c_k are nonincreasing with k , then (4.4) is nondecreasing with β , which will establish that the fixed point is unique.

Lemma 4.3.3 *When c_k are nonincreasing, (4.4) is nondecreasing with β .*

Proof: Since the binomial distribution for $(n-1)$ “trials” with “success” probability β is the convolution of $(n-1)$ Bernoulli distributions each with success probability β , and the Bernoulli distribution is obviously stochastically increasing with β , it follows that the binomial distribution is stochastically increasing with β . Hence (4.4) is the expectation of an increasing function of k (i.e., $(1-c_k)$) with respect to the binomial distribution that is stochastically increasing with the parameter β . Hence, (4.4) is nondecreasing with β . ■

We now have the following general result.

Theorem 4.3.2 *Let $G(\gamma)$ be a monotone decreasing function and $F(\gamma)$ be one-to-one and decreasing. Then there exists a unique fixed point solution for the system of equations (4.3) when c_k are nonincreasing. Further the fixed point is balanced.*

Proof: Brouwer’s fixed point theorem guarantees the existence of a fixed point solution to (4.3). The uniqueness of the fixed point follows from Lemma 4.3.2 and Lemma 4.3.3. ■

Corollary 4.3.1 *For a single cell IEEE 802.11(e) EDCA WLAN with homogeneous nodes and saturated queues, all sending uplink data to the AP, the fixed point equations have a unique solution.* ■

4.4 Summary

In this chapter, we have proposed a general framework to study single cell IEEE 802.11(e) WLANs with the possibility of frame capture at the receiver. We have incorporated a simplification proposed in [6] to model the attempt process of IEEE 802.11e nodes. This allows us to study a variety of network scenarios including non-homogeneity in the backoff parameters of the nodes (IEEE 802.11e EDCA). We discussed in detail, the traffic scenarios and the channel conditions that can be studied using the proposed framework. One of the main contributions in this chapter is to identify the possibility of multistability due to capture. We show using an example that capture can lead to multistability among the nodes, causing short-term unfairness in the throughput received. We observed that the set of equations characterizing such systems can have multiple solutions, and the balanced fixed point fails to capture the average system behaviour. Hence, we provide sufficient conditions that guarantee a unique solution to the fixed point equations. In particular, we show that for homogeneous IEEE 802.11e nodes in an infrastructure setup with uplink traffic, the fixed point equations have a unique solution.

Part II

Topics in Resource Optimization in Wireless Networks

Chapter 5

Spatial Reuse and Cooperative Communication in Dense Wireless Networks

5.1 Introduction

We consider a wireless network comprising n nodes confined to a two dimensional region of fixed area A (independent of n). Such networks are called dense or fixed SNR networks, because, the attenuation between any transmitter-receiver pair is lower bounded by a positive quantity, which is a function of the diameter of the area A . Source-destination (s-d) pairs are chosen randomly (as in the Gupta-Kumar random traffic model, [50]) and the s-d pairs communicate by sharing the common wireless channel. For an average power constraint p at a node, we consider a total network average power constraint, \bar{P} , where $\bar{P} = np$. For a realistic interference and path loss model, we are interested in the scaling of the aggregate end-to-end throughput between the s-d pairs with respect to the network power constraint \bar{P} , and the number of nodes n .

Using a far-field path loss model of $\frac{1}{d^\alpha}$ for every transmitter-receiver separation of d , in [50], Gupta and Kumar showed that the end-to-end throughput of dense wireless networks scales as $\Theta\left(n^{\frac{1}{2}}\right)$. It was first observed in [49] that $\Theta\left(n^{\frac{1}{2}}\right)$ scaling is not feasible

in realistic scenarios, as the far-field path loss model (used in [50]) assumes a channel power gain greater than unity for very small d . In our work, we study the scaling laws of dense wireless networks with a realistic path loss model and observe that it depends not only on the number of nodes (n), but also on the network power constraint (\bar{P}). Our main result is that the end-to-end throughput of dense networks scales only as $\Theta(\log(\bar{P}))$ (or as $\Theta(\log(n))$, when $\bar{P} = np$ for a fixed p), due to interference from simultaneous transmitters and bounded distance between any transmitter-receiver pair. This contrasts with the $\Theta(n^{\frac{1}{2}})$ scaling achievable for an extended network, where the size of the network scales as n (see for e.g., [50] and [9]). Viewed differently, the logarithmic scaling of the aggregate end-to-end throughput follows from the fact that the maximum achievable bit-rate in the network scales only as $\Theta(\log(\bar{P}))$ or $\Theta(\log(n))$, and not as $\Theta(n)$ (as in extended networks).

The logarithmic scaling, for n tending to infinity, or, for very large \bar{P} , is achieved using direct communication between the source-destination pairs, without any spatial reuse. However, better scaling results are achievable for small and moderate \bar{P} , by using spatial reuse, multihopping or other communication techniques. For the path loss model of $\frac{1}{d^\alpha}$, [50] showed that spatial reuse and multihopping achieves an end-to-end throughput of $\Theta(n^{\frac{1}{2}})$. A recent result, [60], achieved $\Theta(n^{\frac{2}{3}})$ throughput using cooperative communication techniques, in a rich scattering environment. The idea is to organize nodes into clusters which would communicate data for the nodes in a single hop. Using the cooperative communication technique proposed in [60], and by implementing a hierarchy, [9] obtained a $\Theta(n)$ throughput for dense wireless networks. The above results (as reported in [50], [60] and [9]) are not feasible for a realistic path loss scenario, and the scaling fails when the nodes become sufficiently close. While it is true that the scaling does not hold for n tending to infinity, we are interested in understanding the feasibility of the communication strategies for sufficiently large n (when the path loss model of $\frac{1}{d^\alpha}$ still holds). For such a scenario (when the path loss model still holds for the given area A and a node density n), we observe that the amount of spatial reuse feasible in the network is limited by the diameter of the network. In fact, we show that the spatial reuse achievable in the

network is inversely proportional to the end-to-end path loss in the network. This puts a restriction on the gains achievable using cooperative communication techniques discussed in [60] and [9], as they rely on direct communication over long distances in the network. We observe that, while spatial reuse and multihopping (as reported in [50]) can provide throughput enhancements for sufficiently large n , even in realistic scenarios, cooperative communication gains (as reported in [60] and [9]) may not be achievable.

5.1.1 Outline of the Chapter

In Section 5.2, we define the dense wireless network model, the realistic interference and path loss model and the objective function. In Section 5.3, we show that the aggregate throughput of a dense network scales only as $\Theta(\log(\bar{P}))$ or $\Theta(\log(n))$. We discuss the feasibility of spatial reuse and cooperative communication for practical dense wireless networks in Section 5.4. We summarize the results in Section 5.5.

5.2 Network Model and Assumptions

We consider a wireless network comprising n nodes distributed over a two dimensional region of fixed area A .

- $\frac{n}{2}$ source-destination pairs are formed in the network, with each node belonging to a distinct s-d pair. The s-d pairs are chosen randomly such that the mean s-d pair distance is $O(1)$, with respect to the diameter of the network.
- The s-d pairs communicate by sharing the common wireless channel. The channel power gain between any transmitter-receiver pair is assumed fixed, and is determined by the path loss between them.
- We consider a total network average power constraint \bar{P} , accounting only for the transmit power of all the nodes in the network. Further, the nodes have an individual average power constraint p , that is related to \bar{P} as $\bar{P} = np$. In our work, we assume

that p is fixed for a given scenario, and hence, the network power constraint \bar{P} scales as n . We do not model a maximum power constraint per node.

- We assume that the system is slotted and nodes communicate over the slots of fixed duration. We consider a realistic physical model of interference (SINR based) in the network. When the nodes use single user decoding transceivers, we assume that the bit rate achieved between a transmitter and a receiver is given by Shannon's formula, $C = \log_2(1 + \text{SINR})$ bits per symbol.
- When the nodes communicate cooperatively, we assume that the nodes are synchronized without any additional overheads.

5.2.1 Path Loss Model

In [50], it was assumed that the power gain between a transmitter and a receiver scaled with the distance d as $\frac{1}{d^\eta}$, where $\eta > 2$ is the path loss exponent. While this holds true for far-field distances, the above model is not appropriate when the receiver is very close to the transmitter. In our work, we use a generalized model in which the channel power gain between a node pair (i, j) is $\alpha_{i,j}$, where $0 < \alpha_A \leq \alpha_{i,j} \leq 1$; α_A is the minimum channel power gain between any transmitter-receiver pair, and is related to the diameter of the network, d_A , as $\alpha_A = \frac{1}{d_A^\eta}$. The assumption $\alpha_{i,j} \leq 1$ implies that a receiver cannot receive power more than the power transmitted.

5.2.2 Objective

Our objective is to study the scaling of the aggregate end-to-end throughput of the dense wireless network for the interference and path loss model discussed above. We study the scaling laws for two different network power constraint regimes - large \bar{P} (in terms of \bar{P}) and moderate \bar{P} (in terms of n). We consider spatial reuse, multihopping and cooperative communication as the strategies used in the network.

5.3 Scaling Laws for Large \bar{P}

In this section, we will obtain scaling laws of dense wireless networks for large \bar{P} . Suppose that the source-destination pairs are chosen arbitrarily (instead of randomly, as stated earlier), such that the s-d pairs are chosen as close as possible. Then, the aggregate end-to-end throughput achievable in this scenario is the same as the maximum bit rate achievable in the network, or the maximum spatial reuse feasible. Clearly, the bit rate achieved in this scenario, upper bounds the bit rate achieved for the random traffic model (notice that the first phase in every communication strategy studied in [50], [60] and [9] is spatial reuse). Now, we assume that the nodes use single user decoding receivers, treating every simultaneous transmission (other than the intended one) as interference. We will now upper bound the bit rate achievable in this scenario.

5.3.1 An Upper Bound on the Network Throughput

Consider a slot t , when node i , $1 \leq i \leq n$, transmits with power $P_i(t)$, and the transmit powers are such that they satisfy a network power constraint, $\sum_{i=1}^n P_i(t) \leq \bar{P}(t)$. For ease of notation, we will omit the index t now, and include it again later (at the end of this section). The SINR achievable (in slot t) at the receiver of the transmitter i is bounded above by

$$SINR \leq \frac{\alpha_i P_i}{N + \sum_{\{j \neq i\}} \alpha_j P_j}$$

where α_i and α_j are the constant gains at the receiver from the transmitters i and j and N is the noise power. Then, it follows from the path loss model, that the SINR is bounded above by

$$SINR \leq \frac{P_i}{N + \alpha_A \sum_{\{j \neq i\}} P_j}$$

(since $0 \leq \alpha_A \leq \alpha \leq 1$). For an allocated total network power of \bar{P} , an optimal power allocation (that maximizes throughput) must satisfy $\sum_i P_i = \bar{P}$. Hence, using the equality

$\sum_i P_i = \bar{P}$ in the above expression, we have,

$$SINR \leq \frac{P_i}{N + \alpha_A(\bar{P} - P_i)}$$

Now, the maximum throughput achievable in the network is bounded above by

$$C(\alpha_A) := \sum_i \log \left(1 + \frac{P_i}{N + \alpha_A(\bar{P} - P_i)} \right)$$

We denote the above expression as $C(\alpha_A)$, denoting the dependence on the parameter α_A . We will now maximize $C(\alpha_A)$ by optimizing the above expression for P_i , i.e., we will maximize $C(\alpha_A)$ subject to the power constraint $\sum_i P_i \leq \bar{P}$.

5.3.2 The Optimization Problem

Define $f(P) := \log \left(1 + \frac{P}{N + \alpha_A(\bar{P} - P)} \right)$. Then the optimization problem we are interested in can be written as

$$\max \sum_i f(P_i) \tag{5.1}$$

subject to the power constraint

$$\sum_i P_i \leq \bar{P}$$

Lemma 5.3.1 $f(P)$ is monotone increasing with P for $0 \leq P \leq \bar{P}$. ■

Proof: See Appendix 5.6.1. ■

Lemma 5.3.2 Suppose that $2\alpha_A - 1 > 0$. Then, f is convex in P for $0 \leq P \leq \bar{P}$. Further, the solution for the optimization problem in (5.1) is to allot all the power to a single transmitter. And the optimal value for the objective function in (5.1) is $f(\bar{P})$.

Proof: See Appendix 5.6.1. ■

Lemma 5.3.3 Suppose that $2\alpha_A - 1 < 0$. For large \bar{P} , there exists a P' , $0 \leq P' \leq \bar{P}$, such that $f(P)$ is concave upto P' and convex thereafter.

Proof: See Appendix 5.6.1. ■

$f(\bar{P}) = \log\left(1 + \frac{\bar{P}}{N}\right)$, increases to infinity with \bar{P} , and $f'(0) = \frac{1}{N + \alpha_A \bar{P}}$, decreases to 0 as \bar{P} increases. Now, observe that, for large \bar{P} , we have $f'(0) \leq \frac{f(\bar{P})}{\bar{P}}$. The following lemma upper bounds $f(P)$ for all $0 \leq P \leq \bar{P}$.

Theorem 5.3.1 *Suppose that $2\alpha_A - 1 < 0$. For large \bar{P} , $f(P) \leq \frac{f(\bar{P})}{\bar{P}}P$ for all $0 \leq P \leq \bar{P}$.*

Proof: See Appendix 5.6.1. ■

Define $g(P) := f(\bar{P})\frac{P}{\bar{P}}$. Let $\{\tilde{P}_i\}$ be an optimal solution for the optimization problem (5.1). From Theorem 5.3.1 and the definition of $g(\cdot)$, we have,

$$\sum_i f(\tilde{P}_i) \leq \sum_i g(\tilde{P}_i)$$

whenever \bar{P} is large enough. Observe that,

$$\sum_i g(\tilde{P}_i) = \sum_i \frac{f(\bar{P})}{\bar{P}} \tilde{P}_i = \frac{f(\bar{P})}{\bar{P}} \sum_i \tilde{P}_i = \frac{f(\bar{P})}{\bar{P}} \bar{P} = f(\bar{P})$$

This implies that for large \bar{P} , $f(\bar{P}) \geq \sum_i f(P_i)$, for any power allocation, or $f(\bar{P})$ is the optimal solution. Thus, $\log\left(1 + \frac{\bar{P}}{N}\right)$ is an upper bound on the network throughput. This implies that the maximum achievable bit rate for a dense network with arbitrary s-d pairs is $O(\log(\bar{P}))$.

Remark:

1. We have only shown that for a per slot network power constraint $\bar{P}(t)$, the aggregate bit rate scales as $\log(\bar{P}(t))$. It is now straightforward to extend the above results to a sequence of network power constraints $\{\bar{P}(t), t = 1, 2, \dots\}$ which satisfy $\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \bar{P}(i) \leq \bar{P}$.
2. We have shown that the aggregate bit rate of an arbitrary network scales as $O(\log(\bar{P}))$ (this is also the maximum bit rate achievable in the network). Hence, the aggregate end-to-end throughput of a random network (as defined in Section 5.2) can scale only as $O(\log(\bar{P}))$.

3. Also, observe that the above results depend only on α_A and \bar{P} , but are independent of the number of the nodes in the network (and hence, on the spacing between the nodes).
4. We call all \bar{P} greater than the threshold (from Theorem 5.3.1) for which the logarithmic scaling holds, as large power regime. And networks with total power constraint \bar{P} lesser than this threshold are called moderate power networks.
5. For an extended network, where the network size scales with n , the path loss from the farthest node decreases to 0. For $\eta > 2$, [50] showed that the cumulative interference from simultaneous transmitters can then be bounded, thus achieving $\Theta(n)$ aggregate bit rate with spatial reuse. Using multihopping strategy, [50] achieved $\Theta\left(n^{\frac{1}{2}}\right)$ end-to-end throughput for extended networks.
6. Consider a simple TDM scheme, where each node transmits with power $\bar{P} = np$, to its intended destination in its slot. Since a node gets access to the channel once in every n slots, the average power per node is $\frac{\bar{P}}{n} = p$. And the achieved bit rate in the proposed scheme scales as $\log(n)$. This proves the achievability of $\Theta(\log(n))$ scaling for dense wireless networks.

The following theorem summarizes the above arguments.

Theorem 5.3.2 *The aggregate end-to-end throughput of a dense wireless network scales as $\Theta(\log(\bar{P}))$, where \bar{P} is the network average power constraint. In terms of the number of nodes, n , the maximum achievable throughput of a dense network scales as $\Theta(\log(n))$.*

■

5.4 Spatial Reuse and Cooperative Communication

For dense wireless networks, [50], [60] and [9] achieved $\Theta\left(n^{\frac{1}{2}}\right)$, $\Theta\left(n^{\frac{2}{3}}\right)$ and $\Theta(n)$ end-to-end throughput respectively, by using a far field path loss model (with a path loss of $\frac{1}{d^\eta}$ for any transmitter-receiver separation of d). As observed in [49], this model requires power amplification by the channel for sufficiently small values of d , and hence, is not practical.

The scaling fails when the nodes become sufficiently close, i.e., when n tends to infinity. In this section, we are interested in understanding the feasibility of the communication strategies discussed in [50] (spatial reuse and multihopping), [60] (spatial reuse, multihopping and cooperative communication) and [9] (spatial reuse, multihopping, cooperative communication and hierarchy) for sufficiently large n , when the path loss model of $\frac{1}{d^n}$ still holds.

For example, consider a $1\text{Km} \times 1\text{Km}$ planar area, with a million nodes arranged in a square grid with a minimum spacing of 1 meter between them. For a carrier frequency of 3 GHz, the carrier wavelength is around $0.1m$ much smaller than the node separation of 1m. The path loss model holds for this deployment, and hence, we could expect that such results as spatial reuse (of $\Theta(n)$), multihopping (of $\Theta\left(n^{\frac{1}{2}}\right)$) and cooperative communication (of $\Theta\left(n^{\frac{2}{3}}\right)$ or $\Theta(n)$) hold approximately (for e.g., with some probability). Observing that spatial reuse is essential to every communication strategy (studied in [50], [60] and [9]), in this context, we will study the feasibility of $\Theta(n)$ spatial reuse in the network, and the impact it has on using cooperative communication techniques.

The following simple calculations given below show that in order to support a spatial reuse of $\Theta(n)$, the network size must be at least as large as $\Theta\left(n^{\frac{1}{n}}\right)$. Let us fix the SINR requirement for point-to-point communication to β , independent of the number of nodes and the dimensions of the network. Suppose that all the transmissions involve constant transmit power p . Let $S(n, A)$ denote the spatial reuse achievable in the network with n nodes while supporting a SINR of β . Then, this implies that the maximum interference gain observed at any receiver needs to be bounded above, as seen below.

$$\beta \leq \text{SINR} \leq \left(\frac{p}{N + p(S(n, A) - 1)\alpha_A} \right)$$

This implies that

$$(S(n, A) - 1)\alpha_A \leq \gamma$$

for some constant, γ , independent of n or p . Absorbing the constants and simplifying the

expression, we have,

$$S(n, A)\alpha_A \leq 1$$

Hence, to achieve a spatial reuse of $\Theta(n)$, we require,

$$\Theta(n)\alpha_A \leq 1 \tag{5.2}$$

In terms of d_A , we have,

$$\Theta(n)\frac{1}{d_A^\eta} \leq 1$$

or,

$$\Theta(n) \leq d_A^\eta$$

The above expression implies that the total spatial reuse feasible in the network is bounded by the dimensions of the network. In other words, to support a spatial reuse of $\Theta(n)$, the dimensions of the network should scale at least as $\Theta\left(n^{\frac{1}{\eta}}\right)$, or the area A should be as large as $\Theta\left(n^{\frac{2}{\eta}}\right)$.

Viewed differently, the end-to-end path loss between any source-destination pair (according to the random traffic model) will scale at least as $\Theta\left(\frac{1}{n}\right)$ (from Equation (5.2)), when the spatial reuse scales as $\Theta(n)$. The cooperative communication models described in [60] and [9] (see Figure 5.1), require the source cluster (the biggest cluster containing the source node) and the destination cluster (the biggest cluster containing the destination node) to communicate directly, using cooperative communication involving the nodes in the cluster. For a spatial reuse of $\Theta(n)$, and the number of nodes in the cluster M ($M < \Theta(n)$, for e.g., in [60], $M = n^{\frac{2}{3}}$), an upper bound on the maximum bit rate achievable in the cooperative communication phase is,

$$n \log \left(1 + \frac{p\alpha_A}{N} \right) \leq n \log \left(1 + \frac{p}{\Theta(n)N} \right)$$

where we have modeled the cooperative communication phase as comprising of n parallel independent channels, with only the path loss between a transmitter-receiver pair in the cluster. Observe that the throughput, as given by the above expression does not scale as

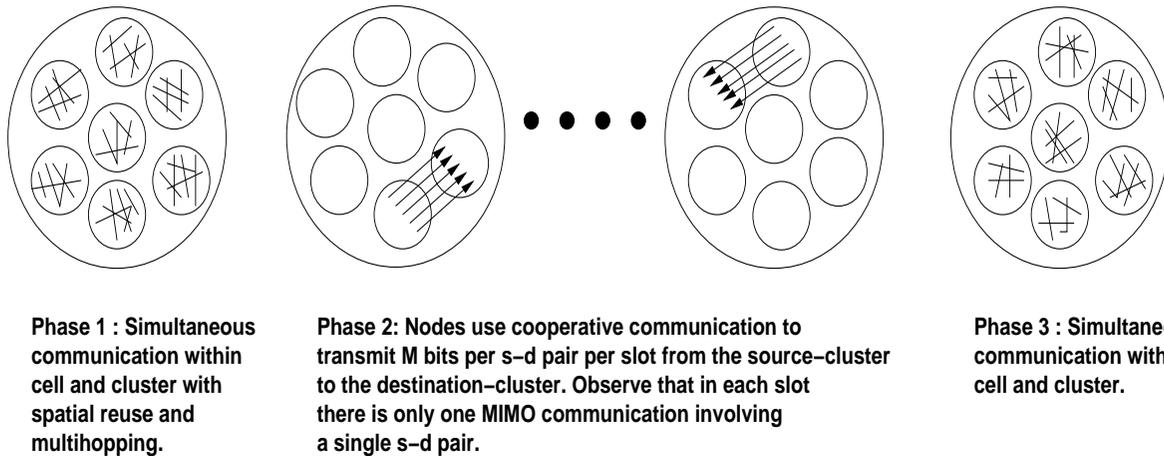


Figure 5.1: Cooperative communication strategy in [60]: The first phase of any cooperative communication strategy is spatial reuse and multihopping. Nodes cluster among themselves (with M nodes per cluster) and communicate information within their cluster. In the second phase of the cooperative communication strategy, clusters communicate directly with each other. There is no spatial reuse, however, M communications happen simultaneously (between one cluster and the other). In the third phase of the communication, the nodes in the cluster use spatial reuse and multihopping again to transfer the data to the intended destination node in the cluster.

$\Theta(n)$. In fact, it is bounded above by a constant. The key observation is that the path loss that permits spatial reuse in the channel is so restrictive that it is unable to support long distance MIMO communications. It is easy to verify from the above formulation that, for any power allocation with a total network power of np , the achievable throughput using cooperative communication (as reported in [60] and [9]) is bounded by a constant, independent of n .

Returning to the example of a million nodes arranged in a grid in a $1\text{Km} \times 1\text{Km}$ planar area, we see that, while spatial reuse and multihopping may increase the performance of the system (as compared to direct transmissions involving the s-d pairs), cooperative communication does not. In other words, while spatial reuse and multihopping can co-exist, even in a practical scenario, for sufficiently large n , spatial reuse and cooperative communication (as reported in [60] and [9]) cannot. The direct communication between the source-destination clusters should be avoided in order to enhance the throughput in realistic scenarios. However, by restricting the communication distance between the clusters, we lose throughput due to multihopping costs.

5.5 Summary

The important feature of a dense network, as compared to an extended network is the positive interference due to a simultaneous transmission any where in the network. We have observed that this implies that the scaling results are a function of both the number of nodes and the network power. More specifically, for large power networks, we observe that the achievable throughput scales only as $\Theta(\log(\bar{P}))$, irrespective of the number of nodes in the network. However, for moderate \bar{P} , when spatial reuse may be efficient, we showed that spatial reuse is restricted by the network size, which affects the gains achievable using cooperative communication techniques.

5.6 Appendix

5.6.1 Proofs of Theorems and Lemmas

Proofs for Section 5.3

Lemma 5.3.1 : $f(P)$ is monotone increasing with P for $0 \leq P \leq \bar{P}$.

Proof: Differentiating $f(P)$ with respect to P , we have,

$$\begin{aligned}
 f'(P) &= \frac{1}{1 + \frac{P}{N + \alpha_A \bar{P} - \alpha_A P}} \left(\frac{1}{N + \alpha_A \bar{P} - \alpha_A P} + \frac{\alpha_A P}{(N + \alpha_A \bar{P} - \alpha_A P)^2} \right) \\
 &= \frac{1}{N + \alpha_A \bar{P} - \alpha_A P + P} \left(1 + \frac{\alpha_A P}{(N + \alpha_A \bar{P} - \alpha_A P)} \right) \\
 &= \frac{1}{N + \alpha_A \bar{P} - \alpha_A P + P} \left(\frac{N + \alpha_A \bar{P} - \alpha_A P + \alpha_A P}{N + \alpha_A \bar{P} - \alpha_A P} \right) \\
 &= \left(\frac{1}{N + \alpha_A \bar{P} + P - \alpha_A P} \right) \left(\frac{(N + \alpha_A \bar{P})}{N + \alpha_A \bar{P} - \alpha_A P} \right)
 \end{aligned}$$

Clearly, $f'(P) \geq 0$ for all $0 \leq P \leq \bar{P}$. Hence, $f(P)$ is monotone increasing for all $0 \leq P \leq \bar{P}$. ■

Differentiating $f'(P)$ with respect to P again, we have,

$$f''(P) = \frac{d}{dP} \left(\frac{1}{K + (1 - \alpha_A)P} \frac{1}{K - \alpha_A P} \right)$$

We have used the substitution $K := N + \alpha_A \bar{P}$ in the above expression. Also, we are interested only in the sign of $f''(P)$, hence, we have ignored $(N + \alpha_A \bar{P})$ in the numerator as well.

$$\begin{aligned} f''(P) &= \frac{1}{K - \alpha_A P} \left(\frac{-1(1 - \alpha_A)}{(K + (1 - \alpha_A)P)^2} \right) + \frac{1}{K + (1 - \alpha_A)P} \left(\frac{-1 \times -\alpha_A}{(K - \alpha_A P)^2} \right) \\ &= \frac{1}{(K - \alpha_A P)(K + (1 - \alpha_A)P)} \left(\frac{-(1 - \alpha_A)}{K + (1 - \alpha_A)P} + \frac{\alpha_A}{K - \alpha_A P} \right) \end{aligned}$$

Clearly, $K - \alpha_A P = N + \bar{P} - \alpha_A P \geq 0$ (for $0 \leq P \leq \bar{P}$) and $K + (1 - \alpha_A)P \geq 0$. Hence, we will concentrate only on the terms inside the braces,

$$\begin{aligned} f''(P) &= \frac{-(1 - \alpha_A)}{K + (1 - \alpha_A)P} + \frac{\alpha_A}{K - \alpha_A P} \\ &= \frac{-(K - \alpha_A P)(1 - \alpha_A) + \alpha_A(K + (1 - \alpha_A)P)}{(K + (1 - \alpha_A)P)(K - \alpha_A P)} \end{aligned}$$

Ignoring the denominator (which is always positive), we have,

$$\begin{aligned} f''(P) &= -[K - K\alpha_A + \alpha_A^2 P - \alpha_A P] + [\alpha_A K + \alpha_A P - \alpha_A^2 P] \\ &= 2\alpha_A K - 2\alpha_A^2 P + 2\alpha_A P - K \\ &= (2\alpha_A - 1)K + 2\alpha_A P(1 - \alpha_A) \end{aligned}$$

Clearly, $1 \geq \alpha_A$. Now, if $2\alpha_A - 1 > 0$, then we see that the above expression is positive for all $0 \leq P \leq \bar{P}$. Hence, $f'' \geq 0$, or, the function $f(P)$ is convex increasing.

Lemma 5.3.2 : Suppose that $2\alpha_A - 1 > 0$. Then, f is convex in P for $0 \leq P \leq \bar{P}$. Further, the solution for the optimization problem in (5.1) is to allot all the power to a single transmitter. And the optimal value for the objective function in (5.1) is $f(\bar{P})$. ■

Now suppose that $2\alpha_A - 1 < 0$. Then, $f''(0) < 0$ and for large enough \bar{P} , $f''(\bar{P}) > 0$, i.e., for large \bar{P} , there would exist a $P' \leq \bar{P}$ such that $f''(P') \leq 0$ for all $0 \leq P \leq P'$ and $f''(P) \geq 0$ for all $P' \leq P \leq \bar{P}$. The following lemma summarizes the idea.

Lemma 5.3.3 : Suppose that $2\alpha_A - 1 < 0$. For large \bar{P} , there exists a P' , $0 \leq P' \leq \bar{P}$, such that $f(P)$ is concave upto P' and convex there after. ■

From the definition of $f(\cdot)$, we see that, $f(\bar{P}) = \log\left(1 + \frac{\bar{P}}{N}\right)$ increases to infinity with \bar{P} . Also, $f'(0) = \frac{1}{N + \alpha_A \bar{P}}$, decreases to 0 as \bar{P} increases. Further, for large \bar{P} , we see that $f'(0) \leq \frac{f(\bar{P})}{\bar{P}}$.

Theorem 5.3.1 : Suppose that $2\alpha_A - 1 < 0$. For large \bar{P} , $f(P) \leq \frac{f(\bar{P})}{\bar{P}} P$.

Proof: From Lemma 5.3.3, we know that $f(P)$ is concave upto P' . Hence, $f(P) \leq f(0) + f'(0)P$ for $0 \leq P \leq P'$. Since $f(0) = 0$, we have, $f(P) \leq f'(0)P$. For large \bar{P} , we have $f'(0) \leq \frac{f(\bar{P})}{\bar{P}}$ (from the previous arguments). Hence, for $0 \leq P \leq P'$,

$$f(P) \leq \frac{f(\bar{P})}{\bar{P}} P$$

In the region $P' \leq P \leq \bar{P}$, $f(P)$ is convex increasing, hence, we have,

$$f(P) - f(P') \leq \frac{(f(\bar{P}) - f(P'))}{\bar{P} - P'} (P - P')$$

Simplifying the above expression, we have,

$$f(P) \leq f(P') + f(\bar{P}) \frac{(P - P')}{\bar{P} - P'} - f(P') \frac{(P - P')}{\bar{P} - P'}$$

Or,

$$f(P) \leq f(\bar{P}) \frac{(P - P')}{\bar{P} - P'} + f(P') \left(1 - \frac{(P - P')}{\bar{P} - P'}\right)$$

Or,

$$f(P) \leq f(\bar{P}) \frac{(P - P')}{\bar{P} - P'} + f(P') \frac{\bar{P} - P}{\bar{P} - P'}$$

Substituting $f(P') \leq f'(0)P'$, we have,

$$f(P) \leq f(\bar{P}) \frac{(P - P')}{\bar{P} - P'} + f'(0)P' \frac{\bar{P} - P}{\bar{P} - P'}$$

For large \bar{P} , $f'(0) \leq \frac{f(\bar{P})}{\bar{P}}$. Substituting again, we get,

$$f(P) \leq f(\bar{P}) \frac{(P - P')}{\bar{P} - P'} + \frac{f(\bar{P})}{\bar{P}} P' \frac{\bar{P} - P}{\bar{P} - P'}$$

Or,

$$f(P) \leq f(\bar{P}) \left(\frac{(P - P')}{\bar{P} - P'} + \frac{P'}{\bar{P}} \frac{\bar{P} - P}{\bar{P} - P'} \right)$$

Simplifying the above expression, we have,

$$f(P) \leq f(\bar{P}) \frac{P}{\bar{P}}$$

for all $P' \leq P \leq \bar{P}$, which completes the proof. ■

Chapter 6

Power Control and Routing for a Single Cell, Dense Wireless Network

6.1 Introduction

In Chapter 5, we studied the capacity of dense wireless networks and showed that single cell operation is optimal for large network power constraints. Spatial reuse was ineffective and direct transmission between the source and the destination nodes is the throughput optimal strategy. Spatial reuse leads to interference, which is perceived as a loss of throughput. This makes it sub-optimal for low power networks as well (especially when energy overheads are associated with communication). Further, scheduling simultaneous transmissions using a distributed medium access protocol is very cumbersome. Hence, in this chapter, we consider a dense wireless network operated as a single cell. We are interested in a throughput maximization problem subject to a network power constraint.

We consider a scenario in which there is a large number of stationary nodes (e.g., hundreds of nodes) confined to a small area (e.g., spatial diameter 30m), and organised into a multihop ad hoc wireless network. We assume that, traffic in the network is homogeneous and data packets are sent between source-destination pairs by multihop relaying with single user decoding and forwarding of packets, i.e., signals received from nodes other than the intended transmitter are treated as interference. A distributed

multiaccess contention scheme is used in order to schedule transmissions; for example, the CSMA/CA based distributed coordination function (DCF) of the IEEE 802.11 standard for wireless local area networks (WLANs). We assume that all nodes can decode all the contention control transmissions (i.e., there are no hidden nodes), and only one successful transmission takes place at any time in the network. In this sense we say that we are dealing with a *single cell* scenario. Thus our work in this chapter can be viewed as an extension of the performance analysis presented in Chapter 3. We further assume that, during the exchange of contention control packets, pairs of communicating nodes are able to estimate the channel fade between them and are thus able to perform power control per transmission.

There is a natural tradeoff between using high power and long hop lengths (single hop direct transmission between the source-destination pair), versus using low power and shorter hop lengths (multihop communication using intermediate nodes), with the latter necessitating more packets to be transported in the network. The objective of the present chapter is to study optimal routing, in terms of the hop length, and optimal power control for a fading channel, when the network (described above) is used in a multihop mode. Our objective is to maximise a certain measure of network transport capacity (measured in bit-meters per second; see Section 6.3), subject to a network power constraint. A network power constraint determines, to a first order, the lifetime of the network.

Situations and considerations such as those that we study could arise in a dense ad hoc wireless sensor network. Ad hoc wireless sensor networks are now being studied as possible replacements for wired measurement networks in large factories. For example, a distillation column in a chemical plant could be equipped with pressure and temperature sensors and valve actuators. The sensors monitor the system and communicate the pressure and temperature values to a central controller which in turn actuates the valves to operate the column at the desired operating point. Direct communication between the sensors and actuators is also a possibility. Such installations could involve hundreds of devices, organised into a single cell ad hoc wireless network because of the physical proximity of the nodes. There would be many flows within the network and there would

be multihopping. We wish to address the question of optimal organisation of such an ad hoc network so as to maximise its transport capacity subject to a power constraint. The power constraint relates to the network life-time and would depend on the application. In a factory situation, it is possible that power could be supplied to the devices, hence large power would be available. In certain emergencies, “transient” sensor networks could be deployed for situation management; we use the term “transient” as these networks are supposed to exist for only several minutes or hours, and the devices could be disposable. Such networks need to have large throughputs, but, being transient networks, the power constraint could again be loose. On the other hand sensor networks deployed for monitoring some phenomenon in a remote area would have to work with very small amounts of power, while sacrificing transport capacity. Our formulation aims at providing insights into optimal network operation in a range of such scenarios.

6.1.1 Preview of Contributions

We motivate the definition of the transport capacity of the network as the product of the aggregate throughput (in bits per second) and the hop distance (in meters). For random spatio-temporal fading, we seek the power control and the hop distance that jointly maximizes the transport capacity, subject to a network average power constraint. For a fixed data transmission time strategy (discussed in Section 6.2.2), we show that the optimal power allocation function has a water pouring form (Section 6.4.1). At the optimal operating point (hop distance and power control) the network throughput (Θ_{opt} , in bits per second) is shown to be a fixed quantity, depending only on the contention mechanism and fading model, but independent of the network power constraint (Section 6.4.2). Further, we show that the optimal transport capacity is of the form $d_{opt}(\bar{P}_t) \times \Theta_{opt}$, with d_{opt} scaling as $\bar{P}_t^{\frac{1}{\eta}}$, where \bar{P}_t is the available time average transmit power, and η is the power law path loss exponent (Theorem 6.4.2). Finally, we provide a condition on the fading density that leads to a simple characterisation of the optimal hop distance (Section 6.4.3).

6.1.2 Motivation for Single Cell Operation

In this context (a dense, ad hoc wireless network), the seminal paper by Gupta and Kumar [50] would suggest that each node should communicate with neighbours as close as possible while maintaining network connectivity. This maximises network transport capacity (in bit-meters per second), while minimising network average power. It has been observed by Dousse and Thiran in [49], that if, unlike [50], a practical model of bounded received power for finite transmitter power is used, then the increasing interference with an increasing density of simultaneous transmitters is not consistent with a minimum SINR requirement at each receiver. More recently, El Gamal and Mammen [4] have shown that, if the transceiver energy and communication overheads at each hop is factored in, then the operating regime studied in [50] is neither energy efficient nor delay optimal. Fewer hops between the transmitter and receiver (and hence, less spatial reuse) reduce the overhead energy consumption and lead to a better throughput-delay tradeoff.

While optimal operation of the network might suggest using some spatial reuse (finite, as discussed above), coordinating simultaneous transmissions (in a distributed fashion), in a constrained area, is extremely difficult and the associated time, energy and synchronisation overheads have to be accounted for. In view of the above discussion, in this chapter, we assume that the medium access control (MAC) is such that only one transmitter-receiver pair communicates at any time in the network.

6.1.3 Outline of the Chapter

In Section 6.2 we describe the network model and in Section 6.3 we motivate the objective. We study the transport capacity of a single cell multihop wireless network operating in the fixed transmission time mode, in Section 6.4. Section 6.5 summarizes the results in the chapter.

6.2 Network Model and Assumptions

There is a dense collection of immobile nodes that use multiaccess multihop radio communication with single user decoding and packet forwarding to transport packets between various source-destination pairs.

- All nodes use the same contention mechanism with the same parameters (e.g., all nodes use IEEE 802.11 DCF with the same backoff parameters).
- We assume that nodes send control packets (such as RTS/CTS in IEEE 802.11) with a constant power (i.e., power control is not used for the control packets) during contention, and these control packets are decodable by *every* node in the network. As in IEEE 802.11, this can be done by using a low rate, robust modulation scheme and by restricting the diameter of the network. This is the “single cell” assumption and implies that there can be only one successful ongoing transmission at any time.
- During the control packet exchange, each transmitter learns about the channel “gain” to its intended receiver, and decides upon the power level that is used to transmit its data packet. For example, in IEEE 802.11, the channel gain to the intended receiver could be estimated during the RTS/CTS control packet exchange. Such channel information can then be used by the transmitter to do power control. In our work, we assume that such channel estimation and power control is possible on a transmission-by-transmission basis.
- In this work, we model only an average power constraint and not a peak power constraint.
- Saturation assumption : We assume that the traffic is homogeneous in the network and all the nodes have data to send at all times; these could be locally generated packets or transit packets.

Data packets are sent between source-destination pairs by multihop relaying. We do not restrict to straight line paths and permit arbitrary routes between the source and the

destination nodes. For example, in [16], the authors study a load balancing strategy in dense multihop wireless networks with arbitrary traffic requirements. In our work, we assume that the path between the source and the destination nodes is fixed a priori and we optimize only over the hop length taken over the path. The dense network assumption permits us to vary the hop length arbitrarily over the path. Based on the dense network and traffic homogeneity assumption, we further make the following assumption.

- The nodes self-organise so that all hops are of length d , i.e., a one hop transmission always traverses a distance of d meters. This hop distance, d , will be one of our optimisation variables.

For a random node deployment, the hop distance that maximizes the system throughput need not be the same for every node and every flow. However, the approximation holds good for a homogeneous network with large number of nodes. Further, it will be practically infeasible to optimize every hop in a dense setup with hundreds of nodes.

6.2.1 Channel Model: Path Loss, Fading and Transmission Rate

The channel gain between a transmitter-receiver pair for a hop is assumed to be a function of the hop length (d) and the multipath fading “gain” (h). The path loss for a hop distance d is given by $\frac{1}{d^\eta}$, where η is the path loss exponent, chosen depending on the propagation characteristics of the environment (see, for e.g., [64]). This variation of path loss with d holds for $d > d_0$, the far field reference distance; we will assume that this inequality holds (i.e., $d > d_0$), and will justify this assumption in the course of the analysis (see Theorem 6.4.2).

We assume a flat and slow fading channel with additive white Gaussian noise of power σ^2 . We assume that for each transmitter-receiver pair, the channel gain due to multipath fading may change from transmission to transmission, but remains constant over any packet transmission duration. Since successive transmissions can take place between randomly selected pairs of nodes (as per the outcome of the distributed contention mechanism) we are actually modeling a spatio-temporal fading process. We assume that this

fading process is stationary in space and time with some given marginal distribution H . Let the cumulative distribution of H be $A(h)$ (with a p.d.f. $a(h)$), which by our assumption of spatio-temporal stationarity of fading is the same for all transmitter-receiver pairs and for all transmissions. We assume that the channel coherence time, τ_c , applicable to all the links in the network, upper bounds every data transmission duration in the network. Further, we assume that H and τ_c are independent of the hop distance d .

When a node transmits to another node at a distance d (in the transmitting antenna's far field), using transmitter power P , with channel power gain due to fading, h , then we assume that the transmission rate given by Shannon's formula is achieved over the transmission burst; i.e., the transmission rate is given by

$$C = W \log \left(1 + \frac{hP\alpha}{\sigma^2 d^\eta} \right)$$

where W is the signal bandwidth and α is a constant accounting for any fixed power gains between the transmitter and the receiver. Note that this requires that the transmitter has available several coding schemes of different rates, one of which is chosen for each channel state and power level.

6.2.2 Fixed Transmission Time Strategy

We consider a fixed transmission time scheme, where all data transmissions are of equal duration, T ($< \tau_c$) secs, independent of the bit rate achieved over the wireless link. This implies that the amount of data that a transmitter sends during a transmission opportunity is proportional to the achieved physical link rate. Upon a successful control packet exchange, the channel (between the transmitter, that "won" the contention, and its intended receiver) is reserved for a duration of T seconds independent of the channel state h . This is akin to the "TxOP" (transmission opportunity) mechanism in the IEEE 802.11 standard. Thus, when the power allocated during the channel state h is $P(h)$, $C(h)T$ bits are sent across the channel, where $C(h) = W \log \left(1 + \frac{P(h)h\alpha}{\sigma^2 d^\eta} \right)$. When $P(h) = 0$, we assume that the channel is left idle for the next T seconds. The transmitter does not

relinquish the channel immediately, and the channel reserved for the transmitter-receiver pair (for example, by the RTS/CTS signalling) is left empty for the duration of T seconds.

The optimality of a fixed transmission time scheme, for throughput, as compared to a fixed packet length scheme has been shown in Appendix 6.6.4. When using fixed packet lengths, a transmitter may be forced to send the entire packet even if the channel is poor, thus taking longer time and more power. On the other hand, in a fixed transmission time scheme, we send more data when the channel is good and limit our inefficiency when the channel is poor.

6.3 Multihop Transport Capacity

Let d denote the common hop length and $\{P(h)\}$ a power allocation policy, with $P(h)$ denoting the transmit power used when the channel state is h . We take a simple model for the random access channel contention process. The channel goes through successive contention periods. Each period can be either an idle slot, or a collision period, or a successful transmission with probabilities p_i, p_c and p_s respectively. Under the node saturation assumption, the aggregate bit rate carried by the system, $\Theta_T(\{P(h)\}, d)$, for the hop distance d and power allocation $\{P(h)\}$, is given by (see [6])

$$\Theta_T(\{P(h)\}, d) := \frac{p_s(\int_0^\infty L(h) dA(h))}{p_i T_i + p_c T_c + p_s(T_o + T)} \quad (6.1)$$

where $L(h) = C(h)T$, and, T_i, T_c and T_o are the average time overheads associated with an idle slot, collision and data transmission. For e.g., in IEEE 802.11 with the RTS/CTS mechanism being used, a collision takes a fixed time independent of the data transmission rate. We note that p_i, p_s, p_c, T_i, T_o , and T_c depend only on the parameters of the distributed contention mechanism (MAC protocol) and the channel, and not on any of the decision variables that we consider (i.e., d and $\{P(h)\}$).

With $\Theta_T(\{P(h)\}, d)$ defined as in (6.1), we consider $\Theta_T(\{P(h)\}, d) \times d$ as our measure of transport capacity of the network. This measure can be motivated in several ways. $\Theta_T(\{P(h)\}, d)$ is the rate at which bits are transmitted by the network nodes. When

transmitted successfully, each bit traverses a distance d . Hence, $\Theta_T(\{P(h)\}, d) \times d$ is the rate of spatial progress of the flow of bits in the network (in bit-meters per second). Viewed alternatively, it is the weighted average of the end-to-end flow throughput with respect to the distance traversed. Suppose that a flow i covers a distance D_i with $\frac{D_i}{d}$ hops (assumed to be an integer for this argument). Let $\beta_i \Theta_T(\{P(h)\}, d)$ be the fraction of throughput of the network that belongs to flow i . Then, $\frac{\beta_i \Theta_T(\{P(h)\}, d)}{\frac{D_i}{d}}$ is the end-to-end throughput for flow i and $\frac{\beta_i \Theta_T(\{P(h)\}, d)}{\frac{D_i}{d}} \times D_i = \beta_i \Theta_T(\{P(h)\}, d) \times d$ is the end-to-end throughput for flow i in bit-meters per second. Summing over all the flows, we have $\Theta_T(\{P(h)\}, d) \times d$, the aggregate end-to-end flow throughput in bit-meters per second.

With the above motivation, our aim in this work is to maximise the quantity $\Theta_T(\{P(h)\}, d) \times d$ over the hop distance d and over the power control $\{P(h)\}$, subject to a *network average power constraint*, \bar{P} . We use a network power constraint that accounts for the energy used in data transmission as well as the energy overheads associated with communication. The network average power, $\mathcal{P}(\{P(h)\})$, is given by,

$$\mathcal{P}(\{P(h)\}) := \frac{p_i E_i + p_c E_c + p_s (E_o + T \int_0^\infty P(h) dA(h))}{p_i T_i + p_c T_c + p_s (T_o + T)} \quad (6.2)$$

E_i , E_c and E_o correspond to the energy overheads associated with an idle period, collision and successful transmission. Thus, E_i denotes the total energy expended in the network over an idle slot, E_c denotes the total average energy expended by the colliding nodes, as well as the idle energy of the idle nodes, and E_o denotes the average energy expended in the successful contention negotiation between the successful transmitter-receiver pair, the receive energy at the receiver (in the radio and in the packet processor), and the idle energy expended by all the other nodes over the time $T_o + T$. We note that E_i , E_c and E_o depend only on the contention mechanism and not on the decision variables d and $\{P(h)\}$.

6.4 Optimising the Transport Capacity

For a given $\{P(h)\}$ and d , and the corresponding throughput $\Theta_T(\{P(h)\}, d)$, the transport capacity in bit-meters per second, which we will denote by $\psi(\{P(h)\}, d)$, is given by

$$\psi(\{P(h)\}, d) := \Theta_T(\{P(h)\}, d) \times d$$

Maximizing $\psi(\cdot, \cdot)$ involves optimizing over d , as well as $\{P(h)\}$. However, we observe that, it would not be possible to vary d with fading, as routes cannot vary at the fading time scale. Hence, we propose to optimize first over $\{P(h)\}$ for a given d , and then optimize over d , i.e., we seek to solve the following problem,

$$\max_d \max_{\{\{P(h)\}:\mathcal{P}(\{P(h)\})\leq\bar{P}\}} \psi(\{P(h)\}, d) \quad (6.3)$$

For a given d and power allocation $\{P(h)\}$, define the time average transmission power, $\bar{P}_t(\{P(h)\}, d)$, and the time average overhead power, \bar{P}_o , as

$$\begin{aligned} \bar{P}_t(\{P(h)\}, d) &:= \frac{p_s(\int_0^\infty P(h) dA(h))T}{p_i T_i + p_c T_c + p_s(T_o + T)} \\ \bar{P}_o &:= \frac{p_i E_i + p_c E_c + p_s E_o}{p_i T_i + p_c T_c + p_s(T_o + T)} \end{aligned}$$

Observe that \bar{P}_o does not depend on $\{P(h)\}$ and d . Now, the network power constraint can be viewed as

$$\bar{P}_t(\{P(h)\}, d) \leq \bar{P} - \bar{P}_o$$

where the right hand side is independent of $\{P(h)\}$ or d . $\bar{P}_t := \bar{P} - \bar{P}_o$, is the *time average* transmitter power constraint for the network.

6.4.1 Optimization over $\{P(h)\}$ for a fixed d

Consider the optimization problem (from (6.3))

$$\max_{\{\{P(h)\}:\mathcal{P}(\{P(h)\})\leq\bar{P}\}} \psi(\{P(h)\}, d) \quad (6.4)$$

The denominators of $\Theta_T(\cdot, \cdot)$ in (6.1) and of \mathcal{P} in (6.2) are independent of d and the power control $\{P(h)\}$. Thus, with d fixed, the optimization problem simplifies to maximizing $\int_0^\infty L(h) dA(h)$ or,

$$\int_0^\infty \log \left(1 + \frac{P(h)h\alpha}{\sigma^2 d^\eta} \right) dA(h)$$

subject to the power constraint,

$$\int_0^\infty P(h) dA(h) \leq \bar{P}'_t$$

where \bar{P}'_t is given by,

$$\bar{P}'_t := \frac{(p_i T_i + p_c T_c + p_s (T_o + T))}{p_s T} \bar{P}_t$$

\bar{P}'_t is the average transmit power constraint averaged only over the transmission periods (successful contention slots).

Without a peak power constraint, this is a well-known problem whose optimal solution has the water-pouring form (see [5]). The optimal power allocation function $\{P(h)\}$ is given by

$$P(h) = \left(\frac{1}{\lambda} - \frac{d^\eta \sigma^2}{h\alpha} \right)^+$$

where λ is obtained from the power constraint equation

$$\int_{\frac{\lambda \sigma^2 d^\eta}{\alpha}}^\infty a(h) P(h) dh = \bar{P}'_t$$

The optimal power allocation is a nonrandomized policy, where a node transmits with

power $P(h)$ every time the channel is in state h (whenever $P(h) > 0$), or leaves the channel idle for h such that $P(h) = 0$.

6.4.2 Optimization over d

By defining $\xi(h) := \frac{P(h)}{d^\eta}$, the problem of maximising the throughput over power controls, for a fixed d , becomes

$$\max \int_0^\infty \log \left(1 + \frac{\alpha h}{\sigma^2} \xi(h) \right) a(h) dh$$

subject to

$$\int_0^\infty \xi(h) a(h) dh \leq \frac{\bar{P}_t'}{d^\eta}$$

Observe that \bar{P}_t' and d influence the optimization problem only as $\frac{\bar{P}_t'}{d^\eta}$. Denoting by $\Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right)$ the optimal value of this problem, the problem of optimisation over the hop-length, d , now becomes

$$\max_d d \times \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right) \quad (6.5)$$

Theorem 6.4.1 *In the problem defined by (6.5), the objective $d \times \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right)$, when viewed as a function of d , is continuously differentiable. Further, when the channel fading random variable, H , has a finite mean ($\mathbf{E}(\mathbf{H}) < \infty$), then*

1. $\lim_{d \rightarrow 0} d \times \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right) = 0$ and,
2. if in addition, $\eta \geq 2$, $\frac{1}{h^2} a \left(\frac{1}{h} \right)$ is continuously differentiable and $\mathbf{P}(\mathbf{H} > h) = \mathbf{O} \left(\frac{1}{h^2} \right)$ for large h , then, $\lim_{d \rightarrow \infty} d \times \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right) = 0$,

Proof: The proofs of continuous differentiability of $d \times \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right)$, 1) and 2) are provided in Appendix 6.6.2. ■

Remarks 6.4.1

1. Under the conditions proposed in Theorem 6.4.1, it follows that $d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$ is bounded over $d \in [0, \infty)$ and achieves its maximum in $d \in (0, \infty)$.
2. When the objective function (6.5) is unbounded, the optimal solution occurs at $d = \infty$ (follows from the continuity results).
3. We note that, in practice, $\eta \geq 2$.

Let d_0 be the far field reference distance (discussed in Section 6.2.1).

Theorem 6.4.2 *The following hold for the problem in (6.5),*

1. *Without the constraint $d > d_0$, the optimum hop distance d_{opt} scales as $(\bar{P}_t')^{\frac{1}{\eta}}$.*
2. *There is a value $\bar{P}_{t\min}'$ such that, for $\bar{P}_t' > \bar{P}_{t\min}'$, $d_{opt} > d_0$, and the optimal solution obeys the scaling shown in 1).*
3. *For $\bar{P}_t' > \bar{P}_{t\min}'$, the optimum power control $\{P(h)\}$ is of the water pouring form and scales as \bar{P}_t' .*
4. *For $\bar{P}_t' > \bar{P}_{t\min}'$, the optimal transport capacity scales as $(\bar{P}_t')^{\frac{1}{\eta}}$.*

Proof:

1. Let d_{opt} be optimal for $\bar{P}_t' > 0$. We claim that, for $x > 0$, $x^{\frac{1}{\eta}}d_{opt}$ is optimal for the power constraint $x\bar{P}_t'$. For suppose this was not so, it would mean that there exists $d > 0$ such that

$$\left(x^{\frac{1}{\eta}}d_{opt} \Gamma\left(\frac{x\bar{P}_t'}{(x^{\frac{1}{\eta}}d_{opt})^\eta}\right)\right) < d \Gamma\left(\frac{x\bar{P}_t'}{d^\eta}\right)$$

or, equivalently,

$$\left(d_{opt} \Gamma\left(\frac{\bar{P}_t'}{d_{opt}^\eta}\right)\right) < x^{-\frac{1}{\eta}}d \Gamma\left(\frac{\bar{P}_t'}{(x^{-\frac{1}{\eta}}d)^\eta}\right)$$

which contradicts the hypothesis that d_{opt} is optimal for \bar{P}_t' .

2. With the path loss model $\frac{P}{d^\eta}$, we see that for $d < d_0$, the received power is scaled more than the transmitted power P , due to the factor $\frac{1}{d^\eta}$, and an d_0^η factor in α , i.e., the model over-estimates the received power and the transport capacity. Hence, the achievable transport capacity for $d < d_0$ is definitely less than $d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$. The result now follows from the scaling result in 1).
3. It follows from 1) that, if \bar{P}_t' scales by a factor x , then the optimum d scales by $x^{\frac{1}{\eta}}$, so that, at the optimum, $\frac{\bar{P}_t'}{d^\eta}$ is unchanged. Hence the optimal $\{\xi(h)\}$ is unchanged, which means that $\{P(h)\}$ must scale by x . The water pouring form is evident.
4. Again, by 1) and 2), if \bar{P}_t' scales by a factor x , then the optimum d scales by $x^{\frac{1}{\eta}}$, so that, at the optimum, $\frac{\bar{P}_t'}{d^\eta}$ is unchanged. Thus $\Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$ is unchanged, and the optimal transport capacity scales as the optimum d , i.e., by the factor $x^{\frac{1}{\eta}}$.

■

Remarks 6.4.2

The above theorem yields the following observations for the fixed transmission time model.

1. As an illustration, with $\eta = 3$, in order to double the transport capacity, we need to use 2^3 times the \bar{P}_t' . This would result in a considerable reduction in network lifetime, assuming the same battery energy.
2. We observe that as the power constraint \bar{P}_t' scales, the optimal bit rate carried in the network, $\Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$, stays constant, but the optimal transport capacity increases since the optimal hop length increases. Further, because of the way the optimal power control and the optimal hop length scale together, the nodes transmit at the same *physical bit rate* in each fading state; see the proof of Theorem 6.4.2 part 3).

■

6.4.3 Characterisation of the Optimal d

By the results in Theorem 6.4.1 we can conclude that the optimal solution of the maximisation in (6.5) lies in the set of points for which the derivative of $d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$ is zero. For a

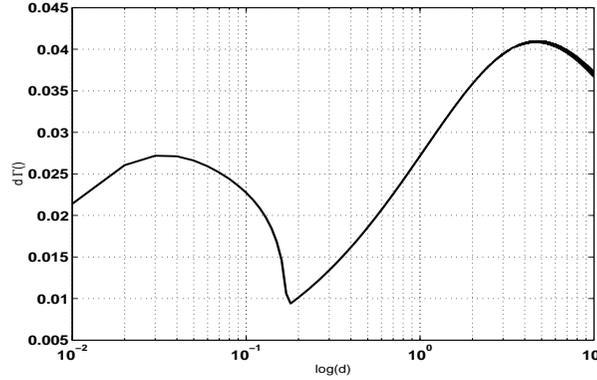


Figure 6.1: Plot of $d \times \Gamma\left(\frac{1}{d^3}\right)$ (linear scale) vs. d (log scale) for a channel with two fading states h_1, h_2 . The fading gains are $h_1 = 100$ and $h_2 = 0.5$, with probabilities $a_{h_1} = 0.01 = 1 - a_{h_2}$. The function has 3 non-trivial stationary points.

fixed \bar{P}_t' , define $\pi(d) := \frac{\bar{P}_t'}{d^\eta}$. Differentiating $d \times \Gamma(\pi(d))$, we obtain, (see Appendix 6.6.1)

$$\frac{\partial}{\partial d}(d \Gamma(\pi(d))) = \Gamma(\pi(d)) - \eta\pi(d)\lambda(\pi(d))$$

where $\lambda(\pi)$ is the Lagrange multiplier for the optimisation problem that yields $\Gamma(\pi(d))$. Since d appears only via $\pi(d)$, we can view the right hand side as a function only of π . We are interested in the zeros of the above expression. Clearly, $\pi = 0$ is a solution. The solution $\pi = 0$ corresponds to the case $d = \infty$; However, we are interested only in solutions of d in $(0, \infty)$, and hence, we seek positive solutions of π of

$$\Gamma(\pi) - \eta\pi\lambda(\pi) = 0 \tag{6.6}$$

Remarks 6.4.3 *In Appendix 6.6.1, we consider a continuously distributed fading random variable H with p.d.f. $a(h)$. The analysis can be done for a discrete valued fading distribution as well, and we provide this analysis in Appendix 6.6.3. The following example then illustrates that, in general, the function $\Gamma(\pi) - \eta\pi\lambda(\pi) = 0$ can have multiple solutions. Consider a fading distribution that takes two values: $h_1 = 100$ and $h_2 = 0.5$, with probabilities $a_{h_1} = 0.01 = 1 - a_{h_2}$. Figure 6.1 plots $d \times \Gamma\left(\frac{1}{d^3}\right)$ for the system with $\eta = 3$. Notice that there are 3 stationary points other than the trivial solution $d = \infty$ (which is not shown in the figure). Also, the maximising solution is not the first stationary point*

(the stationary point close to 0). If, on the other hand, $a_{h_1} = 0.001 = 1 - a_{h_2}$, we again have 3 stationary points, but the optimal solution now is the first stationary point. ■

More generally, and still pursuing the discrete case, let \mathcal{H} denote the set of fading states when the fading random variable is discrete with a finite number of values; $|\mathcal{H}|$ denotes the cardinality of \mathcal{H} .

Theorem 6.4.3 *There are at most $2|\mathcal{H}| - 1$ stationary points of $d \Gamma(\pi(d))$ in $0 < d < \infty$.*

Proof: See Appendix 6.6.3 for the related analysis and the proof. ■

We conclude from the above discussion that it is difficult to characterise the optimal solution when there are multiple stationary points. Hence we seek conditions for a unique positive stationary point, which must then be the maximising solution. In Appendix 6.6.1, we have shown that the equation characterising the stationary points, $\Gamma(\pi) - \eta\pi\lambda(\pi) = 0$, can be rewritten as

$$\int_0^1 (\log(y) - \eta(y - 1)) \frac{\lambda^2}{y^2} f\left(\frac{\lambda}{y}\right) dy = 0 \quad (6.7)$$

for $f(x) := a \left(\frac{\sigma^2 x}{\alpha}\right) \frac{\sigma^2}{\alpha}$, the density of the random variable $X := \frac{\alpha H}{\sigma^2}$. Notice that π does not appear in this expression. The solution directly yields the Lagrange multiplier of the throughput maximisation problem for the optimal value of hop length. The following theorem guarantees the existence of at most one solution of (6.7).

Theorem 6.4.4 *If for any $\lambda_1 > \lambda_2 > 0$, $\frac{f(\frac{\lambda_2}{y})}{f(\frac{\lambda_1}{y})}$ is a strictly monotonic decreasing function of y , then the objective function $d \times \Gamma\left(\frac{\bar{P}_t'}{d^n}\right)$ has at most one stationary point d_{opt} , $0 < d_{opt} < \infty$.*

Proof: The proof follows from Lemmas 6.6.1, and 6.6.2 in Appendix 6.6.1. ■

Corollary 6.4.1 *If H has an exponential distribution and $\eta \geq 2$, then the objective in the optimisation problem of (6.5) has a unique stationary point $d_{opt} \in (0, \infty)$, which achieves the maximum.*

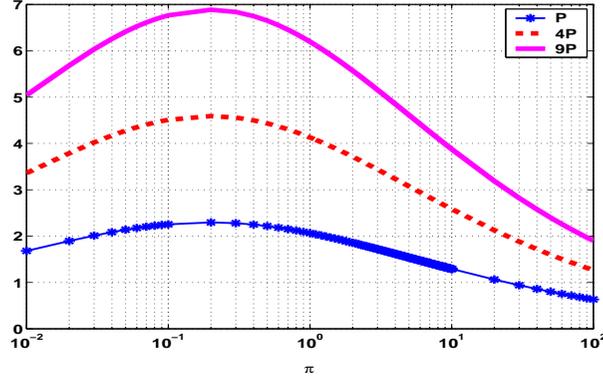


Figure 6.2: Plot of $d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)$ (linear scale) vs. $\pi (= \frac{\bar{P}_t'}{d^\eta})$ (log scale) for a fading channel (with exponential distribution). We consider 3 power levels (\bar{P}_t' , $4\bar{P}_t'$ and $9\bar{P}_t'$) and $\eta = 2$. The function has a unique optimum π_{opt} ($\pi_{opt} \approx 0.2$) for all the 3 cases.

Proof: $a(h)$ is of the form $\mu e^{-\mu h}$. From Theorem 6.4.1, we see that $\lim_{d \rightarrow 0} d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right) = 0$ and $\lim_{d \rightarrow \infty} d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right) = 0$. And, the monotonicity hypothesis in Theorem 6.4.4 holds for $a(h)$. ■

Remarks 6.4.4 1. Hence, for $\eta \geq 2$, for the Rayleigh fading model there exists a unique stationary point which corresponds to the optimal operating point.

2. For $\bar{P}_t' > \bar{P}_{t\min}'$, and for the conditions in Theorem 6.4.1 and 6.4.4, let π_{opt} denote the unique stationary point of (6.6). Then define $\Gamma(\pi_{opt}) = \Theta_{opt}$. It follows from Theorem 6.4.2 that the optimal transport capacity takes the form $\left(\frac{\bar{P}_t'}{\pi_{opt}}\right)^{\frac{1}{\eta}} \Theta_{opt}$, where Θ_{opt} depends on $a(h)$ and the MAC parameters but not on \bar{P} (or \bar{P}_t).

3. Figure 6.2 numerically illustrates our results for the Rayleigh fading distribution and $\eta = 2$. Scaling \bar{P}_t' by 4 scales the transport capacity from 2.3 to 4.6, i.e., by $4^{\frac{1}{2}} = \sqrt{4}$ and similarly for scaling \bar{P}_t' by 9.

The uniqueness result guarantees that a distributed implementation of the optimization problem, if it converges, shall converge to the unique stationary point, which is the optimal solution.

6.5 Summary

In this chapter we have studied a problem of optimal power control and self-organisation in a single cell, dense, ad hoc multihop wireless network. The self-organisation is in terms of the hop distance used when relaying packets between source-destination pairs.

We formulated the problem as one of maximising the transport capacity of the network subject to an average power constraint. We showed that, for a fixed transmission time scheme, there corresponds an intrinsic aggregate packet carrying capacity at which the network operates at the optimal operating point, independent of the average power constraint. We also obtained the scaling law relating the optimal hop distance to the power constraint, and hence relating the optimal transport capacity to the power constraint (see Theorem 6.4.2). Because of the way the power control and the optimal hop length scale, the optimal physical bit rate in each fading state is invariant with the power constraint. In Theorem 6.4.4, we provide a characterisation of the optimal hop distance for cases in which the fading density satisfies a certain monotonicity condition.

One motivation for our work is the optimal operation of sensor networks. If a sensor network is supplied with external power, or if the network is not required to have a long life-time, then the value of the power constraint, \bar{P} , can be large, and a long hop distance will be used, yielding a large transport capacity. On the other hand, if the sensor network runs on batteries and needs to have a long life-time then \bar{P} would be small, yielding a small hop length. In either case, the optimal aggregate bit rate carried by the network would be the same.

In [62], the authors study the problem of developing a distributed algorithm for nodes to adapt themselves towards the optimal operating point. They first propose a distance discretization technique in which the hop distance on the critical geometric graph is used as a distance measure. Using the distance approximation, they then develop a distributed algorithm aimed to maximize the transport capacity of the network in the sense of our framework presented in this chapter.

6.6 Appendix

6.6.1 Stationary Points of $d \times \Gamma(\pi(d))$

Recall that we defined $\pi(d) := \frac{\bar{P}_t'}{d^n}$. Further, $\Gamma(\pi(d))$ was defined as

$$\Gamma(\pi(d)) := \max \int_0^\infty \log \left(1 + \frac{\alpha h P(h)}{\sigma^2 d^n} \right) a(h) dh \quad (6.8)$$

where the maximum is over all power controls $\{P(h)\}$ satisfying the constraint

$$\int_0^\infty \frac{P(h)}{d^n} a(h) dh \leq \pi(d) \quad (6.9)$$

For ease of notation, let us use the substitution $x := \frac{\alpha h}{\sigma^2}$. Write $\xi(x) := \xi\left(\frac{\alpha h}{\sigma^2}\right) = \frac{P(h)}{d^n}$ and $f(x) := a\left(\frac{\sigma^2 x}{\alpha}\right) \frac{\sigma^2}{\alpha}$. Note that $f(\cdot)$ is the probability density of the random variable $X := \frac{\alpha H}{\sigma^2}$. Then, equations (6.8) and (6.9) can be rewritten as

$$\Gamma(\pi) = \max \int_0^\infty \log(1 + x\xi(x)) f(x) dx$$

and

$$\int_0^\infty \xi(x) f(x) dx \leq \pi$$

This optimisation problem is one of maximising a convex functional of $\{\xi(x)\}$, subject to a linear constraint. The optimal solution of the problem has water-pouring form, and the optimal solution is given by,

$$\xi(x) = \left(\frac{1}{\lambda(\pi)} - \frac{1}{x} \right)^+$$

where $\lambda(\pi)$ is obtained from

$$\int_{\lambda(\pi)}^\infty \left(\frac{1}{\lambda(\pi)} - \frac{1}{x} \right) f(x) dx = \pi$$

Further, the derivative of the optimum value $\Gamma(\pi)$, w.r.t. π , i.e., $\frac{\partial \Gamma(\pi)}{\partial \pi} = \lambda(\pi)$ (see Aubin [27]).

Let us now reintroduce the dependence on d , and consider the problem of maximizing $d \times \Gamma(\pi(d))$ over d . Differentiating $d \times \Gamma(\pi(d))$ w.r.t. d , we get,

$$\begin{aligned}
 \frac{\partial}{\partial d}(d \Gamma(\pi(d))) &= \Gamma(\pi(d)) + d \frac{\partial}{\partial d} \Gamma(\pi(d)) \\
 &= \Gamma(\pi(d)) + d \frac{\partial \Gamma}{\partial \pi}(\pi(d)) \times \frac{\partial \pi(d)}{\partial d} \\
 &= \Gamma(\pi(d)) + d \Gamma'(\pi(d)) \times \frac{-\eta \bar{P}_t'}{d^{n+1}} \\
 &= \Gamma(\pi(d)) - \eta \pi(d) \Gamma'(\pi(d))
 \end{aligned}$$

where $\Gamma'(\pi) := \frac{\partial \Gamma(\pi)}{\partial \pi}$. Substituting $\Gamma'(\pi) = \lambda(\pi)$, we have,

$$\frac{\partial}{\partial d}(d \Gamma(\pi(d))) = \Gamma(\pi(d)) - \eta \pi(d) \lambda(\pi(d)) \quad (6.10)$$

The stationary points of $d \times \Gamma(\pi(d))$ are now obtained by equating the right hand side of (6.10) to zero. Note that since d appears in this equation only as $\pi(d)$, we need only study the roots of the equation

$$\Gamma(\pi) - \eta \pi \lambda(\pi) = 0 \quad (6.11)$$

We now proceed to obtain a characterisation of the stationary points. Substituting the optimal solution in the expression of $\Gamma(\pi)$ and $\lambda(\pi)$, and suppressing the argument π in $\lambda(\pi)$, we get,

$$\Gamma(\pi) = \int_{\lambda}^{\infty} \log\left(\frac{x}{\lambda}\right) f(x) dx \quad (6.12)$$

with λ being given by

$$\pi = \int_{\lambda}^{\infty} \left(\frac{1}{\lambda} - \frac{1}{x}\right) f(x) dx \quad (6.13)$$

Using the substitution $z = \frac{1}{x}$, $l = \frac{1}{\lambda}$, and defining $g(z) = \frac{1}{z^2} f\left(\frac{1}{z}\right)$, (6.12) and (6.13) becomes,

$$\Gamma(\pi) = \int_0^l \log\left(\frac{l}{z}\right) g(z) dz \quad (6.14)$$

with l (actually, $l(\pi)$) being given by

$$\pi = \int_0^l (l - z) g(z) dz \quad (6.15)$$

We note that $g(\cdot)$ is the density of the random variable $Z := \frac{1}{X} = \frac{\sigma^2}{\alpha H}$.

We will use the following definitions for convenience. For a function $t(\cdot)$ of the random variable Z , define the operators $E_l(\cdot)$ and $G_l(\cdot)$ as

$$\begin{aligned} E_l(t(Z)) &:= \frac{\int_0^l t(z) g(z) dz}{\int_0^l g(z) dz} \\ G_l(t(Z)) &:= \int_0^l t(z) g(z) dz \end{aligned}$$

Lemma 6.6.1 *The roots of (6.11) are equivalent to the roots of the equation*

$$\eta G_l\left(\frac{Z}{l} - 1\right) = G_l\left(\log\left(\frac{Z}{l}\right)\right) \quad (6.16)$$

with l then being given by (6.15).

Proof: Using the definitions of $E_l(\cdot)$ and $G_l(\cdot)$, (6.14) and (6.15) simplify to

$$\Gamma(\pi) = \log(l)P(Z \leq l) - G_l(\log(Z)) \quad (6.17)$$

$$\pi = lP(Z \leq l) - G_l(Z) \quad (6.18)$$

(6.18) provides the l (actually $l(\pi)$) to be substituted in (6.17). Substituting for $\Gamma(\pi)$ (from (6.17)), and for l (from (6.18)), into (6.11), dividing across by $P(Z \leq l)$, and using the definition of $E_l(\cdot)$, we have,

$$\begin{aligned} \log\left(\frac{\pi + G_l(Z)}{P(Z \leq l)}\right) - E_l(\log(Z)) - \frac{\eta\pi}{\pi + G_l(Z)} &= 0 \\ \log\left(\frac{\pi}{P(Z \leq l)} + E_l(Z)\right) - E_l(\log(Z)) - \frac{\eta\pi}{\pi + G_l(Z)} &= 0 \end{aligned}$$

$$\log \left[\left(\frac{\pi}{G_1(Z)} + 1 \right) E_1(Z) \right] + \log \left(e^{-E_1(\log(Z))} \right) - \frac{\eta\pi}{\pi + G_1(Z)} = 0$$

Rearranging terms, we get,

$$\log \left(\frac{\pi + G_1(Z)}{G_1(Z)} \right) + \log \left(E_1(Z) e^{-E_1(\log(Z))} \right) - \frac{\eta\pi}{\pi + G_1(Z)} = 0$$

Denote $b_l := \log \left(E_1(Z) e^{-E_1(\log(Z))} \right)$. Then, we have,

$$\log \left(\frac{\pi + G_1(Z)}{G_1(Z)} \right) + b_l - \frac{\eta\pi}{\pi + G_1(Z)} = 0$$

From (6.18), we have

$$\frac{G_1(Z)}{\pi + G_1(Z)} = \frac{G_1(Z)}{lP(Z \leq l)} = \frac{E_1(Z)}{l}$$

which, with the previous equation, yields

$$\log \left(\frac{l}{E_1(Z)} \right) + b_l - \eta \left(1 - \frac{E_1(Z)}{l} \right) = 0$$

Recall that l is actually $l(\pi)$. We now find that π appears in the equation only as $l(\pi)$. Hence we can view this as an equation in the variable $l(= \frac{1}{\lambda})$. Rearranging terms, we get

$$-\log \left(\frac{E_1(Z)}{l} \right) + \eta \frac{E_1(Z)}{l} = -(b_l - \eta)$$

Exponentiating both sides, and substituting back for b_l , yields

$$\frac{E_1(Z)}{l} e^{-\eta \frac{E_1(Z)}{l}} = E_1(Z) e^{-E_1(\log(Z))} e^{-\eta}$$

On cancelling $E_1(Z)$, and transposing terms, we next obtain

$$e^{-\eta \left(\frac{E_1(Z)}{l} - 1 \right)} = e^{-E_1(\log(\frac{Z}{l}))}$$

or,

$$e^{-\eta(\mathbb{E}_l(\frac{Z-1}{l}))} = e^{-\mathbb{E}_l(\log(\frac{Z}{l}))}$$

Taking log on both sides, we have,

$$\eta \mathbb{E}_l \left(\frac{Z-1}{l} \right) = \mathbb{E}_l \left(\log \left(\frac{Z}{l} \right) \right)$$

In terms of $G_l(\cdot)$, this is equivalent to

$$\eta G_l \left(\frac{Z-1}{l} \right) = G_l \left(\log \left(\frac{Z}{l} \right) \right)$$

which is the desired result. ■

We next address the question of a *unique* positive solution of (6.16). The following lemma guarantees the existence of a unique positive solution, when $f(\cdot)$, the density of $\frac{\alpha H}{\sigma^2}$, satisfies a certain monotonicity condition.

Lemma 6.6.2 (6.16) has at most one positive solution if for any $0 < l_1 < l_2$, $\frac{f(\frac{1}{yl_2})}{f(\frac{1}{yl_1})}$ is a strictly monotone decreasing function of y .

Proof: Expanding $G_l(\cdot)$, (6.16) becomes,

$$\eta \int_0^l \left(\frac{z}{l} - 1 \right) g(z) dz - \int_0^{\frac{1}{l}} \log \left(\frac{z}{l} \right) g(z) dz = 0$$

Rewriting the equation in terms of $f(\cdot)$, we have,

$$\int_0^l \left(\eta \left(\frac{z}{l} - 1 \right) - \log \left(\frac{z}{l} \right) \right) \frac{1}{z^2} f \left(\frac{1}{zl} \right) dz = 0$$

Using a substitution $y = \frac{z}{l}$ in the above equation, we get,

$$\int_0^1 (\log(y) - \eta(y-1)) \frac{1}{y^2 l^2} f \left(\frac{1}{yl} \right) dy = 0 \quad (6.19)$$

Define $c(y) := (\log(y) - \eta(y-1)) \frac{1}{y^2}$ and $b_l(y) := f \left(\frac{1}{yl} \right)$. We are now interested in a

positive l that solves

$$\int_0^1 c(y)b_l(y)dy = 0$$

Observe that $\lim_{y \rightarrow 0} c(y) = -\infty$ and $c(1) = 0$. Further, there exists a unique y' such that $c(y) \leq 0$ for all $0 \leq y \leq y'$ and $c(y) \geq 0$ for all $y' \leq y \leq 1$. Since $b_l(y) \geq 0$ for all y and l , we have $c(y)b_l(y) \leq 0$ for all $0 \leq y \leq y'$ and $c(y)b_l(y) \geq 0$ for all $y' \leq y \leq 1$. In particular,

$$\int_0^{y'} c(y)b_l(y)dy \leq 0$$

$$\int_{y'}^1 c(y)b_l(y)dy \geq 0$$

Consider l_1, l_2 such that $0 < l_1 < l_2$. By hypothesis, $\frac{b_{l_2}(y)}{b_{l_1}(y)}$ is a strictly monotone decreasing function of y . Hence, $\frac{c(y)b_{l_2}(y)}{c(y)b_{l_1}(y)}$ is also a strictly monotone decreasing function of y . We then have,

$$\frac{\int_0^{y'} |c(y)|b_{l_2}(y)dy}{\int_0^{y'} |c(y)|b_{l_1}(y)dy} = \frac{\int_0^{y'} |c(y)|\frac{b_{l_2}(y)}{b_{l_1}(y)}b_{l_1}(y)dy}{\int_0^{y'} |c(y)|b_{l_1}(y)dy} > \frac{b_{l_2}(y')}{b_{l_1}(y')}$$

And,

$$\frac{\int_{y'}^1 c(y)b_{l_2}(y)dy}{\int_{y'}^1 c(y)b_{l_1}(y)dy} = \frac{\int_{y'}^1 c(y)\frac{b_{l_2}(y)}{b_{l_1}(y)}b_{l_1}(y)dy}{\int_{y'}^1 c(y)b_{l_1}(y)dy} < \frac{b_{l_2}(y')}{b_{l_1}(y')}$$

Hence,

$$\frac{\int_0^{y'} |c(y)|b_{l_2}(y)dy}{\int_0^{y'} |c(y)|b_{l_1}(y)dy} > \frac{\int_{y'}^1 c(y)b_{l_2}(y)dy}{\int_{y'}^1 c(y)b_{l_1}(y)dy}$$

Interchanging terms, we get,

$$\frac{\int_0^{y'} |c(y)|b_{l_2}(y)dy}{\int_{y'}^1 c(y)b_{l_2}(y)dy} > \frac{\int_0^{y'} |c(y)|b_{l_1}(y)dy}{\int_{y'}^1 c(y)b_{l_1}(y)dy}$$

i.e., the ratio of the negative area of the integral to the positive area of the integral is a strictly monotonic function of l . Hence, as l increases, the integral (6.19) can cross 0 at most once, or, there exists at most one (non-trivial) solution for (6.19). ■

6.6.2 Proof of Theorem 6.4.1

In this section, we will use the variables and equations from the discussion in Appendix 6.6.1.

Lemma 6.6.3 $d \times \Gamma\left(\frac{\bar{P}_t'}{d^n}\right)$ is continuously differentiable with respect to d .

Proof: Recall that $\pi := \frac{\bar{P}_t'}{d^n}$. $\Gamma(\pi)$ and $\lambda(\pi)$ (equations (6.12) and (6.13)) are continuous function of π , and π itself is a continuous function of d . Hence, from (6.10), we see that $d \times \Gamma\left(\frac{\bar{P}_t'}{d^n}\right)$ is a continuously differentiable function of d . ■

Lemma 6.6.4 If H (or equivalently $X := \frac{H\alpha}{\sigma^2}$) has a finite mean, then $\lim_{d \rightarrow 0} d \times \Gamma\left(\frac{\bar{P}_t'}{d^n}\right) = 0$.

Proof: Consider (6.15)

$$\int_0^l (l-z)g(z)dz = \pi$$

where l is in fact $l(\pi)$. Talking l outside the integral, we get,

$$l \int_0^l \left(1 - \frac{z}{l}\right) g(z) dz = \pi$$

Rewriting the integral as an expectation, we have, $l \mathbf{E}_{\mathbf{Z}}\left(\mathbf{1} - \frac{\mathbf{Z}}{l}\right)^+ = \pi$ or $\mathbf{E}_{\mathbf{Z}}\left(\mathbf{1} - \frac{\mathbf{Z}}{l}\right)^+ = \frac{\pi}{l}$. Using Monotone Convergence Theorem, we get,

$$\lim_{l \rightarrow \infty} \mathbf{E}_{\mathbf{Z}}\left(\mathbf{1} - \frac{\mathbf{Z}}{l}\right)^+ \uparrow \mathbf{1}$$

or,

$$\lim_{l \rightarrow \infty} \frac{\pi}{l} = 1$$

From (6.15), we see that, $l \rightarrow \infty$ as $\pi \rightarrow \infty$ ($d \rightarrow 0$). Hence, we have,

$$\lim_{\pi \rightarrow \infty} \frac{l(\pi)}{\pi} = 1 \tag{6.20}$$

Now, consider the following limit, $\lim_{d \rightarrow 0} d \times \Gamma(\pi(d))$, or equivalently, $\lim_{\pi \rightarrow \infty} \pi^{-\frac{1}{\eta}} \Gamma(\pi)$. We know that,

$$\pi^{-\frac{1}{\eta}} \Gamma(\pi) \geq 0$$

From (6.14), we have,

$$\pi^{-\frac{1}{\eta}} \Gamma(\pi) = \pi^{-\frac{1}{\eta}} \mathbf{E}_{\mathbf{Z}} \left(-\log \left(\frac{\mathbf{Z}}{\mathbf{I}(\pi)} \right) \right)^+$$

Expanding the term inside the expectation, we have,

$$= \pi^{-\frac{1}{\eta}} \mathbf{E}_{\mathbf{Z}} \left(\log \left(\frac{\mathbf{1}}{\mathbf{Z}} \right) + \log \left(\frac{\mathbf{I}(\pi)}{\pi} \right) + \log(\pi) \right)^+$$

Using the inequality $\log \left(\frac{1}{z} \right) \leq \frac{1}{z}$ (for $z \geq 0$) in the above inequality, we get,

$$\leq \pi^{-\frac{1}{\eta}} \mathbf{E}_{\mathbf{Z}} \left(\frac{\mathbf{1}}{\mathbf{Z}} + \log \left(\frac{\mathbf{I}(\pi)}{\pi} \right) + \log(\pi) \right)^+$$

$\mathbf{E}_{\mathbf{Z}} \left(\frac{\mathbf{1}}{\mathbf{Z}} \right) < \infty$ (follows from the definition $Z := \frac{1}{X}$ and the hypothesis on $\mathbf{E}\mathbf{X}$), $\eta > 0$ and from (6.20), we have the right hand side of the above expression $\rightarrow 0$ as $\pi \rightarrow \infty$, which implies that $\lim_{\pi \rightarrow \infty} \pi^{-\frac{1}{\eta}} \Gamma(\pi) = 0$, or

$$\lim_{d \rightarrow 0} d \times \Gamma(\pi(d)) = 0$$

■

Lemma 6.6.5 *Let $\eta \geq 2$, $\frac{1}{x^2} f \left(\frac{1}{x} \right)$ be continuously differentiable and $\lim_{x \rightarrow 0} \frac{1}{x^2} f \left(\frac{1}{x} \right) = 0$. Then $\frac{\partial}{\partial d} \left(d \times \Gamma \left(\frac{\bar{P}_k'}{d^\eta} \right) \right) \leq 0$ as $d \rightarrow \infty$.*

Proof: From (6.10) and the discussion in the proof of Lemma 6.6.1, we have,

$$\begin{aligned} \frac{\partial}{\partial d} (d \Gamma(\pi(d))) &= \Gamma(\pi(d)) - \eta \pi(d) \lambda(\pi(d)) \\ &= \kappa \int_0^l \left(\eta \left(\frac{z}{l} - 1 \right) - \log \left(\frac{z}{l} \right) \right) \frac{1}{z^2} f \left(\frac{1}{z} \right) dz \end{aligned}$$

where $\kappa \geq 0$. Using a substitution $y = \frac{z}{l}$, we get,

$$\frac{\partial}{\partial d}(d \Gamma(\pi(d))) = \kappa \int_0^1 (\eta(y-1) - \log(y)) \frac{1}{y^2 l^2} f\left(\frac{1}{yl}\right) dy \quad (6.21)$$

Define $b(y) := \eta(y-1) - \log(y)$. For $\eta > 1$, there exists a y' (depending on η) such that $b(y) \geq 0$ for $0 \leq y \leq y'$ and $b(y) \leq 0$ for $y' \leq y \leq 1$, also $b(1) = 0$. Then, in (6.21), we see that,

$$\begin{aligned} \int_0^{y'} (\eta(y-1) - \log(y)) \frac{1}{y^2} f\left(\frac{1}{yl}\right) dy &\geq 0 \\ \int_{y'}^1 (\eta(y-1) - \log(y)) \frac{1}{y^2} f\left(\frac{1}{yl}\right) dy &\leq 0 \end{aligned}$$

Further,

$$\int_0^1 (\eta(y-1) - \log(y)) dy = 1 - \frac{\eta}{2}$$

For $\eta \geq 2$, the integral $\int_0^1 b(y) dy$ is non-positive.

Let $g(y) := \frac{1}{y^2} f\left(\frac{1}{y}\right)$. Then $g(y)$ is continuously differentiable function and $\lim_{y \rightarrow 0} g(y) = 0$ (by hypothesis). Define y_0 as

$$y_0 := \sup\{y : g(z) = 0, 0 \leq z \leq y\}$$

If $y_0 > 0$, then, we see that for l sufficiently small,

$$\int_0^{y'} (\eta(y-1) - \log(y)) \frac{1}{y^2 l^2} f\left(\frac{1}{yl}\right) dy = 0$$

This is because for sufficiently small l , $\frac{1}{y^2} f\left(\frac{1}{yl}\right) = 0$ for $0 \leq y \leq y'$. Hence, $\lim_{d \rightarrow \infty} \frac{\partial}{\partial d} \left(d \times \Gamma\left(\frac{\bar{P}_t'}{d^n}\right) \right) \leq 0$.

If $y_0 = 0$, we then have $g'(y) \geq 0$ in a small neighbourhood of 0 (since g is continuously differentiable by hypothesis). Hence, the function $g(y)$ is a monotonic increasing function in an ϵ neighbourhood of 0, i.e., $g(0) < g(y) \leq g(y') \leq g(\epsilon)$ for all $0 < y < y' < \epsilon$. Hence for all sufficiently small l , $\frac{1}{y^2} f\left(\frac{1}{yl}\right)$ is a monotone increasing function of y in $[0, 1]$. Hence,

in (6.21), we have,

$$\begin{aligned} & \int_0^{y'} (\eta(y-1) - \log(y)) \frac{1}{y^2 l^2} f\left(\frac{1}{yl}\right) dy + \int_{y'}^1 (\eta(y-1) - \log(y)) \frac{1}{y^2 l^2} f\left(\frac{1}{yl}\right) dy \leq \\ & \left(\frac{1}{y'l}\right)^2 f\left(\frac{1}{y'l}\right) \int_0^{y'} (\eta(y-1) - \log(y)) dy + \left(\frac{1}{y'l}\right)^2 f\left(\frac{1}{y'l}\right) \int_{y'}^1 (\eta(y-1) - \log(y)) dy \\ & = \left(\frac{1}{y'l}\right)^2 f\left(\frac{1}{y'l}\right) \left(1 - \frac{\eta}{2}\right) \end{aligned}$$

The final expression is non-positive for $\eta \geq 2$. Thus, $\frac{\partial}{\partial d} \left(d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right)\right) \leq 0$ as $d \rightarrow \infty$. ■

Lemma 6.6.6 *Let $\eta \geq 2$ and $\frac{1}{x^2} f\left(\frac{1}{x}\right)$ be continuously differentiable. If for large x , $P(X > x) = O\left(\frac{1}{x^2}\right)$ (or equivalently for $H = \frac{\sigma^2 X}{\alpha}$), then $\lim_{d \rightarrow \infty} d \times \Gamma\left(\frac{\bar{P}_t'}{d^\eta}\right) = 0$.*

Proof: Let $P(X > x) = O\left(\frac{1}{x^2}\right)$ for large x . i.e.,

$$\int_x^\infty f(x) dx = O\left(\frac{1}{x^2}\right)$$

Using a substitution $z = \frac{1}{x}$, we have,

$$\int_0^z \frac{1}{z^2} f\left(\frac{1}{z}\right) dz = O(z^2)$$

Define $g(z) := \frac{1}{z^2} f\left(\frac{1}{z}\right)$. Then,

$$\int_0^z g(z) dz = O(z^2) \tag{6.22}$$

Since $g(z) \geq 0$ and continuous (by hypothesis), we have, $g(0) = 0$. Suppose not, then, we have $g(z) \geq \epsilon$ for all $0 \leq z < \delta$ for some δ . Then,

$$\int_0^z g(z) dz \geq \epsilon z$$

for all $z \leq \delta$, which is a contradiction to (6.22). Hence $\lim_{z \rightarrow 0} g(z) = 0$ or $\lim_{z \rightarrow 0} \frac{1}{z^2} f\left(\frac{1}{z}\right) = 0$.

We know from (6.10) that

$$\frac{\partial}{\partial d}(d \Gamma(\pi(d))) = \Gamma(\pi(d)) - \eta\pi(d)\lambda(\pi(d))$$

Now from Lemma 6.6.5, we see that, for $\eta \geq 2$, and for $d \rightarrow \infty$,

$$\Gamma(\pi(d)) - \eta\pi(d)\lambda(\pi(d)) \leq 0$$

In other words,

$$\Gamma(\pi(d)) \leq \eta\pi(d)\lambda(\pi(d))$$

Multiplying by d on both the sides, we have,

$$d \Gamma(\pi(d)) \leq \eta\pi(d)\lambda(\pi(d))d = \eta \frac{\bar{P}_t'}{d^{\eta-1}} \lambda(\pi(d)) \quad (6.23)$$

Since $\frac{\partial}{\partial d} \left(d \Gamma \left(\frac{\bar{P}_t'}{d^\eta} \right) \right) \leq 0$ as $d \rightarrow \infty$, the function $d \Gamma(\pi(d))$ is monotonic decreasing for $d \rightarrow \infty$. Also $d \Gamma(\pi(d)) \geq 0$. Suppose that, $\lim_{d \rightarrow \infty} d \Gamma(\pi(d)) \neq 0$, it implies that $\lim_{d \rightarrow \infty} d \Gamma(\pi(d)) \geq \epsilon > 0$, which, using (6.23), implies that $\frac{\lambda(\pi(d))}{d^{\eta-1}} \geq \epsilon$ or as $d \rightarrow \infty$

$$\lambda(\pi(d)) \geq \epsilon d^{\eta-1} \quad (6.24)$$

From (6.13), we have,

$$\int_{\lambda}^{\infty} \left(\frac{1}{\lambda} - \frac{1}{x} \right) f(x) dx = \frac{\bar{P}_t'}{d^\eta}$$

ignoring the negative term, we have,

$$\frac{1}{\lambda} \int_{\lambda}^{\infty} f(x) dx \geq \frac{\bar{P}_t'}{d^\eta}$$

or,

$$\int_{\lambda}^{\infty} f(x) dx \geq \frac{\bar{P}_t'}{d^\eta} \lambda$$

Substituting from (6.24), we have,

$$\int_{\lambda}^{\infty} f(x)dx \geq \frac{\bar{P}_t'}{d^{\eta}} \epsilon d^{\eta-1} = \bar{P}_t' \epsilon \frac{1}{d} \quad (6.25)$$

But we have

$$\int_{\lambda}^{\infty} f(x)dx = \mathbf{P}(X > \lambda) = \mathcal{O}\left(\frac{1}{\lambda^2}\right) \leq \mathcal{O}\left(\frac{1}{d^{2\eta-2}}\right) \quad (6.26)$$

where the last inequality follows from (6.24). For $\eta \geq 2$, (6.25) and (6.26) yields a contradiction. Hence, $\lim_{d \rightarrow \infty} d \times \Gamma\left(\frac{\bar{P}_t'}{d^{\eta}}\right) = 0$. ■

6.6.3 Discrete Fading States

The optimization problem (6.4) for the discrete fading state case, simplifies to

$$\begin{aligned} \max \quad & \sum_{h \in \mathcal{H}} a_h \ln \left(1 + \left(\frac{\alpha h}{\sigma^2} \right) \frac{P(h)}{d^{\eta}} \right) \\ \text{subject to} \quad & \sum_{h \in \mathcal{H}} a_h P(h) \leq \bar{P}_t' \end{aligned} \quad (6.27)$$

For notational convenience, let us index the set of fading states, \mathcal{H} , in descending order by the index i , $1 \leq i \leq |\mathcal{H}|$, i.e., $h_1 > h_2 > h_3 > \dots$. Further, denote

$$a_{h_i} = a_i, \quad x_i = \frac{\alpha h_i}{\sigma^2}, \quad \text{and} \quad \xi_i = \frac{P(h_i)}{d^{\eta}}$$

Also, denote

$$\Pi = \frac{\bar{P}_t'}{d^{\eta}}$$

We will later recall that, for each power constraint \bar{P}_t' , Π is a function of d . Using this new notation and change of variables, we obtain the problem

$$\begin{aligned} \max \quad & \sum_i a_i \ln(1 + x_i \xi_i) \\ \text{subject to} \quad & \sum_i a_i \xi_i \leq \Pi \end{aligned} \quad (6.28)$$

We have the maximisation of a concave mapping from $\mathbf{R}^{|\mathcal{H}|}$ to \mathbf{R} subject to a linear

constraint. The KKT conditions are necessary and sufficient, and the following “water pouring” form of the optimal solution is well known. There exists $\lambda(\Pi) > 0$, such that, for $1 \leq i \leq |\mathcal{H}|$,

$$\xi_i = \left(\frac{1}{\lambda(\Pi)} - \frac{1}{x_i} \right)^+$$

with $\lambda(\Pi)$ being given by

$$\sum_{\{i: \frac{x_i}{\lambda(\Pi)} > 1\}} a_i \left(\frac{1}{\lambda(\Pi)} - \frac{1}{x_i} \right) = \Pi$$

Defining, for $1 \leq k \leq |\mathcal{H}|$,

$$p_k = a_1 + a_2 + \cdots + a_k, \text{ and } \alpha_k = \sum_{i=1}^k \frac{a_i}{x_i}$$

and $\Pi_0 = 0, \Pi_{|\mathcal{H}|} = \infty$, the Lagrange multiplier, $\lambda(\Pi)$, is given by

$$\lambda(\Pi) = \left(\frac{1}{p_k} (\alpha_k + \Pi) \right)^{-1} \quad (6.29)$$

for $\Pi_{k-1} < \Pi \leq \Pi_k$ when $1 \leq k \leq |\mathcal{H}| - 1$, and for $\Pi_{|\mathcal{H}|-1} < \Pi < \infty$ when $k = |\mathcal{H}|$. Here the break-points $\Pi_k, 1 \leq k \leq |\mathcal{H}| - 1$, are obtained by equating the values of $\lambda(\Pi)$ on either sides of the break-points, and are expressed as

$$\Pi_k = \left(\frac{\frac{\alpha_{k+1}}{p_{k+1}} - \frac{\alpha_k}{p_k}}{\frac{1}{p_k} - \frac{1}{p_{k+1}}} \right)$$

The denominator of this expression is clearly > 0 , and a little algebra shows that, since $x_{k+1} > x_i, 1 \leq i \leq k$, the numerator is also > 0 .

For each Π , let us denote the optimal value of the problem defined by (6.28) by $\Gamma(\Pi)$. We infer that

$$\frac{\partial \Gamma}{\partial \Pi} = \lambda(\Pi)$$

Now, fixing the power constraint \bar{P}_t' , and reintroducing the dependence on d , we recall

that $\Pi(d) = \frac{\bar{P}_t'}{d^\eta}$, and hence conclude that

$$\frac{\partial \Gamma}{\partial d} = \lambda(\Pi(d)) \left(\frac{-\eta \bar{P}_t'}{d^{\eta+1}} \right)$$

Define $d_0 = \infty$, $d_{|\mathcal{H}|} = 0$, and, for $1 \leq k \leq |\mathcal{H}| - 1$, define

$$d_k^\eta = \bar{P}_t' \cdot \left(\frac{\frac{1}{p_k} - \frac{1}{p_{k+1}}}{\frac{\alpha_{k+1}}{p_{k+1}} - \frac{\alpha_k}{p_k}} \right)$$

Note that $0 = d_{|\mathcal{H}|} < d_{|\mathcal{H}|-1} < \dots < d_2 < d_1 < d_0 = \infty$. Now, substituting for $\lambda(\Pi(d))$ from (6.29) and integrating, yields the following result

Theorem 6.6.1 *For given \bar{P}_t' , the optimal value $\Gamma(d)$ of the problem defined by (6.27) has the following characterisation.*

1. *The derivative of $\Gamma(d)$ w.r.t. d is given by*

$$\frac{\partial \Gamma}{\partial d} = \frac{1}{d} \left(\frac{-\eta p_k \bar{P}_t'}{\alpha_k d^\eta + \bar{P}_t'} \right) \quad (6.30)$$

for $d_k \leq d < d_{k-1}$ when $1 \leq k \leq |\mathcal{H}| - 1$, and for $0 < d < d_{|\mathcal{H}|-1}$ when $k = |\mathcal{H}|$.

2. *$\frac{\partial \Gamma}{\partial d}$ is a negative, continuous and increasing function of d . In particular $\Gamma(d)$ is a decreasing, and convex function of d .*
3. *The function $\Gamma(d)$ is given by*

$$\Gamma(d) = p_k \ln \left(\alpha_k + \frac{\bar{P}_t'}{d^\eta} \right) \gamma_k \quad (6.31)$$

for $d_k \leq d < d_{k-1}$ when $1 \leq k \leq |\mathcal{H}| - 1$, and for $0 < d < d_{|\mathcal{H}|-1}$ when $k = |\mathcal{H}|$, with the constants of integration γ_k being given as follows.

$$\gamma_1 = \frac{1}{\alpha_1} = \frac{x_1}{a_1}$$

and, for $2 \leq k \leq \mathcal{H}$, γ_k is obtained recursively as

$$\gamma_k = \frac{\left(\left(\alpha_{k-1} + \frac{\bar{P}_t'}{d_{k-1}^\eta} \right) \gamma_{k-1} \right)^{\left(\frac{p_{k-1}}{p_k} \right)}}{\alpha_k + \frac{\bar{P}_t'}{d_{k-1}^\eta}}$$

Proof: (6.31) is obtained by integrating the derivative in (6.30) over each segment of its definition. The integration constants γ_k are obtained by equating $\Gamma(d)$ on either sides of the break-points of the argument d . ■

Optimisation over d

Using Theorem 6.6.1, we conclude that we need to look at the stationary points of $\Gamma(d)d$. To this end, consider the solutions of

$$\Gamma(d) + d \Gamma'(d) = 0$$

Reintroducing the variable $\Pi = \frac{\bar{P}_t'}{d^\eta}$, and canceling p_k , we need the solutions of

$$\ln \left(1 + \frac{\Pi}{\alpha_k} \right) \alpha_k \gamma_k - \frac{\eta \Pi}{\alpha_k + \Pi} = 0$$

for $\Pi_{k-1} < \Pi \leq \Pi_k$ when $1 \leq k \leq |\mathcal{H}| - 1$, and for $\Pi_{|\mathcal{H}|-1} < \Pi < \infty$ when $k = |\mathcal{H}|$, with the break-points $\Pi_k, 1 \leq k \leq |\mathcal{H}|$, as given earlier. Let us write $\frac{\Pi}{\alpha_k + \Pi} = 1 - \frac{1}{1 + \frac{\Pi}{\alpha_k}}$, define $b_k = \ln \alpha_k \gamma_k$ (observe that $b_1 = 0$), and, for given k , use the new variable

$$y = \frac{1}{1 + \frac{\Pi}{\alpha_k}}$$

Note that, for $0 < \Pi < \infty$, $1 > y > 0$. Define $\delta_k = \frac{1}{1 + \frac{\Pi_k}{\alpha_k}}$. Then we seek the solutions of

$$\ln \frac{1}{y} + b_k - \eta(1 - y) = 0$$

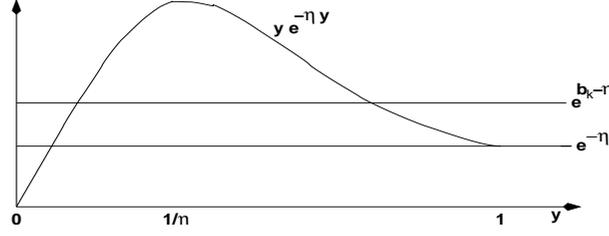


Figure 6.3: The stationary points of $\Gamma(d)d$ lie among the intersections of the curve $ye^{-\eta y}$ and lines $e^{b_k - \eta}$, $1 \leq k \leq |\mathcal{H}|$, in the interval $0 < y < 1$. Here the plot is drawn for $\eta = 3$.

for $\delta_k \leq y < \delta_{k-1}$, for each k , $1 \leq k \leq |\mathcal{H}|$; note that $\delta_0 = 1$, and $\delta_{|\mathcal{H}|} = 0$. The equations can be written more simply as

$$e^{b_k - \eta} = ye^{-\eta y},$$

and are depicted in Figure 6.3. At this point we can conclude the following

Theorem 6.6.2 1. $\lim_{d \rightarrow 0} \Gamma(d)d = 0$

2. There are at most $2|\mathcal{H}| - 1$ stationary points of $\Gamma(d)d$ in $0 < d < \infty$.

Proof: 2) follows from the arguments just before the theorem statement, since each line $e^{(b_k - \eta)}$, for $2 \leq k \leq |\mathcal{H}|$, has at most two intersections with $ye^{-\eta y}$, in $0 < y < 1$, and $e^{-\eta}$ has only one such intersection. ■

6.6.4 Fixed Transmission Time vs Fixed Packet Size

In this section, we will formally establish that fixed transmission time schemes are more throughput efficient compared to fixed packet size schemes, for a given average power constraint. We will prove this result in a general framework, without explicitly modelling the underlying MAC, the power control schemes used or the channel fading distribution.

Data Transmission Model: In a fixed transmission time scheme, all data transmissions (with positive rate) are of a fixed amount of time T , independent of the channel state h and the power used. Earlier, in our work (see Section 6.2.2), we assumed that, when the channel fade is poor (and hence $P(h) = 0$), the channel is left idle for the next T seconds. Further, the optimal power control policy for such a system was found to be a non-randomized policy, where a node transmits with constant power $P(h)$ every time

the channel is in state h (see Section 6.4.1). Here, we will allow the possibility of the channel *being relinquished when bad* with a fixed time overhead $\leq T$. We consider a spatio-temporal fading process with successive transmitter-receiver pairs being selected by a distributed multiaccess contention mechanism. Hence, relinquishing the channel might improve throughput, as successive fade levels might have little correlation. The optimal policy for such a MAC could be a randomized policy. Hence, we will allow a randomized power control, i.e., for a channel state h , the transmitter chooses a power P_h according to some distribution. In a fixed packet size scheme, all data transmissions (with positive rate) carry a fixed amount of data L independent of the channel state h and the power control used. Here as well, we will allow the possibility of a randomized power control and the possibility of relinquishing the channel with a fixed time overhead (when the channel fade is poor).

Optimality Criterion: The *throughput* optimality of a data transmission scheme is established either by comparing the energy required to send a certain amount of bits in a given time or by comparing the amount of bits sent with a given amount of energy in a given time. (We will discuss more about this optimality criterion in Remark 6.6.1). We study a data transmission scheme by considering two data transmissions of positive rates, in some arbitrary channel states with gains h_1 and h_2 and with applied powers P_{h_1} and P_{h_2} . We do not make any assumption on the probabilities of h_1 and h_2 , and about the power control policy which yields the powers P_{h_1} and P_{h_2} .

For a given power control scheme (h, P_h) , we will then assume that the transmission rate given by Shannon's formula is achieved over the transmission burst; i.e., the transmission rate is given by

$$C_h = W \log(1 + hP_h)$$

We have absorbed the factor $\frac{\alpha}{\sigma^2 d^\eta}$ in to the term h (since d is fixed in this discussion). Hence, the time durations taken to transmit the L bits during the channel states h_1 and h_2 (with the powers P_{h_1} and P_{h_2}) are given by $T_{h_1} := \frac{L}{W \log(1+h_1 P_{h_1})}$ and $T_{h_2} := \frac{L}{W \log(1+h_2 P_{h_2})}$.

Then, the total time occupied by these two transmissions is

$$T_P = \frac{L}{W \log(1 + h_1 P_{h_1})} + \frac{L}{W \log(1 + h_2 P_{h_2})} \quad (6.32)$$

spending an amount of energy equal to

$$E_P = \frac{L P_{h_1}}{W \log(1 + h_1 P_{h_1})} + \frac{L P_{h_2}}{W \log(1 + h_2 P_{h_2})} \quad (6.33)$$

Define $L_P := 2 \times L$ as the amount of bits sent in time T_P using an energy E_P in channel states h_1 and h_2 .

Lemma 6.6.7 *Let $h_1 > h_2$. For a fixed packet size scheme, if P_{h_1} and P_{h_2} are applied powers during channel states h_1 and h_2 , then having $h_1 P_{h_1} \geq h_2 P_{h_2}$ is throughput optimal.*

Proof: Suppose that $h_1 P_{h_1} < h_2 P_{h_2}$. Then,

$$\log(1 + h_1 P_{h_1}) < \log(1 + h_2 P_{h_2})$$

Find power controls \tilde{P}_{h_1} and \tilde{P}_{h_2} such that

$$\log(1 + h_1 P_{h_1}) = \log(1 + h_2 \tilde{P}_{h_2}) \quad (6.34)$$

$$\log(1 + h_2 P_{h_2}) = \log(1 + h_1 \tilde{P}_{h_1}) \quad (6.35)$$

or, equivalently,

$$h_1 P_{h_1} = h_2 \tilde{P}_{h_2} \quad (6.36)$$

$$h_2 P_{h_2} = h_1 \tilde{P}_{h_1} \quad (6.37)$$

With the power control scheme $(h_1, \tilde{P}_{h_1}), (h_2, \tilde{P}_{h_2})$, the total time occupied in the transmissions of $2 \times L$ bits during the channel states h_1 and h_2 is,

$$T_{\tilde{P}} = \frac{L}{W \log(1 + h_1 \tilde{P}_{h_1})} + \frac{L}{W \log(1 + h_2 \tilde{P}_{h_2})}$$

$$= T_P$$

(from (6.34) and (6.35)). Now, consider the energy spent to transmit these $2 \times L$ bits, i.e.,

$$E_{\tilde{P}} = \frac{L\tilde{P}_{h_1}}{W \log(1 + h_1\tilde{P}_{h_1})} + \frac{L\tilde{P}_{h_2}}{W \log(1 + h_2\tilde{P}_{h_2})}$$

Substituting for \tilde{P}_{h_1} and \tilde{P}_{h_2} from (6.36) and (6.37), we have,

$$E_{\tilde{P}} = \frac{1}{h_1} \frac{Lh_2P_{h_2}}{W \log(1 + h_2P_{h_2})} + \frac{1}{h_2} \frac{Lh_1P_{h_1}}{W \log(1 + h_1P_{h_1})}$$

Rearranging the terms, we have,

$$\begin{aligned} E_{\tilde{P}} &= \frac{1}{h_2} \frac{Lh_1P_{h_1}}{W \log(1 + h_1P_{h_1})} + \frac{1}{h_1} \frac{Lh_2P_{h_2}}{W \log(1 + h_2P_{h_2})} \\ &< \frac{1}{h_1} \frac{Lh_1P_{h_1}}{W \log(1 + h_1P_{h_1})} + \frac{1}{h_2} \frac{Lh_2P_{h_2}}{W \log(1 + h_2P_{h_2})} \\ &= \frac{LP_{h_1}}{W \log(1 + h_1P_{h_1})} + \frac{LP_{h_2}}{W \log(1 + h_2P_{h_2})} \\ &= E_P \end{aligned}$$

where the inequality follows from the fact that

$$\begin{aligned} &\frac{Lh_1P_{h_1}}{W \log(1 + h_1P_{h_1})} \left(\frac{1}{h_2} - \frac{1}{h_1} \right) \\ &< \frac{Lh_2P_{h_2}}{W \log(1 + h_2P_{h_2})} \left(\frac{1}{h_2} - \frac{1}{h_1} \right) \end{aligned}$$

since $h_1 > h_2$ and $h_1P_{h_1} < h_2P_{h_2}$ (by assumption) and the fact that $\frac{x}{\log(1+x)}$ is strictly monotone increasing.

It follows that an optimal power control must have $h_1P_{h_1} \geq h_2P_{h_2}$. ■

Remark: From Lemma 6.6.7, we see that, when $h_1 > h_2$, $C_{h_1} := W \log(1 + h_1P_{h_1}) \geq W \log(1 + h_2P_{h_2}) =: C_{h_2}$, or equivalently, $T_{h_1} \leq T_{h_2}$. ■

We will now provide a comparison of the fixed packet scheme with a fixed transmission time scheme and show the optimality of the fixed transmission time schemes. The comparison is done under the following assumption.

- The channel has the same marginal fading distribution, whenever sampled by a transmitter, for either schemes. This is a reasonable assumption as we consider spatio-temporal fading, with successive transmissions from possibly different source-destination pairs chosen by the distributed multiaccess contention scheme.

For the fixed packet size scheme, $L_P := 2 \times L$ bits were transmitted in $T_P (= T_{h_1} + T_{h_2})$ time (see (6.32)) with an amount of energy equal to E_P (see (6.33)), in two channel samples h_1 and h_2 . A reasonable comparison would be to find the throughput of a fixed transmission time scheme for a total duration of T_P seconds involving two data transmissions with channel samples h_1 and h_2 of equal duration $T = \frac{T_P}{2}$ and a total energy of E_P . We will assume that P_{h_1} and P_{h_2} , the power used for the fixed packet size scheme are such that $T_{h_1} \leq T_{h_2}$ (see Lemma 6.6.7). Hence, we have $T_{h_1} \leq T \leq T_{h_2}$, or, the fixed transmission time scheme spends relatively more time on a better channel. Clearly, its throughput is better than the fixed packet size scheme for the same energy constraint, as seen below.

Let $P_{t_{h_1}}$ and $P_{t_{h_2}}$ be the optimal power control for the fixed transmission time strategy such that

$$E_T := P_{t_{h_1}} T + P_{t_{h_2}} T = P_{h_1} T_{h_1} + P_{h_2} T_{h_2} = E_P$$

We have,

$$L_P = 2L = T_{h_1} W \log(1 + h_1 P_{h_1}) + T_{h_2} W \log(1 + h_2 P_{h_2})$$

Expanding the left hand side, we have,

$$\begin{aligned} 2L &= T_{h_1} W \log(1 + h_1 P_{h_1}) + (T_{h_2} - T) W \log(1 + h_2 P_{h_2}) \\ &+ T W \log(1 + h_2 P_{h_2}) \end{aligned}$$

Using $h_1 > h_2$, we get,

$$\begin{aligned}
2L &\leq T_{h_1} \log(1 + h_1 P_{h_1}) + (T_{h_2} - T) \log(1 + h_1 P_{h_2}) \\
&\quad + T \log(1 + h_2 P_{h_2}) \\
&\leq T \log(1 + h_1 P_{t_{h_1}}) + T \log(1 + h_2 P_{t_{h_2}}) \\
&=: L_T
\end{aligned}$$

where the last inequality follows from the fact that $(h_1, P_{t_{h_1}})$ and $(h_2, P_{t_{h_2}})$ is the optimal power control scheme for the fixed transmission time scheme with time $T_P (= 2 \times T)$ and energy $E_T (= E_P)$.

Remarks 6.6.1 For $L(t)$ defined as the amount of bits sent up to time t , and $E(t)$ defined as the total energy spent up to time t , the average throughput (Θ) and the average power (\bar{P}) of the system are, in general, defined as

$$\begin{aligned}
\Theta &:= \liminf_{t \rightarrow \infty} \frac{L(t)}{t} \\
\bar{P} &:= \limsup_{t \rightarrow \infty} \frac{E(t)}{t}
\end{aligned}$$

Under additional assumptions on the fading process and the power control scheme used, the expressions are simplified as an ensemble average (for example, see (6.1) and (6.2) for a fixed transmission time scheme). In this section, the optimality of the schemes have been shown directly, by comparing the amount of bits transmitted for a particular sample of channel for a given amount of time and energy, or by comparing the amount of energy used to transmit a given amount of bits for a particular sample of channel in a given amount of time. For example, the argument provided here directly translates to an argument with the ensemble average for the discrete fading case. This approach is not only straightforward, but also is very general. ■

Chapter 7

Network Coding and Power Control for a Two Link Wireless Network

7.1 Introduction

In the previous chapters, we have considered a store-and-forward strategy to route packets between nodes. It is now well known that (see [57]) (network) coding enhances throughput and delay characteristics of networks (including wired networks) as compared to the traditional store-and-forward routing strategy. Motivated by the results in [57] and in the literature, in this chapter, we study a power minimization problem using a network coding strategy for a simple two link wireless network.

Consider a two link, slotted wireless network with bidirectional traffic as shown in Figure 7.1. Packets flow from Node 1 to Node 2 and from Node 2 to Node 1 and are routed via the central Node 0. Data packets from either route are queued in Node 0 before transmission. We assume that Node 0 has sufficient buffer space and the packets are queued until transmission, i.e., there are no packet losses due to buffer overflow in

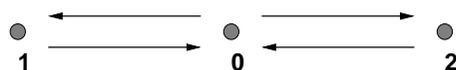


Figure 7.1: A two link wireless network with bidirectional traffic between nodes 1 and 2. Data packets from Node 1 to Node 2 and from Node 2 to Node 1 are routed via the central Node 0.

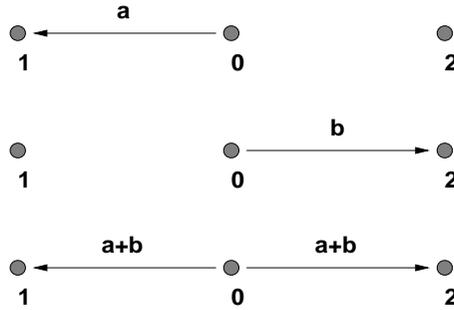


Figure 7.2: Node 0 queues packets from both the routes in its buffer and can network code the packets before transmission. This permits three different transmission strategies for Node 0 in a slot; data transmission to Node 1 ($0 \rightarrow 1$), data transmission to Node 2 ($0 \rightarrow 2$) or (network coded packet) broadcast to Nodes 1 and 2 ($0 \rightarrow 1, 2$).

the network. We make the following assumptions about radio transmission and reception. Nodes 1 and 2 can transmit a packet to Node 0, though not simultaneously. Node 0 can transmit a (unicast) packet to either Node 1 or to Node 2. We also permit the central node to “network code” packets belonging to the two routes (queued in its buffer) and broadcast a coded packet containing data for either of the nodes, on both the links, in a slot (see Figure 7.2). This simple strategy (*network coding and broadcasting*) is known to improve throughput as well as reduce the average power requirement.

The idea of network coding and broadcasting can be easily understood using the following example. Let Node 1 and Node 2 have packets a and b to be sent to their destinations, Node 2 and 1 respectively. We will assume that the capacity of the wireless channel is 1 packet per slot on every link. Consider the simple store-and-forward scheduling strategy. In the first slot, packet a is transmitted from Node 1 to Node 0; in the second slot, packet b is transmitted from Node 2 to Node 0; in the third slot, packet a is transmitted from Node 0 to Node 2 and in the fourth slot, packet b is transmitted from Node 0 to Node 1. The store-and-forward strategy requires four slots to communicate two packets to their destinations. Observe that in the third slot, Node 0 has both the packets a and b in its buffer. Suppose that the central node XORs packets a and b and broadcasts a coded packet $a \oplus b$ during the third slot, intended for both nodes. Also, suppose that Nodes 1 and 2 maintain a copy of packets a and b originated from them. Assuming that nodes 1 and 2 *can both decode* the network coded broadcast packet $a \oplus b$ from Node 0, Node 1

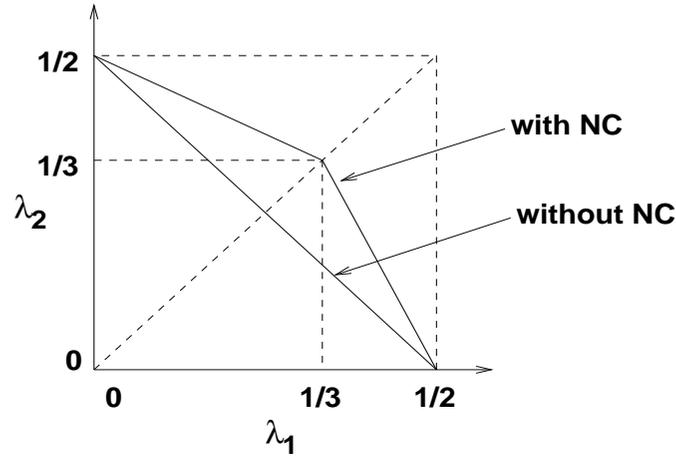


Figure 7.3: Capacity region of a two link wireless network (with 1 packet per slot wireless links and) with bidirectional traffic. Observe that network coding enhances the capacity region as compared to store-and-forward strategy.

(resp., 2) can then decode its packets of interest i.e., b (a) by XORing the received packet $a \oplus b$ with the packet a (b). Clearly, the network coding and broadcasting strategy saves 1 slot, thereby improving the throughput of the system. Figure 7.3 shows the capacity region of the example network with and without network coding. Observe that network coding enhances the capacity region of the network as compared to the store-and-forward strategy. In this work, we assume that the central node is capable of network coding, and further, the terminal nodes store their transmitted packets sufficiently long for them to decode any network coded packet.

In this chapter, we apply the idea of network coding and broadcasting strategy for wireless channels with fading. We assume that the channel state evolves according to a stationary Markov process and the scheduler has complete information about the channel fade gains at the beginning of every slot. We make the further assumption that nodes 1 and 2 communicate data packets to Node 0 through an independent channel, and study only the downlink problem at the central Node 0 (For example, $1 \rightarrow 0$ and $2 \rightarrow 0$ could be the uplink of a frequency division duplex cellular system). The packet arrivals at Node 0 (from Nodes 1 and 2) are modeled as stationary Markov processes. Now, in a slot, for any channel fade gain (h_1, h_2) , where h_1 and h_2 are the instantaneous gains on links $0 \rightarrow 1$ and $0 \rightarrow 2$ respectively (letting $h_1 > h_2$), the scheduler is confronted with the choice

of transmitting at a higher rate (a function of the channel fade gain h_1) to user 1 or to network code and transmit at a lower rate (restricted by the fading gain h_2), decodable by both the users. In this work, we study this tradeoff and obtain the optimal power allocation and coding strategy that *minimizes the average power required to stabilize the arrival processes*.

Situations such as these could arise in downlink OFDMA wireless schedulers if two mobiles on an OFDMA carrier wish to communicate with each other. Suppose that users 1 and 2 are involved in a bidirectional communication via base-station 0. The uplink and downlink traffic in OFDMA MAC are scheduled in different time slots and the optimization problem for the uplink/downlink traffic are handled separately. In such a scenario, the base station 0 can optimize the downlink resources by network coding packets of users 1 and 2 queued in its buffer. Gaming and other point-to-point applications contribute to peer-to-peer intracell traffic, and our work aims to study the resource allocation problem in such scenarios.

7.1.1 Literature Survey

The seminal work in [57] has sparked significant interest in network coding, both for wired and wireless networks. While there is considerable literature on network coding for single source multicast networks (see references in [3]), little is known about multisource networks, especially with unicast traffic. Also, most work on unicast traffic restrict network coding to simple models, like bidirectional traffic on links or restricted activation sets per slot ([74], [51], [58]).

In [51], the authors propose a scheduling strategy for general wireless networks by restricting network coding to bidirectional traffic, i.e., network coding is used along those linear paths that carry traffic on either direction. [51] extended the backpressure technique proposed in [37] and [42] (for networks without network coding), and achieved a $\Omega\left(\frac{1}{v}\right)$ tradeoff for delay for an excess average power of $O(v)$. In our work, we study a two-link wireless network with bidirectional traffic. We focus on the optimal power control and network coding policy that minimizes the average power required to stabilize the queues.

This work can be considered as an extension to the work of Goldsmith and Varaiya [5] for a single point-to-point fading wireless channel. Further, we extend the power-delay tradeoff results for wireless networks, studied in [55], [47] and [51], and obtain the optimal tradeoff for the network coding scenario.

7.1.2 Outline of the Chapter

In Section 7.2, we discuss the network model, transmission strategies and the transmission rate region. In Section 7.3, we study the capacity region of the network and the minimum average power required to support the given arrival rate. We discuss briefly, the power benefits of network coding in Section 7.3.3. In Section 7.4, we study the power-delay tradeoff achievable with network coding. Finally, we summarize the results in Section 7.5.

7.2 Network Model

We consider a two link slotted wireless network with bidirectional traffic. We assume that the nodes are synchronized and the slots are of a fixed length, T seconds. Packets flow from Node 1 to Node 2 and from Node 2 to Node 1 and are routed via the central Node 0. In this work, we assume that the packets from Node 1 and Node 2 reach Node 0 through a separate channel, and we focus only on the scheduling problem at the central node 0. Packet arrivals into the buffer at Node 0 (for Nodes 1 and 2) are modeled as stationary Markov processes. Let $\mathbf{A}(t) := (A_1(t), A_2(t))$ represent the vector of data arrivals to the queues at Node 0, corresponding to nodes 1 and 2, during the slot t . We define the average arrival rates as $(\lambda_1, \lambda_2) := (\mathbb{E}[A_1(t)], \mathbb{E}[A_2(t)])$. The packets are queued in an infinite buffer at Node 0 until transmission. Let $\mathbf{H}(t) := (H_1(t), H_2(t))$ represent the channel state vector corresponding to links $0 \rightarrow 1$ and $0 \rightarrow 2$ (respectively) during the slot t . The channel state corresponds to the power gain due to fading over the links, and we assume that the power gains remain constant during a slot. Further, we assume that the number of channel states is finite. Denoting the set of channel states by \mathcal{H} , we then define $\pi(\mathbf{h})$ as the probability of the channel being in state $\mathbf{h} = (h_1, h_2)$.

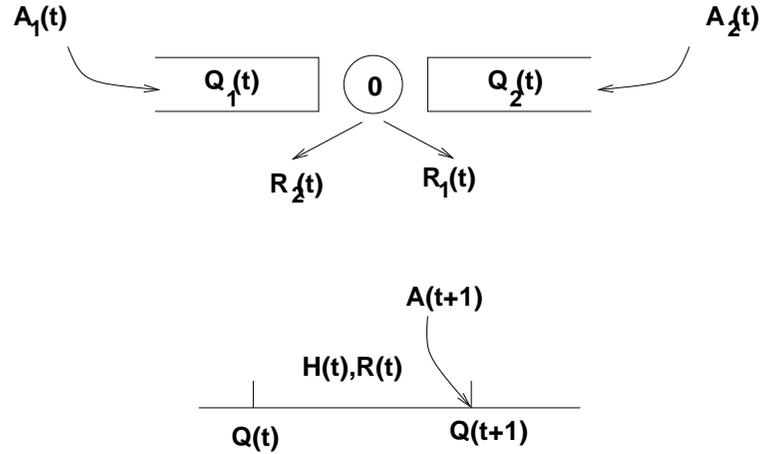


Figure 7.4: The scheduling problem at the central Node 0. $(A_1(t), A_2(t))$ represents the vector of data arrivals (for nodes 1 and 2 respectively) to the queues at Node 0 and $(Q_1(t), Q_2(t))$ corresponds to their queue lengths. $(H_1(t), H_2(t))$ is the channel power gain in the links $0 \rightarrow 1$ and $0 \rightarrow 2$ respectively, and $(R_1(t), R_2(t))$ is the allotted transmission rate (over the links) during the slot t . The queue dynamics for the system is given by, for $i = 1, 2$, $Q_i(t+1) = (Q_i(t) - R_i(t))^+ + A_i(t+1)$.

Let $\mathbf{Q}(t)$ denote the vector of data queued in the buffer of the source Node 0 at the beginning of slot t . For $\mathbf{R}(t)$, the transmission rate vector used in the slot t , the queue dynamics of the network is given by, for $i = 1, 2$,

$$Q_i(t+1) = \max \{Q_i(t) - R_i(t), 0\} + A_i(t+1)$$

The transmission rate vector, $\mathbf{R}(t) = (R_1(t), R_2(t))$, depends on the power gain $\mathbf{H}(t)$, the power allocated to the links and the transmission strategy employed. We do not restrict $R_i(t)$ (and hence $Q_i(t)$) to take integer values. In the remainder of this section, we will discuss two example transmission strategies for the bidirectional network employing different coding strategies.

7.2.1 Transmission Strategies and Rate Vectors

The feasible rate vector depends on the scheduling strategy and the coding strategy employed. Let (h_1, h_2) be the channel fade gain in a slot, and let $h_1 \geq h_2$. Define $C(h, p; \sigma^2)$ as the maximum bit rate achievable over a single link with channel power

gain h , transmit power p and noise power σ^2 . In general, we expect $C(h, p; \sigma^2)$ to be a concave non-decreasing function in p and non-decreasing in h . For e.g., $C(h, p; \sigma^2)$ could be $\log\left(1 + \frac{hp}{\sigma^2}\right)$ or $\sqrt{\frac{hp}{\sigma^2}}$. In practice, we can choose $C(h, p; \sigma^2)$ such that, the bit error rate is small and we ignore any residual errors as we assume that the application can tolerate such error rates.

The transmitter at node 0 now has the following three simple transmission options in a slot.

- **Schedule link $0 \rightarrow 1$** : The scheduler allots power p_1 and transmits $C(h_1, p_1; \sigma^2) \cdot T$ bits over link $0 \rightarrow 1$ in the slot, i.e., $(r_1, r_2) = (C(h_1, p_1; \sigma^2), 0)$.
- **Schedule link $0 \rightarrow 2$** : The scheduler allots power p_2 and transmits $C(h_2, p_2; \sigma^2) \cdot T$ bits over link $0 \rightarrow 2$ in the slot, i.e., $(r_1, r_2) = (0, C(h_2, p_2; \sigma^2))$.
- **XOR packets and broadcast over links $0 \rightarrow 1$ and $0 \rightarrow 2$** : The scheduler XORs a packet from each of the two flows, and broadcasts the coded packet with a common power p_{nc} in the slot. We will assume that $C(\min(h_1, h_2), p_{nc}; \sigma^2) \cdot T$ bits can be transmitted over either link in the slot, i.e.,

$$(r_1, r_2) = (C(\min(h_1, h_2), p_{nc}; \sigma^2), C(\min(h_1, h_2), p_{nc}; \sigma^2))$$

We have used $\min(h_1, h_2)$ as the worst case power gain for the broadcast packet, to ensure that the coded packet is received correctly over both the links. Thus, in this scheme, for a network coded packet, the minimum fading gain determines the maximum bit rate achievable.

Figure 7.5 shows the transmission rate vectors (bold continuous lines) associated with the simple transmission strategies discussed above. Also for $h_1 \geq h_2$, the network coding strategy dominates the second strategy, schedule link $0 \rightarrow 2$, i.e., the rate vectors achievable using the network coding strategy encompasses the rate vectors achievable using the strategy, schedule link $0 \rightarrow 2$.

Alternatively, the scheduler can employ a coding strategy similar to that used in broadcast channels ([65]). The transmitter codes data for the link $0 \rightarrow 1$ with power

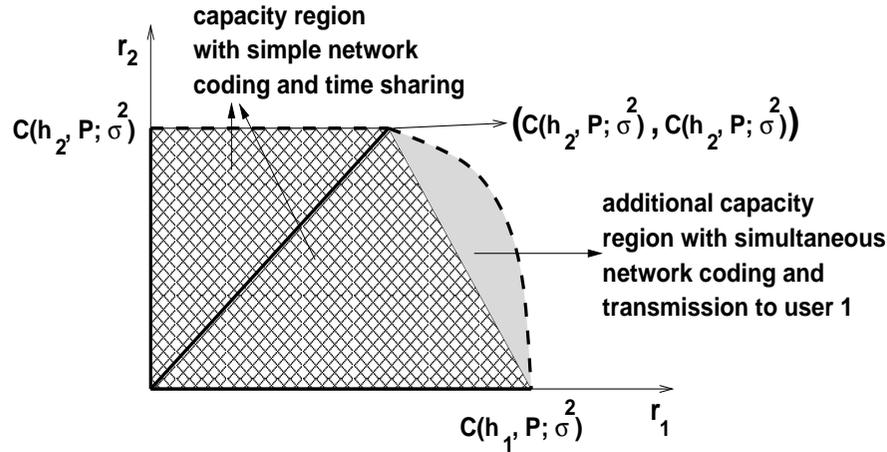


Figure 7.5: Transmission rate region for the simple transmission scheme, and the broadcast and coding scheme. We have fixed the channel state as (h_1, h_2) (with $h_1 > h_2$) and we have assumed a transmit power p , $p \in [0, P]$. The dark continuous lines correspond to the rate vector with the simple transmission scheme, i.e., $(C(h_1, p; \sigma^2), 0)$, $(0, C(h_2, p; \sigma^2))$ and $(C(\min(h_1, h_2), p; \sigma^2), C(\min(h_1, h_2), p; \sigma^2))$ for $p \in [0, P]$. The dark dashed line correspond to the rate vector with the broadcast strategy, i.e., $(C(h_2, p_2; h_2 p_1 + \sigma^2) + C(h_1, p_1; \sigma^2), C(h_2, p_2; h_2 p_1 + \sigma^2))$ for $p_1 \in [0, P]$ and $p_2 = P - p_1$. The shaded region is the rate region and is achieved by time sharing of the rate vectors.

p_1 (over a channel with power gain h_1) and common data with power p_{nc} (restricted by channel gain h_2) decodable by both the users. We note here that, since $h_1 \geq h_2$, the data coded with p_{nc} is decodable at User 1 as well. User 2 can now decode the common data at a rate $r_2 = C(h_2, p_{nc}; h_2 p_1 + \sigma^2)$; The power allocated for transmission to node 1, i.e., p_1 , is perceived as noise by user 2. User 1, seeing a better channel (h_1), can also decode the common data. Further, it can decode the data aimed to it at a rate $C(h_1, p_1; \sigma^2)$ by canceling the interference due to the common data transmitted with power p_{nc} (see [65]). Effectively, user 1 sees a combined rate $r_1 = C(h_2, p_{nc}; h_2 p_1 + \sigma^2) + C(h_1, p_1; \sigma^2)$. Figure 7.5 shows the transmission rate vectors achievable with the broadcast strategy (bold dashed lines). While the broadcast strategy increases the system complexity, the feasible rate region enlarges as well. Observe that the broadcast strategy improves the transmission rate region as compared with the simple transmission schemes discussed earlier (shaded region within the dashed bold lines).

7.3 Capacity Region

In this section, we will define the capacity region of the wireless network following the framework developed in [42] and [44]. Define a power allocation policy in a slot, \mathbf{p} , as $\mathbf{p} = (p_1, p_2, p_{nc})$, where p_1, p_2 and p_{nc} are the powers allotted, respectively, to data for users 1, 2, and the network coded data (common data). For every transmission strategy employed, we assume that, $\mathbf{p} \in \Pi$, a compact set, which defines the set of admissible power vectors in a slot. For example, Π could possibly define a maximum power constraint over the slot, i.e., $\Pi_{P_{\max}} := \{\mathbf{p} : p_1 + p_2 + p_{nc} \leq P_{\max}\}$. The feasible rate vector is now denoted by $(r_1, r_2) = \mathbf{r}(\mathbf{h}, \mathbf{p})$, for $\mathbf{h} \in \mathcal{H}$ and $\mathbf{p} \in \Pi$, where $\mathbf{r}(\mathbf{h}, \mathbf{p})$ could correspond to any transmission and coding strategy (similar to those discussed in Section 7.2.1).

Fixing a transmission scheme (e.g., one of the two strategies in Section 7.2.1), define $\boldsymbol{\mu}(\mathbf{h})$, the rate region for channel state \mathbf{h} , as the convex hull of the feasible rate vectors (r_1, r_2) in a slot, i.e.,

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{h}) &:= \text{ConvexHull}\{\mathbf{r}(\mathbf{h}, \mathbf{p}) : \mathbf{p} \in \Pi\} \\ &= \left\{ (r_1, r_2) : (r_1, r_2) \leq \sum_i \alpha_i \mathbf{r}(\mathbf{h}, \mathbf{p}_i) \text{ for } \mathbf{p}_i \in \Pi \text{ and } \sum_i \alpha_i \leq 1 \right\} \end{aligned}$$

$\mathbf{r}(\mathbf{h}, \mathbf{p})$ for $\mathbf{p} \in \Pi$ defines the set of all rate vectors available in a slot, and $\boldsymbol{\mu}(\mathbf{h})$ defines the set of all rate vectors achievable over a time sharing policy in a slot for a given channel state \mathbf{h} . In Figure 7.5, the three bold lines correspond to the rate vectors $\mathbf{r}(\mathbf{h}, \mathbf{p})$ and the shaded region corresponds to the rate region $\boldsymbol{\mu}(\mathbf{h})$. We will now use the definition of $\boldsymbol{\mu}(\mathbf{h})$ and define the capacity region of the wireless network, as the set of all data rates $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, for which some power allocation and coding strategy exists to stabilize the queues. We begin with introducing the notion of stability.

Definition 7.3.1 (from [42]) Define $e_i(M)$ as,

$$e_i(M) := \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \Pr(Q_i(\tau) > M)$$

The queues at node 0 are defined to be stable if $e_i(M) \rightarrow 0$ as $M \rightarrow \infty$ for all $i = 1, 2$. ■

The capacity region of the wireless network with a fading channel is now defined as follows.

Definition 7.3.2 (from [42]) *For an ergodic channel process $\mathbf{H}(t)$ with marginal p.m.f. $\pi(\mathbf{h})$, $\mathbf{h} \in \mathcal{H}$, and a compact admissible power set Π , define the rate region Γ by*

$$\Gamma := \sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \boldsymbol{\mu}(\mathbf{h})$$

■

The following theorem relates the stability of any arrival rate vector $\boldsymbol{\lambda}$ in terms of the capacity region Γ .

Theorem 7.3.1 (from [42]) *A necessary condition for stability is $\boldsymbol{\lambda} \in \Gamma$. Further, when arrivals and channel state variations are Markov modulated on a finite state space, a sufficient condition for stability is that $\boldsymbol{\lambda}$ is strictly in the interior to Γ .* ■

7.3.1 Minimizing the Average Power Requirement

In the previous section, we defined the rate region and the capacity region of the bidirectional wireless network shown in Figure 7.4. We will now obtain the minimum average power required to support a given rate vector $\boldsymbol{\lambda}$. Let $g(\mathbf{p})$ be a nonnegative and continuous function of the power vector \mathbf{p} (e.g., $g(\mathbf{p})$ could be $p_1 + p_2 + p_{nc}$). Now, define the average power cost as,

$$g_{avg} := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \mathbf{E} \{g(\mathbf{p}(\tau))\}$$

where $\mathbf{E} \{g(\mathbf{p}(t))\}$ is the expected power cost in slot t . It was shown in [44] that when $g(\mathbf{p})$ and $\mathbf{r}(\mathbf{h}, \mathbf{p})$ are continuous in the power vector \mathbf{p} , then the minimum average power cost g_{avg}^* required to support an arrival rate $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ which is stabilizable, is given by the solution to the following optimization problem.

$$\begin{aligned} \text{Minimize} & : g_{avg} = \sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \sum_{k=1}^L \gamma_k^{\mathbf{h}} g(\mathbf{p}_k^{\mathbf{h}}) \\ \text{subject to} & : \sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \sum_{k=1}^L \gamma_k^{\mathbf{h}} \mathbf{r}(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) \geq \boldsymbol{\lambda} \end{aligned} \quad (7.1)$$

$$\text{where } \forall \mathbf{h} \in \mathcal{H}, \quad \mathbf{p}_k^{\mathbf{h}} \in \Pi, \gamma_k^{\mathbf{h}} \in \mathcal{R}, \gamma_k^{\mathbf{h}} \geq 0, \sum_{k=1}^L \gamma_k^{\mathbf{h}} = 1$$

This formulation is understood as follows. The variables in the above optimization problem are $\gamma_k^{\mathbf{h}}$ and $p_k^{\mathbf{h}}$, for all $\mathbf{h} \in \mathcal{H}$ and $1 \leq k \leq L$. For a given channel state \mathbf{h} , $\gamma_k^{\mathbf{h}}$, $1 \leq k \leq L$ correspond to a set of auxiliary variables used to achieve every point in the rate region $\boldsymbol{\mu}(\mathbf{h})$. A power $p_k^{\mathbf{h}}$ is allotted for a fraction $\gamma_k^{\mathbf{h}}$ of slots during which the channel gain is \mathbf{h} . The average power allocated for the channel state \mathbf{h} is then given by $\sum_{k=1}^L \gamma_k^{\mathbf{h}} g(p_k^{\mathbf{h}})$ and the average rate (vector) achieved for the channel state \mathbf{h} is $\sum_{k=1}^L \gamma_k^{\mathbf{h}} \mathbf{r}(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}})$. From the Caratheodory theorem for convex hulls, we know that we need only a finite number of auxiliary variables per channel state \mathbf{h} to achieve every rate vector in $\boldsymbol{\mu}(\mathbf{h})$ (i.e., $L < \infty$). In our examples, we have $L = 3$ for the simple transmission strategies and for the broadcast strategy.

7.3.2 Optimal Power Allocation

We will now obtain the power allocation scheme that minimizes the average power for a given arrival rate $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$. The rate constraints in the optimization problem can be expanded as,

$$\sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \sum_{k=1}^L \gamma_k^{\mathbf{h}} r_1(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) \geq \lambda_1 \quad (7.2)$$

$$\sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \sum_{k=1}^L \gamma_k^{\mathbf{h}} r_2(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) \geq \lambda_2 \quad (7.3)$$

Using the Lagrangian approach, we can now relax the rate constraints in (7.1) and simplify the optimization problem as given below.

$$\text{Minimize} : \sum_{\mathbf{h} \in \mathcal{H}} \pi(\mathbf{h}) \sum_{k=1}^L \gamma_k^{\mathbf{h}} \left(g(\mathbf{p}_k^{\mathbf{h}}) - \beta_1 r_1(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) - \beta_2 r_2(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) \right) \quad (7.4)$$

$$\text{subject to} : \forall \mathbf{h} \in \mathcal{H}, \quad \mathbf{p}_k^{\mathbf{h}} \in \Pi, \gamma_k^{\mathbf{h}} \geq 0, 1 \leq k \leq L \text{ and } \sum_{k=1}^L \gamma_k^{\mathbf{h}} = 1$$

Observe that the optimization problem can be solved via a term-by-term minimization of $\sum_{k=1}^L \gamma_k^{\mathbf{h}} (g(\mathbf{p}_k^{\mathbf{h}}) - \beta_1 r_1(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) - \beta_2 r_2(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}))$ for each $\mathbf{h} \in \mathcal{H}$, i.e.,

$$\begin{aligned} \text{Minimize} & : \sum_{k=1}^L \gamma_k^{\mathbf{h}} (g(\mathbf{p}_k^{\mathbf{h}}) - \beta_1 r_1(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}}) - \beta_2 r_2(\mathbf{h}, \mathbf{p}_k^{\mathbf{h}})) & (7.5) \\ \text{subject to} & : \mathbf{p}_k^{\mathbf{h}} \in \Pi, \gamma_k^{\mathbf{h}} \geq 0 \quad \forall 1 \leq k \leq L \text{ and } \sum_{k=1}^L \gamma_k^{\mathbf{h}} = 1 \end{aligned}$$

Theorem 7.3.2 For $g(\mathbf{p}^{\mathbf{h}}) := p_1^{\mathbf{h}} + p_2^{\mathbf{h}} + p_{nc}^{\mathbf{h}}$, the sum power used when the channel state is \mathbf{h} , the optimization problem in (7.5) has a minimizer.

Proof: The objective function is continuous and the domain is a compact set. Hence, there exists a minimum and a minimizer for the optimization problem. ■

The optimal power allocation policy is now obtained by choosing the parameters β_1 and β_2 such that the average rate requirements (7.2) and (7.3) are satisfied.

An interesting feature of the optimal power allocation policy with network coding is that the set of channel states in which positive power is allotted is not an ordered set with respect to the rates (unlike a water-filling solution studied in [5]). Consider two rate vectors (λ_1, λ_2) and (ν_1, ν_2) such that $(\lambda_1, \lambda_2) \leq (\nu_1, \nu_2)$. Then it is not necessarily true that the set of channel states in which positive power is allotted for (λ_1, λ_2) is a subset of the set of channel states in which positive power is allotted for (ν_1, ν_2) . For example, consider a two state channel, $\mathcal{H} := \{(7, 0), (5, 5)\}$ and let $\pi(7, 0) = \frac{1}{2} = \pi(5, 5)$. Let $C := \log(1 + hp)$. Now, we see that, for an arrival rate $(x, 0)$, for x sufficiently small, it is optimal to allot positive power only to channel state $(7, 0)$. However, for an arrival rate (x, x) , it would be optimal to allot positive power only to the channel state $(5, 5)$ and perform network coding and allocate no power to the channel state $(7, 0)$.

7.3.3 Power Benefits of Network Coding

So far, we have seen that network coding enhances the set of rates that can be supported for a given transmission power constraint. Viewed alternately, network coding provides power savings as well. The following example shows that the power savings with network

coding can be arbitrarily large. Consider a constant channel gain (no fade, $h = 1$) system with the capacity function $C(h, p; \sigma^2)$ given by $C(h, p; \sigma^2) = \log\left(1 + \frac{hp}{\sigma^2}\right)$. For a given arrival rate vector (λ, λ) , define p_λ as the minimum average power required to stabilize the queues. The transmission power is minimized by network coding all the packets, and in our example, we have, $\lambda = \log\left(1 + \frac{p_\lambda}{\sigma^2}\right)$ or $p_\lambda = \sigma^2(e^\lambda - 1)$. Suppose that we choose not to network code. Then, the minimum power required to support the arrival rate, \tilde{p}_λ , is given by, $\tilde{p}_\lambda = \sigma^2(e^{2\lambda} - 1)$. For large λ , $\frac{\tilde{p}_\lambda}{p_\lambda} \approx e^\lambda$, which is unbounded. We note that the above bound is strictly a function of the capacity curve $C(h, p; \sigma^2)$. For example, when $C(h, p; \sigma^2) = \sqrt{\frac{hp}{\sigma^2}}$, a similar analysis will show that $\frac{\tilde{p}_\lambda}{p_\lambda} = 4$, i.e., the power gain with network coding is bounded.

In fact, such bounds hold true for arbitrary arrival rates and channel processes. For example, consider an average arrival rate (λ_1, λ_2) and an ergodic channel process with the transmission rate function $C(h, p; \sigma^2) = \sqrt{\frac{hp}{\sigma^2}}$. Let $p_{(\lambda_1, \lambda_2)}, p_{\lambda_1}$ and p_{λ_2} be the minimum average powers required to support the arrival rates $(\lambda_1, \lambda_2), (\lambda_1, 0)$ and $(0, \lambda_2)$. Trivially, $p_{(\lambda_1, \lambda_2)} \geq \max(p_{\lambda_1}, p_{\lambda_2})$. Let $\tilde{p}_{(\lambda_1, \lambda_2)}$ be the minimum power required to support the arrival rates without network coding. Then, again, $\tilde{p}_{(\lambda_1, \lambda_2)} \leq \max(4p_{\lambda_1}, 4p_{\lambda_2})$; The bound is achievable by allotting half the time slots exclusively for both the users. Hence, $\frac{\tilde{p}_{(\lambda_1, \lambda_2)}}{p_{(\lambda_1, \lambda_2)}} \leq 4$, or, the power gain with network coding is bounded again.

In practice, when the arrival rates are small, the set of channel states that are network coded are less. Hence, the benefit of network coding could be small. However, with increasing rates, the number of channel states that are network coded increases as well. Then, network coding becomes beneficial as it provides substantial power gains.

7.4 Power - Delay Tradeoff

In the previous section, we obtained the minimum average power required to achieve stability when the arrival rates are within the capacity region. It is well known that (see [55]) with minimum power, infinite delay has to be allowed in the system (e.g., for i.i.d. arrival process and i.i.d. channel process). Now, one can tradeoff the average power required to support the arrival process with the average queueing delay in the system. In

this section, we will study this tradeoff for our network scenario.

Define $Q(t)$ as the aggregate queue length in the system at time t (see Figure 7.4), $Q(t) := Q_1(t) + Q_2(t)$. We now define the time average of the mean aggregate queueing delay as,

$$D_{avg} = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_1^t \frac{\mathbb{E}[Q(t)]}{(\lambda_1 + \lambda_2)}$$

In [55], for a single downlink wireless fading channel with arrival rate λ into the queue, it was shown that, for i.i.d. arrival process and i.i.d. channel process, the asymptotic optimal average queueing delay (D_{avg}) scales as $\Omega\left(\frac{1}{\sqrt{v}}\right)$ for an excess average power of $O(v)$, i.e., for an average power of $P_\lambda + O(v)$, where P_λ is the minimum power required to support the arrival rate λ , the average queueing delay scales as $\Omega\left(\frac{1}{\sqrt{v}}\right)$. A buffer threshold based policy was used to achieve the optimal tradeoff. The results were extended to multi-user downlink wireless fading channels in [45]. However, the results in [45] are not applicable here (for the network coding case), because, the proof (in [45]) requires the Jacobian determinant of the minimum power cost function at the point of interest (λ_1, λ_2) to be non-zero. In the remainder of this section, we will study the optimal power-delay tradeoff achievable with network coding (for a two link wireless network) and show that the average queueing delay can scale only as $\Omega\left(\frac{1}{v}\right)$, and not as $\Omega\left(\frac{1}{\sqrt{v}}\right)$ (recall that for v close to 0, $\frac{1}{v} > \frac{1}{\sqrt{v}}$). We will motivate the tradeoff by considering a simpler model, without channel fade and with Poisson arrival of packets. We conjecture that the results can be extended to more complex scenarios as well.

Consider the wireless network in Figure 7.4, without channel fading and with Poisson arrival of packets with rate (λ, λ) . Let P_λ be the minimum power required to support the arrival rate $(\lambda, 0)$ (and $(0, \lambda)$). Then, clearly, P_λ is the minimum power required to support the arrival rate (λ, λ) as well. We achieve this minimum power by network coding every packet in the system. Packets of either traffic are queued until they can be network coded. This delay component is independent of the transmission delay and we define this delay as the *network-coding* delay.

From [55], we know that the optimal delay tradeoff for arrival rates $(\lambda, 0)$ and $(0, \lambda)$ is $\Omega\left(\frac{1}{\sqrt{v}}\right)$. Delay for the single user traffic is attributed to the transmission delay (or the

transmission rate). In our scenario, with an arrival rate (λ, λ) into both the queues, delay in the network comprises two components, transmission delay and *network-coding* delay. Without loss of generality, in this example, we will assume that the transmission delay is zero, and consider only the network-coding delay. Suppose that the transmission rate function is linear in the allotted power, i.e., $C(p) = kp$, where p is the allotted power and $k > 0$ is a constant. Then, the optimal policy would be to transmit at the maximum rate $C_{max} = C(p_{max})$ at all the time. The policy is optimal because the energy spent in transmitting a packet is independent of the bit rate when $C(p)$ is a linear function of p . Hence, delay is minimized by maximizing the transmission rate. Now, when the maximum transmission rate is unbounded (e.g., when $p_{max} = \infty$), we do not have any transmission delay in the network. In this example, we will assume that there is no transmission delay, and consider only the network-coding delay.

We will assume that every packet transmission consumes 1 unit of energy. When the scheduler network codes, then only $\frac{1}{2}$ unit of energy is consumed per packet transmission. Consider any scheduling policy, and let $f(v)$ be the fraction of packets that are not network coded with the scheduling policy. Then, the average power required to support the arrival rate (λ, λ) is given by $2\lambda \left((1 - f(v))\frac{1}{2} + f(v) \right) = 2\lambda(\frac{1}{2} + f(v))$, i.e., the excess average power used is $O(f(v))$. Hence, if the excess average power used is $O(v)$, then the fraction of uncoded packets is also $O(v)$. Thus, a constraint on average power directly translates into a constraint on network coding, which governs the queueing delay in the system.

As we have assumed that the transmission delay is zero, we do not gain anything by making the packets wait in both the queues. Hence, when there are packets in both the queues, we will serve them immediately. In other words, only one of the buffers has positive queue at any time. Now, the system evolution can be represented by a single variable Q , which represents the queue status, e.g., $Q = Q_1 - Q_2$. Define $\{\mu_i, -\infty \leq i \leq \infty\}$, as the service rate chosen when the queue size is i ; μ_i is the rate of departure of packets from state i . Clearly, we have $\mu_i \geq \lambda$ for all i , because every packet arrival into the other queue is counted as a service. In addition, we could decrease the queueing delay by

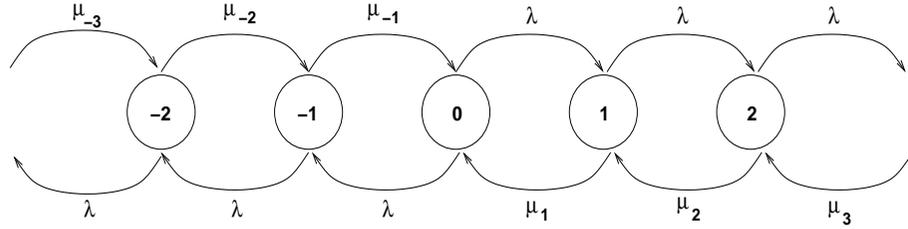


Figure 7.6: Continuous Time Markov Chain depicting the state evolution of the example system with network coding and zero transmission delay. The state represents the difference ($Q_1 - Q_2$) of the queues in the two buffers. Here, $\mu_i \geq \lambda$ for all i .

increasing the service rate μ_i beyond λ . μ_i can be viewed as a control parameter in state i which determines the fraction of packets that are network coded. For example, when $\mu_i = \lambda$, all the packets are network coded. When $\mu_i > \lambda$, $\mu_i - \lambda$ is the departure rate of uncoded packets (when the system is in state i). Figure 7.6 shows the state evolution of the Markov chain Q when the service rate is $\{\mu_i, -\infty \leq i \leq \infty\}$. For $\pi(i)$ the steady state distribution of Q , the mean queue length of the system is now given by,

$$\sum_{i=-\infty}^{\infty} |i| \pi(i) \quad (7.6)$$

and the fraction of packets that are not network coded is given by

$$\frac{1}{2\lambda} \sum_{i=-\infty}^{\infty} \pi(i) (\mu_i - \lambda) \quad (7.7)$$

Our problem is to minimize the queueing delay (7.6) subject to a constraint on the fraction of uncoded packets (7.7), equivalently, the excess average power. It is now easy to show that the optimal tradeoff for delay is $\Omega\left(\frac{1}{v}\right)$ when the excess average power is $O(v)$. We have motivated the proof in Appendix 7.6.1 by restricting to exponential service rates $\{\mu_i, -\infty \leq i \leq \infty\}$. While we have considered a simple network model (no channel fade) and zero transmission delay in this example, we note that the results can be extended to more complex scenarios. In [51], the authors propose a scheduling scheme which achieves $\Omega\left(\frac{1}{v}\right)$ tradeoff for delay for fading channels and concave transmission rate functions. Our contribution is to prove the optimality of the tradeoff achieved in [51].

7.5 Summary

In this chapter, we have studied a network coding and power allocation problem for a two link slotted wireless network with bidirectional traffic. We assumed that the central node can network code data packets belonging to the two routes (involved in bidirectional communication) and focussed on the downlink scheduling problem at the central node. In this framework, we defined the capacity region of the network and obtained the optimal power allocation policy and the network coding strategy that minimizes the average power requirement to support a given arrival process. Finally, we studied the asymptotic optimal tradeoff achievable between the average transmission power and the average queueing delay in the presence of network coding.

7.6 Appendix

7.6.1 Delay is $\Omega\left(\frac{1}{v}\right)$ if Excess Power is $O(v)$

All order notations in this section are with respect to $v \rightarrow 0$. We will use the following results in this section. For $v > 0$, define $\rho := \frac{\lambda}{\lambda+v}$, where $\lambda > 0$. Then as $v \rightarrow 0$, we have,

$$\begin{aligned}\sum_{i=0}^{\infty} \rho^i &= \frac{1}{1-\rho} = \frac{\lambda+v}{v} = \Theta\left(\frac{1}{v}\right) \\ \rho^{\frac{1}{v}} &= \left(\frac{\lambda}{\lambda+v}\right)^{\frac{1}{v}} = \Theta(1)\end{aligned}$$

and,

$$\sum_{i=0}^{\frac{1}{v}} \rho^i = \sum_{i=0}^{\infty} \rho^i - \sum_{i=\frac{1}{v}+1}^{\infty} \rho^i = \frac{1}{1-\rho} - \rho^{\frac{1}{v}+1} \frac{1}{1-\rho} = \frac{1}{1-\rho} (1 - \rho^{\frac{1}{v}+1}) = \Theta\left(\frac{1}{v}\right) \Theta(1) = \Theta\left(\frac{1}{v}\right)$$

where we have assumed that $\frac{1}{v}$ is an integer.

Consider any scheduling policy, and let $f(v)$ be the average fraction of packets that are not network coded with the scheduling policy. Then, the average power required to support the arrival rate (λ, λ) with the scheduling policy is $2\lambda \left((1-f(v))\frac{1}{2} + f(v) \right) =$

$\lambda(1+2f(v))$. Hence, if the excess average power used is $O(v)$, then the fraction of uncoded packets is also $O(v)$. Hence, hereafter, we will use these terms (fraction of uncoded packets and excess average power) interchangeably.

For the network shown in Figure 7.4, without channel fade and with Poisson arrival of packets, we will now show that, average packet delay is $\Omega\left(\frac{1}{v}\right)$ if the excess average power is $O(v)$. We will prove the result by contradiction, i.e., we will show that there does not exist any scheduling policy such that the average delay is $\Theta\left(\frac{1}{v^{1-\epsilon}}\right)$, $\epsilon > 0$, if the excess average power is $O(v)$. We will restrict to exponential rate controls (i.e., $\{\mu_i, -\infty \leq i \leq \infty\}$ are exponential), and hence, the CTMC shown in Figure 7.6 represents the system evolution.

Suppose that there exists a rate allocation policy $\{\tilde{\mu}_i, -\infty \leq i \leq \infty\}$, with an excess average power $O(v)$ and average queueing delay, $\Theta\left(\frac{1}{v^{1-\epsilon}}\right) = \frac{k}{v^{1-\epsilon}}$, $k > 0$. We will assume that $\tilde{\mu}_i = \tilde{\mu}_{-i}$, and $\tilde{\mu}_i$ are monotone increasing in $|i|$. Also, we will assume that $\tilde{\mu}_i = \lambda + \Theta(v^{1-\epsilon})$ for large $|i|$. A rough justification is the following, we require $\tilde{\mu}_i \geq \lambda + \Theta(v^{1-\epsilon})$ for large $|i|$ to support $\Theta\left(\frac{1}{v^{1-\epsilon}}\right)$ queueing delay. Further, for the target average queueing delay scaling, we can minimize the average power if we set $\tilde{\mu}_i = \lambda + \Theta(v^{1-\epsilon})$.

Let $\{\tilde{\pi}(i), -\infty \leq i \leq \infty\}$ be the stationary distribution of the queue obtained from solving the CTMC shown in Figure 7.6, and

$$\tilde{\pi}(i) = \frac{1 + \sum_{j=1}^{|i|} \prod_{k=1}^j \tilde{\rho}_k}{1 + 2 \sum_{j=1}^{\infty} \prod_{k=1}^j \tilde{\rho}_k}$$

where $\tilde{\rho}_i = \frac{\lambda}{\tilde{\mu}_i}$. We will first show that the delay constraint, $\frac{k}{v^{1-\epsilon}}$, requires the denominator of $\tilde{\pi}$ to be equal to $\Theta\left(\frac{1}{v^{1-\epsilon}}\right)$. Then, we will show that the network coding constraint (equivalently the excess power constraint) requires the denominator of $\tilde{\pi}$ to be $\Omega\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$, which yields the contradiction.

As we have assumed that $\sum_i |i| \tilde{\pi}(i)$, the average queueing delay, to be equal to $\frac{k}{v^{1-\epsilon}}$, we require that the probability the queue exceeds $\frac{2k}{v^{1-\epsilon}}$ is less than $\frac{1}{2}$, i.e.,

$$1 - \sum_{i=-2\frac{k}{v^{1-\epsilon}}}^{2\frac{k}{v^{1-\epsilon}}} \tilde{\pi}(i) \leq \frac{1}{2}$$

or,

$$\sum_{i=-2\frac{k}{v^{1-\epsilon}}}^{2\frac{k}{v^{1-\epsilon}}} \tilde{\pi}(i) \geq \frac{1}{2}$$

Rewriting in terms of $\tilde{\rho}_i$, we have,

$$\frac{1 + 2 \sum_{i=1}^{2k} \prod_{j=1}^i \tilde{\rho}_j}{1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j} \geq \frac{1}{2}$$

Upper bounding $\tilde{\rho}_j$ by 1 (since $\tilde{\mu}_j \geq \lambda$ for all j) in the numerator of the above inequality, we have,

$$\frac{1 + 2 \left(\frac{2k}{v^{1-\epsilon}} \right)}{1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j} \geq \frac{1}{2}$$

Rearranging terms, we get

$$2 \left(1 + \frac{4k}{v^{1-\epsilon}} \right) \geq 1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j$$

or,

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j = O \left(\frac{1}{v^{1-\epsilon}} \right)$$

for small v . But, we have assumed that $\tilde{\mu}_i \leq \lambda + \Theta(v^{1-\epsilon})$ for all i . Hence, we require,

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j = \Omega \left(\frac{1}{v^{1-\epsilon}} \right)$$

as well. From the above two expressions, we conclude that,

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j = \Theta \left(\frac{1}{v^{1-\epsilon}} \right) \quad (7.8)$$

We have assumed that $\lambda \leq \tilde{\mu}_i \leq \lambda + \Theta(v^{1-\epsilon})$ for all i . Further, we know that $\tilde{\mu}_i = \lambda + \Theta(v^{1-\epsilon})$ for large $|i|$. Hence, there exists a N_v such that $\tilde{\mu}_i \geq \lambda + \Theta(v^{1-\frac{\epsilon}{2}})$ for all $|i| \geq N_v$. Now, consider the constraint on the fraction of uncoded packets,

$$O(v) = \frac{2}{2\lambda} \sum_{i=1}^{\infty} \pi(i) (\tilde{\mu}_i - \lambda)$$

$$\begin{aligned}
 &\geq \frac{2}{2\lambda} \sum_{i=N_v}^{\infty} \pi(i)(\tilde{\mu}_i - \lambda) \\
 &\geq \frac{2}{2\lambda} \sum_{i=N_v}^{\infty} \pi(i)\Theta\left(v^{1-\frac{\epsilon}{2}}\right) \\
 &= \Theta\left(v^{1-\frac{\epsilon}{2}}\right) \frac{2}{2\lambda} \sum_{i=N_v}^{\infty} \pi(i)
 \end{aligned}$$

Since $O(v)$ is smaller than $\Theta\left(v^{1-\frac{\epsilon}{2}}\right)$ for $v \rightarrow 0$, we require that $2 \sum_{i=N_v}^{\infty} \pi(i) < \Omega(1)$.

Rewriting the condition in terms of $\tilde{\rho}_i$, we need to have

$$2 \frac{\prod_{j=1}^{N_v} \tilde{\rho}_j \left(1 + \sum_{i=N_v+1}^{\infty} \prod_{j=N_v+1}^i \tilde{\rho}_j\right)}{1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j} < \Omega(1)$$

Substituting from (7.8) in the denominator of the above expression, we get,

$$2 \frac{\prod_{j=1}^{N_v} \tilde{\rho}_j \left(1 + \sum_{i=N_v+1}^{\infty} \prod_{j=N_v+1}^i \tilde{\rho}_j\right)}{\Theta\left(\frac{1}{v^{1-\epsilon}}\right)} < \Omega(1) \quad (7.9)$$

Consider the summation series in the numerator of the left hand side expression of the above equation, i.e., $1 + \sum_{i=N_v+1}^{\infty} \prod_{j=N_v+1}^i \tilde{\rho}_j$. Since $\tilde{\mu}_i \leq \lambda + \Theta(v^{1-\epsilon})$ for all i , we have,

$$1 + \sum_{i=N_v+1}^{\infty} \prod_{j=N_v+1}^i \tilde{\rho}_j = \Omega\left(\frac{1}{v^{1-\epsilon}}\right)$$

Substituting in (7.9) and simplifying the expression, we need,

$$\prod_{j=1}^{N_v} \tilde{\rho}_j < \Omega(1) \quad (7.10)$$

Now, $\tilde{\mu}_i \leq \lambda + \Theta(v^{1-\frac{\epsilon}{2}})$ for all $i \leq N_v$. Hence,

$$\left(\frac{\lambda}{\lambda + \Theta(v^{1-\frac{\epsilon}{2}})}\right)^{N_v} \leq \prod_{j=1}^{N_v} \tilde{\rho}_j$$

But, $\left(\frac{\lambda}{\lambda + \Theta(v^{1-\frac{\epsilon}{2}})}\right)^{N_v} = \Omega(1)$ for $N_v = O\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$. Hence, for $\prod_{j=1}^{N_v} \tilde{\rho}_j < \Omega(1)$, we need $N_v > O\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$.

Now consider the summation series in the left hand side of the equation (7.8).

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j = 1 + 2 \sum_{i=1}^{N_v-1} \prod_{j=1}^i \tilde{\rho}_j + 2 \sum_{i=N_v}^{\infty} \prod_{j=N_v}^i \tilde{\rho}_j \geq 1 + 2 \sum_{i=1}^{N_v-1} \prod_{j=1}^i \tilde{\rho}_j$$

Lower bounding $\tilde{\rho}_j$ by $\frac{\lambda}{\lambda + \Theta(v^{1-\frac{\epsilon}{2}})}$ for all $i \leq N_v$, we have,

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j \geq 1 + 2 \sum_{i=1}^{N_v-1} \left(\frac{\lambda}{\lambda + \Theta(v^{1-\frac{\epsilon}{2}})} \right)^i$$

But, $\sum_{i=1}^{N_v} \left(\frac{\lambda}{\lambda + \Theta(v^{1-\frac{\epsilon}{2}})} \right)^i = \Theta\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$ for $N_v > O\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$. Hence, we have

$$1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j \geq \Theta\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$$

But we know from (7.8) that $1 + 2 \sum_{i=1}^{\infty} \prod_{j=1}^i \tilde{\rho}_j = \Theta\left(\frac{1}{v^{1-\epsilon}}\right)$ which is strictly smaller than $\Theta\left(\frac{1}{v^{1-\frac{\epsilon}{2}}}\right)$, which gives a contradiction. Hence, there does not exist a $\epsilon > 0$ such that the average delay is $\Theta\left(\frac{1}{v^{1-\epsilon}}\right)$, or, the optimal delay is only $\Omega\left(\frac{1}{v}\right)$ for an excess average power of $O(v)$.

Chapter 8

Delay Optimal Scheduling in a Two Hop Vehicular Relay Network

8.1 Introduction

In the previous chapters, we have studied resource optimization problems for stationary wireless nodes. In this chapter, we consider a mobile multihop wireless network scenario and study a delay minimization problem. We characterize the queueing delay (at the buffer of the source node) as well as the transit delay (in the mobile nodes).

We consider a scheduling problem in a wireless network where vehicles are used as store-and-forward relays. A stationary source node wishes to communicate a file to a stationary destination node, located beyond its communication range. In the absence of any fixed infrastructure connecting the source node and the destination node, we study the possibility of data transfer by relaying the packets of the file using vehicles passing by. For example, the source and the destination of the file transfer could be located by the side of a road. At random times, vehicles equipped with radio transceivers (and willing to serve as relays) pass by the source node towards the destination node, with speed v sampled from a known probability distribution. The source node communicates packets of the file to the destination node, by using these vehicles as relays. In this chapter, we assume that the vehicles communicate with the source node and the destination node

only, and they do not communicate among themselves. Thus, all packet communication between the source and the destination involves two hops. Packet delays in this network scenario comprise two components; queueing delay at the source node (before the packet is relayed) and transit delay in the relay vehicle (until the packet is delivered). The transit delay is assumed to be a function of the speed of the vehicle, and we assume that the source node has full knowledge of the vehicle speed at the time of relaying.

As the decision to relay a packet is taken only once (by the source node), there arises a natural tradeoff between the queueing delay of the packet and its transit delay. A small average queueing delay leads to large average transit delay and vice versa. Our objective is to study the source node's sequential decision problem of transferring packets of the file to vehicles as they pass by, with the objective of minimizing the average packet delay in the network. We study both the finite file size case and the infinite file size case. In the finite file size case, we are interested in minimizing the expected delay to transfer the file (queueing delay plus the transit delay). In the infinite file size case, we study the optimal tradeoff achievable between the average queueing delay of the packets at the source node buffer and the average transit delay of the packets in the relay vehicle.

The above situation would arise in a wireless data network where there is limited access to backbone infrastructure and vehicles are used as a means of communication (as a relay). The fixed infrastructure of the wireless data network would consist of data posts located at various points along the roads. The data posts, in general, need not communicate directly with each other and we envision the possibility of using vehicles passing by to route packets among the data posts. Some of the data posts are connected to the backbone network, which in turn is connected to the Internet (we would call such data posts as gateway-data posts). Except for a few gateway-data posts, the network requires very little infrastructure, i.e., the data posts can be deployed anywhere along the roads, even without any direct connectivity to any other data post or a fixed network. The data posts function both like routers as well as (WLAN) access points in hot-spots. End users communicate via the data posts, and the packets are routed appropriately to their destination using the vehicles. The advantage of the proposed architecture as compared

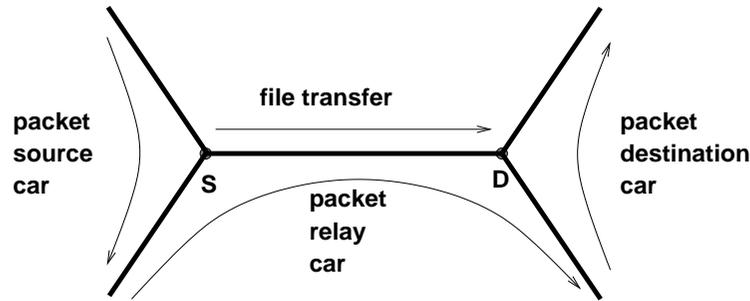


Figure 8.1: A vehicular relay network: “Source car” has z data packets to communicate to the “destination car”. The packets are first transferred from the source car to the stationary node S . Node S communicates the packets to another stationary node D using a “relay car” that passes by the nodes S and D . Finally, the packets are transferred from the stationary node D to the “destination car”.

to a traditional wireless network is that, we maintain network connectivity among the data posts using the vehicles and hence, there is no interference or the need of network design. Figure 8.1 shows an example scenario. A “source car” wishes to transfer a file to a “destination car”, beyond its communication range. The packets are first transferred from the source car to a stationary node S . Then, the file is communicated to another stationary node D using a “relay car”, which passes by nodes S and D . The file is finally transferred from the stationary node D to the destination car, when the destination car passes by node D .

Delay in such networks could be quite large. For example, a source node may have to wait for few tens of seconds to a few minutes, before a vehicle arrives. Similarly, the relay vehicle might cover the distance between the source and the destination node in a few minutes to even hours. The TCP protocol developed for the Internet will not work with such large end-to-end delays. Our vehicular network is an example of a Delay Tolerant Network (or Disruption Tolerant Network/DTN). The main characteristics of DTN are, intermittent connection, large delay and asymmetric data rates. The transmission opportunities could be random and can include large intermediate delays and arbitrary rates. The protocols suited to the Internet (with delay tolerance of a few tens of seconds) are not appropriate here and we require new protocols and strategies for the efficient operation of such networks. A number of recent research works focus on the performance, optimization

and operation of such delay tolerant networks. For example, in [32], the authors propose a MaxProp protocol for effective routing of messages, by prioritizing the schedule of packets transmitted to other peers. These priorities are based on the network properties like path likelihoods to peers. In this chapter, we restrict ourselves to a scheduling problem in a simple one hop communication scenario between two data posts, a source node (S) and a destination node (D). The file (at the source node) is assumed to be partitioned into fixed length fragments. These fragments are not necessarily IP packets and can be of arbitrary size depending on the bit rates achievable between the source/destination node and the relay vehicles ([36] discusses the idea of aggregating data into “bundles” (of fixed size) in DTNs). In this work, for ease, we will refer to such fragments only as packets. We have assumed that the source and the destination nodes (S and D) are stationary. Extensions to mobile source and destination nodes and to arbitrary networks are not studied here and could be considered as future work.

8.1.1 A Survey of the Literature

In this chapter, we are interested in the throughput and delay performance of a vehicular relay network. In [41], Grossglauser and Tse studied the capacity of mobile ad hoc networks with n random source-destination pairs. They showed that mobility increases the capacity of ad hoc networks ($O(n)$) as compared to fixed networks ($O(\sqrt{n})$). The optimal throughput-delay tradeoff for the scenario was later studied in [43]. We consider a single source-destination pair, and assume that passing by vehicles (governed by a distribution) are used to relay packets between them. The traffic is assumed to be along a road connecting the source and the destination, and our framework is very similar to the vehicular network problems studied in [73], [52] and [8]. The focus of such works is emergency applications. The objective is to communicate emergency or traffic information to avoid accidents and traffic congestion. The vehicles use a broadcasting strategy to communicate packets among themselves. Delay, and not throughput, is the most important performance measure in such problems, and hence, they tend to be throughput inefficient.

In [35], the authors study the bandwidth in road to vehicle communications using simulations. They compare the performance of the system for different medium access protocols used in the road-to-vehicle communication. We also study the bandwidth in the road to vehicle communication, however, in terms of the distribution of the interarrival times and the vehicle speeds. Further, we study the delay characteristics of the system as well.

The idea of using stationary nodes as data posts can be easily motivated from the literature. In [31], the authors study the idea of using vehicles to carry and forward data packets to the destination. Using a vehicle mobility model, the next vehicle to forward the packet is chosen such that the packet delay is minimized. Special purpose ferrying nodes were suggested in [70] to improve the delay characteristics of such mobile network. In our framework, we have assumed fixed nodes (data posts like S and D), instead of mobile nodes as suggested in [70], for storing and forwarding the data packets. [15] proposed the idea of INFOSTATIONS, with isolated pockets of high bandwidth connectivity, for future data and messaging services. Our data posts act both as INFOSTATIONS as well as routers, and the vehicles serve the purpose of relay nodes as well.

8.1.2 Outline of the Chapter

In Section 8.2, we describe the network model, and in Section 8.3, the optimization problem. We study the expected total delay minimization problem for the finite file case in Section 8.4. In Section 8.5, for the infinite file case, we study the asymptotically optimal tradeoff achievable between the queueing delay and the transit delay of the packets. Section 8.6 concludes the chapter.

8.2 The Network Model

A stationary source node intends to transfer a file to a stationary destination node, located at a distance s metres away (i.e., a road of length s connects the two). At random times, vehicles that drive in the direction of the destination node enter the communication range of the source node. We assume that the interarrival times of the vehicles are i.i.d. and have

a c.d.f. $\mathcal{I}(\cdot)$ *known* to the source node. The i^{th} vehicle that enters the communication range of the source node travels at an average speed V_i towards the destination node, where $\{V_i, i \geq 1\}$ are assumed to be an i.i.d. sequence with c.d.f. $V(\cdot)$ (*known* to the source node). We assume that the i^{th} vehicle (with speed V_i) takes $\frac{s}{V_i}$ seconds to cover the distance between the source and the destination. Further, we make a simplifying assumption that the interarrival times of the vehicles are distributed independent of the speed of the vehicles. As the i^{th} vehicle enters the communication range of the source node, the source node gets information about this event as well as about the speed of the vehicle, V_i .

In this work, we assume that the source node can relay at most a single packet to a vehicle, when it passes the coverage area of the source, independent of the vehicle speed. We note that the time spent by the vehicles in the communication range of the source/destination will depend on the speed of the vehicle. This may permit the vehicles to relay different amounts of data (depending on their speed). However, when the data posts (such as S and D in Figure 8.1) are located at a junction (e.g., a traffic signal or a crossing), the vehicle speeds in the coverage region will have little correlation with the average speed V_i achieved over the distance s to the destination. Hence, we restrict to single packet (of arbitrary but fixed size) per relay vehicle model. Packet relay to vehicles (with different speeds) is achieved either by coding packets at different data rates or by restricting the set of vehicles to which communication is permitted.

Due to randomness in the interarrival times and in the vehicular speeds, it is possible that the different packets of a file arrive at the destination node in a random order. We assume that the destination node has the capability to reassemble these packets. Also, it is possible that more than one vehicle is in the coverage region of the source or the destination node at any given point of time. We assume that both the source and the destination nodes are capable of simultaneous communication with vehicles, possibly in different, non-interfering bands.

8.3 The Optimization Problem

Packet delay in this network comprises queueing delay at the source node, transmission delay (depends on the bit rate at which data is transmitted to/from the relay vehicle) and transit delay in the relay vehicle. In this work, we assume that the transmission delay is very small as compared to the transit delay and study only the queueing delay and the transit delay of the packets. We consider the following two versions of the scheduling problem in this chapter.

1. **Finite File Case:** The source node has a single file with z packets ($1 \leq z < \infty$). As we relay only one packet per vehicle, this requires that the source node use z vehicles to relay the file. Starting with z packets at $t = 0$, the problem is to sequentially select z vehicles, so as to minimize the expected time until the file is completely transferred to the destination. We note here that the random file transfer completion time is the maximum of the sojourn times of the z packets in the network, and not just the sojourn time of the last transmitted packet. When minimizing the expected sojourn time of the file (queueing delay plus the transit delay), if a slow vehicle arrives, it might be optimal to ignore it and to wait until the next vehicle arrives in the hope that it would be faster. In fact we would wish it to be sufficiently faster so as to compensate for the extra waiting time. In this work, we study this tradeoff and find the delay minimizing policy.
2. **Infinite File Case:** We assume that the source node has infinitely many packets to communicate to the destination node. More precisely, there is a packet arrival process at the source node, independent of the vehicle arrival process. Packets are queued at the source buffer until transmission, and our aim is to study the impact of the scheduling policy on the queueing delay and the transit delay of the packets. There is a natural tradeoff between the queueing delay and the transit delay of the packets. We can minimize the queueing delay by choosing every vehicle (slow or fast) as a relay, thereby increasing the transit delays of the packets. Similarly, by relaying only to fast vehicles, we can decrease the transit delay of the packets while

increasing the queueing delay at the source buffer. We first obtain the maximum throughput sustainable in the network for a given transit delay constraint. Using the throughput vs transit-delay curve, we then characterize the asymptotically optimal tradeoff between the queueing delay and the transit delay of the packets in the network.

8.4 Finite File Size

In this section, we study the case where the source node has a file comprising z packets ($1 \leq z < \infty$) to communicate to the destination node. We are interested in minimizing the expected delay in transferring the file completely to the destination. Each instant at which a vehicle enters the communication range of the source node, a decision has to be made on using the new vehicle as a relay. Starting with the first vehicle to arrive¹, the decision problem evolves over vehicle arrival instants $\{T_1 = 0, T_2, \dots\}$ with vehicle speeds $\{V_1 = v, V_2, \dots\}$. As discussed in Section 8.2, $\{T_k, k \geq 1\}$ are renewal instants with i.i.d. interarrival times $\{\mathcal{I}_{k+1} = T_{k+1} - T_k, k \geq 1\}$ and the vehicle speeds $\{V_k, k \geq 1\}$ are i.i.d. with c.d.f. $V(\cdot)$. Let X_k denote the residual number of packets of the file, at the source node, at time instant T_k (with $X_1 = z$). Define Y_k as the action chosen at time instant T_k , i.e., the number of packets relayed to the k^{th} vehicle. In our context, we have, $Y_k \in \{0, 1\}$. When $X_k = 0$, $Y_k = 0$ is the only permissible action.

We are interested in the total delay incurred in transferring the file. Let \mathcal{T}_i denote the random time at which the i^{th} packet of the file arrives at the destination. Then, the random delay of the file transfer is given by $\max(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_z)$. Suppose that the source node had relayed $l < z$ packets up to time T_k . Then, the partial cost (delay) incurred due to the transfer of l packets is $\max(\mathcal{T}_1, \dots, \mathcal{T}_l)$. Define $\{D_k, k \geq 1\}$ as the residual delay (after T_k) for the last of any previously relayed data to reach the destination, i.e., $D_k := (\max(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_l) - T_k)^+$. Clearly, $T_k + D_k$ lower bounds the time required to complete the file transfer. Further, the decision to relay a packet at T_k is clearly

¹the mean time until the arrival of the first vehicle is a fixed quantity. Hence, this need not be accounted for in the optimization.

influenced by the value of D_k . For example, when $\frac{s}{V_k}$ is such that $\frac{s}{V_k} < D_k$, then, a packet relay by vehicle k does not increase the file transfer delay, because, $\max(\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_l) = \max(\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_{l+1})$ in this case. Also, D_k together with T_k contains all the necessary information required to define the partial and total cost functions at time T_k . Hence, we define (X_k, V_k, D_k) as the system state. The system state dynamics is given by,

$$\begin{aligned} X_{k+1} &= X_k - Y_k \\ D_{k+1} &= \left(\max \left\{ D_k, I_{\{Y_k > 0\}} \frac{s}{V_k} \right\} - \mathcal{I}_{k+1} \right)^+ \end{aligned}$$

Define $R_k(x_k, v_k, d_k, y_k)$, the single stage cost associated with the system state (x_k, v_k, d_k) and action y_k as,

$$R_k(x_k, v_k, d_k, y_k) = \begin{cases} 0 & x_k = 0 \\ \mathcal{I}_{k+1} & x_k > 1 \\ \mathcal{I}_{k+1} & x_k = 1 \text{ and } y_k = 0 \\ \max \left\{ d_k, \frac{s}{v_k} \right\} & x_k = 1 \text{ and } y_k = 1 \end{cases} \quad (8.1)$$

R_k is equal to the interarrival time \mathcal{I}_{k+1} , when the file transfer is incomplete. Hence, the partial sum $\sum_{i=1}^k R_i(x_i, v_i, d_i, y_i)$, corresponding to the cost incurred up to time instant T_k , is equal to T_k when the file transfer is incomplete. Now, for a sequence $\{(X_k, V_k, D_k, Y_k), k \geq 1\}$, satisfying the above dynamics, define the total cost function as

$$\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, Y_k)$$

The total cost expression is the random time of delivery of the file to the destination from the moment $T_1 = 0$. Our aim is to minimize the expected value of the above total cost function, i.e., to minimize $\mathbb{E}_{\{\mathcal{I}, V\}} [\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, Y_k)]$, where the expectation is over the interarrival times $\{\mathcal{I}_1, \mathcal{I}_2, \dots\}$ and the vehicular speeds $\{V_1, V_2, \dots\}$. The minimization is over the set of all policies $\{\pi_k, k \geq 1\}$, where π_k , the decision rule for time instant T_k , is a function of $\{(X_j, V_j, D_j), 1 \leq j \leq k\}$ and is a distribution on $\{0, 1\}$.

Given $X_1 = z, V_1 = v$ and $D_1 = 0$, our aim is to solve the following problem,

$$\inf_{\{\pi_k, k \geq 0\}} \mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, Y_k) \middle| X_1 = z, V_1 = v, D_1 = 0 \right] \quad (8.2)$$

The problem has been formulated in a Markov decision process framework with a total cost criterion. Suppose that the speed of the vehicle is continuously distributed (or discrete) within a bounded set, i.e., $V_k \in [v_{min}, v_{max}]$. Further, assume that the interarrival time is continuously distributed (or discrete). Then, $D_k \in [0, \infty)$, and also, the state space of (X_k, V_k, D_k) is a Borel set. The action space Y_k is finite in our case. When $\mathbb{E}[\mathcal{I}]$, the mean interarrival time of vehicles, is finite, the single stage cost function R , and its expectation is bounded as well. Then, from [54], we see that there exists a stationary deterministic Markov policy that achieves the minimum for (8.2), i.e., there exists a optimal policy π^* such that $\pi_k^* \equiv \pi^*(x_k, v_k, d_k) \in \{0, 1\}$. The optimal cost function satisfies the Bellman's dynamic programming equation. Also, the stationary policy which solves the Bellman's dynamic programming equation is an optimal policy.

Theorem 8.4.1 *Let $\tau^*(x, v, d)$ be the optimal value of (8.2). Then, $\tau^*(x, v, d)$ satisfies the following dynamic programming (DP) equation.*

$$\tau^*(x, v, d) := \begin{cases} 0 & x = 0 \\ \min \left\{ \mathbb{E}[\mathcal{I}] + \mathbb{E}_{\mathcal{I}, V}[\tau^*(x, V, (d - \mathcal{I})^+)], \max\left(\frac{s}{v}, d\right) \right\} & x = 1 \\ \min_{y \in \{0, 1\}} \left\{ \mathbb{E}[\mathcal{I}] + \mathbb{E}_{\mathcal{I}, V}[\tau^*(x - y, V, (\max(d, I_{\{y > 0\}} \frac{s}{v}) - \mathcal{I})^+)] \right\} & x > 1 \end{cases} \quad (8.3)$$

The stationary policy that chooses the minimizer of the right hand side expression in (8.3) is an optimal policy. ■

Remark: Define $\pi^*(x, v, d)$ as the stationary policy which minimizes the DP equation (8.3). Notice that $\pi^*(x, v, d) \in \{0, 1\}$, as we relay at most one packet using any relay vehicle. By definition, when $x = 0$, $\tau^*(0, v, d) = 0$ and $\pi^*(0, v, d) = 0$. When $x > 0$, the minimum cost function τ^* is equal to the minimum of the expected cost of the two choices, to relay and to not relay. For example, when $\pi^*(x, v, d) = 0$, $\tau^*(x, v, d)$

is equal to $E[\mathcal{I}] + E_{\mathcal{I},V}[\tau^*(x, V, (d - \mathcal{I})^+)]$, where $E[\mathcal{I}]$ is the mean waiting time before another vehicle arrives and, $E_{\mathcal{I},V}[\tau^*(x, V, (d - \mathcal{I})^+)]$ is the minimum expected delay with the random interarrival time \mathcal{I} and the random vehicle speed V (i.e., when the system state is $(x, V, (d - \mathcal{I})^+)$). ■

8.4.1 One Shot Problem ($z = 1$)

Consider the special case, $z = 1$. The Bellman equation (8.3) simplifies to

$$\tau^*(1, v) = \min \left\{ E[\mathcal{I}] + E[\tau^*(1, V)], \frac{s}{v} \right\} \quad (8.4)$$

and $\tau^*(0, v) = 0$. From Theorem 8.4.1 and (8.4), we see that it is optimal to transmit (i.e., $y = 1$) when, $\frac{s}{v}$, the cost of relaying using the current vehicle is less than the expected future cost, i.e., when

$$\frac{s}{v} \leq E[\mathcal{I}] + E[\tau^*(1, V)] \quad (8.5)$$

Observe that the expression on the right hand side does not depend on v . Hence, the optimal policy is a threshold policy. Suppose that the vehicle speeds are continuously distributed within a bounded set, $[v_{min}, v_{max}]$. Also, suppose that there exists a $v^* \in [v_{min}, v_{max}]$, which satisfies (8.5) with equality, i.e.,

$$\frac{s}{v^*} = E[\mathcal{I}] + \int_{u=v_{min}}^{v_{max}} \tau^*(1, u) dV(u) \quad (8.6)$$

For any $v \geq v^*$, $\frac{s}{v} \leq \frac{s}{v^*}$. Hence, we see that it is optimal to relay if a vehicle arrives with speed $v \geq v^*$ and the optimal cost-to-go value is $\frac{s}{v}$. Equivalently, if a vehicle arrives with speed $v \leq v^*$, then it is optimal not to relay using this vehicle. The optimal cost-to-go value is the same (independent of v) and is equal to $E[\mathcal{I}] + \int_{u=v_{min}}^{v_{max}} \tau^*(1, u) dV(u)$. From (8.6), we see that $\tau^*(1, v) = \tau^*(1, v^*) = \frac{s}{v^*}$ for all $v \leq v^*$. Substituting for τ^* (for the two

cases) in (8.6) and expanding the integral, we have,

$$\begin{aligned} \frac{s}{v^*} &= \mathbb{E}[\mathcal{I}] + \int_{u=v_{min}}^{v^*} \tau^*(1, u) dV(u) + \int_{v^*}^{v_{max}} \tau^*(1, u) dV(u) \\ &= \mathbb{E}[\mathcal{I}] + \left(\frac{s}{v^*}\right) \left(1 - \int_{v^*}^{v_{max}} dV(u)\right) + \int_{v^*}^{v_{max}} \left(\frac{s}{u}\right) dV(u) \end{aligned}$$

Cancelling $\frac{s}{v^*}$, we get,

$$\frac{s}{v^*} \int_{v^*}^{v_{max}} dV(u) = \mathbb{E}[\mathcal{I}] + \int_{v^*}^{v_{max}} \left(\frac{s}{u}\right) dV(u) \quad (8.7)$$

Rearranging (8.7), we have,

$$\frac{s}{v^*} = \frac{1}{\int_{v^*}^{v_{max}} dV(u)} \left(\mathbb{E}[\mathcal{I}] + \int_{v^*}^{v_{max}} \frac{s}{u} dV(u) \right)$$

Remarks 8.4.1 *The optimal threshold v^* has a simple interpretation. v^* is an optimal threshold speed when the time required for the vehicle to reach the destination ($\frac{s}{v^*}$) is equal to the expected time we need to wait for a vehicle which can be used as a relay (with speed greater than v^*) and the subsequent time to travel to the destination.*

8.4.2 Piecewise Transmission Problem ($z > 1$)

The source node has $z, z > 1$ packets to relay to the destination node. Due to randomness in the interarrival times and the vehicle speeds, the z packets may reach the destination out-of-order. The optimal policy and the minimum cost are, hence, a function of the remaining packets to transmit (x) and the residual delay (d). The following theorem characterizes the optimal cost-to-go function (τ^*) and the optimal policy (π^*) for the piecewise transmission problem.

Theorem 8.4.2 *Suppose $z > 1$. Then, from (8.3), we can conclude the following.*

1. $\tau^*(x, v, d)$ is a non-decreasing function of x and d , and a non-increasing function of v .

2. The optimal policy $\pi^*(x, v, d)$ is a threshold type policy with respect to v , with the threshold depending on x and d .

Proof: See Appendix 8.7.1. ■

Remark: $\tau^*(x, v, d)$ is the minimum expected delay for relaying x packets to the destination, with the initial vehicle speed v and residual delay d . Clearly, we would expect τ^* to be a non-decreasing function of x (the delay can only increase with the number of residual packets). Also, the transit delay associated with a vehicle ($\frac{x}{v}$) is a decreasing function of v . Hence, we would expect τ^* to be a non-increasing function of v . We showed that the optimal policy $\pi^*(x, v, d)$ is a threshold policy with respect to v . We conjecture that the threshold is a non-increasing function of x and of d . ■

Solving the DP (8.3), for the general case $z > 1$, is numerically cumbersome. Hence, we will study a sub-optimal scheduling policy, $\tilde{\pi}$, which is easy to characterize. Also, we will bound the difference in cost between the sub-optimal policy ($\tilde{\pi}$) and the optimal policy (π^*). For any scheduling policy π , define $\mathcal{T}_1^\pi, \mathcal{T}_2^\pi, \dots, \mathcal{T}_z^\pi$ as the random times (after $T_1 = 0$) at which packets $1, 2, \dots, z$ reach the destination node. Our original objective is to solve the following optimization problem.

$$\inf_{\pi} \left\{ \mathbb{E}_{\{\mathcal{I}, V\}} [\max(\mathcal{T}_1^\pi, \mathcal{T}_2^\pi, \dots, \mathcal{T}_z^\pi)] \right\} \quad (8.8)$$

Alternatively, we will solve (8.9) (below) and show that the optimal scheduling policy for (8.9) provides a good approximation to the optimal value of the original problem (8.8). Consider the following modified optimization problem,

$$\inf_{\pi} \left\{ \max \left(\mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_1^\pi], \mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_2^\pi], \dots, \mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_z^\pi] \right) \right\} \quad (8.9)$$

Let $\tilde{\tau}$ be the optimal value of the alternate objective function, i.e.,

$$\tilde{\tau} := \inf_{\pi} \left\{ \max \left(\mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_1^\pi], \mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_2^\pi], \dots, \mathbb{E}_{\{\mathcal{I}, V\}} [\mathcal{T}_z^\pi] \right) \right\}$$

Define $\mathcal{T}_{(w)i}^\pi$ as the random waiting time of the i^{th} packet, at the head-of-queue at the source node. Also, define $\mathcal{T}_{(t)i}^\pi$ as the random transit delay of the i^{th} packet in the relay node. Then, $\mathcal{T}_i^\pi = \sum_{k=1}^i \mathcal{T}_{(w)k}^\pi + \mathcal{T}_{(t)i}^\pi$. Also,

$$\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_i^\pi] = \sum_{k=1}^i \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(w)k}^\pi] + \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(t)i}^\pi]$$

Observe that \mathcal{T}_i^π depends on $\mathcal{T}_j^\pi, 1 \leq j \leq i-1$, only through $\mathcal{T}_{(w)j}^\pi, 1 \leq j \leq i-1$ and does not depend on the transit delay incurred by the previously relayed packets. For any set of values $\{\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(w)i}^\pi], 1 \leq i \leq z\}$, we minimize $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_i^\pi]$ by minimizing $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(t)i}^\pi]$ (the transit delay of the i^{th} packet) subject to $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(w)i}^\pi]$ (the expected waiting time of the i^{th} packet). This will yield a threshold policy in v for every $i, \{v_i, 1 \leq i \leq z\}$, i.e., there exists a threshold speed v_i such that any vehicle with speed $v > v_i$ is used as a relay for the i^{th} packet. Thus, every scheduling policy π can be identified with a collection of z thresholds for speeds (v_1, \dots, v_z) . The expected sojourn time of the i^{th} packet for a policy π can now be written as,

$$\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_i^\pi] = \sum_{j=1}^{i-1} \frac{\mathbf{E}[\mathcal{I}]}{\int_{v_j}^{v_{max}} dV(u)} + \int_{v_i}^{v_{max}} \frac{s}{u} dV(u) \quad (8.10)$$

It is now straightforward to solve the optimization problem (8.9) by minimizing the objective function over the vector of thresholds (v_1, v_2, \dots, v_z) . Let $\tilde{\pi} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_z)$ be the cost minimizing policy. The following theorem shows a monotonicity property of the optimal speed thresholds \tilde{v}_i .

Theorem 8.4.3 *Let $\tilde{\pi} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_z)$ be the cost minimizing policy for the optimization problem (8.9). Then, $\tilde{v}_1 \leq \tilde{v}_2 \leq \dots \leq \tilde{v}_z$.*

Proof: Suppose not, let $\tilde{v}_i > \tilde{v}_j$ where $i < j$. Now, consider a new scheduling policy $\hat{\pi}$ with threshold speeds $(\hat{v}_1, \hat{v}_2, \dots, \hat{v}_z)$ such that $\hat{v}_k = \tilde{v}_k$ when $k \neq i$ and $\hat{v}_i = \tilde{v}_j$. Clearly, $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_k^{\hat{\pi}}] = \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_k^{\tilde{\pi}}]$ for all $1 \leq k < i$. Since, $\hat{v}_i < \tilde{v}_i$, $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(w)i}^{\hat{\pi}}] \leq \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_{(w)i}^{\tilde{\pi}}]$. Hence, $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_k^{\hat{\pi}}] \leq \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_k^{\tilde{\pi}}]$ for all $k > i$. Also, $\mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_i^{\hat{\pi}}] \leq \mathbf{E}_{\{\mathcal{I},V\}}[\mathcal{T}_j^{\tilde{\pi}}]$, since $i < j$

and $\hat{v}_i = \hat{v}_j$ (from (8.10)). Hence,

$$\max\left(\mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_1^{\hat{\pi}}], \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_2^{\hat{\pi}}], \dots, \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\hat{\pi}}]\right) \leq \max\left(\mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_1^{\tilde{\pi}}], \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_2^{\tilde{\pi}}], \dots, \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\tilde{\pi}}]\right)$$

or, $\hat{\pi}$ is an optimal policy with $\hat{v}_1 \leq \hat{v}_2 \leq \dots \leq \hat{v}_z$. \blacksquare

The difference in the expected cost between the scheduling policy $\tilde{\pi}$ and the optimal policy π^* is bounded and is obtained below.

Theorem 8.4.4 $\tau^* \leq \mathbb{E}_{\{\mathcal{I},V\}}[\max(\mathcal{T}_1^{\tilde{\pi}}, \dots, \mathcal{T}_z^{\tilde{\pi}})] \leq \tau^* + \frac{s}{\tilde{v}_1}$, where $\tilde{\pi} := (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_z)$ is the optimal scheduling policy for (8.9).

Proof: $\tilde{\tau} = \max(\mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_1^{\tilde{\pi}}], \dots, \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\tilde{\pi}}]) \geq \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\tilde{\pi}}]$. Also, we have,

$$\begin{aligned} \tau^* &= \inf_{\pi} \left\{ \mathbb{E}_{\{\mathcal{I},V\}}[\max(\mathcal{T}_1^{\pi}, \mathcal{T}_2^{\pi}, \dots, \mathcal{T}_z^{\pi})] \right\} \\ &\geq \inf_{\pi} \left\{ \max\left(\mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_1^{\pi}], \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_2^{\pi}], \dots, \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\pi}]\right) \right\} \\ &= \tilde{\tau} \end{aligned} \tag{8.11}$$

Hence, $\mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\tilde{\pi}}] \leq \tilde{\tau} \leq \tau^*$. Consider the actual delay of the file transfer with the policy $\tilde{\pi}$, i.e., $\mathbb{E}_{\{\mathcal{I},V\}}[\max(\mathcal{T}_1^{\tilde{\pi}}, \dots, \mathcal{T}_z^{\tilde{\pi}})]$. We know that

$$\mathcal{T}_i^{\tilde{\pi}} \leq \mathcal{T}_z^{\tilde{\pi}} + \frac{s}{\tilde{v}_1}$$

holds true for all $1 \leq i \leq z$ and for every sample path of the interarrival times and the vehicle speeds. This is because, $\tilde{v}_1 \leq \tilde{v}_2 \leq \dots \leq \tilde{v}_z$, and $\frac{s}{\tilde{v}_1}$ is the maximum transit delay incurred by any packet. Hence,

$$\max(\mathcal{T}_1^{\tilde{\pi}}, \dots, \mathcal{T}_z^{\tilde{\pi}}) \leq \mathcal{T}_z^{\tilde{\pi}} + \frac{s}{\tilde{v}_1}$$

Taking expectation on both the sides, we get,

$$\mathbb{E}_{\{\mathcal{I},V\}}[\max(\mathcal{T}_1^{\tilde{\pi}}, \dots, \mathcal{T}_z^{\tilde{\pi}})] \leq \mathbb{E}_{\{\mathcal{I},V\}}[\mathcal{T}_z^{\tilde{\pi}}] + \frac{s}{\tilde{v}_1}$$

But we know that $E_{\{I,V\}}[\mathcal{T}_z^{\tilde{\pi}}] \leq \tau^*$. Hence, we have,

$$\tau^* \leq E_{\{I,V\}}[\max(\mathcal{T}_1^{\tilde{\pi}}, \dots, \mathcal{T}_z^{\tilde{\pi}})] \leq \tau^* + \frac{s}{\bar{v}_1}$$

which shows that the average delay of file transfer with the policy $\tilde{\pi}$ is bounded within $\frac{s}{\bar{v}_1}$ of the optimal value τ^* . ■

8.5 Infinite File Size

In this section, we study the case where the source node has infinitely many packets to communicate to the destination node. Packets arrive at the source node according to some point process independent of the vehicle speeds and the vehicle interarrival times. The packets are queued in an infinite buffer at the source node before they are relayed to the destination node using the vehicles. We are interested in the average packet delay in this scenario, at the queue in the source node and in transit in the relay vehicles. There is a natural tradeoff between the queueing delay and the transit delay of the packets. For example, by choosing every vehicle as a relay, we can minimize the queueing delay of the packets, while increasing the transit delay. Similarly, by relaying only to fast vehicles, we can decrease the transit delay of the packets while simultaneously increasing the queueing delay at the source buffer. We are interested in this tradeoff and study it in an asymptotic regime.

Each time a vehicle enters the communication range of the source node, the source node has to make a decision on using the current vehicle (say, with speed v) as a relay. Here again, we assume that only one packet is relayed using any vehicle. Starting with the first vehicle to arrive, the decision problem evolves over vehicle arrival instants $\{T_1 = 0, T_2, T_3, \dots\}$ with vehicle speeds $\{V_1, V_2, V_3, \dots\}$. Let X_k denote the number of packets in the source buffer at time instant T_k and let Y_k denote the number of packets relayed using the k^{th} vehicle. As before, $Y_k \in \{0, 1\}$. Of course, when $X_k = 0$, $Y_k = 0$ is the only

permissible action. The system state dynamics for the problem is now given by,

$$X_{k+1} = (X_k - Y_k)^+ + A_k$$

where A_k is the number of packets arriving at the source buffer during the time interval $\mathcal{I}_{k+1} = T_{k+1} - T_k$. We assume that $\{A_k, k \geq 1\}$ is an i.i.d. sequence with mean $\mathbf{E}[A]$. Also, A_k is assumed to be independent of X_k and Y_k , but may depend on the interarrival time \mathcal{I}_{k+1} . Define $\lambda := \frac{\mathbf{E}[A]}{\mathbf{E}[\mathcal{I}]}$, the rate of packets arriving at the source node per second. Since each vehicle carries at most one packet, we require that $\lambda < 1$.

Define $R_k(x_k, v_k, y_k)$, the single stage cost function associated with the system state (x_k, v_k) and action y_k as

$$R_k(x_k, v_k, y_k) = \frac{1}{\lambda} x_k + \frac{1}{\lambda \mathbf{E}[\mathcal{I}]} \frac{y_k}{v_k}$$

The cost expression comprises of two components: the first term $\frac{x_k}{\lambda}$ corresponds to the queue length (or the queueing delay) and the second term $\frac{1}{\lambda \mathbf{E}[\mathcal{I}]} \frac{y_k}{v_k}$ corresponds with the transit delay. Consider the following average cost expression with $R_k(x_k, v_k, y_k)$ being the single stage cost function,

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \mathbf{E} \left[\sum_{i=1}^k \left(\frac{X_k}{\lambda} + \frac{1}{\lambda \mathbf{E}[\mathcal{I}]} \frac{Y_k}{V_k} \right) \right] \quad (8.12)$$

Note that when the limits exists $\frac{1}{\lambda} \lim_{k \rightarrow \infty} \frac{1}{k} \mathbf{E} \left[\sum_{i=1}^k X_k \right]$ is the average queueing delay in the system, and

$$\lim_{k \rightarrow \infty} \frac{1}{k} \mathbf{E} \left[\sum_{i=1}^k \frac{1}{\lambda \mathbf{E}[\mathcal{I}]} \frac{Y_k}{V_k} \right] = \frac{1}{\lambda \mathbf{E}[\mathcal{I}]} \lim_{k \rightarrow \infty} \frac{1}{k} \mathbf{E} \left[\sum_{i=1}^k \frac{Y_k}{V_k} \right]$$

is the average transit delay of the packets (since $\lim_{k \rightarrow \infty} \frac{1}{k} \mathbf{E} \left[\sum_{i=1}^k Y_i \right] = \lambda \mathbf{E}[\mathcal{I}]$, the mean number of packet arrivals per vehicle arrival). The average delay minimization problem can now be studied by minimizing the cost function (8.12). As we are interested in studying the tradeoff achievable between the queueing delay and the transit delay, we will

introduce an additional parameter β , $\beta > 0$ (essentially a Lagrange multiplier), and study the following modified average cost problem.

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \mathbb{E} \left[\sum_{i=1}^k \left(\frac{X_k}{\lambda} + \beta \frac{1}{\lambda \mathbb{E}[\mathcal{T}]} \frac{Y_k}{V_k} \right) \right]$$

We have ignored the resequencing delay in the above formulation, as we expect the transit delay to be bounded in the optimal solution.

In [55], Berry and Gallager have studied the problem of optimal rate and power control in a fading wireless channel, which has a similar mathematical model. They consider a slotted wireless fading channel, where the data rate achievable over a slot is a function of the channel fade “gain” h during that slot and the power allocated P . Packets arrive into an infinite buffer independent of the queue length and the channel evolution process, and the objective was to study the tradeoff achievable between the queueing delay at the buffer and the average power required to support the arrival process. The single stage cost function in [55] was $\frac{X_i}{\lambda} + \beta P_i$, where X_i is the queue size at the i th slot, P_i is the power allocated during that slot, λ is the average arrival rate and β is a Lagrange multiplier. The optimization problem in [55] was to find a policy π that minimizes the following objective function.

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \mathbb{E} \left[\sum_{i=1}^k \left(\frac{X_i}{\lambda} + \beta P_i \right) \right]$$

Observe that, our problem formulation is very similar to the problem formulation in [55] and hence, the results from [55] can be directly extended to our framework by observing that the average transit delay in our framework is analogous with the average power in [55]. More formally, we will first define a concave throughput - transit delay function $C(d)$ equivalent to the the concave throughput - power function $C(P)$ (capacity of a power constrained fading channel) used in the Berry and Gallager model. The tradeoff then follows from the proof in [55].

Queueing delay is a function of throughput (the rate at which packets are served by the source node), and throughput itself is a function of the transit delay incurred by the packets. Accommodating a large transit delay leads to large throughput and

small queueing delay for the packets. The following definition provides the transit delay constrained throughput for an infinitely backlogged queue.

Definition 8.5.1 Let $C(d)$ be the maximum throughput sustainable for an infinitely backlogged queue, for a given average transit delay constraint d . $C(d)$ is defined as,

$$\begin{aligned}
 C(d) &:= \max_{\pi} \frac{1}{\mathbf{E}[I]} \int \pi(u) dV(u) & (8.13) \\
 \text{with } \pi &\text{ s.t.} & \frac{s}{\int \pi(u) dV(u)} \int \left(\frac{\pi(u)}{u} \right) dV(u) \leq d \\
 \text{and} & & \pi(u) \in [0, 1]
 \end{aligned}$$

where $\pi(u)$ is the fraction of vehicles with speed u which are used as a relay by the source node. ■

Remark: Trivially $C(d)$ is a non-decreasing function of d . Also, for any $d := \alpha d_1 + (1-\alpha)d_2$, where $d_1, d_2 \geq 0$ and $\alpha \in [0, 1]$, we see that $C(d) \geq \alpha C(d_1) + (1-\alpha)C(d_2)$. Hence, $C(d)$ is a concave function of d as well. The throughput maximizing policy π_d^* (of the optimization problem (8.13)) can be shown to be a threshold policy, i.e., there exists a v^* such that for all $v > v^*$, $\pi^*(v) = 1$. ■

Let λ be the average arrival rate of packets into the system. Define d_λ as the minimum average transit delay incurred to support the arrival rate λ , i.e., $C(d_\lambda) = \lambda$ (see Figure 8.2). Trivially, we know that for any scheduling policy with an average transit delay $d \leq d_\lambda$, the queueing delay in the system will be infinite. Consider the throughput maximizing schedule π_d^* , obtained from the optimization problem (8.13) for an average transit delay constraint d . For $d = d_\lambda + \Theta(\delta)$ (where $\delta > 0$), $C(d) = C(d_\lambda) + O(\delta) = \lambda + O(\delta)$. Suppose that we use the fixed schedule π_d^* to relay packets to vehicles, independent of the queue length at the source node buffer. Then, the average transit delay of the packets is $O(d)$. And, the average queue length (or delay) of the packets at the source buffer is $\Omega\left(\frac{1}{\delta}\right)$; This follows from the fact that an upper bound for the average queue length of a G/G/1 queue with arrival rate λ and service rate $\lambda + \Theta(\delta)$ is $\Theta\left(\frac{1}{\delta}\right)$. Now, instead of using a fixed schedule π_d^* , we will, as in [55], use two schedules $\pi_{d_\lambda-\delta}^*$ and $\pi_{d_\lambda+\delta}^*$ depending on the queue size being below and above a predetermined threshold (see Figure 8.4). This

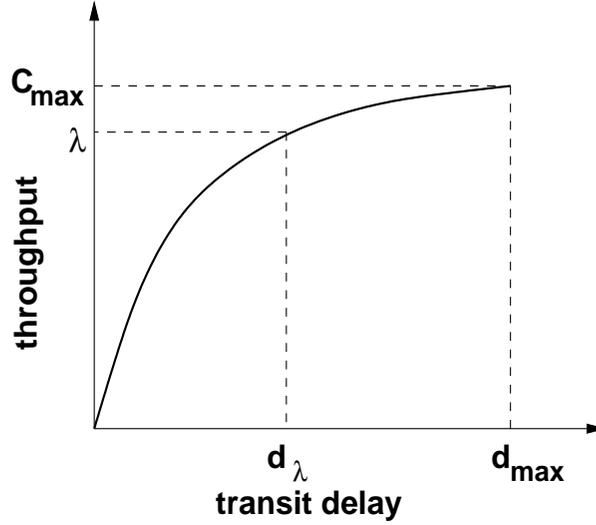


Figure 8.2: Throughput ($C(d)$) vs Transit delay (d) for the vehicular network scenario from (8.13). The maximum transit delay, d_{max} and the maximum throughput achievable, C_{max} are given by $d_{max} = s \int \frac{1}{u} dV(u)$ and $C_{max} = \frac{1}{\mathbb{E}[\mathcal{I}]}$. Given λ , the arrival rate of the packets into the system, define d_λ as the minimum transit delay incurred in supporting the arrival process, and $C(d_\lambda) := \lambda$.

threshold based policy was shown to reduce the average queueing delay to $\Omega\left(\frac{1}{\sqrt{\delta}}\right)$ for an average transit delay of $d_\lambda + O(\delta)$ (see [55]). The following theorem summarizes the above discussion, whose proof follows directly from the tradeoff studied in [55].

Theorem 8.5.1 *Let the packet arrival process A_k and the vehicular speeds V_k be i.i.d. and independent of each other. Define λ as the average arrival rate of packets into the source buffer, and let d_λ be the minimum average transit delay incurred to support the arrival rate λ , i.e., $C(d_\lambda) = \lambda$. Suppose that $C'(d_\lambda) > 0$ and let the packet arrival process A_k be a compact subset of \mathbb{R}^+ . Then, for an excess average transit delay of $O(\delta)$ (i.e., for an average transit delay of $d_\lambda + O(\delta)$), the average queueing delay in the system scales as $\Omega\left(\frac{1}{\sqrt{\delta}}\right)$. ■*

Remark: When $C'(d_\lambda) = 0$, we can achieve a better tradeoff for delay (see [45]). Also, we believe that the compactness requirement on A_k (the number of packet arrivals during a vehicle interarrival time) is not restrictive in practical scenarios.

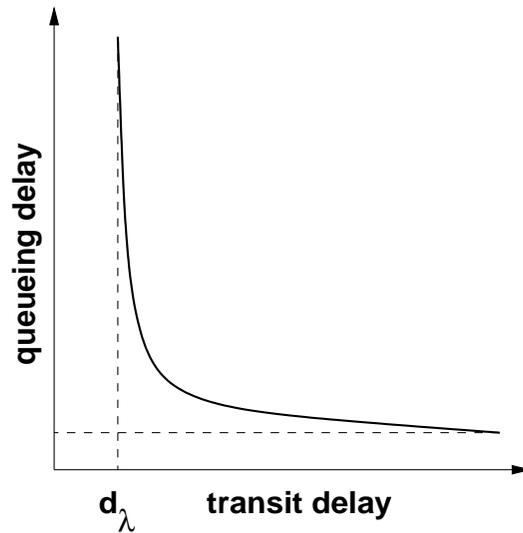


Figure 8.3: Figure plots a characteristic tradeoff achievable between the average queuing delay and the average transit delay when the arrival rate is λ . Observe that the queuing delay approaches infinity as the transit delay constraint approaches d_λ .

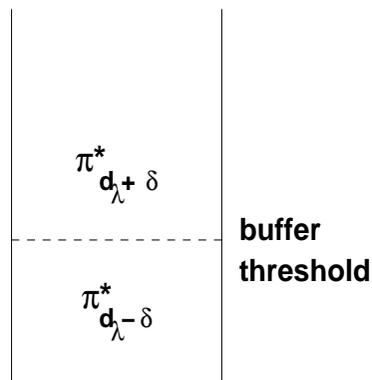


Figure 8.4: Figure shows the buffer threshold based scheduling policy proposed by Berry and Gallager. When the queue in the buffer is less than the threshold, a policy $\pi_{d_\lambda - \delta}^*$ is used and when the queue exceeds the threshold, policy $\pi_{d_\lambda + \delta}^*$ is used to schedule packets.

8.6 Summary

In this chapter, we have studied a scheduling problem in a wireless network, where vehicles are used as relays. A stationary source node has a file to communicate to a stationary destination node, and passing by vehicles are used as relays to transfer the file to the destination. All packet communication involves only two hops, and we are interested in minimizing the average queueing delay and the average transit delay of the packets in the network. We studied both the finite file size case and the infinite file size case. In the finite file size case, we obtained the expected total delay minimizing schedule using a Markov decision process framework. We also obtained a simple sub-optimal scheduling policy whose average delay is within a known bound from the optimal value (obtained from the MDP formulation). In the infinite file size case, we studied the asymptotically optimal tradeoff achievable between the queueing delay and the transit delay of the packets. By defining the maximum throughput sustainable for a given transit delay constraint, we showed that the average queueing delay of the system scales as $\Omega\left(\frac{1}{\sqrt{\delta}}\right)$ for an excess average transit delay of $O(\delta)$.

8.7 Appendix

8.7.1 Proof of Theorems and Lemmas

Proof of Theorem 8.4.2.1

Proof: Let π be any stationary Markov policy. For the single stage cost function $R_k(x_k, v_k, d_k, y_k)$ given by (8.1), the random total cost function with the scheduling policy π is given by,

$$\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k))$$

And, the expected total cost (or the average delay in delivery) for a file with z packets, initial vehicle speed v and residual delay d is given by,

$$\mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k)) \middle| X_1 = z, V_1 = v, D_1 = d \right]$$

Consider two file sizes z and $z + 1$. We aim to prove that $\tau^*(z, v, d) \leq \tau^*(z + 1, v, d)$. For π , a stationary Markov policy, define π_{-1} , another stationary Markov policy, as

$$\pi_{-1}(x, v, d) = \pi(x + 1, v, d)$$

and $\pi_{-1}(0, v, d) = 0$. Now, it is straightforward to see that

$$\begin{aligned} & \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k)) \middle| X_1 = z + 1, V_1 = v, D_1 = d \right] \\ & \geq \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi_{-1}(X_k, V_k, D_k)) \middle| X_1 = z, V_1 = v, D_1 = d \right] \end{aligned}$$

for every sample path of the interarrival times and the vehicle speeds. In other words, for every policy π , there exists a scheduling policy π_{-1} such that the random delay in delivering a file with $z + 1$ packets (and with policy π) is greater than or equal to the random delay in delivering a file with z packets (and with a policy π_{-1}). Taking expectation over the sample paths, we have,

$$\begin{aligned} & \mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k)) \middle| X_1 = z + 1, V_1 = v, D_1 = d \right] \\ & \geq \mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi_{-1}(X_k, V_k, D_k)) \middle| X_1 = z, V_1 = v, D_1 = d \right] \end{aligned}$$

Now, taking infimum over the set of stationary Markov policies π , we have,

$$\begin{aligned} \tau^*(z + 1, v, d) &= \inf_{\pi} \mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k)) \middle| X_1 = z + 1, V_1 = v, D_1 = d \right] \\ &\geq \inf_{\pi_{-1}} \mathbb{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi_{-1}(X_k, V_k, D_k)) \middle| X_1 = z, V_1 = v, D_1 = d \right] \end{aligned}$$

$$\begin{aligned}
 &\geq \inf_{\pi} \mathbf{E}_{\{\mathcal{I}, V\}} \left[\sum_{k=1}^{\infty} R_k(X_k, V_k, D_k, \pi(X_k, V_k, D_k)) \middle| X_1 = z, V_1 = v, D_1 = d \right] \\
 &= \tau^*(z, v, d)
 \end{aligned}$$

which is our desired result. The proof for the ordering of τ^* with respect to v and d follows similar arguments and is omitted here. ■

Proof of Theorem 8.4.2.2

Proof: We know that $\tau^*(x, v, d)$ satisfies the following DP given in (8.3),

$$\tau^*(x, v, d) := \begin{cases} 0 & x = 0 \\ \min \left\{ \mathbf{E}[\mathcal{I}] + \mathbf{E}_{\mathcal{I}, V}[\tau^*(x, V, (d - \mathcal{I})^+)], \max\left(\frac{s}{v}, d\right) \right\} & x = 1 \\ \min_{y \in \{0, 1\}} \left\{ \mathbf{E}[\mathcal{I}] + \mathbf{E}_{\mathcal{I}, V}[\tau^*(x - y, V, (\max(d, I_{\{y > 0\}} \frac{s}{v}) - \mathcal{I})^+)] \right\} & x > 1 \end{cases}$$

Define $\pi^*(x, v, d)$ as the stationary policy that chooses the minimizer of the right hand expression in the above DP.

Consider the special case $x = 1$. The optimal policy chooses the minimum of

$$\left\{ \mathbf{E}[\mathcal{I}] + \mathbf{E}[\tau^*(1, V)], \frac{s}{v} \right\}$$

Note that the first term is independent of v . And $\frac{s}{v}$ is a decreasing function of v . Hence, we see that the optimal policy chooses to relay whenever

$$\mathbf{E}[\mathcal{I}] + \mathbf{E}[\tau^*(1, V)] \geq \frac{s}{v}$$

or, the optimal policy is a threshold policy with v .

Now, consider the case $x > 1$. The optimal policy chooses the minimum of the following two terms,

$$\left\{ \mathbf{E}[\mathcal{I}] + \mathbf{E}_{\mathcal{I}, V}[\tau^*(x, V, (d - \mathcal{I})^+)], \mathbf{E}[\mathcal{I}] + \mathbf{E}_{\mathcal{I}, V} \left[\tau^* \left(x - 1, V, \left(\max \left(d, \frac{s}{v} \right) - \mathcal{I} \right)^+ \right) \right] \right\}$$

As before, the first term is independent of v . And, the second term is a non-increasing function of v (since $\tau^*(x, v, d)$ is a non-increasing function with d). Hence, we see that the optimal policy is a threshold policy for $x > 1$ as well. ■

Chapter 9

Conclusion

In this thesis, we have studied a number of topics in modeling, analysis and optimization of wireless networks. We have studied performance analysis as well as resource optimization problems in a variety of wireless network scenarios, including WiFi networks, ad hoc (and sensor) networks and vehicular networks.

In the first part of the thesis, we focussed on the performance analysis of IEEE 802.11(e) wireless local area networks. We studied the distributed coordination function (DCF) and the enhanced distributed channel access (EDCA) MAC of the IEEE 802.11(e) standard. We were interested in the saturation throughput performance of a single cell IEEE 802.11(e) WLAN. We modeled both the *pure collision* channel as well as frame capture at the receiver. Our analysis led us to a set of fixed point equations whose solution would characterize the operating point.

In Chapter 2 (and in Chapter 4), we showed that the fixed point equations can have multiple solutions, and in such cases, the system exhibits multistability and short-term unfairness of throughput. Also, the fixed point analysis failed to characterize the average system behaviour when the system has multiple solutions. Multistability was attributed to the backoff parameter values (node response formula) and the capture probabilities in the channel. We then obtained sufficient conditions (in terms of the backoff parameters of the nodes and the number of nodes) under which the fixed point equations have a unique solution (Theorem 2.5.2 in Chapter 2 and Theorem 4.3.1 in Chapter 4). We also showed

that the default parameters of the IEEE 802.11(e) DCF (EDCA) standard satisfied these sufficient conditions.

Then, using the fixed point analysis, in Chapter 3, we studied the throughput differentiation provided by the different backoff parameters, including AIFS and multiple queues per node. We observed that using initial backoff window, in general, a fixed throughput ratio can be achieved. On the other hand, when using p and *AIFS*, the service is significantly biased towards the high priority class, with the differentiation increasing in favour of the high priority class as the load in the system increases. We also observed that the effect of collision priority (due to virtual collision), where there are multiple access categories per node, decreases when the number of nodes increases.

The fixed point approach is simply a heuristic that is found to work well in some cases. Our work suggests where it might not work and where it might work. In a recent work [11], the authors have proved that for random backoff algorithms, when the number of sources grow large, the system is indeed decoupled, providing a theoretical justification of decoupling arguments used in the analysis.

The fixed point framework developed by us for a single cell WLAN with saturated queues has been extended to the case of unsaturated queues and multi-cell scenarios as well. In [61], the throughput performance of an IEEE 802.11e infrastructure WLAN carrying packet telephone calls, streaming video sessions and TCP file downloads has been studied by extending the saturation throughput analysis discussed in Chapters 2 and 3. In [39], the authors study the performance of IEEE 802.11 multi-cell networks comprising interfering co-channel cells using the fixed point approach.

In the second part of the thesis, we studied resource allocation and optimization problems for a variety of wireless network scenarios. In Chapter 5, for a dense wireless network deployed over a small area, and with a network average power constraint, we showed that single cell operation is throughput efficient in the asymptotic regime in which the network average power is made large. We showed that, with a realistic path loss model and a physical interference model (SINR based), the maximum aggregate bit rate among

arbitrary transmit-receive pairs scales only as $\Theta(\log(\bar{P}))$, where \bar{P} is the network average power. Spatial reuse is ineffective and direct transmission between source-destination pairs is the throughput optimal strategy which achieves the $\Theta(\log(\bar{P}))$ scaling. Also, for moderate \bar{P} , when spatial reuse may be efficient, we showed in section 5.4 that spatial reuse is restricted by the network dimensions, which affects the gains achievable using cooperative communication techniques. Our result suggests single cell operation for all practical network scenarios, when the ad hoc networks span a small area of few 1000 sq mts. Single cell is easy to operate and when large network power is available, single cell operation is optimal as well. We did not model a maximum node power constraint in our work. We expect the scaling results to fare poorer with such constraints.

In Chapter 6, for a dense ad hoc wireless network operating as a single cell and with a finite network power constraint, we studied the optimal hop length (routing strategy) and power control (for a fading channel) that maximizes the network aggregate throughput. For a fixed transmission time strategy, we studied the throughput maximizing schedule under homogeneous traffic and MAC assumptions. We showed that there corresponds an intrinsic aggregate packet carrying capacity at which the network operates at the optimal operating point, independent of the average power constraint. We also obtained the scaling law relating the optimal hop distance to the power constraint, and relating the optimal transport capacity to the power constraint (see Theorem 6.4.2). In Theorem 6.4.4 we also provide a characterisation of the optimal hop distance for cases in which the fading density satisfies a certain monotonicity condition. In [62], the authors study the problem of developing a distributed algorithm for nodes to adapt themselves towards the optimal operating point. They first propose a distance discretization technique in which the hop distance on the critical geometric graph is used as a distance measure. Using the distance approximation, they then develop a distributed algorithm aimed to maximize the transport capacity of the network in the sense of our framework presented in Chapter 6. Interesting extensions to the throughput maximization problem include modeling non-homogeneous traffic loads and channel conditions and studying non-saturated queue scenarios.

It is now well understood that in a multihop network, performance can be enhanced by network coding, instead of just forwarding packets. For a two link slotted wireless network employing a network coding strategy and with fading channels, in Chapter 7, we studied the optimal power control and optimal exploitation of network coding opportunities that minimizes the average power required to support a given arrival rate. We also obtained the optimal power-delay tradeoff for the network (with the network coding strategy) and showed that the minimum average queueing delay scales as $\Omega\left(\frac{1}{v}\right)$ for an excess average power of $O(v)$. In our work, we have focussed on a simple two link slotted wireless network. Extensions to tandem networks is straight forward and the power-delay tradeoff results still hold. Future work includes studying the capacity region of arbitrary wireless networks with a network coding strategy and characterizing the network coding delay in such scenarios.

Finally, in Chapter 8, we studied a delay minimization problem in a vehicular relay network scenario. A stationary source node has a file to communicate to a stationary destination node, and passing by vehicles are used as relays to transfer the file to the destination. We considered both the finite file size case and the infinite file size case. In the finite file size case, we obtained the expected total delay minimizing schedule using a Markov decision process framework. We also obtained a simple sub-optimal scheduling policy whose average delay is within a known bound from the optimal value (obtained from the MDP formulation). In the infinite file size case, we studied the asymptotically optimal tradeoff achievable between the queueing delay and the transit delay of the packets. By defining the maximum throughput sustainable for a given transit delay constraint, we showed that the average queueing delay of the system scales as $\Omega\left(\frac{1}{\sqrt{\delta}}\right)$ for an excess average transit delay of $\Theta(\delta)$. In this work, we consider a single source-destination pair with a single route between them. In future extensions, we can study multiple source-destination pairs in an arbitrary network, allowing for different routing options to the relaying nodes. Another interesting problem would be to permit the relay vehicles to carry different amounts of data between the source-destination pair.

Bibliography

- [1] IEEE 802.11 WG, Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, LAN MAN Standards Committee of the IEEE Computer Society, 1999.
- [2] IEEE 802.11 WG, Information technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, 2005.
- [3] Network Coding Bibliography,
<http://www.ifp.uiuc.edu/~koetter/NWC/Bibliography.html>
- [4] Abbas El Gamal and James Mammen, Optimal Hopping in Ad hoc Wireless Networks, Proc of IEEE Infocom, April, 2006.
- [5] Andrea J Goldsmith and Pravin P Varaiya, Capacity of Fading Channels with Channel Side Information, IEEE Transactions on Information Theory, November, 1997.
- [6] Anurag Kumar, Eitan Altman, Daniele Miorandi and Munish Goyal, New Insights from a Fixed Point Analysis of Single Cell IEEE 802.11 WLANs, IEEE/ACM Transactions on Networking, June, 2007.

- [7] Arvind Krishna and Richard O LaMaire, A Comparison of Radio Capture Models and their Effect on Wireless LAN protocols, Third Annual International Conference on Universal Personal Communications, 1994.
- [8] Ashish Sangwan, Venkatesh Ramaiyan and Rajeev Shorey, Reliable Multihop Broadcast Protocols for Inter-Vehicular Communication in a Fading Channel, 2nd International Conference on Communication Systems Software and Middleware, COM-SWARE, January, 2007.
- [9] Ayfer Ozgur, Olivier Leveque and David Tse, How does the Information Capacity of Ad Hoc Networks Scale?, Proceedings of the Forty-fourth Annual Allerton Conference on Communication, Control and Computing, September, 2006.
- [10] Bo Li and Roberto Battiti, Performance Analysis of An Enhanced IEEE 802.11 Distributed Coordination Function Supporting Service Differentiation, International Workshop on Quality of Future Internet Services, QoFIST, October, 2003.
- [11] Charles Bordenave, David McDonald and Alexandre Proutiere, Random Multi-access Algorithms: A Mean Field analysis, Allerton conference on Communication, Control and Computing, 2005.
- [12] Chiew T Lau and Cyril Leung, Capture Models for Mobile Packet Radio Networks, IEEE Transactions on Communications, May, 1992.
- [13] Christopher Ware, Joe Chicharo and Tadeusz Wysocki, Modelling Capture Behaviour in IEEE 802.11 Radio Modems, IEEE International Conference on Telecommunications, ITC, 2001.
- [14] Chunyu Hu, Hwangnam Kimz and Jennifer C Hou, An Analysis of the Binary Exponential Backoff Algorithm in Distributed MAC Protocols, Technical Report No. UIUCDCS-R-2005-2599, July, 2005.

-
- [15] David J Goodman, Joan Borras, Narayanan B Mandayam and Roy D Yates, INFO-STATIONS: A New System Model for Data and Messaging Services, IEEE Vehicular Technology Conference, May, 1997.
- [16] Esa Hyytia and Jorma Virtamo, On Load Balancing in a Dense Wireless Multihop Network, Proceedings of the 2nd Conference on Next Generation Internet Design and Engineering, 2006.
- [17] F Daneshgaran, Massimiliano Laddomada, F Mesiti and M Mondin, Unsaturated Throughput Analysis of IEEE 802.11 in Presence of Non Ideal Transmission Channel and Capture Effects, IEEE Transactions on Wireless Communications, April, 2008.
- [18] Giuseppe Bianchi, Performance Analysis of the IEEE 802.11 Distributed Coordination Function, IEEE Journal on Selected Areas in Communications, March, 2000.
- [19] Giuseppe Bianchi, Ilenia Tinnirello and Luca Scalia, Understanding 802.11e Contention-Based Prioritization Mechanisms and Their Coexistence with Legacy 802.11 Stations, IEEE Network, July/August, 2005.
- [20] Giuseppe Bianchi and Ilenia Tinnirello, Remarks on IEEE 802.11 DCF Performance Analysis, IEEE Communications Letters, August, 2005.
- [21] Hesham El Gamal, On the Scaling Laws of Dense Wireless Sensor Networks: The Data Gathering channel, IEEE Transactions on Information Theory, March, 2005.
- [22] Hoon Chang, Vishal Misra and Dan Rubenstein, A General Model and Analysis of Physical Layer Capture in 802.11 Networks, Proc of IEEE Infocom, April, 2006.
- [23] Hua Zhu and Imrich Chlamtac, An Analytical Model for IEEE 802.11e EDCF Differential Services, The 12th International Conference on Computer Communications and Networks, ICCCN, 2003.
- [24] Ilenia Tinnirello and Giuseppe Bianchi, On the Accuracy of some common Modeling Assumptions for EDCA analysis, CITSA, July, 2005.

-
- [25] Jae Hyun Kim and Jong Kyu Lee, Capture Effects of Wireless CSMA/CA protocols in Rayleigh and Shadow Fading channels, *IEEE Transactions on Vehicular Technology*, July, 1999.
- [26] Jang-Won Lee, Ao Tang, Jianwei Huang, Mung Chiang, and A Robert Calderbank, Reverse Engineering MAC: A Non-Cooperative Game Model, *IEEE Journal on Selected Areas in Communications*, August, 2007.
- [27] Jean-Pierre Aubin, *Applied Functional Analysis*, John Wiley & Sons, 2000.
- [28] Jeffrey W Robinson and Tejinder S Randhawa, Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function, *IEEE Journal on Selected Areas in Communications*, June, 2004.
- [29] Jens C Arnbak and Wim Van Blitterswijk, Capacity of Slotted Aloha in Rayleigh Fading Channels, *IEEE Journal on Selected Areas in Communications*, February, 1987.
- [30] Jianhua He, Lin Zheng, Zhongkai Yang and Chun Tung Chou, Performance Analysis and Service Differentiation in IEEE 802.11 WLAN, 28th Annual IEEE International Conference on Local Computer Networks, LCN, 2003.
- [31] Jing Zhao and Guohong Cao, VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks, *IEEE Transactions on Vehicular Technology*, Volume 57, Issue 3, 2008.
- [32] John Burgess, Brian Gallagher, David Jensen and Brian Neil Levine, MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks, *Proceedings of IEEE Infocom*, April, 2006.
- [33] Julie Y H Zhao and Okechukwu C Ugweje, Analysis of Capture Probability Performance techniques for Wireless LAN, *IEEE Vehicular Technology Conference, VTC*, 2002.

- [34] Jun Zhao, Zihua Guo, Qian Zhang and Wenwu Zhu, Performance study of MAC for Service Differentiation in IEEE 802.11, Globecom, 2002.
- [35] Kaichi Fujimura and Takaaki Hasegawa, A Study of Bandwidth in Road to Vehicle Communications using Contention type Medium Access Control, IEEE Intelligent Transportation Systems, October, 2003.
- [36] Kevin Fall, A Delay Tolerant Networking Architecture for Challenged Internets, Proceedings of Sigcomm, August, 2003.
- [37] Leandros Tassiulas and Anthony Ephremides, Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks, IEEE Transactions on Automatic Control, December, 1992.
- [38] M Garetto and C F Chiasserini, Performance Analysis of the 802.11 Distributed Coordination Function under Sporadic Traffic, Volume 3462/2005, NETWORKING, 2005.
- [39] Manoj Kumar Panda, Anurag Kumar and S H Srinivasan, Saturation Throughput Analysis of a System of Interfering IEEE 802.11 WLANs, IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM, June, 2005.
- [40] Mart L Molle, A New Binary Logarithmic Arbitration Method for Ethernet, 1994.
- [41] Matthias Grossglauser and David N C Tse, Mobility Increases the Capacity of Ad hoc Wireless Networks, IEEE/ACM Transactions on Networking, August, 2002.
- [42] Michael J Neely, Eytan Modiano and Charles E Rohrs, Dynamic Power Allocation and Routing for Time-Varying Wireless Networks, IEEE Journal on Selected Areas in Communications, January, 2005.
- [43] Michael J Neely and Eytan Modiano, Capacity and Delay Tradeoffs for Ad hoc Mobile Networks, IEEE Transactions on Information Theory, June, 2005.

-
- [44] Michael J Neely, Energy Optimal Control for Time Varying Wireless Networks, IEEE Transactions on Information Theory, July, 2006.
- [45] Michael J Neely, Optimal Energy and Delay Tradeoffs for Multi-User Wireless Downlinks, IEEE Transactions on Information Theory, September, 2007.
- [46] Michele Zorzi and Ramesh R Rao, Capture and Retransmission Control in Mobile Radio, IEEE Journal on Selected Areas in Communications, Volume 12, October, 1994.
- [47] Munish Goyal, Anurag Kumar and Vinod Sharma, Power Constrained and Delay Optimal Policies for Scheduling Transmissions over a Fading Channel, Proc of IEEE Infocom, 2003.
- [48] Nandini Vasudevan, Anurag Kumar and Venkatesh Ramaiyan, Throughput Measurements for Saturated UDP Transfers over IEEE 802.11 WLANs, Technical Report, 2004.
- [49] Olivier Dousse and Patrick Thiran, Connectivity vs Capacity in Dense Ad hoc Networks, Proc IEEE Infocom, March, 2004.
- [50] Piyush Gupta and P R Kumar, The Capacity of Wireless Networks, IEEE Transactions on Information Theory, March, 2000.
- [51] Prasanna Chaporkar and Alexandre Proutiere, Adaptive Network Coding and Scheduling for Maximizing Throughput in Wireless Networks, Mobicom, 2007.
- [52] Qing Xu, Tony Mak, Jeff Ko and Raja Sengupta, Vehicle-to-vehicle Safety Messaging in DSRC, Proceedings of the 1st ACM International Workshop on Vehicular Ad hoc Networks, 2004.
- [53] Rajendra K Jain, Dan-Ming W Chiu and William R Hawe, A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems, DEC Research Report TR-301, September, 1984.

-
- [54] Ralph E Strauch, Negative Dynamic Programming, *The Annals of Mathematical Statistics*, Vol. 37, No. 4, August, 1966.
- [55] Randall A Berry and Robert G Gallager, Communication over Fading Channels with Delay Constraints, *IEEE Transactions on Information Theory*, May, 2002.
- [56] Rima Khalaf, Izhak Rubin and Julian Hsu, Throughput and Delay Analysis of Multihop IEEE 802.11 Networks with Capture, *IEEE International Conference on Communications, ICC*, June, 2007.
- [57] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li and Raymond W Yeung, Network information flow, *IEEE Transactions on Information Theory*, July, 2000.
- [58] Sachin Katti, Hariharan Rahul, Wenjun Hu, Dina Katabi, Muriel Mdard and Jon Crowcroft, XORs in the Air: Practical Wireless Network Coding, *IEEE/ACM Transactions on Networking*, June, 2008.
- [59] Saud A Al-Semari and Guizani M, Channel Throughput of Slotted Aloha in Nakagami Fading Environment, *IEEE International Conference on Communications, ICC*, June, 1997.
- [60] Shuchin Aeron and Venkatesh Saligrama, Wireless Ad hoc Networks: Strategies and Scaling laws for the Fixed SNR Regime, *IEEE Transactions on Information Theory*, June, 2007.
- [61] Sri Harsha, Anurag Kumar and Vinod Sharma, An Analytical Model for Performance Evaluation of Multimedia Applications over EDCA in an IEEE 802.11e WLAN, *Wireless Networks*, October, 2008.
- [62] Srivathsa Acharya, Distributed Self-Organisation in Dense Wireless Ad Hoc Sensor Networks, M.E. Thesis, 2007, Department of ECE, Indian Institute of Science, Bangalore - 12.

- [63] Stefan Mangold, Sunghyun Choi, Peter May, Ole Klein, Guido Hiertz and Lothar Stibor, IEEE 802.11e Wireless LAN for Quality of Service, Proc of European Wireless, February, 2002.
- [64] Theodore S Rappaport, Wireless Communications: Principles & Practice, Prentice Hall, 2002.
- [65] Thomas M Cover and Joy A Thomas, Elements of Information Theory, 2nd Edition, Wiley Series in Telecommunications and Signal Processing.
- [66] Venkatesh Ramaiyan, Anurag Kumar and Nandini Vasudevan, Fixed Point Analysis of the Saturation Throughput of IEEE 802.11 WLANs with Capture, Proceedings National Conference on Communications, January, 2005.
- [67] Venkatesh Ramaiyan, Anurag Kumar and Eitan Altman, Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability, IEEE Transactions on Networking, October, 2008.
- [68] Venkatesh Ramaiyan, Anurag Kumar and Eitan Altman, Jointly Optimal Power Control and Routing for a single Cell, Dense, Ad hoc Wireless Network, Tech Rep. at, <http://ece.iisc.ernet.in/~anurag/papers/anurag/ramaiyan-et-al08singlecell-multihop.pdf.gz>, 2008.
- [69] Wang Li-Chun, Huang Shi-Yen and Chen Anderson, On the Throughput Performance of CSMA-based Wireless Local Area Network with Directional Antennas and Capture Effect : A Cross-layer Analytical Approach, IEEE Wireless Communications and Networking Conference, 2004.
- [70] Wenrui Zhao, Mostafa Ammar and Ellen Zegura, A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad hoc Networks, Proceedings of the 5th ACM International Symposium on Mobile Ad hoc Networking and Computing, 2004.

- [71] Xiaohu Ge, Dongyan and Yaoting Zhu, Throughput Model of IEEE 802.11 Networks with Capture Effect, International Conference on Wireless Communications, Networking and Mobile Computing, September, 2006.
- [72] Xiaolong Li and Qing-An Zeng, Performance Analysis of the IEEE 802.11 MAC protocols over a WLAN with Capture Effect, Proceedings International Conference on Mobile Computing and Ubiquitous Networking, 2005.
- [73] Xue Yang, Jie Liu, Feng Zhao and Nitin H Vaidya, A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning, First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004.
- [74] Yalin Evren Sagduyu and Anthony Ephremides, Network Coding in Wireless Queuing Networks: Tandem Network Case, Proceedings of IEEE International Symposium on Information Theory, ISIT, July, 2006.
- [75] Yang Xiao, Enhanced DCF of IEEE 802.11e to Support QoS, IEEE Wireless Communications and Networking, WCNC, March, 2003.
- [76] Yang Xiao, Backoff based Priority Schemes for IEEE 802.11, Proc. of IEEE International Conference on Communications, ICC, May, 2003.
- [77] Yang Xiao, An Analysis for Differentiated Services in IEEE 802.11 and IEEE 802.11e Wireless LANs, Proceedings of the 24th International Conference on Distributed Computing Systems, ICDCS, 2004.
- [78] Yunli Chen, Qing-An Zeng and Dharma P. Agrawal, Performance Analysis of IEEE 802.11e Enhanced Distributed Coordination Function, The 11th IEEE International Conference on Networks (ICON'03), Pages: 573-578, 2003.
- [79] Yu-Liang Kuo, Chi-Hung Lu, Eric Hsiao-Kuang Wu, Gen-Huey Chen and Yi-Hsien Tseng, Performance Analysis of the Enhanced Distributed Coordination Function in the IEEE 802.11e, IEEE 58th Vehicular Technology Conference, VAC 2003-Fall, Vol. 5, 2003.

-
- [80] Zhen-ning Kong, Danny H K Tsang, Brahim Bensaou and Deyun Gao, Performance Analysis of IEEE 802.11e Contention-Based Channel Access, *IEEE Journal on Selected Areas in Communications*, December, 2004.
- [81] Zoran Hadzi-Velkov and Boris Spasenovski, Capture Effect in IEEE 802.11 Basic Service Area under influence of Rayleigh Fading and Near/Far effect, *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, September, 2002.
- [82] Zoran Hadzi-Velkov and Boris Spasenovski, An Analysis of CSMA/CA Protocol with Capture in Wireless LANs, *Proceedings Wireless Communications and Networks*, March, 2003.