RANDOM ACCESS OVER MULTI-PACKET RECEPTION CHANNELS: A MEAN-FIELD APPROACH

Seyed Shwan Ashrafi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Sumit Roy, Chair

Chen Feng

Archis Ghate

Sreeram Kannan

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

RANDOM ACCESS OVER MULTI-PACKET RECEPTION CHANNELS:
A MEAN-FIELD APPROACH

Seyed Shwan Ashrafi

Chair of the Supervisory Committee:
Sumit Roy
Electrical Engineering

This thesis addresses the problem of Multi-Packet Reception (MPR) for random access scenarios in Wireless LAN type infrastructure network. Our results builds on a recently proposed physical-layer network coding mechanism – Compute-and-Forward (C&F) [1]; we embed this within slotted ALOHA and Carrier Sense Multiple Access (CSMA) protocols for operational scenarios where multiple[1] simultaneous transmissions are likely to occur.

A mean-field approach is used to understand the impact of multi-packet reception in random access networks. By focusing on a special family of MPR channels – the all-or-nothing model, stability conditions are derived for slotted ALOHA and CSMA systems under the above assumption. Interestingly, a number of physical-layer network coding schemes such as compute-and-forward, successive compute-and-forward (SCF) and successive interference cancellation can be viewed as special cases of all-or-nothing symmetric MPR model. In addition, the problem is analyzed under the general symmetric MPR channels which turns out not to be fundamentally different from the all-or-nothing MPR.

The primary outcome is a deeper understanding of the behavior of random access schemes over general symmetric MPR channels. Due to the interaction among users and the challenges emanating from such interactions, random access can be more readily analyzed if the

[1]This is typically an uplink scenario, i.e., from client to Access Point (AP).

state of users are assumed decoupled and an approximate stability region obtained for the system. Nonetheless, the approximate stability region computation based on the decoupling assumption is still complicated for even small (few clients) networks. By a mean-field approach, the random access problem is studied in the large number of clients regime. As a result, the evolution of the system state in the limit can be well approximated by a deterministic dynamical system. The stability conditions for the limiting system then paves the way for a better understanding of a known/observed phenomenon called meta-stability whereby (finite) random-access networks undergo a phase transition: from stability to meta-stability.

In the thesis, we first present results for slotted ALOHA with all-or-nothing MPR model. An approximate stability region is characterized which is then used for system performance evaluation in terms of throughput and delay. These results are then extended to the case of general symmetric MPR channels. Next, the stability of persistent CSMA systems is analyzed similar to slotted ALOHA analysis and throughput and delay results are obtained. Meta-stability is discussed for CSMA systems and system design guidelines are outlined to guarantee certain quality-of-service requirements, i.e., avoidance of meta-stable behavior. By adding a back-off mechanism, the stationary behavior of a saturated system is studied, again through mean-field analysis. Finally, we propose a novel extension to the cooperating multiple AP scenario for dense networks, where we present new algorithms for MPR-capable systems and present some initial analysis.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

The explosive growth in consumer mobile devices continues to fuel concomitant demand for wireless services. The surge of social network phenomenon, video/voice sharing and streaming, and other wireless services available on smart-phones and tablets are typical exemplars of this dramatic growth. New technology trends like the Internet of Things (IoT) — networking of an extra-ordinarily large number of common devices that are hitherto unconnected — will contribute to further overloading of already congested networks.

Since much of such new traffic will flow through access networks such as WiFi, *densification* of wireless local area networks (WLANs) is a common response to this exponential increase in data traffic density. However (as is well known), the interference resulting from simultaneous data transmissions [2] constitutes the fundamental limit to performance in such networks. Current 802.11 WLAN networks treat such simultaneous transmission as *undesired*, yet there is a significant history of ideas and methods that promote multi-packet reception (MPR) as a potential solution to such interference-limited channels. This thesis contributes to the theory and practice of MPR within a random access scenario (broadly conformal to Distributed Coordination Function (DCF) MAC protocol for 802.11 networks). We begin by identifying a suitable MPR model adopted by communication theorists as an abstraction of 802.11 physical layer.

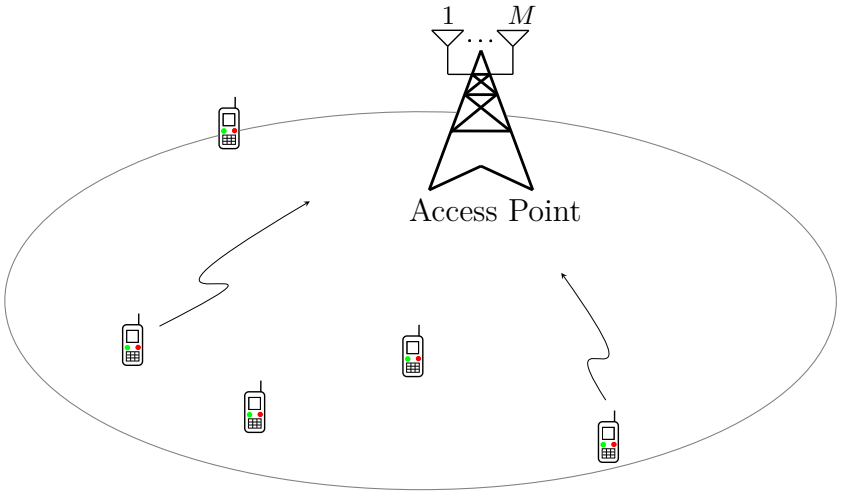

Figure 1.1: A random access scenario where clients communicate with a multiple antenna access point over a shared wireless medium.

1.1 Multi-Packet Reception Channel

Consider a dense 802.11 network scenario where clients associated with a cell, attempt to communicate with an Access Point (AP) over a shared wireless medium as in Fig. 1.1[1]. While the DCF MAC seeks to minimize concurrent transmissions, with dense networks the incidences of hidden nodes increase leading to possibilities for simultaneous uplink transmissions from multiple clients.

By and large, the *collision* model in which concurrent transmissions are viewed as failed transmissions, constitutes the prevailing assumption for random access WLANs. Rather than avoiding such simultaneous transmissions, there has been increasing interest in proposing techniques that exploit such multiple access [3–8], i.e., make use of advanced signal-processing techniques in the physical-layer to allow MPR [9], thereby improving the overall network performance.

Physical-layer Network Coding (PNC) is one such technique in which a wireless user

[1]Consistent with the premise of this thesis, we focus on uplink scenarios, driven by applications where users share data among themselves.

infers the sum of the superimposed packets as opposed to decoding individual data packets. The idea appears to have been independently proposed by Nazer and Gastpar [10], Zhang, Liew, Lam [11] and Popovski and Yomo [12]. Although PNC improves network throughput, it is prone to fading in wireless channels where the amplitude and phase of a transmitted signal undergo distortion. As a consequence, it becomes very difficult to recover the sum of transmitted packets and maintain the potential throughput gain from PNC at the same time. The compute-and-forward (C&F) strategy proposed by Nazer and Gastpar [1] to some extent overcomes this difficulty by incorporating a lattice structure into transmitted signal.

In recent years, compute-and-forward has emerged as a powerful physical-layer network coding technique [10, 13]. It has attracted increasing attention from many fields, including communications [14,15], information theory [16,17], and coding theory [18]. C&F enjoys two significant theoretical advantages: First, it achieves close-to-optimal information-theoretic performance in various important channel models, including interference channels [17], two-way relay channels [16], and Gaussian multi-hop networks [19]. Second, it admits highly efficient encoding and decoding methods, with a favorable performance-complexity trade-off [18].

In multiple access or concurrent transmission scenarios, the AP can take advantage of the broadcast and fading natures of the wireless medium to its benefit. By using a PNC technique such as C&F, the AP can infer (decode) multiple linear combinations of collided packets provided it has enough independent measurements to recover the collided packets. Other PNC techniques such as successive interference cancellation and successive C&F can also be used for this purpose. We discuss these techniques in depth in Chapter 2 for a special family of symmetric MPR channel models called all-or-nothing MPR.

C&F relies on *lattice coding*, where a lattice is a discrete set of points in a (multidimensional) Euclidean space with the fundamental property that the sum and the difference of any two lattice points are themselves lattice points [20]. That is, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are two lattice points, so are $\mathbf{x}_1 + \mathbf{x}_2$ or $\mathbf{x}_1 - \mathbf{x}_2$. In other words, lattices are closed under addition and subtraction. This implies that *every integer linear combination* $a_1\mathbf{x}_1 + a_2\mathbf{x}_2$ $(a_1, a_2 \in \mathbb{Z})$ *is also*

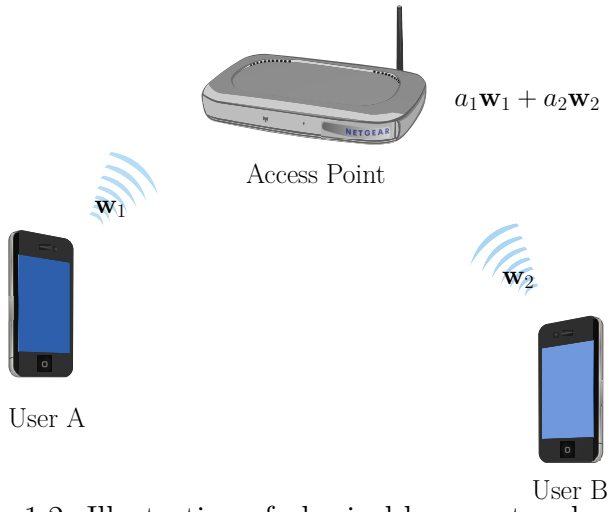

Figure 1.2: Illustration of physical-layer network coding.

a lattice point. Continuing this argument, one can see that every integer linear combination of a set of lattice points is again a lattice point. This is the key property that C&F exploits, as we will see shortly.

Let us explore the key ideas in C&F via an example. Consider a network of three users where users A and B want to transmit their messages ($\mathbf{w}_1$ and $\mathbf{w}_2$ respectively) to an access point over a fading channel as depicted in Fig. 1.2. Users A and B send a coded version of their packets, i.e., $\mathbf{x}_A$ and $\mathbf{x}_B$, simultaneously. The received signal at the access point is

$$\mathbf{y} = h_A\mathbf{x}_A + h_B\mathbf{x}_B + \mathbf{z}$$

where h_A and h_B are complex numbers representing the so-called channel gains from users A and B to the access point and $\mathbf{z}$ is channel noise. The idea is to make the channel gains "close" to integer numbers by appropriate scaling. The AP applies the scaling operation as

$$\alpha\mathbf{y} = \alpha h_A\mathbf{x}_A + \alpha h_B\mathbf{x}_B$$

to steer the effective channel gains αh_A and αh_B towards integer values. Let a_1 and a_2 be the closest integer value to αh_A and αh_B. The AP aims to infer the linear combination

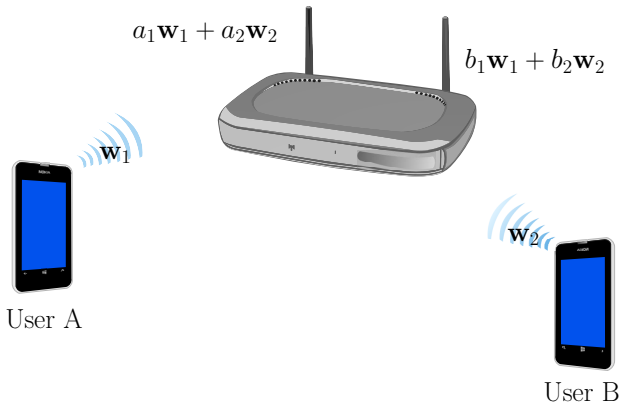

Figure 1.3: C&F and multi-antenna systems.

$a_1\mathbf{w}_A + a_2\mathbf{w}_B$.

By adding one more antenna to the AP it will be possible for the AP to infer a second linear combination $b_1\mathbf{w}_A + b_2\mathbf{w}_B$ where b_1 and b_2 are the closest integer values to channel gains from users A and B to the AP. If the two antennas are sufficiently distanced from each other then the channel gains and hence the integers b_1 and b_2 will be different from those corresponding to the first antenna. With two linearly independent combinations of messages the AP can then recover $\mathbf{w}_1$ and $\mathbf{w}_2$ by solving a system of linear equations.

Despite the popularity of PNC techniques and specifically C&F strategy, there is a critical gap between the theoretical advances of C&F and its real-world use. For instance, most prior work focuses on information-theoretic analysis and encoding-decoding methods, without considering important practical constraints of wireless networks, such as channel variations, decentralized operations, and random data arrival.

1.2 Random Access with MPR: State-of-the-Art

The stability condition of slotted ALOHA (one of the simplest random access protocols) was first introduced and analyzed by Tsybakov and Mikhailov [21] back in 70's. Despite a continuing effort in the past three decades (e.g., [22–25] and in general for parallel queueing networks [26–28]), this problem is still open except for several special cases. Recently, Borde-

nave *et al.* [29, 30] developed a mean-field framework for interacting particles and applied it to slotted ALOHA systems. In particular, they proposed an approximate stability condition and proved that it is asymptotically exact for large systems.

The study of random access protocols with MPR capability dates back to late 80's. In their seminal work [9, 31], Ghez, Verdú and Schwartz introduced a so-called symmetric MPR channel model for slotted ALOHA and characterized the stability condition in the large-systems regime. Following [9, 31], Sant and Sharma studied the finite-user regime by focusing on a special class of the symmetric MPR model [32]. An asymmetric MPR model was analyzed by Naware, Mergen and Tong [33], with a particular focus on the two-user case and the homogeneous case (due to several technical challenges explained in [33]). For more general cases, they provided an inner bound on the stability region. However, as they pointed out, their inner bound is "too difficult if not impossible to evaluate in practice in general." The stability condition for slotted ALOHA with MPR under the *inhomogeneous* case remains largely open in the literature.

The extension from slotted ALOHA to CSMA systems under MPR assumption started from the work of Chan, Berger and Tong [34] in 2004. In particular, the maximum stable throughput and the service delay were derived in [6, 34] for the symmetric MPR model of [9] under the homogenous case. Other works along this direction include [35–41]. Recently, CSMA with MPR has received renewed interest from the research community, mainly due to rapid advances in multiple-input multiple-output (MIMO) technology for wireless LAN. For example, Tan *et al.* [7] developed a CSMA-type system with chain-decoding MPR technique and demonstrated its clear advantage over conventional 802.11 through prototype implementation. Wu *et al.* [36] evaluated the performance of such system design in terms of saturated throughput and service delay. Bae *et al.* [38] studied the optimal transmission probability to maximize the stable throughput for CSMA systems where the access point can decode up to K simultaneous transmissions, which is a special case of the symmetric MPR model.

Despite a large body of work on random access protocols with MPR capability [5, 6, 8, 9, 33–38, 40, 42], some of their fundamental properties are still not well understood. For instance,

most performance analysis of MPR-capable CSMA focuses on the homogeneous case, where all the mobile users are assumed to have identical packet arrival rates and transmission probabilities [5, 6, 34, 35, 37]. In reality, mobile users often have different packet arrival rates, leading to the inhomogeneous case. However, the throughput and delay performances of MPR-capable CSMA for the inhomogeneous case remain largely unknown in the literature.

As a promising new physical-layer technique and a special case of MPR model, we believe that C&F is a good candidate for multiple access scenarios. Most prior work focused on its physical-layer performance (see, e.g., [43–47] for multi-source multi-relay networks with C&F) with a few exceptions [48–51]. The work of Goseling *et al.* pioneered the throughput analysis of random access with C&F for the homogeneous case.

1.3 Meta-Stability of Random Access Networks

Meta-stability is a known phenomenon in random access networks. For instance, in slotted ALOHA and CSMA systems bi-stability has been observed [52–55]. A meta-stable system oscillates between two or more stable states over long time intervals. This is not a desired phenomenon in practical wireless systems as it impedes provisioning of quality of service guarantees. So it is important to understand what causes a random access system – especially in presence of MPR assumption – to undergo a phase transition and become meta-stable. As we will see later, MPR-capable slotted ALOHA and CSMA systems are in general meta-stable with two or more stable states. The dynamical system description of these systems makes it more convenient to study the stability of these systems and obtain conditions under which meta-stability occurs. Specifically, the mean-field approach enables identifying constraints on the physical-layer that avoids meta-stability. In other words, this thesis provides design criteria for lattice code design such that the slotted ALOHA system become globally stable. This is important from a system design perspective for practical wireless system.

1.4 Our Contributions

In this dissertation, we introduce a general framework for the mean-field analysis of MPR-based random access protocols. Specifically we characterize stability conditions and study system performance with a focus on the applicability of these results in practice. Our main contributions are described in the following on a chapter by chapter basis.

1.4.1 Slotted ALOHA with All-or-Nothing Multi-Packet Reception

Very recently, Bordenave *et. al.* [29] proposed an approximation stability condition for slotted ALOHA. Prior to this work, most of the analysis of stability conditions have focused on homogeneous networks, whereas in most real-world applications, networks are inhomogeneous. For instance a network may consist of laptops, smart-phones, tablets with different demands. The mean-field technique has already been successfully applied in the context of queueing networks in the past. We apply mean-field analysis to inhomogeneous slotted ALOHA systems with all-or-nothing symmetric MPR channel model and in fact take a step forward to solve a decades-long open problem. By a sample-path analysis approach, it will be shown that the behavior of slotted ALOHA with MPR model can be described by a dynamical system. Such a dynamical system is composed of a set of coupled ordinary differential equations (ODEs) that approximates the average of the trajectories of the underlying stochastic process.

It turns out that the approximation to slotted ALOHA stability via this approach is very accurate even when the number of users is small. As we will see the MPR-based random access protocols are in general meta-stable systems, i.e., such systems have multiple stable points which makes their performance analysis quite challenging. We overcome this challenge by imposing certain constraints on the physical-layer which makes the system bi-stable rather than meta-stable with more than two stable points.

Build upon the stability results, throughput and delay performance of MPR-capable slotted ALOHA is studied demonstrating the clear advantage of MPR-based systems over

conventional systems as well as the accuracy of our approximations.

1.4.2 CSMA with All-or-Nothing Multi-Packet Reception

Taking a similar mean-field approach, this dissertation extends the analysis to the Carrier Sense Multiple Access (CSMA) protocol under the MPR model assumption. Specifically we study inhomogeneous persistent CSMA systems with all-or-nothing MPR model. We omit the details of the analysis as it is similar to the slotted ALOHA scenario. Instead, we start from the mean-field decoupling assumption. According to the mean-field decoupling assumption, the evolution of the states of queues (users) becomes independent from each other in large-systems regime when the number of queues tends to infinity. It turns out that a CSMA system – depending on its input arrival rates – may end up in three situations: the system is either stable, bi-stable or unstable. In fact, a slotted ALOHA system also behaves similarly. We further study the throughout and delay performance of persistent CSMA systems and compare it to those of slotted ALOHA both under MPR assumption and for C&F, SCF and SIC MPR models.

1.4.3 Extension I: CSMA with Random Back-off

The infrastructure mode of IEEE 802.11 standard has adopted a back-off mechanism to reduce collision in a CSMA system. In such a system users decrease their contention rate in case a collision occurs to avoid repeatedly interfering with other users' transmissions and reduce collision. The persistent CSMA system that we study in Chapter 4 assumes fixed contention rate (i.e., the transmission probability of every user is fixed) and applies a mean-field-based analysis to CSMA system with back-off for a *saturated* scenario, i.e. all users always have a packet to send. The mean-field approximation is adopted for the analysis of two cases, a) finite b) infinite number of back-off stages. We model the behavior of large-systems with dynamical systems for each case; the existence of a stationary probabilities for each back-off stage is studied and verified by simulations.

1.4.4 Extension II: The Multiple AP Case

The work of chapters 3 and 4 is extended to a multi-AP scenario by identifying two technical challenges and proposing novel solutions to address these challenges. In particular, we first introduce an AP cooperation problem to fully recover packets from packet collisions, and develop a distributed algorithm which is computationally efficient. We then introduce a joint channel estimation and active user recovery problem, and propose an algorithm which works based on the notion of spare recovery. In addition, we discuss that the benefits of adding more APs in terms of throughput and delay performance can be characterized using the results in Chapter 4.

The AP cooperation problem is closely related to the network-coding design problem [14, 56]. As we have shown in Chapter 6 (Section 6.2.4), our distributed solution to AP cooperation makes a particular use of the problem structure, and so is more computational efficient than existing solutions to the network-coding design problem.

While the high-level idea of applying sparse recovery techniques to channel estimation problems is not new (e.g., [57]), our application focuses on a dense network with concurrent transmitters, whereas existing approaches focus on a *single* transmitter-receiver pair (where the sparsity comes from the multi-path effect). Moreover, we make use of second-order Reed-Muller codes, leading to a recovery algorithm whose complexity is sub-linear of the number of users.

Chapter 2
MULTI-PACKET RECEPTION CHANNEL MODELS

Conventionally, random-access protocols are studied based on the so-called collision channel model: a transmission is successful if and only if a single user transmits. Motivated by the advances in multiuser detection, Ghez, Verdú and Schwartz introduced a symmetric multi-packet reception (MPR) channel model for the slotted ALOHA protocol (one of the simplest random-access protocols) [9,31]. In their channel model, the number of successfully received packets in each time slot depends exclusively on the number of simultaneous transmissions. Specifically, given that n packets are being transmitted in one slot, the number of correctly received packets is a random variable specified by conditional probabilities $\{q_{k,n}\}_{k=0}^{n}$ where

$$q_{k,n} = \mathbb{P}\left(k \text{ packets are correctly received} | n \text{ are transmitted}\right).$$

The above symmetric MPR model was later extended by Naware, Mergen and Tong to a general asymmetric setting where the set of successfully received packets in a time slot depends on the set of concurrent transmitters in that slot [33]. For instance, for the two-user case, the asymmetric MPR model in [33] is completely specified by the following conditional probabilities:

$$\begin{aligned} q_{\{i\},\{i\}} &= \mathbb{P}(\text{user } i \text{ is successful} \,|\, \text{only user } i \text{ transmits}) \\ q_{\{i\},\{1,2\}} &= \mathbb{P}(\text{only user } i \text{ is successful} \,|\, \text{both users transmit}) \\ q_{\{1,2\},\{1,2\}} &= \mathbb{P}(\text{both users are successful} \,|\, \text{both users transmit}) \end{aligned}$$

where $i \in \{1, 2\}$.

In this chapter, we discuss our system model assumptions including channel model and

MPR techniques. In particular, we describe the symmetric MPR model and study a special case called "all-or-nothing" MPR which is still general enough to include a number of MPR techniques used in the subsequent chapters.

2.0.1 Channel Model

We use a standard block-fading multiple-access channel model [58] as illustrated in Fig. 2.1, where each user is equipped with a single antenna and the AP is equipped with M antennas. We assume block-level synchronization. Over a block length of b symbols, when there are n active users communicating to the AP, the received signal at the j-th antenna at the AP can be written as

$$\mathbf{y}_j = \sum_{\ell=1}^{n} h_{j,\ell}\mathbf{x}_\ell + \mathbf{z}_j \,.$$

Here, $\mathbf{x}_\ell \in \mathbb{C}^b$ is the transmitted signal (which is a lattice point or codeword) at the ℓth active user subject to the average power constraint $\frac{1}{b}E\|\mathbf{x}_\ell\|^2 \leq P$, $h_{j,\ell}$ is the channel fading coefficient between the ℓth active user and jth antenna of the AP, and $\mathbf{z}_j$ represents the additive white Gaussian noise vector with i.i.d. $\mathcal{CN}(0,1)$ entries (where $\mathcal{CN}(0,a)$ denotes a complex Gaussian random variable with independent zero-mean, variance $a/2$, Gaussian random variables as its real and imaginary parts).

In the matrix form, the AP observes a channel-output matrix matrix $\mathbf{Y} \in \mathbb{C}^{M\times b}$

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}\,,$$

where $\mathbf{H} \in \mathbb{C}^{M\times n}$ is the channel matrix whose entry in the j-th row and ℓ-th column is $h_{j,\ell}$, $\mathbf{X} \in \mathbb{C}^{n\times b}$ is the channel-input matrix with $\mathbf{x}_\ell$ as its ℓth row, and $\mathbf{Z}$ is the noise matrix with $\mathbf{z}_j$ as its jth row. We assume that the channel-coefficient matrix $\mathbf{H}$ is known to the receiver but unknown to the transmitters. We statistically model $\mathbf{H}$ to be i.i.d. with $\mathcal{CN}(0,1)$ entries, i.e., the richly scattered Rayleigh-fading environment[1].

[1]This assumption can be extended to other statistical models of $\mathbf{H}$.

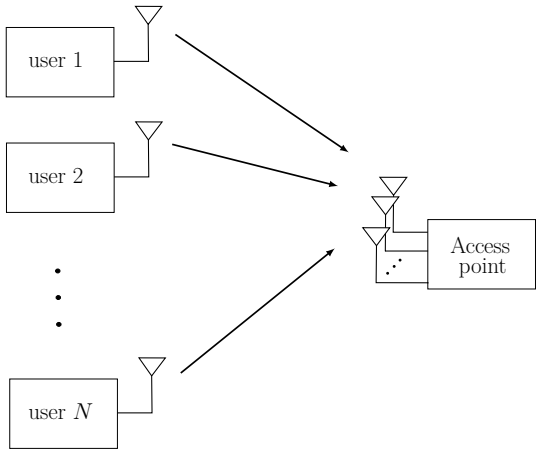

Figure 2.1: A random access wireless channel where N single-antenna users compete to gain access to a channel and transmit packets to a AP with M antennas.

2.0.2 Symmetric MPR Model

We now introduce the so-called symmetric MPR model as a special case of MPR channels defined in [9].

Definition 1(Symmetric MPR) Given that n packets are being transmitted in a block time, the probability of successful recovery of exactly k packets is defined as

$$q_{k,n} = \Pr(k \text{ packets are correctly received } | \, n \text{ packets are transmitted}) \, .$$

These k packets are chosen, uniformly at random, from n concurrent transmitters.

As a special case, if $q_{k,n} = 0$ for all $k \in \{1, \ldots, n-1\}$, we have an "all-or-nothing" symmetric MPR model where the receiver either recovers every simultaneously transmitted packets or none of them.

Definition 2(All-or-Nothing Symmetric MPR) Given that n packets are being transmitted in a block time, the probability of successful recovery of these n packets is defined

as

$$q_n = \mathbb{P}(n \text{ packets are correctly received } | \, n \text{ packets are transmitted}).$$

Clearly, the "all-or-nothing" symmetric MPR model generalizes the classical collision channel model which has $q_1 = 1$ and $q_n = 0$ for all $n > 1$. It also contains the K-user MPR model used in [59] as a special case which assumes that $q_n = 1$ for all $n \leq K$ and $q_n = 0$ for all $n > K$, where K is a threshold determined by practical constraints (such as the speed of Walsh-Hadamard transform [60]). We call this threshold the MPR capability.

Interestingly, the all-or-nothing symmetric MPR model is still general enough to include a number of promising MPR techniques—such as successive interference cancellation (SIC), compute-and-forward (C&F), and successive compute-and-forward (SCF)—as special cases.

2.0.3 Asymmetric MPR Model

The symmetric MPR model can be generalized to asymmetric MPR model. Suppose that each user belongs to one of V possible classes $\mathcal{V} = \{1, \cdots, V\}$. We say a set of $(n_1, \cdots, n_V)$ packets is being transmitted in a slot, if for all $v \in \mathcal{V}$, n_v packets are being transmitted. Similarly, we say a set of $(k_1, \cdots, k_V)$ packets is correctly received, if for all $v \in \mathcal{V}$, exactly k_v packets are correctly received.

Definition 3(Asymmetric MPR Channel Model) Given that a set of $(n_1, \cdots, n_V)$ packets is being transmitted in a time slot, the probability of successful recovery of a set of $(k_1, \cdots, k_V)$ packets is defined as

$$q_{(k_1,\cdots,k_V),(n_1,\cdots,n_V)} = \mathbb{P}\Big((k_1, \cdots, k_V) \text{ packets are correctly received} \Big| \\ (n_1, \cdots, n_V) \text{ packets are transmitted}\Big).$$

For all $v \in \mathcal{V}$, these k_v packets are chosen, uniformly at random, from n_v concurrent transmitters of class-v.

Example 1(Two-User Case) Consider a random-access network of two classes of users,

where each class has only one user. Then, the asymmetric MPR channel model defined above is completely specified by five conditional probabilities: $q_{(1,0),(1,0)}$, $q_{(0,1),(0,1)}$, $q_{(1,0),(1,1)}$, $q_{(0,1),(1,1)}$, $q_{(1,1),(1,1)}$.

In this dissertation, we present the results for both all-or-nothing and general symmetric MPR model. We will see that the analysis of general symmetric MPR is not fundamentally different than the analysis of all-or-nothing MPR. The simulations and numerical results are mainly focused on all-or-nothing MPR.

2.1 C&F as an MPR

We give a brief summary of the C&F and SCF schemes proposed in [61], with a particular focus on the symmetric rates and complex-valued channel models. We refer the readers to [61, 62] for more details on C&F.

As a new MPR technique, C&F enables a receiver to recover simultaneously transmitted packets via decoding linear equations [1, 17–19, 63, 64]. Compared to other MPR techniques, C&F achieves close-to-optimal performance [62, 63] under single-user decoding, making it particularly attractive for practical applications.

2.2 An Overview of C&F

Let us begin with a special case in which all of the channel gains are integers. In this case, the receiver simply decodes $\sum_{\ell=1}^{n} h_\ell \mathbf{x}_\ell$ from $\mathbf{y}$ by using a (standard) lattice decoder, since $\sum_{\ell=1}^{n} h_\ell \mathbf{x}_\ell$ is a lattice point. The decoding is correct if the noise $\mathbf{z}$ is sufficiently small.

In general, however, the channel gains are not integers; they are general real numbers. In this case, the idea is to make them "close" to integer numbers by appropriate scaling. Specifically, the receiver applies a scaling operation $\alpha\mathbf{y} = \sum_{\ell=1}^{n} \alpha h_\ell \mathbf{x}_\ell + \alpha\mathbf{z}$ to steer the effective channel gains $\{\alpha h_\ell\}$ towards integer values, without unduly amplifying the noise $\mathbf{z}$. Let a_ℓ be the closest integer to αh_ℓ, for $\ell = 1, \ldots, n$. Then, once again, the receiver decodes

$\sum_{\ell=1}^{n} a_\ell \mathbf{x}_\ell$ from $\alpha\mathbf{y}$ by using a (standard) lattice decoder. Note that

$$\begin{aligned}\alpha\mathbf{y} &= \sum_{\ell=1}^{n} a_\ell \mathbf{x}_\ell + \sum_{\ell=1}^{n} (\alpha h_\ell - a_\ell)\mathbf{x}_\ell + \alpha\mathbf{z} \\ &= \sum_{\ell=1}^{n} a_\ell \mathbf{x}_\ell + \mathbf{n}\,,\end{aligned}$$

where $\mathbf{n} = \sum_{\ell=1}^{n} (\alpha h_\ell - a_\ell)\mathbf{x}_\ell + \alpha\mathbf{z}$ is the effective noise. Thus, the decoding is correct if the effective noise $\mathbf{n}$ is sufficiently small.

It turns out that the decoding performance can be characterized formally by the following *computation rate* $R_{\mathsf{comp}}(\mathbf{a}, \alpha; \mathbf{h})$ for any coefficient vector $\mathbf{a} = (a_1, \ldots, a_n)^T$ and any scalar α [1]:

$$R_{\mathsf{comp}}(\mathbf{a}, \alpha; \mathbf{h}) = \frac{1}{2} \log \left(\frac{P}{\alpha^2 + P\|\alpha\mathbf{h} - \mathbf{a}\|^2} \right), \tag{2.1}$$

where P is the power constraint and $\mathbf{h} = (h_1, \ldots, h_L)^T$. Roughly speaking, the higher the computation rate, the more reliable the decoding is. In other words, the computation rate $R_{\mathsf{comp}}(\mathbf{a}, \alpha; \mathbf{h})$ provides a guideline for the choice of the coefficient vector $\mathbf{a}$ and the scalar α. In particular, for any given coefficient vector $\mathbf{a}$, the best computation rate $R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h})$ is attained at the optimal $\alpha^* = \frac{P\mathbf{h}^T\mathbf{a}}{1+P\|\mathbf{h}\|^2}$ and is given by

$$R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h}) = \frac{1}{2} \log \left(\frac{1 + P\|\mathbf{h}\|^2}{(1 + P\|\mathbf{h}\|^2)\|\mathbf{a}\|^2 - P(\mathbf{h}^T\mathbf{a})^2} \right).$$

The main result of C&F says that the receiver can decode an integer combination $\sum_{\ell=1}^{n} a_\ell \mathbf{x}_\ell$ with coefficient vector $\mathbf{a}$ as long as the computation rate $R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h})$ is larger than the message rate R, i.e., $R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h}) > R$. Hence, if there are multiple coefficient vectors whose computation rates are larger than R, the receiver can recover multiple integer combinations of the transmitted signals.

Finally, once an integer combination $\sum_{\ell=1}^{n} a_\ell \mathbf{x}_\ell$ of transmitted signals is decoded, it can be mapped to a linear combination $\sum_{\ell=1}^{n} a_\ell \mathbf{w}_\ell$ of the corresponding packets. The details can be found in [18].

To summarize, with C&F the receiver is able to decode one or more linear combinations of the form $\sum_{\ell=1}^{n} a_\ell \mathbf{w}_\ell$ from the collision $\mathbf{y} = \sum_{\ell=1}^{n} h_\ell \mathbf{x}_\ell + \mathbf{z}$. The coefficients of these linear combinations are carefully chosen such that their computation rates are larger than the message rate.

2.3 All-or-Nothing MPR channel vs Conventional Channel

We now discuss C&F, SCF, and SIC as special cases of all-or-nothing MPR model and provide concrete examples for each case.

2.3.1 Case Studies of Symmetric MPR Model

Case Study 1: MPR via C&F

As our first case study, we explain how MPR can be achieved via C&F technique. Recall that the received signal at the AP is $\mathbf{Y} = \mathbf{HX} + \mathbf{Z}$. Rather than decoding $\mathbf{X}$ from $\mathbf{Y}$, the AP (which employs C&F technique) will first decode integer-linear combinations $\mathbf{AX}$ and then invert these linear combinations to recover the original signals $\mathbf{X}$. Here, $\mathbf{A} \in \mathbb{C}^{n\times n}$ is a target (invertible) integer-valued matrix of which each row $\mathbf{a}_\ell$ corresponds to an integer combination.

Specifically, the AP chooses an equalizing filter matrix $\mathbf{B} \in \mathbb{C}^{n\times M}$ and computes

$$\begin{aligned}\mathbf{Y}_{\mathsf{eff}} &= \mathbf{BY} \\ &= \mathbf{AX} + (\mathbf{BH} - \mathbf{A})\mathbf{X} + \mathbf{BZ} \\ &= \mathbf{AX} + \mathbf{Z}_{\mathsf{eff}}\,,\end{aligned}$$

where $\mathbf{Z}_{\mathsf{eff}} \triangleq (\mathbf{BH}-\mathbf{A})\mathbf{X}+\mathbf{BZ}$ is the *effective* noise matrix. In other words, C&F transforms the original multiple access channel into n point-to-point sub-channels

$$\mathbf{y}_{\mathsf{eff},\ell} = \mathbf{a}_\ell \mathbf{X} + \mathbf{z}_{\mathsf{eff},\ell},\ \ell = 1, \ldots, n\,,$$

where $\mathbf{y}_{\mathsf{eff},\ell}$ and $\mathbf{z}_{\mathsf{eff},\ell}$ are the ℓth rows of $\mathbf{Y}_{\mathsf{eff}}$ and $\mathbf{Z}_{\mathsf{eff}}$, respectively.

As shown in [62], the optimal choice of the equalizing matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \mathbf{A}\mathbf{H}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1},$$

and an optimal choice of the integer matrix $\mathbf{A}$ can be found through lattice-reduction algorithms[2]. The resulting achievable symmetric rate $R_{\mathsf{sym}}^{\mathsf{CF}}(\mathbf{H})$ is[3]

$$R_{\mathsf{sym}}^{\mathsf{CF}}(\mathbf{H}) = \min_{\ell=1,\ldots,n} \log \left(\frac{\mathsf{SNR}}{\sigma_{\mathsf{eff},\ell}^2} \right), \tag{2.2}$$

where $\sigma_{\mathsf{eff},\ell}^2$ is the ℓth diagonal entry of the matrix $\mathsf{SNR}\ \mathbf{A}(\mathbf{I} + \mathsf{SNR}\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T$.

The optimal choice if $\mathbf{A}$ can also be obtained from the so-called *dominant solutions* defined in [18]. Let

$$R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_\ell) = \log \left(\frac{\mathsf{SNR}}{\sigma_{\mathsf{eff},\ell}^2} \right),$$

Definition 4(Dominant Solutions) A set of linearly independent coefficient vectors $\{\mathbf{a}_j^*\}$ satisfying $R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_1^*) \geq \cdots \geq R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_n^*)$ is called a *dominant solution* if, for any linearly independent coefficient vectors $\{\mathbf{a}_j\}$ satisfying $R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_1) \geq \cdots \geq R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_n)$, we have $R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_j^*) \geq R_{\mathsf{comp}}(\mathbf{H}, \mathbf{a}_j)$ for all $j = 1, \ldots, n$.

By definition, a dominant solution $\{\mathbf{a}_j^*\}$ indeed maximizes the message rate. Also, a dominant solution $\{\mathbf{a}_j^*\}$ can be computed by a greedy algorithm as in [18, Section VIII].

Example 2(MPR via C&F) Consider the following received signal

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} -0.3034 - 0.3673i & -0.0056 + 0.4963i \\ 0.4512 - 0.8633i & -0.9261 + 0.5284i \\ -0.0608 - 0.6401i & -0.1765 + 0.5561i \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{bmatrix},$$

---

[2]We note that when the number of active users is small (< 5), there exist some very efficient lattice-reduction algorithms. See, e.g., [65] for details.

[3]All the $\log(\cdot)$ functions in this section are in base 2 .

with $\mathsf{SNR} = 15\,\text{dB}$. That is, the AP (which is equipped with 3 antennas) would like to recover two original signals $\mathbf{x}_1$ and $\mathbf{x}_2$. Instead of decoding $\mathbf{x}_1$ and $\mathbf{x}_2$ directly, the AP decides to decode two integer-linear combinations:

$$\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix} = \begin{bmatrix}1+i & -1-i\\ -2i & -1+2i\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}.$$

In order to do so, the AP applies the following equalizing matrix

$$\begin{aligned}\mathbf{B} &= \mathbf{A}\mathbf{H}^T\left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^T\right)^{-1}\\ &= \begin{bmatrix}-0.1709+0.6350i & 0.2802+0.9764i & -0.0544-0.2305i\\ 0.4580+0.1405i & 1.0795-0.8104i & 0.7780+0.4329i\end{bmatrix},\end{aligned}$$

and computes

$$\begin{bmatrix}\mathbf{y}_{\mathsf{eff},1}\\ \mathbf{y}_{\mathsf{eff},2}\end{bmatrix} = \mathbf{B}\mathbf{Y} = \begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix} + \begin{bmatrix}\mathbf{z}_{\mathsf{eff},1}\\ \mathbf{z}_{\mathsf{eff},2}\end{bmatrix},$$

where

$$\begin{aligned}\begin{bmatrix}\mathbf{z}_{\mathsf{eff},1}\\ \mathbf{z}_{\mathsf{eff},2}\end{bmatrix} &= \begin{bmatrix}0.1102+0.1176i & 0.0482+0.1659i\\ -0.0701-0.0327i & -0.0218-0.0964i\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\\ &\quad + \begin{bmatrix}-0.1709+0.6350i & 0.2802+0.9764i & -0.0544-0.2305i\\ 0.4580+0.1405i & 1.0795-0.8104i & 0.7780+0.4329i\end{bmatrix}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\end{bmatrix}.\end{aligned}$$

This gives rise to 2 point-to-point sub-channels $\mathbf{y}_{\mathsf{eff},j} = \mathbf{u}_j + \mathbf{z}_{\mathsf{eff},j}$, where the effective noise variances can be computed as

$$\sigma^2_{\mathsf{eff},1} = 3.2852 \text{ and } \sigma^2_{\mathsf{eff},2} = 3.3421\,,$$

by using the fact that the noise vector $\mathbf{z}_j$ consists of i.i.d. $\mathcal{CN}(0,1)$ entries and $\mathsf{SNR} = 15\,\text{dB}$.

Hence, the achievable symmetric rate is

$$\begin{aligned} R_{\mathsf{sym}}^{\mathsf{CF}}(\mathbf{H}) &= \min\left\{\log\left(\frac{\mathsf{SNR}}{3.2852}\right), \log\left(\frac{\mathsf{SNR}}{3.3421}\right)\right\} \\ &= 3.2421\,, \end{aligned}$$

and the two combinations can be decoded correctly as long as the message rate R is less than the achievable symmetric rate. Finally, the AP recovers the original signals $\mathbf{x}_1$ and $\mathbf{x}_2$ by inverting the matrix $\mathbf{A}$. That is, $\mathbf{X} = \mathbf{A}^{-1}(\mathbf{AX})$, where

$$\mathbf{A}^{-1} = \begin{bmatrix} -1+2i & 1+i \\ 2i & 1 \end{bmatrix}.$$

Let q_n denote the probability that these n linear combinations (with coefficient vectors $\{\mathbf{a}_j^*\}$) are successfully decoded. For any given statistical model of $\mathbf{H}$, the success probability q_n can be computed as

$$q_n = \Pr\left(R < R_{\mathsf{sym}}^{\mathsf{CF}}(\mathbf{H})\right), \tag{2.3}$$

where R is the message rate of the transmitted packets. In fact, the decoding performance of C&F described in (2.3) can be further improved. Intuitively, after the receiver has correctly decoded $n-1$ linear combinations with coefficient vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$, the receiver can use these linear combinations as side information to decode the nth linear combination with coefficient vector $\mathbf{a}_n$. This leads us to successive C&F which will be discussed in the following section.

Case Study 2: MPR via Successive C&F

Successive C&F combines ideas from classical SIC and C&F. Similar to C&F, successive C&F first recovers a set of integer-linear combinations $\mathbf{AX}$ and then find the original signals $\mathbf{X}$. Rather than decoding each combination $\mathbf{a}_\ell\mathbf{X}$ in parallel, successive C&F decodes these

combinations one at a time and makes use of already decoded combinations in subsequent decoding steps.

For any invertible integer matrix $\mathbf{A}$, the matrix $\mathbf{A}(\mathbf{I}+\mathsf{SNR}\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T$ admits a Cholesky decomposition

$$\mathbf{A}(\mathbf{I}+\mathsf{SNR}\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T = \mathbf{L}\mathbf{L}^T,$$

where $\mathbf{L} \in \mathbb{C}^{n\times n}$ is a lower triangular matrix with strictly positive diagonal entries. As shown in [62], the effective noise matrix $\mathbf{Z}_{\mathsf{eff}}$ can be written as

$$\mathbf{Z}_{\mathsf{eff}} = \sqrt{\mathsf{SNR}}\,\mathbf{L}\mathbf{W}\,,$$

for some matrix $\mathbf{W}$ with unit generalized covariance matrix. Hence, the resulting achievable symmetric rate $R_{\mathsf{sym}}^{\mathsf{SCF}}(\mathbf{H})$ is

$$R_{\mathsf{sym}}^{\mathsf{SCF}}(\mathbf{H}) = \min_{\ell=1,\ldots,n} \log\left(\frac{\mathsf{SNR}}{\mathsf{SNR}L_{\ell,\ell}^2}\right), \tag{2.4}$$

where $L_{\ell,\ell}$ is the ℓth diagonal entry of the matrix $\mathbf{L}$.

Example 3(MPR via Successive C&F) With the same setup as described in Example 2, the effective noise matrix can be written as

$$\begin{bmatrix}\mathbf{z}_{\mathsf{eff},1}\\ \mathbf{z}_{\mathsf{eff},2}\end{bmatrix} = \sqrt{\mathsf{SNR}}\begin{bmatrix}0.3223 & 0\\ -0.1496-0.1301i & 0.2577\end{bmatrix}\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix},$$

for some matrix $\mathbf{W}$ with unit generalized covariance matrix. Hence, the point-to-point subchannels can be expressed as

$$\begin{bmatrix}\mathbf{y}_{\mathsf{eff},1}\\ \mathbf{y}_{\mathsf{eff},2}\end{bmatrix} = \mathbf{B}\mathbf{Y} = \begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix} + \sqrt{\mathsf{SNR}}\begin{bmatrix}0.3223\mathbf{w}_1\\ (-0.1496-0.1301i)\mathbf{w}_1+0.2577\mathbf{w}_2\end{bmatrix}.$$

Rather than decoding $\mathbf{u}_1$ and $\mathbf{u}_2$ in parallel, the AP first decodes $\mathbf{u}_1$, and then recovers $\mathbf{w}_1$

and cancels its contribution to the second sub-channel. In this way, the AP creates a less noisy sub-channel

$$\mathbf{y}'_{\mathsf{eff},2} = \mathbf{u}_2 + \sqrt{\mathsf{SNR}}\, 0.2577\mathbf{w}_2\,,$$

where the variance of the effective noise is reduced from to 3.3421 (see Example 2) to $\mathsf{SNR}\times 0.2577^2 = 2.1000$. Hence, the achievable symmetric rate is

$$\begin{aligned} R_{\mathsf{sym}}^{\mathsf{SCF}}(\mathbf{H}) &= \min\left\{\log\left(\frac{\mathsf{SNR}}{3.2852}\right), \log\left(\frac{\mathsf{SNR}}{2.1000}\right)\right\} \\ &= 3.2669\,. \end{aligned}$$

This improves the performance of C&F whose achievable symmetric rate is 3.2421.

Similarly, the success probability for successive C&F can be computed as

$$q_n = \mathbb{P}(R < R_{\mathsf{sym}}^{\mathsf{SCF}}(\mathbf{H}))\,.$$

Case Study 3: MPR via SIC

As pointed out in [62], if the integer matrix $\mathbf{A}$ is chosen to be a permutation matrix, successive C&F in the previous case study reduces to the well-known SIC (and each permutation matrix corresponds to a corner point in the capacity region). Hence, successive C&F, in general, offers higher achievable symmetric rates than SIC.

Example 4 Under the setup described in Example 2, if the integer matrix $\mathbf{A}$ is chosen to be

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

then the achievable symmetric rate is $\min\{1.3848, 5.7946\} = 1.3848$, according to the for-

Table 2.1: Success probabilities q_L's for SIC, CF, SCF, and JD under independent Rayleigh-fading environment and different message rates.

success prob. q_L	SNR (dB)	message rate R	M	SIC	C&F	SCF	JD
q_1	6	1	1	0.78	0.78	0.78	0.78
q_2				0.46	0.45	0.57	0.60
q_1	15	2	1	0.91	0.91	0.91	0.91
q_2				0.31	0.61	0.66	0.80
q_1				0.98	0.98	0.98	0.98
q_2	15	3	2	0.88	0.92	0.93	0.95
q_3				0.32	0.70	0.81	0.91

mulas presented in Case Study 2. Similarly, if the integer matrix is chosen to be

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

then the achievable symmetric rate is $\min\{1.4998, 5.6796\} = 1.4998$. Therefore, the achievable symmetric rate for SIC is

$$R_{\mathsf{sym}}^{\mathsf{SIC}}(\mathbf{H}) = \max\{1.3848, 1.4998\} = 1.4998\,.$$

This corresponds to the decoding order of first recovering $\mathbf{x}_2$ and then $\mathbf{x}_1$.

The success probability q_n for SIC can be computed in a similar way as before, using the formula

$$q_n = \mathbb{P}(R < R_{\mathsf{sym}}^{\mathsf{SIC}}(\mathbf{H}))\,.$$

Table 2.1 provides the values of q_n for SIC, CF, SCF, assuming independent Rayleigh-fading environment. As a comparison, the performance of joint decoding (JD) is also provided, where the symmetric rate $R_{\mathsf{sym}}^{\mathsf{JD}}(\mathbf{H})$ lies in the boundary of the capacity region for any given channel matrix $\mathbf{H}$.

As shown in Table 2.1, q_1 remains the same for all the schemes under the three configurations. This is because when $n = 1$ (i.e., there is only one active user), the channel model

reduces to a standard point-to-point channel. Also, it is observed that SCF significantly outperforms SIC in terms of q_2 and q_3, especially in the high-SNR regime. This is because SIC is a very special case of SCF. Moreover, the performance of SCF is close to that of JD even in the low-SNR regime.

2.3.2 **Implementation Considerations.**

At first glance, the complexity of SCF appears to be significantly higher than that of SIC due to the use of lattice codes and lattice decoding. It turns out that SCF has essentially the same complexity as SIC for certain lattice codes constructed from convolutional codes and LDPC codes. Specifically, as shown in [18, Appendix G], a slightly modified version of the Viterbi decoder can be used to implement a lattice decoder for lattice codes constructed through convolutional codes. In addition, as demonstrated in [66], a family of high-performance lattice codes can be constructed via (non-binary) LDPC codes and decoded using an iterative message-passing algorithm whose complexity is essentially linear in the lattice dimension.

When a practical lattice code $\mathcal{C}$ described above is used, the previous symmetric rates (2.2) and (2.4) no longer hold, since they are information-theoretic bounds based on asymptotically (as the lattice dimension goes to infinity) good lattice codes. In this case, we can use the following formulas as suggested in [67, Chapter 8] to estimate the achievable symmetric rates:

$$R_{\mathsf{sym}}^{\mathsf{CF}}(\mathbf{H}) = \min_{\ell=1,\ldots,L} \log\left(\frac{\mathsf{SNR}}{\sigma_{\mathsf{eff},\ell}^2}\right) - \log\left(2\pi e G(\mathcal{L}')\right) - \log\left(\frac{\mu(\mathcal{L}, P_e)}{2\pi e}\right), \tag{2.5}$$

and

$$R_{\mathsf{sym}}^{\mathsf{SCF}}(\mathbf{H}) = \min_{\ell=1,\ldots,L} \log\left(\frac{\mathsf{SNR}}{\mathsf{SNR} L_{\ell,\ell}^2}\right) - \log\left(2\pi e G(\mathcal{L}')\right) - \log\left(\frac{\mu(\mathcal{L}, P_e)}{2\pi e}\right), \tag{2.6}$$

where $\mathcal{L}$ and $\mathcal{L}'$ are the coding lattice and shaping lattice for the lattice code $\mathcal{C}$, respectively. Here, $G(\mathcal{L}')$ is the normalized second moment for the shaping lattice $\mathcal{L}'$. (For example, $G(\mathcal{L}') = 1/12$ if the shaping lattice $\mathcal{L}'$ is a hypercube.) Also, $\mu(\mathcal{L}, P_e)$ is the normalized volume-to-noise ratio (which is closely related to the coding gain) for a given error probability P_e in nearest neighbor decoding of the lattice $\mathcal{L}$. Note that (2.5) and (2.6) apply to an

arbitrary lattice code instead of asymptotically-good lattice codes, where $\log\left(2\pi e G(\mathcal{L}')\right)$ and $\log\left(\frac{\mu(\mathcal{L},P_e)}{2\pi e}\right)$ represent the shaping loss and coding loss, respectively. We refer our readers to [67, chapters 3, 8] for details.

Like many other advanced signal-processing techniques, SCF requires symbol-level synchronization and block-level synchronization. Such synchronizations can be achieved by leveraging some recent work on coherent transmission (see, e.g., [68]). In addition, the requirement of block-level synchronization can be partly relaxed by using the technique developed in [69]. Finally, channel estimation is another important step of implementing SCF. We believe that it can be achieved through a combination of conventional channel estimation with blind C&F technique proposed in [70], which is beyond the scope of this dissertation. For more details regarding the implementation aspects of physical layer network coding techniques see [71–74].

Chapter 3

SLOTTED ALOHA WITH ALL-OR-NOTHING MULTI-PACKET RECEPTION

In this chapter, we take a step forward towards the decade-long open problem of characterizing the stability region of slotted ALOHA with general symmetric MPR model. Building on the work by Bordenave *et. al.* [29] and by exploiting the mean-field technique, we derive an accurate approximation for the stability region of slotted ALOHA in inhomogeneous networks where users' transmission probabilities and arrival rates are not necessarily identical and belong to V different classes. It is shown that the approximate stability region is exact asymptotically when the number of users tends to infinity. In finite regime case, the accuracy is verified through simulations and the notion of homogeneous direction [29]. As applications of our approximate stability region for MPR-capable slotted ALOHA systems, simple expressions are derived for network throughput, service delay and packet delay. Finally, meta-stability of such systems is discussed and a physical-layer design criterion is obtained to avoid meta-stability so that the global stability of the system is guaranteed.

3.1 Motivation

Although the MPR models we discussed in Chapter 2 have attracted considerable research attention over the past decades, some of their fundamental properties are still not well understood. For example, the stability region of slotted ALOHA with MPR capability is largely open in the literature except for a few special cases summarized in Table 3.1, and the delay analysis of slotted ALOHA with MPR is mostly focused on the scenario where all the users have the same packet arrival rate (also referred to as the symmetric-arrival scenario). An in-depth understanding of the stability region and delay performance of these

Table 3.1: Summary of three special cases for which the stability regions are known.

MPR Capability	Packet Arrivals	System Size	Buffer Size	Reference
symmetric	asymmetric	∞	1	[9]
symmetric	symmetric	finite users	∞	[33]
asymmetric	asymmetric	two users	∞	[33]

MPR channel models is of particular importance nowadays, since these MPR models capture the effects of multiuser detection and space-time processing that have been widely deployed in emerging generation of IEEE 802.11 protocols such as 802.11ac [5].

3.2 System Model

We consider a random-access network as illustrated in Fig. 3.1, where N mobile users contend to transmit packets to an access point (AP). We assume that time is slotted and a slot duration equals to the packet transmission time. Each user belongs to one of the V possible classes $\mathcal{V} = \{1, \cdots, V\}$, where all users in the same class have *identical* arrival rate and transmission probability that will be defined shortly. The number of class-v users is denoted by N_v. Clearly, we have $N = \sum_{v=1}^{V} N_v$. This V-class model captures user heterogeneity in real-world wireless networks, corresponding to the so-called asymmetric-arrival scenario.

Each user is equipped with an infinite buffer for storing packets in a FIFO manner. Packets arrive at the buffer of a class-v user according to a Bernoulli process with rate $\lambda_v^{(N)}$. That is, at each time slot, a new packet arrives into the buffer of a class-v user with probability $\lambda_v^{(N)}$. The arrival processes are assumed to be independent across users. If a class-v user's buffer is nonempty at the beginning of a time slot, it transmits a packet with probability $p_v^{(N)}$ in that slot[1].

At the beginning of each slot, if user i's buffer is nonempty, it transmits a packet with probability $p_i^{(N)}$ (independent of the behavior of other users in the system). At the end of

[1]We drop the superscript (N) for notational convenience as necessary.

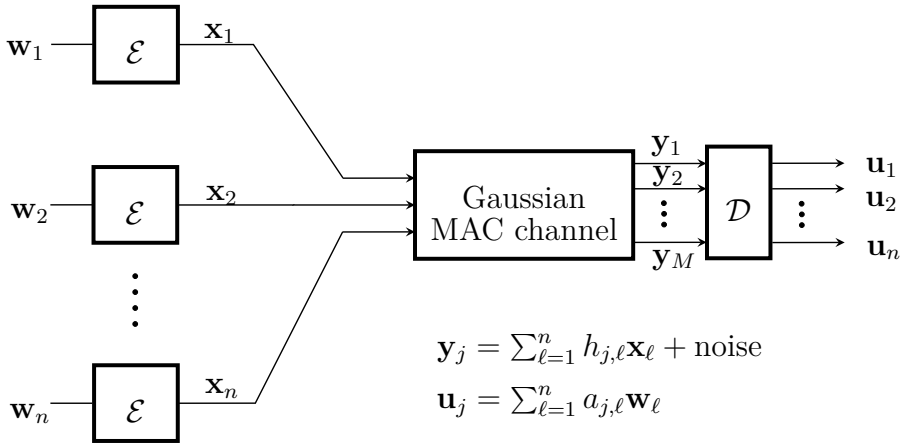

Figure 3.1: A random access wireless channel where n single-antenna active users transmit their packets simultaneously to an M-antenna AP with the MPR capability K.

each slot, the AP tries to recover the transmitted packets using an MPR technique, e.g. C&F, SCF or SIC as explained in Chapter 2. In this chapter, we focus on the all-or-nothing symmetric MPR model which is discussed in details in Chapter 2.

3.2.1 System State

Let us fix N. Denote by $X_i^{(N)}(m)$ the number of packets in user i's buffer (queue) at the beginning of slot m, where $i \in \{1, \ldots, N\}$. The state of the system at slot m is given by

$$\boldsymbol{X}^{(N)}(m) \triangleq \left(X_1^{(N)}(m), \ldots, X_N^{(N)}(m)\right) .$$

Clearly, $\left\{\boldsymbol{X}^{(N)}(m) : m = 0, 1, \ldots\right\}$ is an N-dimensional discrete-time Markov chain.

Note that the evolution of the system is class symmetric, in the sense that it does not depend on the specific identity of users besides their classes and queue lengths. Hence, we can use a new discrete-time Markov chain $\boldsymbol{Q}_v^{(N)}(m) \triangleq \left(Q_{v,0}^{(N)}(m), Q_{v,1}^{(N)}(m), \cdots\right)$ to describe the system evolution, where $Q_{v,\ell}^{(N)}(m)$ is the fraction of users from class-v with at least ℓ

packets in their buffers at slot m. That is

$$Q_{v,\ell}^{(N)}(m) = \frac{1}{N}\sum_{i=1}^{N} 1_{\left\{X_i^{(N)}(m)\geq \ell\right\}},$$

where $1_{\{.\}}$ is the indicator function.

$$W_{v,\ell}^{(N)}(m) \triangleq \sum_{i=\ell}^{\infty} Q_{v,i}^{(N)}(m)\,.$$

While the $\left\{\boldsymbol{X}^{(N)}(m)\right\}$ representation has a more intuitive interpretation, the $\left\{Q_{v,\ell}^{(N)}(m)\right\}$ and $\left\{W_{v,\ell}^{(N)}(m)\right\}$ representations are more convenient to cope with, especially in deriving stability regions. For this reason, we will use the $\left\{Q_{v,\ell}^{(N)}(m)\right\}$ and $\left\{W_{v,\ell}^{(N)}(m)\right\}$ representations in the rest of this chapter.

3.2.2 Stability Region

If the Markov chain $\left\{Q_{v,\ell}^{(N)}(m)\right\}$ has a stationary distribution for the arrival-rate vector $\boldsymbol{\lambda}^V \triangleq (\lambda_1^{(N)}, \ldots, \lambda_V^{(N)})$, we say that the system is *stable* for $\boldsymbol{\lambda}^V$. The *stability region* Λ^V is defined as the set of vectors $\boldsymbol{\lambda}^V$ such that the system is stable for $\boldsymbol{\lambda}^V$.

3.2.3 Notation

The following notations and definitions will be used throughout this chapter. Let $\mathbb{Z}_+$ be the set of non-negative integers. Define

$$\mathcal{Q} \triangleq \left\{Q_v \in [0,1]^{\mathbb{Z}_+} : 1 = Q_{v,0} \geq Q_{v,1} \geq \cdots \geq 0\,, \forall\, v \in \mathcal{V}\right\}.$$

Denote by $\mathbb{D} \equiv \mathbb{D}([0,T];\mathbb{R}^{\mathbb{Z}_+})$ the space of right-continous-with-left-limit (RCLL) functions $\mathbf{x} : [0,T] \longmapsto \mathbb{R}^{\mathbb{Z}_+}$ associated with the metric $d(\cdot,\cdot)$

$$d(\mathbf{x},\mathbf{y}) \triangleq \|\mathbf{x}-\mathbf{y}\|_T \qquad \forall\ \mathbf{x},\mathbf{y} \in \mathbb{D}\,,$$

where $\|\mathbf{x}\|_T \triangleq \sup_{t\in[0,T]} \|\mathbf{x}(t)\|$ and $\|\mathbf{x}(t)\|^2 = \sum_{i=0}^{\infty} \frac{x_i^2(t)}{2^i}$ is weighted L_2 norm. Define

$$\mathcal{N}^V \triangleq \left\{(n_1, \cdots, n_V) \in \mathbb{Z}^V : \forall v \in \mathcal{V}, 0 \le n_v \le N_v\right\}.$$

For a positive integer K, define

$$\mathcal{N}_K^V \triangleq \left\{(n_1, \cdots, n_V) \in \mathbb{Z}^V : n_1 + \cdots + n_V \le K, \forall v \in \mathcal{V}, 0 \le n_v \le N_v\right\}.$$

When $K \ge N$, $\mathcal{N}_K^V$ is precisely $\mathcal{N}^V$. Define

$$C_n \triangleq q_{1,n} + 2q_{2,n} + \cdots + nq_{n,n}\,,$$

i.e., the expected number of packets correctly decoded at the AP.

3.3 Main Results

3.3.1 Approximate Stability Region

Define

$$p(n_1, \ldots, n_V; \boldsymbol{\rho}^{(N)}) \triangleq \prod_{v=1}^{V} \binom{N_v}{n_v} \left(\rho_v^{(N)} p_v^{(N)}\right)^{n_v} \left(1 - \rho_v^{(N)} p_v^{(N)}\right)^{N_v - n_v}.$$

$p\left(n_1, \ldots, n_V; \boldsymbol{\rho}^{(N)}\right)$ is the probability that exactly n_v class-v users are contending for the channel under the independence assumption for all $v \in \mathcal{V}$. Furthermore, $\rho_v^{(N)}$ is the non-empty (utilization) probability of a class-v user.

Definition 5(Approximate Stability Region under Symmetric MPR) The approximate stability region $\hat{\Lambda}^V$ is the region lying below one of V boundaries $\partial_u \hat{\Lambda}^V$ defined by

$$\partial_u \hat{\Lambda}^V \triangleq \left\{\boldsymbol{\lambda}^V : \exists\, \boldsymbol{\rho}^{(N)} \in \partial_u [0,1]^V \text{ such that } \forall v,\ \lambda_v^{(N)} = P_v\left(\boldsymbol{\rho}^{(N)}\right)\right\},$$

where

$$P_v\left(\boldsymbol{\rho}^{(N)}\right) = \sum_{(n_1,\cdots,n_V)\in\mathcal{N}^V} R_v(n_1,\cdots,n_V)p\left(n_1,\ldots,n_V;\boldsymbol{\rho}^{(N)}\right), \tag{3.1}$$

where

$$R_v(n_1,\cdots,n_V) = \frac{1}{N_v}\sum_{0\le k_v\le n_v} k_v \sum_{\substack{0\le k_u\le n_u\\ \forall u\in\mathcal{V}\setminus\{v\}}} q_{(k_1+\cdots+k_V),(n_1+\cdots+n_V)} \tag{3.2}$$

is the average throughput of a user assuming that $(n_1,\cdots,n_V)$ packets are transmitted.

Definition 6(Approximate Stability Region under Asymmetric MPR) The approximate stability region $\hat{\Lambda}^V$ is the region lying below one of V boundaries $\partial_u\hat{\Lambda}^V$ defined by

$$\partial_u\hat{\Lambda}^V \triangleq \left\{\boldsymbol{\lambda}^V : \exists\,\boldsymbol{\rho}^{(N)}\in\partial_u[0,1]^V \text{ such that } \forall v,\ \lambda_v^{(N)} = P_v\left(\boldsymbol{\rho}^{(N)}\right)\right\},$$

where

$$P_v\left(\boldsymbol{\rho}^{(N)}\right) = \sum_{(n_1,\cdots,n_V)\in\mathcal{N}^V} R_v(n_1,\ldots,n_V)p(n_1,\ldots,n_V;\boldsymbol{\rho}^{(N)}). \tag{3.3}$$

Similarly

$$R_v(n_1,\cdots,n_V) = \frac{1}{N_v}\sum_{1\le k_v\le n_v} k_v \sum_{\substack{0\le k_u\le n_u\\ \forall u\in\mathcal{V}\setminus\{v\}}} q_{(k_1,\cdots,k_V),(n_1,\cdots,n_V)} \tag{3.4}$$

is the average throughput of a class-v user assuming that $(n_1,\cdots,n_V)$ packets are transmitted.

Interestingly, our approximate stability region $\hat{\Lambda}^V$ turns out to be *exact* for the two-user case and the symmetric-MPR symmetric-arrival case discussed extensively in [33].

Example 5(Two-User Case) When $V=2$ and $N_1=N_2=1$, the boundaries $\partial_v\hat{\Lambda}^2$ in

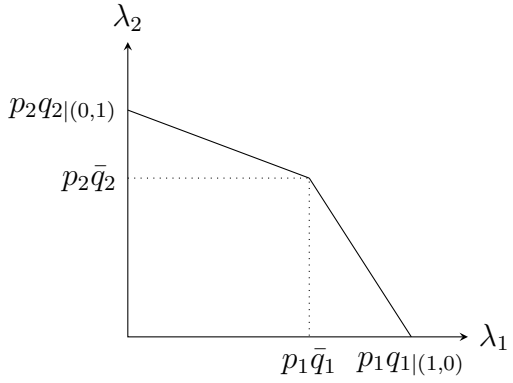

Figure 3.2: Stability region for the two-user case with a fixed transmission probability vector (p_1, p_2).

Definition 6 are the line segments given by

$$\partial_1\hat{\Lambda}^2 = \left\{(q_{1|(1,0)}p_1, 0) + \left((q_{1|(1,1)} - q_{1|(1,0)})p_1, \bar{q}_2\right) p_2\rho : \rho \in [0,1]\right\},$$
$$\partial_2\hat{\Lambda}^2 = \left\{(0, q_{2|(0,1)}p_2) + \left(\bar{q}_1, (q_{2|(1,1)} - q_{2|(0,1)})p_2\right) p_1\rho : \rho \in [0,1]\right\},$$

where $\bar{q}_1 \triangleq q_{1|(1,0)}(1-p_2) + q_{1|(1,1)}p_2$ and $\bar{q}_2 \triangleq q_{2|(0,1)}(1-p_1) + q_{2|(1,1)}p_1$. Clearly, $\hat{\Lambda}^2$ is a polygon with four vertices $(0,0)$, $(p_1q_{1|(1,0)}, 0)$, $(p_1\bar{q}_1, p_2\bar{q}_2)$, and $(0, p_2q_{2|(0,1)})$, as illustrated in Fig. 3.2. This is precisely the stability region Λ^2 derived in [33].

Example 6(Homogeneous Case) When $V = 1$, the boundary in Definition 5 is a single point

$$\frac{1}{N}\sum_{n=1}^{N} C_n \binom{N}{n} p^n (1-p)^{N-n}.$$

This is precisely the stability region for the homogeneous case [33].

3.3.2 Interpretation of the Approximate Stability Region

The main idea behind the mean-field approximation is that users' queues evolve independently of each other. It further assumes that the queue-length evolution processes of the same class are statistically identical. We justify the validity of these assumptions later in

this chapter. Under these assumptions, the queue evolution of a class-v user is a discrete-time random process with packet arrival rate $\lambda_v^{(N)}$ and utilization probability $\rho_v^{(N)}$, i.e., the probability that the queue is non-empty in a limiting time slot. Then, the average throughput (i.e., the average number of packets transmitted successfully in a time slot) of a class-v user is given by (3.1).

The approximate stability region is then formed by the union of the surfaces $\partial_v \hat{\Lambda}^V$ for all $v \in \mathcal{V}$. In fact $\partial_v \hat{\Lambda}^V$ corresponds to all the arrival rate vectors for which $\rho_v^{(N)} = 1$ and thus the class-v users are saturated.

3.3.3 Asymptotic Exactness

We study a sequence of systems indexed by N with arrival rates $\boldsymbol{\lambda}^N = \left(\lambda_1^{(N)}, \cdots, \lambda_V^{(N)}\right)$ and transmission probabilities $\boldsymbol{p}^N = \left(p_1^{(N)}, \cdots, p_V^{(N)}\right)$ such that

$$N\lambda_v^{(N)} \to \lambda_v \quad \text{and} \quad Np_v^{(N)} \to p_v \qquad \text{as} \quad N \to \infty\,.$$

The $1/N$ scaling is necessary to ensure a non-zero throughput per user. For any fixed N, we let N_v be the number of class-v users and assume that

$$\frac{N_v}{N} \to \beta_v, \text{ as } N \to \infty\,.$$

That is, the fraction of class-v users converges to β_v. Define $\gamma_v \triangleq \beta_v \rho_v p_v$ and $\gamma \triangleq \sum_v \gamma_v$. The asymptotic stability region as $N \to \infty$ has the following form.

Definition 7 The asymptotic stability region Λ^V is the region lying below one of V boundaries $\partial_u \Lambda^V$ defined by

$$\partial_u \Lambda^V \triangleq \left\{\boldsymbol{\lambda}^V : \exists\, \boldsymbol{\rho} \in \partial_u [0,1]^V \text{ such that } \forall v,\ \lambda_v = \bar{\lambda}_v(\boldsymbol{\rho})\right\},$$

where

$$\bar{\lambda}_v(\boldsymbol{\rho}) = e^{-\gamma} \sum_{(n_1,\cdots,n_V)\in\mathcal{N}^V} \prod_{u=1}^{V} \frac{\gamma_u^{n_u}}{n_u!} \bar{R}_v(n_1,\cdots,n_V)\,. \tag{3.5}$$

where

$$\bar{R}_v(n_1,\cdots,n_V) = \frac{1}{\beta_v} \sum_{1\le k_v\le n_v} k_v \sum_{\substack{0\le k_u\le n_u \\ \forall u\in\mathcal{V}\setminus\{v\}}} q_{(k_1,\cdots,k_V),(n_1,\cdots,n_V)}\,.$$

We will show that the asymptotic stability region is exact.

For the all-or-nothing symmetric MPR model, (3.5) simplifies to

$$\bar{\lambda}_v(\boldsymbol{\rho}) = \rho_v p_v \left(q_1 + \frac{q_2}{1!}\gamma + \cdots + \frac{q_K}{(K-1)!}\gamma^{K-1} \right) e^{-\gamma}\,. \tag{3.6}$$

Define

$$\chi(\gamma) \triangleq q_1 + \frac{q_2}{1!}\gamma + \frac{q_3}{2!}\gamma^2 + \cdots + \frac{q_K}{(K-1)!}\gamma^{K-1}, \tag{3.7}$$

We call (3.7) the *characteristic polynomial* of all-or-nothing MPR slotted ALOHA protocol.

3.4 Mean-Field Asymptotic Analysis

The dynamic of MPR-based slotted ALOHA random access depends essentially on the random interaction amongst N queues which makes the characterization of the stability region quite challenging. The *mean-field approximation* or *fluid model approximation* is a technique where a stochastic process corresponding to a system of interacting particles is analyzed in the limit when N is large. By properly scaling time and the parameters involved, it can be shown that the behavior (usually empirical measure) of the stochastic systems approaches that of a (deterministic) dynamical system, as $N \to \infty$. The solution of such a dynamical system – typically a set of ordinary differential equations (ODEs) – is called *fluid solution*. The dynamical system can then be studied in lieu of the direct analysis of the underlying complex stochastic process. When the empirical measure converges to a probability distri-

bution, the states of any finite number of queues become independent from each other and the so-called *decoupling* holds.

Mean field approach has been successfully applied in the context of communications and queueing networks [29, 53, 75–78] where the decoupling assumption is justified. We use the technique to analyze MPR-capable slotted ALOHA and characterize an accurate approximation for its stability region. Roughly speaking, this is a three-step procedure. First we derive the dynamical system corresponding to the Markov chain of the system when N approaches infinity. Following the characterization of the dynamical system, we analyze its stability behavior. Finally, we make connections between the stability conditions for this limiting system and those of the finite system for large enough N.

Recall that the empirical tail measure of user i from class-v who has $X_i^{(N)}(m)$ packets at time slot m is computed as

$$Q_{v,\ell}^{(N)}(m) = \frac{1}{N}\sum_{i=1}^{N} 1_{\left\{X_i^{(N)}(m)\geq \ell\right\}}.$$

Define the scaled empirical tail measure as $Q_{v,\ell}^{(N)}(\lfloor Nt \rfloor)$ for some $t \in \mathbb{R}_+$. Using mean-field approximation we show that

$$\lim_{N\to\infty} Q_{v,\ell}^{(N)}\left(\lfloor Nt \rfloor\right)/\beta_v \to Q_{v,\ell}(t)\,,$$

where $Q_{v,\ell}(t)$ is the proportion of class-v users with at least ℓ packets in their buffers at time t. Let $W_{v,\ell}^{(N)}(t)$ be the scaled process corresponding to $W_{v,\ell}^{(N)}(m)$ with $m = \lfloor Nt \rfloor$. We proceed by deriving a dynamical system description for $Q_{v,\ell}(t)$.

3.4.1 A Dynamical System Description for All-or-Nothing Slotted ALOHA

Denote by $\boldsymbol{\nu} = \left\{Q_v^{(N)}(t)\right\}_{v\in\mathcal{V}}$ the empirical measure of the slotted ALOHA system at time slot $\lfloor Nt \rfloor$. At time slot $\lfloor Nt \rfloor$, the possible state transitions for user i are a packet arrival into, or a packet departure from its queue. So, if user i is in the state (v, ℓ) then the probability

of transition to the state $(v, \ell+1)$ is[2]

$$\lambda_v^{(N)}(\boldsymbol{\nu})/N = \lambda_v/N + o(1/N)\,. \tag{3.8}$$

For large enough N we have

$$\bar{\lambda}_v(\boldsymbol{\nu}) = \lambda_v\,. \tag{3.9}$$

Next we compute the departure probability of a queue. Recall that the AP – equipped with M antennas – can resolve up to K collisions. Hence, the state of user i becomes $(v, \ell-1)$ only if user i transmits a packet *or* user i and no more than $K-1$ other users transmit simultaneously. We now write the departure probability for user i from class v as follows:

$$\begin{aligned}
\mu_{v,\ell}^{(N)}(\boldsymbol{\nu})/N = 1_{\{\ell>0\}} p_i^{(N)} \Bigg(& q_1 \prod_{j\neq i} \left(1-p_j^{(N)}\right)^{1_{\left\{X_j^{(N)}>0\right\}}} \\
& + \sum_{j_1\neq i} q_2 1_{\left\{X_{j_1}^{(N)}>0\right\}} p_{j_1}^{(N)} \prod_{j\neq i,j_1} \left(1-p_j^{(N)}\right)^{1_{\left\{X_j^{(N)}>0\right\}}} + \cdots \\
& + \sum_{\substack{j_1<j_2<\cdots<j_{K-1} \\ j_1,\cdots,j_{K-1}\neq i}} q_K \prod_{k=1}^{K} 1_{\left\{X_{j_k}^{(N)}>0\right\}} p_{j_k}^{(N)} \prod_{j\neq j_1,\cdots,j_{K-1}} \left(1-p_j^{(N)}\right)^{1_{\left\{X_j^{(N)}>0\right\}}} + o(1/N) \Bigg)
\end{aligned}$$

$$\begin{aligned}
= 1_{\{\ell>0\}} p_i^{(N)} \prod_{j\neq i} \left(1-p_j^{(N)}\right)^{1_{\left\{X_j^{(N)}>0\right\}}} & \left(q_1 + q_2 \sum_{j_1\neq i} \frac{p_{j_1}^{(N)}}{1-p_{j_1}^{(N)}} 1_{\left\{X_j^{(N)}>0\right\}} + \cdots \right. \\
\left. + q_K \sum_{\substack{j_1<j_2<\cdots<j_{K-1} \\ j_1,\cdots,j_{K-1}\neq i}} \prod_{k=1}^{K} \frac{p_{j_k}^{(N)}}{1-p_{j_k}^{(N)}} 1_{\left\{X_{j_k}^{(N)}>0\right\}} + o(1/N) \right)
\end{aligned}$$

[2] $g(N) = o(f(N))$ means $\frac{g(N)}{f(N)} = 0$ as $N \to \infty$.

$$= 1_{\{\ell>0\}} \frac{p_v/N}{1-p_v/N} \prod_{u=1}^{V} (1-p_u/N)^{N\beta_u^{(N)} Q_{u,1}^{(N)}}$$
$$\left(q_1 + q_2 \sum_{u=1}^{V} \binom{N\beta_u^{(N)} Q_{u,1}^{(N)}}{1} \frac{p_u/N}{1-p_u/N} + \cdots \right.$$
$$\left. + q_K \sum_{n_1+\cdots+n_V=K-1} \prod_{u=1}^{V} \binom{N\beta_u^{(N)} Q_{u,1}^{(N)}}{n_u} \left(\frac{p_u/N}{1-p_u/N} \right)^{n_u} + o(1/N) \right),$$

where $\beta_v^{(N)} \triangleq N_v/N$ and $Q_{v,1}^{(N)}$ is the proportion of class-v users with non-empty queues. Let β_v and $Q_{v,1}$ be the limiting values corresponding to $\beta_v^{(N)}$ and $Q_{v,1}^{(N)}$ for class-v users when N is large. Let

$$\gamma(t) = \sum_v \beta_v p_v Q_{v,1}(t). \tag{3.10}$$

For large enough N we get

$$\bar{\mu}_{v,\ell}(\boldsymbol{\nu}) = 1_{\{\ell>0\}} p_v \left(q_1 + \frac{q_2}{1!}\gamma(t) + \frac{q_3}{2!}\gamma^2(t) + \cdots + \frac{q_K}{(K-1)!}\gamma^{K-1}(t) \right) e^{-\gamma(t)}. \tag{3.11}$$

With the transition probabilities at hand and from Theorem 3.4.3, we obtain the following fluid model as a deterministic description of the limiting slotted ALOHA system with symmetric MPR channel when $N \to \infty$.

$$\begin{aligned} Q_{v,\ell}(0) &= Q_{v,\ell}^0 \quad \forall v \in \mathcal{V}, \ \ \ell \geq 1, \\ \frac{\mathrm{d}}{\mathrm{d}t} Q_{v,\ell}(t) &= \lambda_v \Big(Q_{v,\ell-1}(t) - Q_{v,\ell}(t) \Big) + p_v \chi(\gamma(t)) e^{-\gamma(t)} \Big(Q_{v,\ell+1}(t) - Q_{v,\ell}(t) \Big). \end{aligned} \tag{3.12}$$

Note that $\gamma(t)$ can be interpreted as an average attempt probability of the network. We further observe that the dynamical system (3.12) can be viewed as the Kolmogorov equation of a single server queue with state-dependent (through $\gamma(t)$) service rate

$$p_v \chi(\gamma(t)) e^{-\gamma(t)}$$

and arrival rate λ_v. The solution to the fluid model can then be used to approximate the sample paths of $\boldsymbol{Q}_v^{(N)}(t)$ for sufficiently large values of N. There are two main drift terms in (3.12) which we provide intuitions for each before rigorously establishing the validity of the mean-field approximation.

- $\lambda_v\Big(Q_{v,\ell-1}(t) - Q_{v,\ell}(t)\Big)$: This term corresponds to the arrival process. Upon arrival of a packet to a class-v queue with $\ell-1$ packets, the number of class-v queues with at least ℓ packets increases by one. Hence, the value of the process $Q_{v,\ell}^{(N)}$ increases by $1/N$ if and only if there is an arrival to a class-v queue with exactly $\ell-1$ packets. The number of class-v queues with $\ell-1$ packets is $N\left(Q_{v,\ell-1}^{(N)}(\lfloor Nt\rfloor) - Q_{v,\ell}^{(N)}(\lfloor Nt\rfloor)\right)$ and the arrival rate of each user is λ_v/N. Taking N to infinity, the increment of $Q_{v,\ell}(t)$ due to an arrival will be $\lambda_v\Big(Q_{v,\ell-1}(t) - Q_{v,\ell}(t)\Big)$.

- $p_v\chi(\gamma(t))e^{-\gamma(t)}\Big(Q_{v,\ell}(t) - Q_{v,\ell+1}(t)\Big)$: This term represents a successful departure of a packet from a class-v user. The argument is analogous to that of the first term.

We observe that, the fluid model is essentially comprised of an ODE for each class $v \in \mathcal{V}$. However, these ODEs are not independent of each other and are coupled through the variable $\gamma(t)$ which in turn is a function of the non-empty probabilities of all classes, i.e., $Q_{v,1}(t)$ for all $v \in \mathcal{V}$.

From (3.12), it is easy to see that the following ODE describes the evolution of a class-v queue workload which is defined as $W_v(t) = W_{v,1}(t)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}W_v(t) = \lambda_v - p_v\chi(\gamma(t))e^{-\gamma(t)}Q_{v,1}(t) \quad \forall v \in \mathcal{V}\,. \tag{3.13}$$

Multiplying (3.13) by β_v and summing over all classes we get the total workload as

$$\frac{\mathrm{d}}{\mathrm{d}t}W(t) = \sum_v \beta_v\lambda_v - \gamma(t)\chi(\gamma(t))e^{-\gamma(t)}. \tag{3.14}$$

The dynamics of workload evolution (3.14) will be used in following section to analyze the

stability of the limiting system (3.12). Define $\zeta(\gamma) \triangleq \gamma\chi(\gamma)e^{-\gamma}$. From (3.14) the workload drift function is defined as

$$\Delta(\gamma) \triangleq \sum_v \beta_v \lambda_v - \zeta(\gamma)\,. \tag{3.15}$$

To make the analysis tractable, we assume that $\zeta(\gamma)$ is a unimodal function.

Proposition 3.4.1 Let γ^* be the maximizer of $\zeta(\gamma)$. If $\sum_v \beta_v\lambda_v < \zeta(\gamma^*)$ then the workload drift function $\Delta(\gamma)$ admits invariant states.

Proof. When $\zeta(\gamma)$ is a unimodal function, it has exactly one maximum. In such a case, the line $\sum_v \beta_v\lambda_v$ and $\zeta(\gamma)$ intersect at two distinct points.

□

3.4.2 Mean-Field Convergence for All-or-Nothing Slotted ALOHA

The justification of mean-field approximation or fluid model is the main focus of the remainder of this chapter. Throughout a series of theorems we show that the process $Q_{v,\ell}^{(N)}(t)$ converges to the invariant solution of the fluid model (3.12). Fig. 3.3 summarizes the relationships between the convergence of $Q_{v,\ell}^{(N)}(t)$ to the solution of the fluid model over a finite time horizon and then the convergence of the fluid solution to the fluid limit, i.e., invariant solution (Theorem 3.4.3 and Theorem 3.4.2). It also shows the convergence of the sequence of steady state distributions (Theorem 3.4.5 and Theorem 3.4.4). Combining Theorem 3.4.3 (convergence to fluid model solution) and Theorem 3.4.2 (convergence to fluid limit) implies that the finite system process $Q_{v,\ell}^{(N)}$ can be approximated by the fluid limit over a finite time horizon. That is

$$\lim_{t\to\infty}\lim_{N\to\infty} \boldsymbol{Q}_v^{(N)}(t) = \boldsymbol{Q}_v^I \quad \forall\, v \in \mathcal{V} \qquad \text{in distribution}$$

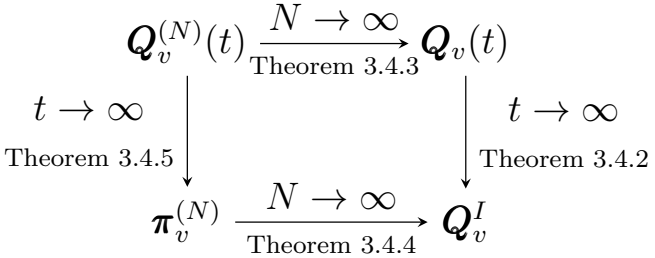

Figure 3.3: Mean field convergence diagram: for large N the stochastic process $\boldsymbol{Q}_v^{(N)}(t)$ converges to the fluid model solution $\boldsymbol{Q}_v(t)$ which in turn converges to the fluid limit $\boldsymbol{Q}_v^I$ in the steady state. Also $\boldsymbol{Q}_v^{(N)}(t)$ admits a stationary distribution $\boldsymbol{\pi}_v^{(N)}$ and the sequence of such stationary distributions is convergent to $\boldsymbol{Q}_v^I$.

In general, we cannot switch the order of limits over t and N. However, we show that for our problem the following relation holds.

$$\lim_{N\to\infty}\lim_{t\to\infty} \boldsymbol{Q}_v^{(N)}(t) = \boldsymbol{Q}_v^I \quad \forall\, v \in \mathcal{V} \qquad \text{in distribution} \tag{3.16}$$

In other words, (3.16) implies that the sequence of steady state distributions – if they exist – concentrates around the fluid limit in the limit when N grows large. Theorem 3.4.4 shows the existence of the steady state distributions and in Theorem 3.4.3 we prove their convergence to the fluid limit.

We start from the dynamical system (3.12) by deriving the conditions that guarantee its globally stability. Let $q \triangleq \{q_i\}_{i=1}^K$. Furthermore, let $\underline{\gamma}(\lambda, q)$ and $\overline{\gamma}(\lambda, q)$ be the distinct roots of the equation $\Delta(\gamma) = 0$ in $(0, \gamma^*)$ and $(\gamma^*, +\infty)$. Finally, let $\underline{\boldsymbol{Q}}_v$ and $\overline{\boldsymbol{Q}}_v$ be the corresponding tail distributions, respectively. We then have the following result for the stability of the limiting system.

Theorem 3.4.2(Global Stability of Limiting System (3.12)) The necessary and sufficient conditions for the global stability of the dynamical system representation of slotted

ALOHA with symmetric MPR are

$$\text{(i)} \quad \sum_v \beta_v \lambda_v < \gamma^* \chi(\gamma^*) e^{-\gamma^*} \tag{3.17}$$

$$\text{(ii)} \quad \sum_v \beta_v p_v < \overline{\gamma}(\lambda, q) \tag{3.18}$$

$$\text{(iii)} \quad \lambda_v < p_v \chi(\underline{\gamma}(\lambda, q)) e^{-\underline{\gamma}(\lambda, q)} \quad \forall v \in \mathcal{V}\,. \tag{3.19}$$

Proof. See Appendix A.2. □

Remark 1 Note that if any of the above conditions are violated the system is either unstable or not *globally* stable. For instance, if (3.17) is violated the system never converges as $\frac{\mathrm{d}}{\mathrm{d}t} W(t) > 0$ and total workload becomes unbounded. If $\sum_v \beta_v p_v > \overline{\gamma}(\lambda, q)$ then $\gamma(\lambda, q)$ may converge to either $\underline{\gamma}(\lambda, q)$ or $\overline{\gamma}(\lambda, q)$ and the system is not globally stable. $\underline{\gamma}(\lambda, q)$ and $\overline{\gamma}(\lambda, q)$ can be achieved if the queues are initially empty or initialized by the stationary probability for the queue-lengths, respectively. Finally, if (3.19) is not satisfied for some $v \in \mathcal{V}$, then the corresponding queue-length evolution and thus the system is unstable.

Remark 2 In the proof of the sufficient and necessary conditions in Theorem 3.4.2 we assume that $\zeta(\gamma)$ has a unique maximum. This assumption is essential in making it possible for us to derive a simple representation of the stability region. In addition, as we will see in Section 3.5 this assumption guarantees that the system will not be metastable and gives a design criteria to avoid metastability.

Theorem 3.4.3(Convergence in the Transient Regime) Consider a sequence of random access systems as the number of queues N tends to infinity. Assume that for any $v \in \mathcal{V}$, $\boldsymbol{Q}_v^{(N)}(0) \to \boldsymbol{Q}_v^0$ in probability. Then for any finite time horizon $T < \infty$ we have

$$\lim_{N \to \infty} \mathbb{P}\left(\left\| \boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t) \right\|_T > \epsilon \right) = 0 \quad \forall\, \epsilon > 0\,,$$

where $\boldsymbol{Q}_v(t)$ is the unique solution of the dynamical system, i.e., fluid solution.

Proof. See Appendix A.5. □

The steady state distribution of the sequence of systems is a point mass distribution concentrated at the fixed point of the dynamical system.

Theorem 3.4.4(Convergence in the Stationary Regime) Let $\left\{\boldsymbol{\pi}_v^{(N)}\right\}_{v\in\mathcal{V}}$ be the steady state distributions (invarient measures) of the process $\left\{\boldsymbol{Q}_v^{(N)}(t)\right\}$. Assume that $\left\{\boldsymbol{\pi}_v^{(N)}\right\}_{v\in\mathcal{V}}$ is tight. Then for all $v \in \mathcal{V}$ the sequence $\boldsymbol{\pi}_v^{(N)}$ converges in distribution to $\boldsymbol{Q}_v^I$, the fixed point of the dynamical system.

Proof. See Appendix A.6. □

Theorem 3.4.5(Asymptotic Exactness) Assume that $\sum_{v\in\mathcal{V}} \beta_v p_v < 1$. Fix any sufficiently small $\epsilon > 0$. There exists some N_ϵ such that for all $N > N_\epsilon$:

(i) if $\boldsymbol{\lambda}^N + \epsilon \mathbf{1}^N \in \hat{\Lambda}^N$, then $\boldsymbol{\lambda}^N \in \Lambda^N$;

(ii) if $\boldsymbol{\lambda}^N - \epsilon \mathbf{1}^N \notin \hat{\Lambda}^N$, then $\boldsymbol{\lambda}^N \notin \Lambda^N$.

and the Markov process $\left\{\boldsymbol{Q}_v^{(N)}(m)\right\}_{m\in\mathbb{N}}$ is positive recurrent with the invariant distribution $\left\{\boldsymbol{\pi}_v^{(N)}\right\}_{v\in\mathcal{V}}$. Moreover, $\left\{\boldsymbol{\pi}_v^{(N)}\right\}_{v\in\mathcal{V}}$ is tight.

Proof. See Appendix A.1. □

3.5 Meta-Stability

Bistability of slotted ALOHA systems is a well-known fact [52–55]. MPR-capable slotted ALOHA systems exhibit bistability and in general are metastable, i.e, the system can have multiple stable states. A metastable Markov chain has a stationary distribution that is concentrated on a few relatively disjoint subsets of its state space. On short time scales, such a Markov chain behaves like a stationary Markov chain. In the long run, however, a metastable Markov chain only occasionally makes transitions between such stable subsets.

In order to see the bistability of slotted ALOHA systems let us look at the workload drift function (3.14). In Fig. 3.4 total (aggregate) arrival rate ($\lambda = \sum_v \beta_v \lambda_v$) and $\zeta(\gamma)$ are depicted for a 2-class network. Observe that for a given pair of input arrival rates, the

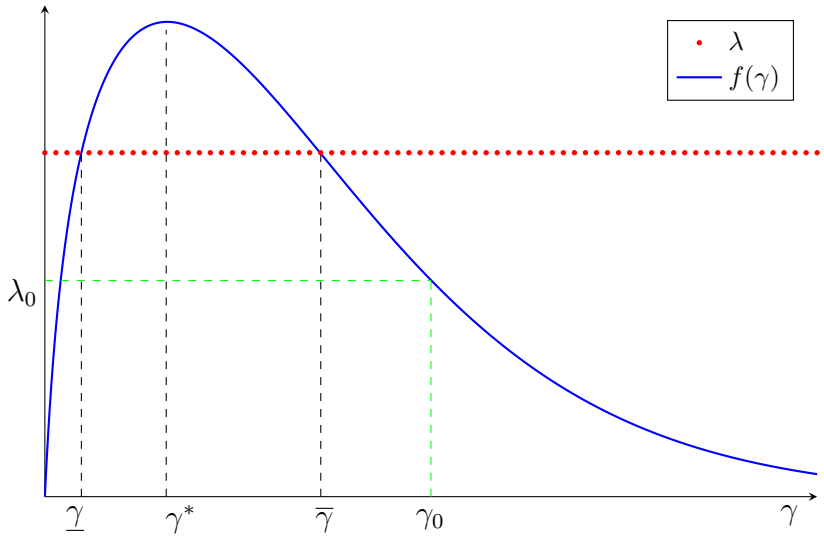

Figure 3.4: The unimodal function $f(\gamma)$ and its intersection with total input arrival rate.

equation $f(\gamma) = \lambda$ may have two, one or no solutions corresponding to a bistable, stable or unstable system as illustrated in Fig. 4.3. Let γ^* be the maximizer of $\zeta(\gamma)$. Clearly $\gamma \leq \gamma_0 \triangleq \sum_v \beta_v p_v$. When $\lambda > \lambda_0$ and $\gamma_0 > \gamma^*$ then the slotted ALOHA system can potentially become bistable. When $\lambda < \lambda_0$, the system is globally stable. Finally, if $\gamma_0 < \gamma^*$ and $\lambda > \lambda_0$ the system becomes unstable.

Next, we explain how to avoid metastability based on the approximate stability region. As pointed out in [53,79] metastability is a highly undesirable property for a communications network. With metastability, the state of a network fluctuates – over long periods of time – between different stable states. Such long oscillations make it very challenging to predict the average network performance in terms of throughput and delay. As a consequence, a certain desired quality-of-service cannot be guaranteed.

The following theorem provides a solution for inhomogeneous slotted ALOHA systems with symmetric MPR model to avoid metastability. In particular, if the MPR technique is carefully designed, then the network is proven to be globally stable, and so metastability can be completely eliminated. This leads to an important design criterion, especially when

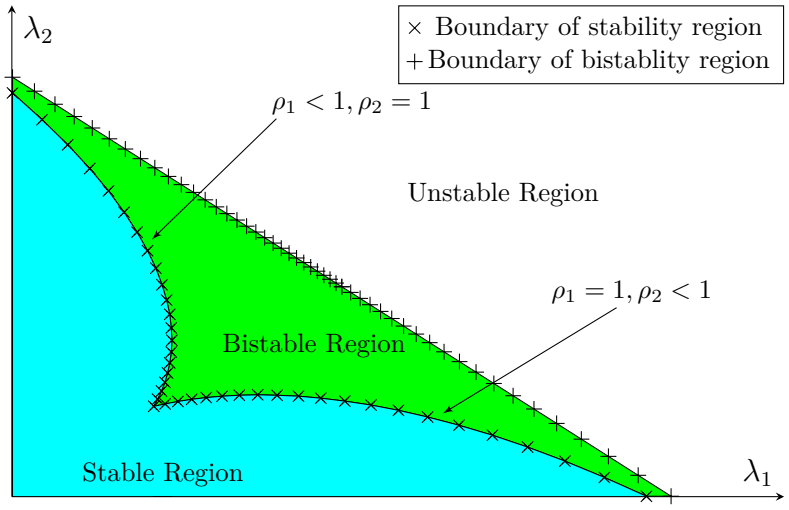

Figure 3.5: Illustration of stable, bistable and unstable regions for slotted ALOHA.

quality-of-service guarantee is in great need.

Theorem 3.5.1 In the inhomogeneous slotted ALOHA system with symmetric MPR model, metastability can be avoided if

$$\sum_{v \in \mathcal{V}} N_v p_v^{(N)} < \gamma^*$$

and

$$q_1 \leq 2q_2 \leq \cdots \leq Kq_K \,. \tag{3.20}$$

Proof. See Appendix A.3. □

3.6 Accuracy of Approximate Stability Region

We showed that the approximate stability region is asymptotically exact. For finite N, we conduct simulations to show that the approximation is indeed accurate even for $N = 3$. Moreover, we apply the notion of *homogeneous directions* proposed by [29] to show that

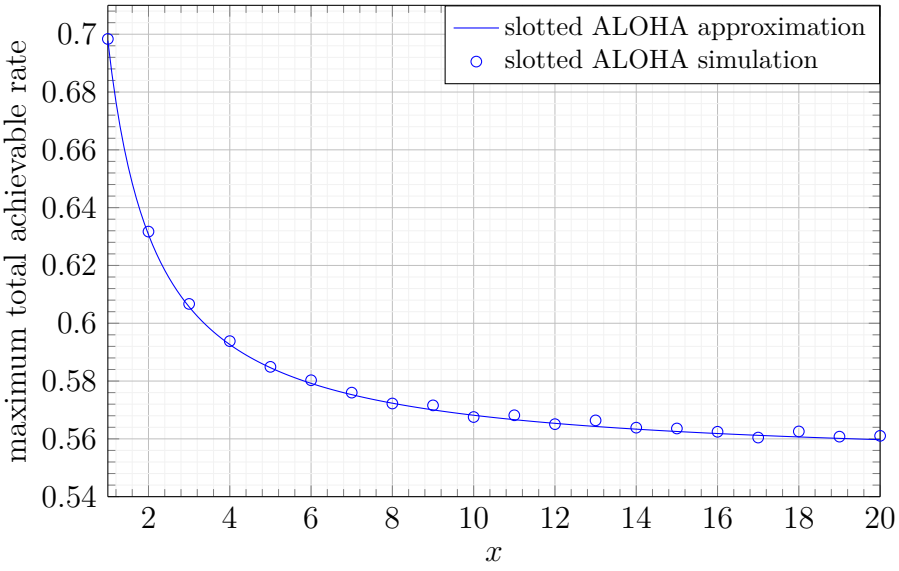

Figure 3.6: Maximum total achievable rate of a network with $N = 3$, $p_1 = p_2 = p_3 = 1/3$.

the approximate and true stability regions coincide in many scenarios where the decoupling assumption holds. We do this by way of two examples.

Example 7 Consider a network with $N = 3$ users with transmission probabilities $p_1 = 1/3$, $p_2 = 1/3$, $p_3 = 1/3$ and arrival rates $\lambda_1 = \lambda$, $\lambda_2 = \lambda/x$ and $\lambda_3 = \lambda$. Assume that x varies in the range $[1, 20]$. After some algebraic manipulation, it can be shown that the approximate stability condition for slotted ALOHA with all-or-nothing symmetric MPR is

$$\sum_{i=1}^{3} \lambda_i < \frac{4}{9}\left(q_1 + \frac{q_2}{2}\right) \frac{2x+1}{2x + \frac{q_1 - \frac{q_2}{2}}{q_1+q_2}}$$

Fig. 3.6 shows the above stability condition compared with simulation result for slotted ALOHA. We find a considerably good match between the curves confirming the accuracy of the approximation for small N.

Example 8 For the $N = 3$ user network we now assume that the transmission probabilities area $p_1 = 1/4$, $p_2 = 1/2$, $p_3 = 1/4$ and the arrival rates are set the same as those in example

11. We let x vary between 1 and 20. Let

$$x_0 = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} \frac{q_1 + q_2\left(\frac{p_1}{1-p_1} + \frac{p_2}{1-p_2}\right)}{q_1 + 2q_2\frac{p_1}{1-p_1}}\,.$$

For the transmission probabilities given above, $x_0 = \frac{q_1+4/3q_2}{3q_1+2q_2}$. It can then be shown that the approximate stability condition is

$$\sum_{i=1}^{3} \lambda_i < \begin{cases} \frac{3}{16}(q_1 + q_2/3)\frac{2x+1}{2x+\frac{q_1-2/3q_2}{3q_1+2q_2}} & \text{if } x \leq \frac{q_1+4/3q_2}{3q_1+2q_2} \\ \frac{(x+1/2)(q_2x+q_1+q_2)}{(x+1)^2} & \text{otherwise}\,. \end{cases}$$

In other words, for $x \leq x_0$ users 1 and 3 are saturated and for $x > x_0$ user 2 is saturated.

3.7 Applications

In this section, we aim to apply the approximate stability region to analyze the performance of slotted ALOHA with all-or-nothing symmetric MPR channel. Simple expression are derived for the network throughput, service delay and packet delay of MPR-capable slotted ALOHA systems.

We start with the throughput of a 2-class network. Consider a 2-class network where users in class-v ($v \in \{1, 2\}$) have the same transmission probability p_v and the same arrival rate λ_v. We assume that the AP's decoding threshold is $K = 2$. Then, the boundaries $\partial_v\hat{\Lambda}^N$ of the approximate stability region $\hat{\Lambda}^N$ can be rewritten as

$$\partial_v\hat{\Lambda}^N \triangleq \left\{(\lambda_1, \lambda_2) : \exists\, \boldsymbol{\rho} \in \partial_v[0, 1]^2 \text{ s.t. } \forall v,\ \lambda_v = R_v(\boldsymbol{\rho})\right\},$$

where

$$R_v(\boldsymbol{\rho}) = \frac{\rho_v p_v}{1 - \rho_v p_v} \prod_u (1 - \rho_u p_u)^{N_u} \left(q_1 - q_2\frac{\rho_v p_v}{1 - \rho_v p_v} + q_2 \sum_u N_u \frac{\rho_u p_u}{1 - \rho_u p_u}\right). \tag{3.21}$$

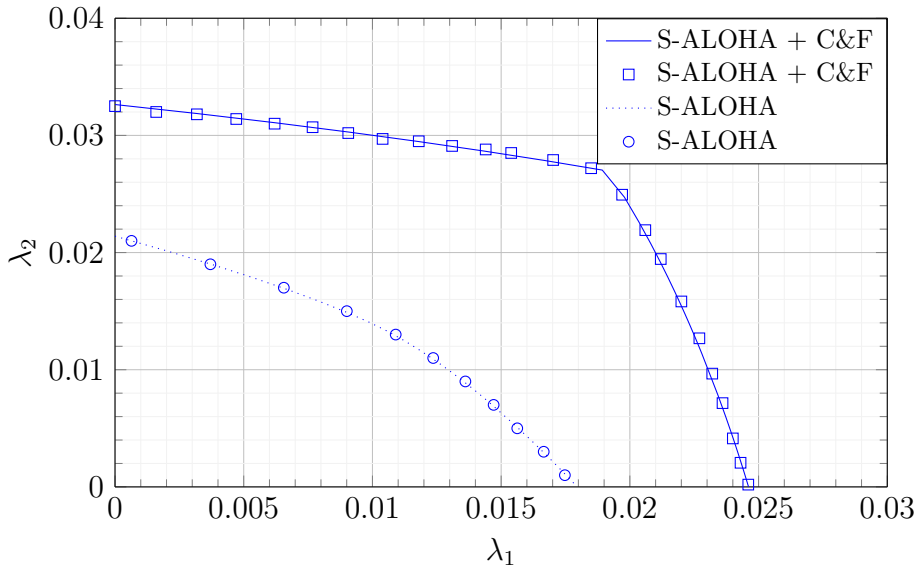

Figure 3.7: Feasible rates for transmission probabilities $p_1 = 0.028$, $p_2 = 0.039$, $q_1 = 0.96$, $q_2 = 0.89$ with $N_1 = N_2 = 15$.

3.7.1 Throughput Analysis

Here, we aim to characterize the throughput slotted ALOHA with all-or-nothing MPR model. We begin with an example where the maximum feasible rates for the two-class network with $K = 2$ is discussed. Then we present the throughput result.

Example 9 Suppose that $\lambda_1 = w\lambda_2$ for some $w > 0$. Then, for any fixed w, we can compute the (approximately) maximum feasible rates $(\lambda_1^*, \lambda_2^*)$ as follows.

Let $w_0 \triangleq R_1(1,1)/R_2(1,1)$. When $w < w_0$, class-2 users are first saturated as we increase the arrival rates (λ_1, λ_2) with $\lambda_1 = w\lambda_2$. In this case, we have $\lambda_1^* = R_1(\rho_1, 1)$ and $\lambda_2^* = \bar{R}_2(\rho_1, 1)$, where ρ_1 is the unique solution of $R_1(\rho_1, 1)/R_2(\rho_1, 1) = w$ in $(0, 1)$. Similarly, when $w \geq w_0$, class-1 users are first saturated with $\lambda_1^* = R_1(1, \rho_2)$ and $\lambda_2^* = R_2(1, \rho_2)$, where ρ_2 is the unique solution of $R_1(1, \rho_2)/R_2(1, \rho_2) = w$ in $(0, 1)$.

Fig. 3.7 depicts the region of feasible rates whose boundary corresponds to the collection of maximum feasible rates. We validate its accuracy by comparing the boundary with several

simulated maximum feasible rates. As expected, our analytical result matches the simulation very well. Using slotted ALOHA with symmetric MPR, in addition, we achieve a much larger set of feasible rates compared to conventional slotted ALOHA.

Next we derive the aggregate throughput of the general case of a V-class network.

Proposition 3.7.1 The aggregate throughput of the inhomogeneous slotted ALOHA with all-or-nothing symmetric MPR in a V-class network is given by

$$R_{\text{MPR-ALOHA}} = \sum_{v=1}^{V} N_v R_v \left(\boldsymbol{\rho}^{(N)}\right) + o(1)\,, \tag{3.22}$$

where the $o(1)$ term is understood as $N \to \infty$. In particular, the saturated throughput is given by $\sum_v N_v R_v(\mathbf{1})$, where $\mathbf{1}$ is an all-one vector of length V.

3.7.2 Delay Analysis

We now turn our attention to study the delay performance. There are two types of delays of particular interest, namely, *service delay* and *total delay* (packet delay). The service delay of a packet is defined as the time it takes for this packet to be successfully decoded at the AP after it reaches the head of the queue, i.e., expected delivery time of a head of line (HOL) packet. The total delay of a packet is defined as the time it takes for the packet to be successfully decoded after it arrives in the queue. Clearly, the total delay equals to the service delay plus the queuing delay. We begin by giving an example for a 2-class slotted ALOHA network.

Example 10 Suppose that the equation

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} R_1(\rho_1, \rho_2) \\ R_2(\rho_1, \rho_2) \end{bmatrix}$$

has a unique solution (ρ_1^*, ρ_2^*) in $(0,1) \times (0,1)$. Then ρ_v^* approximates the probability that a class-v user has a non-empty buffer in the steady state. Hence, for a class-v user with a

non-empty buffer, the success probability $p_{\text{succ},v}$ of sending a packet can be approximated as

$$p_v^{\text{SUCC}} = \frac{p_v}{1-\rho_v^* p_v} \prod_u (1-\rho_u^* p_u)^{N_u} \left(q_1 - q_2 \frac{\rho_v^* p_v}{1-\rho_v^* p_v} + q_2 \sum_u N_u \frac{\rho_u^* p_u}{1-\rho_u^* p_u} \right) + o(1)\,.$$

In particular, we have $p_v^{\text{SUCC}} = R_v(\rho_1^*, \rho_2^*)/\rho_v^*$. Hence, the average packet delay for a class-v user is approximately given by

$$T_{v,\text{ALOHA}} = \frac{1-\lambda_v}{p_v^{\text{SUCC}} - \lambda_v} + o(1)\,. \tag{3.23}$$

Next, we present the service delay and packet delay of slotted ALOHA with all-or-nothing symmetric MPR model for the general case of V-class networks. Denote by $D_{v,\text{ALOHA}}$ the service delay of a HOL packet of a class-v user. Now consider a V-class network and assume that the system is stable. We have the following result for the service delay.

Proposition 3.7.2 The service delay of a class-v user in the slotted ALOHA system with all-or-nothing symmetric MPR is

$$D_{v,\text{ALOHA}} = \frac{\rho_v^*}{\lambda_v} + o(1) \quad \text{for} \ \ v \in \mathcal{V}\,. \tag{3.24}$$

Proof. Suppose that $\rho_v^* \in (0,1)$ for all $v \in \mathcal{V}$ is the unique solution to $\lambda_v = R_v(\boldsymbol{\rho})$. A class-$v$ user is then stable with the arrival rate λ_v and we can apply Little's law to obtain

$$\rho_v^* = \lambda_v D_{v,\text{ALOHA}} + o(1)\,.$$

This proves (3.24). □

Now let $T_{v,\text{ALOHA}}$ be the total delay experienced by a packet in a slotted ALOHA system with all-or-nothing symmetric MPR channel. Total delay is characterized in the following theorem.

Theorem 3.7.3 The total delay of a class-v packet in a slotted ALOHA system with

all-or-nothing symmetric MPR model is

$$T_{v,\text{ALOHA}} = \frac{\frac{1}{\lambda_v} - 1}{\frac{1}{\rho_v^*} - 1} + o(1)\,. \tag{3.25}$$

Proof. See Appendix A.4. □

Remark 3 In order to apply Propositions 3.7.1, 3.24 and Theorem 3.7.3, we need to obtain the value of the utilization probabilities $\boldsymbol{\rho}^*$ for any given $\{\lambda_v\}$ and $\{p_v\}$ for the finite system. As we will discuss in details in chapter 4, our strategy is to approximate the utilization probabilities of the finite system with that of the limiting utilization probabilities.

Fig. 3.8 demonstrates the packet delay as the arrival rate changes in a 2-class network with $N = 30$ ($N_1 = 20$ users in class-1) at $\mathsf{SNR} = 15$dB. We see that the analytical delay very well agrees with the simulation. Note that, in contrast to the existing work, the delay performance is presented for each class separately. In addition, it can be seen that CSMA with SCF is stable for a larger range of arrival rates.

3.8 Extension to General Symmetric MPR Model

The all-or-nothing symmetric MPR model we have discussed so far assumes that $q_{k,n} = 0$ for $k < n$. This assumption can be relaxed and it turns out that the analysis of the general symmetric MPR where $q_{k,n} > 0$ for all $k \leq n$ boils down to that of all-or-nothing symmetric MPR.

Previously we saw that the average throughput per a class-v user can be computed from (3.2). We now give an alternative formula for 3.2. Suppose that there are n concurrent transmissions and n_v of them belong to class-v. Let

$$\bar{q}_n \triangleq \frac{C_n}{n}\,,$$

which is the success probability of a given active user. The average throughput per class is

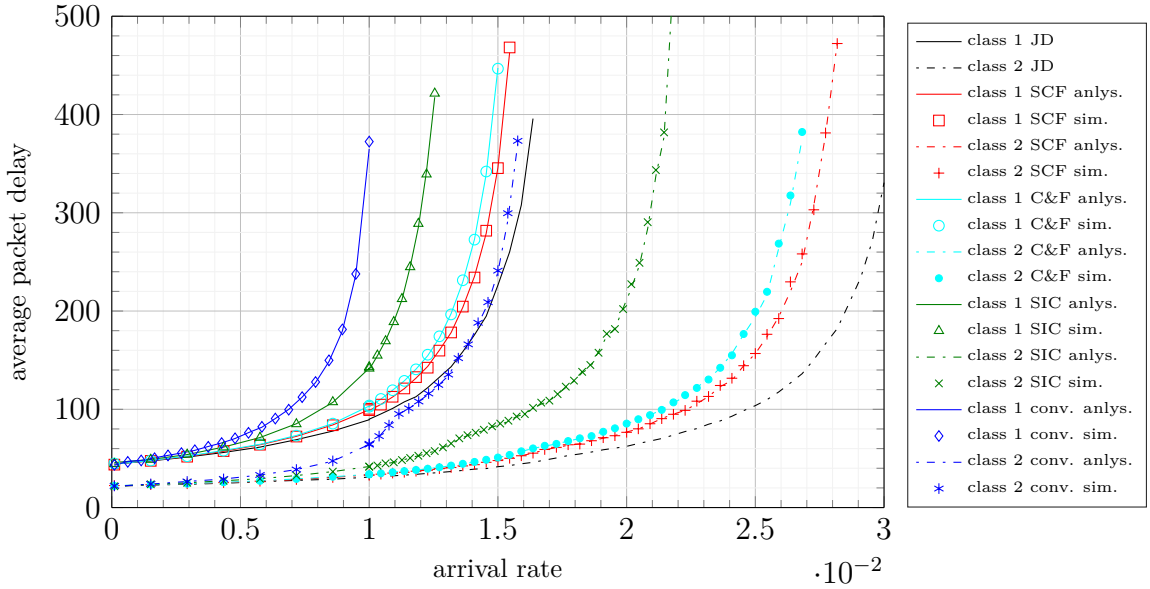

Figure 3.8: Average total delay (packet delay) performance of inhomogeneous persistent CSMA schemes for a 2-class network at $\mathsf{SNR} = 15\,\mathrm{dB}$ with message rate $= 1$, $N_1 = 20$, $N_2 = 10$, $p_1 = 1/40$, $p_2 = 1/20$, $K = 1$.

$C_n n_v/n$. Hence the average throughput per class-v user is

$$R_v(n_1, \cdots, n_v) = \frac{C_n}{n}\frac{n_v}{N_v} = \bar{q}_L \frac{n_v}{N_v}\,. \tag{3.26}$$

For the all-or-nothing MPR model 3.2 simplifies to

$$R_v(n_1, \cdots, n_V) = q_n \frac{n_v}{N_v}\,. \tag{3.27}$$

Comparing (3.26) with (3.26), it becomes evident that general symmetric model can be viewed as the all-or-nothing symmetric MPR with where q_n is replaced by $\bar{q}_n$. Therefore the approximate stability region for general symmetric MPR model is the region lying below one of V boundaries $\partial_u \hat{\Lambda}^V$ defined by

$$\partial_u \hat{\Lambda}^V \triangleq \left\{ \boldsymbol{\lambda}^V : \exists\, \boldsymbol{\rho}^{(N)} \in \partial_u [0,1]^V \text{ such that } \forall v,\ \lambda_v^{(N)} = P_v\left(\boldsymbol{\rho}^{(N)}\right) \right\},$$

where

$$P_v\left(\boldsymbol{\rho}^{(N)}\right) = \sum_{(n_1,\cdots,n_V)\in\mathcal{N}^V} R_v(n_1,\cdots,n_V)p\left(n_1,\ldots,n_V;\boldsymbol{\rho}^{(N)}\right), \tag{3.28}$$

and

$$R_v(n_1,\cdots,n_V) = \bar{q}_n\frac{n_v}{N_v}. \tag{3.29}$$

When $N\to\infty$, it is easy to verify that the average throughput of a class-v user is given by

$$\lim_{N\to\infty} NP_v\left(\boldsymbol{\rho}^{(N)}\right) = \bar{\lambda}_v(\boldsymbol{\rho}) = \rho_v p_v\left(\bar{q}_1 + \frac{\bar{q}_2}{1!}\gamma + \cdots + \frac{\bar{q}_K}{(K-1)!}\gamma^{K-1}\right)e^{-\gamma}. \tag{3.30}$$

The general symmetric MPR model does not differ from the all-or-nothing MPR model in any fundamental way. The key observation is that the mean-field approach is about the average behavior of the system. Hence, q_n probabilities are essentially replaced by an average probability $\bar{q}_n$ as computed in (3.27). As a result, the expressions for throughput and delay of general symmetric MPR model follow directly from those of the all-or-nothing MPR model.

3.9 Extension to Persistent CSMA with All-or-Nothing MPR

Here we briefly discuss how to extend the previous stability result for slotted ALOHA to persistent CSMA with all-or-nothing MPR model.

3.9.1 Approximate Stability Region for MPR-capable CSMA

Define a super slot as either an idle period or a busy period which is the number of time slots required for the successful transmission of a packet (κ time slots) with its overheads, e.g., an acknowledgment (ACK) packet. We assume that an idle period is exactly one time slot and a busy period consists of τ time slots.

Definition 8(Approximate Stability Region) The approximate stability region $\hat{\Lambda}^N$

is the region lying below one of V boundaries $\partial_u \hat{\Lambda}^N$ defined by

$$R_v\left(\boldsymbol{\rho}^{(N)}\right) = \frac{P_v\left(\boldsymbol{\rho}^{(N)}\right)}{P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right) + \tau\left(1 - P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right)\right)}, \tag{3.31}$$

where

$$P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right) \triangleq \prod_{u=1}^{V}\left(1 - \rho_u^{(N)} p_u^{(N)}\right)^{N_u} \tag{3.32}$$

is the probability that a super slot is idle, and

$$P_v\left(\boldsymbol{\rho}^{(N)}\right) \triangleq \sum_{\substack{n_1+\cdots+n_V \le K \\ 0 \le n_u \le K\ \forall u \in \mathcal{V}}} \frac{n_v}{N_v} q_{n_1+\cdots+n_V} \prod_{u=1}^{V} \binom{N_u}{n_u} \left(\rho_u^{(N)} p_u^{(N)}\right)^{n_u} \left(1 - \rho_u^{(N)} p_u^{(N)}\right)^{N_u - n_u} \tag{3.33}$$

is the probability that a packet of a given class-v user is successfully transmitted in a super slot.

As before, the vector $\boldsymbol{\rho}^{(N)}$ represents the non-empty probabilities of the queues in the steady state, and $P^{\text{IDLE}}(\boldsymbol{\rho}^{(N)})$ is the system idle probability. The above approximate stability region $\hat{\Lambda}^N$ turns out to be *exact* for the two-user case ($N = 2$) and the homogeneous case ($p_1 = \cdots = p_V$ and $\lambda_1 = \cdots = \lambda_V$) [33]. Furthermore, similar to the slotted ALOHA case, the above approximate stability region $\hat{\Lambda}^N$ tends to be exact as $N \to \infty$.

Fig. 3.9 compares the approximate stability region of CSMA with slotted ALOHA for the same network as in Fig. 3.9. The advantage of carrier sensing is reflected in the expansion of the stability region compared to slotted ALOHA as evidently higher rates are feasible by CSMA. In addirion, our analytical result matches the simulation quite well.

In the next chapter, we discuss some applications of approximate stability region of CSMA with all-or-nothing MPR model.

3.10 Summary

In this chapter, we have studied the stability region of slotted ALOHA under all-or-nothing MPR assumption. By exploiting a mean-field technique, we find an approximation for the

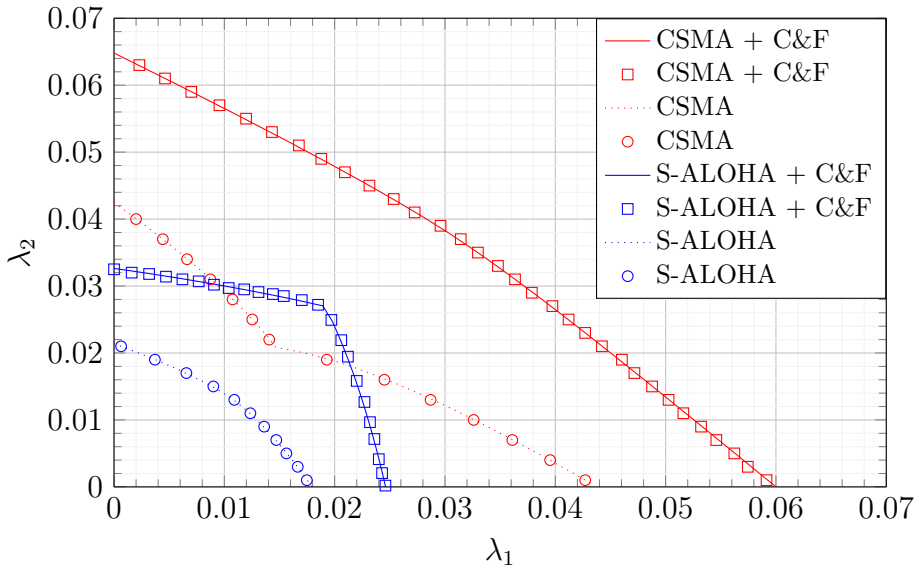

Figure 3.9: Feasible rates for transmission probabilities $p_1 = 0.028$, $p_2 = 0.039$, $q_1 = 0.96$, $q_2 = 0.89$ with $N_1 = N_2 = 15$, and $\kappa = 10$.

stability region of MPR-capable slotted ALOHA which is shown to be asymptotically exact for large systems. It has been found that that the approximate stability region is very accurate for small systems. In addition, simple expressions for throughput, service delay and total packet delay are derived. These performance metrics have not been well understood in the literature due to the complicated nature of the stability region. We have further demonstrated the clear advantage of MPR-capable slotted ALOHA over conventional slotted ALOHA systems.

Chapter 4

CSMA WITH ALL-OR-NOTHING MULTI-PACKET RECEPTION

The problem of Carrier Sense Multiple Access (CSMA) with multi-packet reception (MPR) is studied. Most prior work has focused on the homogeneous case, where all the mobile users are assumed to have identical packet arrival rates and transmission probabilities. The inhomogeneous case remains largely open in the literature. In this chapter, we make a first step towards solving this open problem by deriving throughput and delay expressions for inhomogeneous persistent CSMA, with a particular focus on the all-or-nothing symmetric MPR channel model [9]. This family of MPR models allows us to overcome several technical challenges associated with conventional analysis and to derive accurate throughput and delay expressions in the large-systems regime. As we previously pointed out, this family of MPR models is still general enough to include a number of useful MPR techniques—such as successive interference cancellation (SIC), compute-and-forward (C&F), and successive compute-and-forward (SCF)—as special cases. Based on the throughput and delay expressions, we provide theoretical guidelines for meeting quality-of-service requirements and for achieving global stability; we also evaluate the performances of various MPR techniques, highlighting the clear advantages offered by SCF.

4.1 Motivation

The CSMA protocol was introduced by Kleinrock and Tobagi [80] as a family of random access schemes where prior to any transmission attempt, users make sure that the channel is idle through a sensing mechanism. Specifically, at the beginning of each time slot, users sense the presence of the carrier signal and only attempt a transmission if the channel is

observed in idle state. Hence this scheme is called carrier sense multiple access.

In the previous chapter, we applied mean-field analysis to analyze slotted ALOHA systems with MPR capability. We found an asymptotically exact approximation of the stability region which is quite accurate in the finite user regime as well. Along this line, it is quite natural to ask whether the analysis can be extended to help analyzing carrier sense multiple access protocols.

In this chapter, we go a step further towards solving this open problem by deriving throughput and delay expressions for inhomogeneous CSMA under all-or-nothing symmetric MPR channel model. The use of the all-or-nothing symmetric MPR model not only allows us to overcome several fundamental challenges associated with conventional analysis [5,33], but also leads to simple throughput and delay expressions in the large-systems regime. We further provide theoretical guidelines for practical network design. For example, we show how our throughput and delay results can be used to choose the transmission probabilities in order to meet certain quality-of-service (QoS) requirements. We also prove results to guarantee global stability of MPR-capable CSMA system to avoid meta-stability[1] via a cross-layer design.

Furthermore, we evaluate the performances of various MPR techniques in terms of throughput, packet delay and service delay through both analysis and simulation. Our results highlight the clear advantages of SCF-based CSMA over SIC-based CSMA and conventional CSMA.

4.2 Network and Channel Model

We focus on a random-access network where, as illustrated in Fig. 2.1, N mobile users contend to transmit packets to an access point (AP). Time is assumed to be slotted, i.e., all packet transmissions are slot synchronous, and packets are of constant length requiring κ time slots. Each user belongs to one of V possible classes $\mathcal{V} = \{1, \cdots, V\}$, where all users in the same class have identical arrival rate and transmission probability. This V-class model

[1] A system with multiple stable states is called meta-stable. We give more details on meta-stability in Section 3.5.

captures user heterogeneity in terms of packet arrival rates and transmission probabilities.

Each user is equipped with an infinite buffer for storing packets in a FIFO manner. Packets arrive at the buffer of a class-v user according to a Bernoulli process with rate λ_v. That is, at each time slot, a new packet arrives into the buffer of a class-v user with probability λ_v. The arrival processes are assumed to be independent across users.

In this chapter, we use the channel model as described in Section 2.0.1, and will mainly focus on the all-or-nothing symmetric MPR models. The reader is referred to Chapter 2 for a detailed discussion of symmetric MPR techniques. Similar to slotted ALOHA, the results of this chapter can be extended to the case of the general symmetric MPR model for persistent CSMA. Next, the inhomogeneous persistent CSMA protocol under the symmetric MPR model is introduced.

4.2.1 Inhomogeneous Persistent CSMA with Symmetric MPR

Carrier-Sense Multiple Access (CSMA) with random back-off takes advantage of detecting ongoing transmissions and a back-off mechanism to reduce collisions. Prior to attempting any transmission, a user observes the channel status continuously until it finds the channel idle for a duration of time known as DCF[2] Inter-Frame Spacing (DIFS) period. Users who intend to transmit pick a random back-off counter within a range called contention window (CW) whose size depends on the back-off stage of the user. A user then begins decrementing the back-off counter at the end of every slot that is observed idle.The user with the smallest back-off counter value eventually gains access to channel. After a transmission, the source node awaits an acknowledgment (ACK) from its intended receiver. If the receiver enjoys an error-free decoding, it sends an ACK back to the transmitter after a Short Inter-Frame Spacing (SIFS) period.

In case two or more users pick the same back-off counter value, collision occurs in which case the receiver cannot decode the collided packets and the transmitter will not receive ACK.

[2]Distributed Coordinated Function.

In such a case, the users involving in the collision increase their back-off window (typically via doubling as in binary exponential back-off) to reduce collision probability in subsequent transmission attempts. Once a packet is successfully acknowledged, the back-off stage is set back to minimum.

The CSMA protocol we consider here is very similar to p-persistent CSMA[3] except that the transmission probabilities are user dependent. For this reason, we call this protocol inhomogeneous persistent CSMA. Note that when the packet length is one time slot, persistent CSMA reduces to slotted ALOHA [81, Chapter 4]. This is because carrier sensing will not help reducing collisions when packet duration is one slot. When a transmission is successful all users receive ACK after a SIFS period. Specifically, when there are n users sending packets in a slot, the AP is able to reconstruct all these n packets with probability q_n, as explained in Chapter 2.

Fig. 4.1 illustrates a renewal cycle of the inhomogeneous persistent CSMA. A busy period is comprised of a transmission period followed by SIFS, DIFS durations and ACK packet transmission time. There are also some inactive slots that constitute an idle period when combined together.

4.3 Throughput and Delay Analysis

In this section, we apply the mean-field approximation to study the throughput and delay performance of inhomogeneous CSMA with MPR. For ease of presentation, we will mostly focus on the all-or-nothing symmetric MPR model. The extension to general symmetric MPR will be provided in Section 3.8.

---

[3]In a p-persistent CSMA system, a user avoids transmission when it finds the channel busy; it then persistently monitors the channel and transmits a packet with probability p as soon as the channel becomes idle.

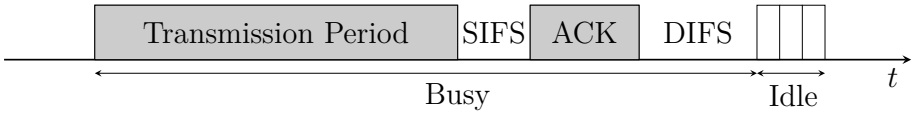

Figure 4.1: Illustration of the busy and idle periods (super slots) in an inhomogeneous persistent CSMA protocol. A transmission period is κ time slots, and a busy period takes τ time slots.

4.3.1 Mean-Field Approximation

Similar to conventional CSMA, the channel can be in either busy or idle state. As illustrated in Fig. 4.1, a busy period is comprised of a transmission period followed by SIFS, DIFS durations and ACK packet transmission time, and an idle period is comprised of a DIFS duration. Without loss of generality, we assume that an idle period (i.e., a DIFS duration) is exactly one time slot and that a busy period consists of τ time slots (since the packets are of constant length requiring κ time slots). We call a busy period or an idle period a *super slot*. Hence a super slot consists of either one single time slot when the channel is in idle state, or τ time slots which is equivalent to a transmission period followed by SIFS, DIFS periods and one ACK.

We are now ready to apply the mean-field approximation. Consider a sequence of systems index by the number of users N. For any fixed N, each class-v user has the same packet arrival rate $\lambda_v^{(N)}$ and the same transmission probability $p_v^{(N)}$. Clearly this is an inhomogeneous network because the arrival rates and transmission probabilities belong to V different classes and are not identical as in a homogeneous network. When we scale the system, we assume that

$$\lambda_v^{(N)} N \to \tilde{\lambda}_v \text{ and } p_v^{(N)} N \to \tilde{p}_v, \text{ as } N \to \infty\,.$$

That is, the arrival rate $\lambda_v^{(N)}$ scales like $\frac{\tilde{\lambda}_v}{N}$, and the transmission probability $p_v^{(N)}$ scales like $\frac{\tilde{p}_v}{N}$. (Note that both $\tilde{\lambda}_v$ and $\tilde{p}_v$ can be larger than 1, since they are no longer probabilities.)

Similarly, for any fixed N, we let N_v be the number of class-v users and assume that

$$\frac{N_v}{N} \to \beta_v, \text{ as } N \to \infty\,.$$

That is, the fraction of class-v users converges to β_v. Finally, we assume that the duration of a time slot scales like $\frac{1}{N}$.

The mean-field approximation proceeds as follows. Assume that users' queues evolve independently of each other, and further assume that the queue-length evolution processes of the same class are statistically identical. These assumptions can be validated in the large-systems limit as we did in Chapter 3. Under these assumptions, the queue evolution of a class-v user is a discrete-time random process with packet arrival rate $\lambda_v^{(N)}$ and utilization probability $\rho_v^{(N)}$ (i.e., the probability that the queue is non-empty in a limiting time slot). Let

$$\boldsymbol{\rho}^{(N)} \triangleq \left(\rho_1^{(N)}, \ldots, \rho_V^{(N)}\right)\,.$$

Then, the average throughput (i.e., the average number of packets transmitted successfully in a time slot) of a class-v user is given by

$$R_v\left(\boldsymbol{\rho}^{(N)}\right) = \frac{P_v\left(\boldsymbol{\rho}^{(N)}\right)}{P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right) + \tau\left(1 - P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right)\right)}\,, \tag{4.1}$$

where

$$P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right) \triangleq \prod_{u=1}^{V}\left(1 - \rho_u^{(N)} p_u^{(N)}\right)^{N_u} \tag{4.2}$$

is the probability that a super slot is idle, and

$$P_v\left(\boldsymbol{\rho}^{(N)}\right) \triangleq \sum_{\substack{n_1+\cdots+n_V \le K \\ 0 \le n_u \le K\ \forall u \in \mathcal{V}}} \frac{n_v}{N_v} q_{n_1+\cdots+n_V} \prod_{u=1}^{V} \binom{N_u}{n_u} \left(\rho_u^{(N)} p_u^{(N)}\right)^{n_u} \left(1 - \rho_u^{(N)} p_u^{(N)}\right)^{N_u - n_u} \tag{4.3}$$

is the probability that a packet of a given class-v user is successfully transmitted in a super slot.

To better understand Eq. (4.1), one shall notice that the numerator $P_v\left(\boldsymbol{\rho}^{(N)}\right)$ corresponds to the average number of packets transmitted successfully in a super slot for a class-v user and the denominator

$$P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right)+\tau\left(1-P^{\text{IDLE}}\left(\boldsymbol{\rho}^{(N)}\right)\right)$$

corresponds to the average length of a super slot. Therefore, Eq. (4.1) indeed gives the average throughput of a class-v user under the mean-field approximation. In particular, if the utilization probability $\rho_v^{(N)}<1$, the queue of a class-v user is stable[4] and $\lambda_v^{(N)}=R_v\left(\boldsymbol{\rho}^{(N)}\right)$.

Let

$$\boldsymbol{\lambda}^V \triangleq \left(\lambda_1^{(N)},\ldots,\lambda_V^{(N)}\right).$$

The approximate stability region is as follows.

Definition 9(Approximate Stability Region under Symmetric MPR) The approximate stability region $\hat{\Lambda}^V$ is the region lying below one of V boundaries $\partial_u\hat{\Lambda}^V$ defined by

$$\partial_u\hat{\Lambda}^V \triangleq \left\{\boldsymbol{\lambda}^V : \exists\,\boldsymbol{\rho}^{(N)}\in\partial_u[0,1]^V \text{ such that } \forall v,\ \lambda_v^{(N)}=R_v(\boldsymbol{\rho}^{(N)})\right\},$$

When $N\to\infty$, it is easy to verify that the (normalized) average throughput of a class-v user admits the simple expression (4.4),

$$\lim_{N\to\infty} NR_v\left(\boldsymbol{\rho}^{(N)}\right)=\frac{\rho_v^{(\infty)}\tilde{p}_v\left(q_1+\frac{q_2}{1!}\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)+\cdots+\frac{q_K}{(K-1)!}\gamma^{K-1}\left(\boldsymbol{\rho}^{(\infty)}\right)\right)e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}}{e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}+\tau\left(1-e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}\right)}, \tag{4.4}$$

where

$$\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\triangleq\sum_{u=1}^{V}\beta_u\tilde{p}_u\rho_u^{(\infty)}. \tag{4.5}$$

---

[4]A random process $X(t)$ is stable if for any $x>0$ the following holds: $\lim_{t\to\infty}\mathbb{P}(X(t)<x)=F(x)$ and $\lim_{x\to\infty}F(x)=1$ where $F(x)$ is the limiting distribution function [33, 82].

Define

$$\chi(\gamma) \triangleq q_1 + \frac{q_2}{1!}\gamma + \frac{q_3}{2!}\gamma^2 + \cdots + \frac{q_K}{(K-1)!}\gamma^{K-1}. \tag{4.6}$$

Then, the (normalized) average throughput of a class-v user can be rewritten as

$$\lim_{N\to\infty} NR_v\left(\boldsymbol{\rho}^{(N)}\right) = \frac{\rho_v^{(\infty)}\tilde{p}_v\chi\left(\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\right)e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}}{e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)} + \tau\left(1 - e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}\right)}. \tag{4.7}$$

Similarly, if the utilization probability $\rho_v^{(\infty)} < 1$, the queue of a class-v user in the large-systems limit is stable, and we have

$$\tilde{\lambda}_v = \frac{\rho_v^{(\infty)}\tilde{p}_v\chi\left(\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\right)e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}}{e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)} + \tau\left(1 - e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}\right)}. \tag{4.8}$$

Recall that the duration of a time slot scales like $\frac{1}{N}$. Thus, the limiting (infinite) system is a collection of V coupled $M/M/1$ queues with arrival rate $\tilde{\lambda}_v$ and service rate

$$\frac{\tilde{p}_v\chi\left(\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\right)e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}}{e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)} + \tau\left(1 - e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}\right)}.$$

The coupling is due to the limiting utilization probabilities $\boldsymbol{\rho}^{(\infty)}$. This proves our main result of mean-field approximation in the following theorem.

Theorem 4.3.1 Under the mean-field approximation, the limiting (infinite) system is a collection of V coupled queues, where the v th queue is $M/M/1$ with arrival rate $\tilde{\lambda}_v$ and service rate

$$\mu_v\left(\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\right) \triangleq \frac{\tilde{p}_v\chi\left(\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)\right)e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}}{e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)} + \tau\left(1 - e^{-\gamma\left(\boldsymbol{\rho}^{(\infty)}\right)}\right)}.$$

We say the limiting system is *globally stable*[5] with respect to $\tilde{\boldsymbol{\lambda}}$ if the following system of equations has a unique solution $\boldsymbol{\rho}^* \in [0,1)^V$:

$$\forall v,\ \tilde{\lambda}_v = \frac{\rho_v \tilde{p}_v \chi\left(\gamma\left(\boldsymbol{\rho}\right)\right) e^{-\gamma(\boldsymbol{\rho})}}{e^{-\gamma(\boldsymbol{\rho})} + \tau\left(1 - e^{-\gamma(\boldsymbol{\rho})}\right)}. \tag{4.9}$$

The *limiting stability region* is defined as the set of vectors $\tilde{\boldsymbol{\lambda}}$ such that the limiting system is stable for $\tilde{\boldsymbol{\lambda}}$. We can characterize the limiting stability region. For simplicity, we assume that the function

$$f(\gamma) \triangleq \frac{\gamma\chi(\gamma)e^{-\gamma}}{e^{-\gamma} + \tau(1 - e^{-\gamma})} = \frac{\gamma\left(q_1 + \frac{q_2}{1!}\gamma + \frac{q_3}{2!}\gamma^2 + \cdots + \frac{q_K}{(K-1)!}\gamma^{K-1}\right)e^{-\gamma}}{e^{-\gamma} + \tau(1 - e^{-\gamma})}$$

is unimodal. In fact, this assumption can be relaxed. We introduce it here to simplify our expressions for the limiting stability region. We notice that this assumption is rather mild. For example, this assumption holds as long as $K \leq 2$. When $K > 2$, this assumption holds as long as $q_1 \leq 2q_2 \leq \cdots \leq Kq_K$, or $q_1 \geq 2q_2 \geq \cdots \geq Kq_K$. From (4.9) we have

$$\begin{aligned}
\lambda &\triangleq \sum_v \beta_v \tilde{\lambda}_v \\
&= \sum_v \beta_v \frac{\rho_v \tilde{p}_v \chi\left(\gamma\right) e^{-\gamma}}{e^{-\gamma} + \tau\left(1 - e^{-\gamma}\right)} \\
&= \frac{\gamma\chi(\gamma)e^{-\gamma}}{e^{-\gamma} + \tau\left(1 - e^{-\gamma}\right)} = f(\gamma).
\end{aligned} \tag{4.10}$$

λ is defined as the total input arrival rate. Let γ^* be the maximizer of f. Define $\gamma_0 \triangleq \sum_v \beta_v \tilde{p}_v$ and let $\lambda_0 = f(\gamma_0)$. We then have the following lemma.

Lemma 4.3.2 The function $f(\gamma) = \frac{\gamma\chi(\gamma)e^{-\gamma}}{e^{-\gamma} + \tau(1-e^{-\gamma})}$ is unimodal if $q_1 \leq 2q_2 \leq \cdots \leq Kq_K$.In particular, $f(\gamma)$ is unimodal if $K \leq 2$.

Proof. See Appendix B.1. □

[5]Note that this definition of global stability is unconventional, since it is based on the computational uniqueness instead of the stochastic behavior of the limiting system. Interestingly, this definition leads to the true stability region based on the stochastic behavior.

Theorem 4.3.3 Suppose that the function $f(\gamma) = \frac{\gamma\chi(\gamma)e^{-\gamma}}{e^{-\gamma}+\tau(1-e^{-\gamma})}$ is unimodal. Then, the limiting stability region $\tilde{\Lambda}$ is the region lying below one of V boundaries $\partial_u\tilde{\Lambda}$ defined by

$$\partial_u\tilde{\Lambda} \triangleq \left\{\tilde{\boldsymbol{\lambda}} \in \mathbb{R}_+^V : \exists\, \boldsymbol{\rho} \in \partial_u[0,1]^V \text{ s.t. } \forall v,\ \tilde{\lambda}_v = \frac{\rho_v \tilde{p}_v \chi\left(\gamma\left(\boldsymbol{\rho}\right)\right) e^{-\gamma(\boldsymbol{\rho})}}{e^{-\gamma(\boldsymbol{\rho})} + \tau\left(1 - e^{-\gamma(\boldsymbol{\rho})}\right)}\right\}, \tag{4.11}$$

where $\partial_u[0,1]^V$ is the set of $\boldsymbol{\rho} \in \mathbb{R}_+^V$ such that $\rho_u = 1$ and $\forall\ w \neq u$, $\rho_w \leq 1$.

Proof. See Appendix B.2. □

Although we define the limiting stability region based on the uniqueness of $\boldsymbol{\rho}^*$, the resulting region turns out to be *exact* in the sense that it is precisely the true stability region of the limiting system. In other words, our mean-field approximation is asymptotically exact as $N \to \infty$.

4.3.2 Accuracy of Approximate Stability Region

For finite N, we conduct simulations to show that the approximate stability region is accurate even for small N. As we did in Chapter 3, we apply the notion of homogeneous directions proposed by [29] to show that the approximate and true stability regions coincide in many scenarios where the mean-field decoupling assumption holds. We do this by way of the following example.

Example 11 Consider a network with $N = 3$ users with transmission probabilities $p_1 = 1/3$, $p_2 = 1/3$, $p_3 = 1/3$ and arrival rates $\lambda_1 = \lambda$, $\lambda_2 = \lambda/x$ and $\lambda_3 = \lambda$. We then vary x in the range $[1, 20]$. After some calculations, it can be shown that the approximate stability condition for persistent CSMA with all-or-nothing symmetric MPR is

$$\sum_{i=1}^{3} \lambda_i < \frac{4}{5\tau+4}\left(q_1 + \frac{q_2}{2}\right) \frac{2x+1}{2x + \frac{9\tau}{5\tau+4}\frac{q_1 - \frac{\tau+8}{18\tau}q_2}{q_1+q_2}}\,.$$

Figure 4.2 shows the above stability condition compared with the simulation result for persistent CSMA. It can be seen that the curves match considerably well confirming the

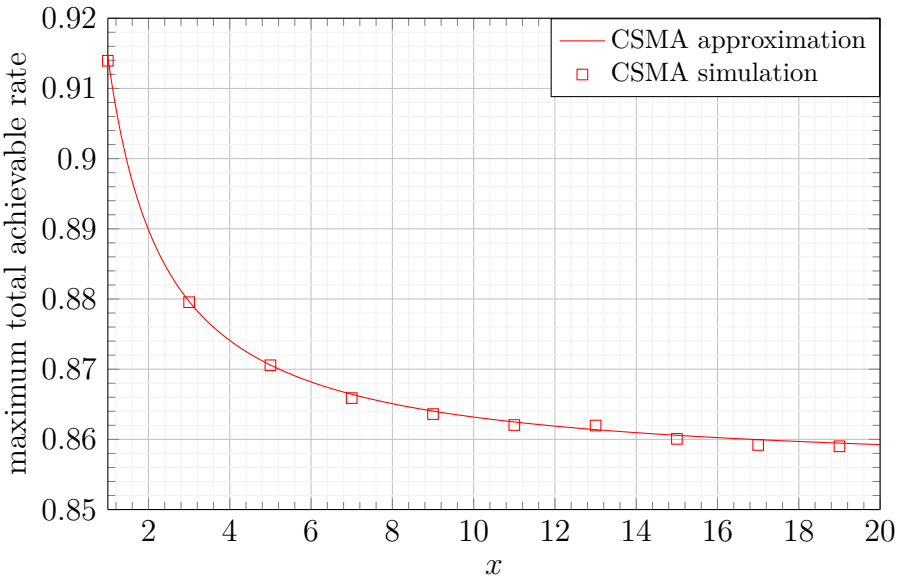

Figure 4.2: Maximum total achievable rate of a network with $N = 3$, $p_1 = p_2 = p_3 = 1/3$ and $\kappa = 5$.

accuracy of the approximation for N small.

4.3.3 Meta-Stability

Theorem 4.3.3 determines the set of arrival rates for which the system is globally stable. Intuitively all arrival rate vectors not in the stability region should constitute the instability region. However, it turns out that this intuition is not accurate and under a certain condition, there exits a third region of arrival rate vectors which make the system bistable. Here we discuss this phenomenon in some details and not only give a sufficient condition for the system to become bistabe, but provide an algorithm that determines whether a given arrival rate vector lies in either stable, bistable or unstable regions.

Specifically, observe that depending on the input arrival rate, the equation $f(\gamma) = \lambda$ may have two, one or no solutions which corresponds to a bistable, stable or unstable system as illustrated in Fig. 4.3. When $\lambda > \lambda_0$ and $\gamma_0 > \gamma^*$ then the persistent CSMA system can potentially become bistable. To see why this is true assume a 2-class network with $\tilde{p}_1 = \tilde{p}_2$

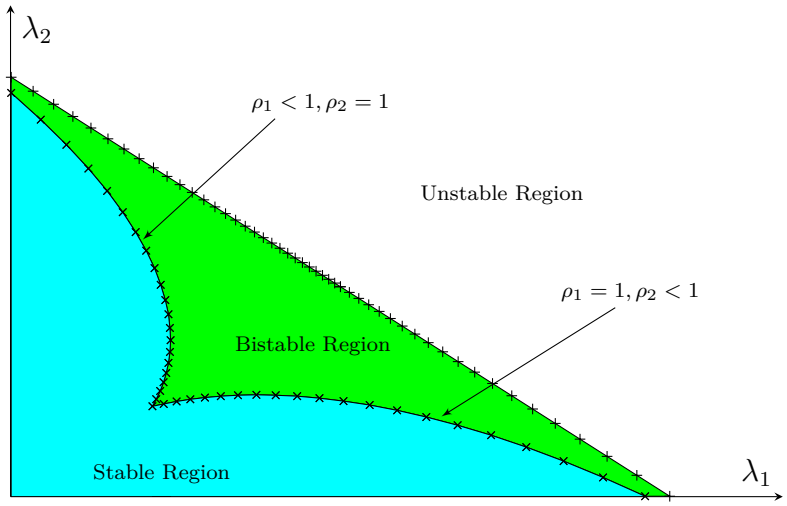

Figure 4.3: Illustration of stable, bistable and unstable regions for persistent CSMA.

and arrival rates $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ such that $\frac{\tilde{\lambda}_2}{\tilde{\lambda}_1} = r > 1$ is fixed. As we increase $\tilde{\lambda}_1$ the two roots $\underline{\gamma}$ and $\overline{\gamma}$ of the equation $f(\gamma) = \lambda$ become smaller and larger respectively. Once the boundary of the stability region is reached $\boldsymbol{\rho} = (\rho_1, 1)$. Further increase of $\tilde{\lambda}_2$ makes $\overline{\gamma}$ smaller such that $\boldsymbol{\rho}_1 = (\rho_1, \rho_2)$ and $\rho_1 < 1$, $\rho_2 < 1$. In other words there will be two stable points.

We now discuss how to compute $\boldsymbol{\rho}^*$ and how to figure out which region a given rate lies. For the rest of this discussion see Fig. 4.4. The procedure presented in Algorithm 1 takes a rate vector as input and determines $\boldsymbol{\rho}^*$ and the state of the system, i.e., the region where the rate vector lies. First observe that γ is a bounded quantity and $\gamma \leq \gamma_0$. When the total input rate $\lambda < \lambda_0$ then the equation $\lambda = f(\gamma)$ has only one root as f is a unimodal function and $\gamma \leq \gamma_0$ should hold. When $\lambda \geq \lambda_0$ and $\gamma_0 \leq \gamma^*$ then $\lambda = f(\gamma)$ has no root and the system is unstable. On the other hand when $\lambda \geq \lambda_0$ and $\gamma_0 > \gamma^*$, the equation $\lambda = f(\gamma)$ has two distinct roots and the system may be bistable or stable depending on whether $\boldsymbol{\rho}_1 = \left(\lambda_1/\mu_1(\underline{\gamma}), \cdots, \lambda_V/\mu_V(\underline{\gamma})\right)$ and $\boldsymbol{\rho}_2 = (\lambda_1/\mu_1(\overline{\gamma}), \cdots, \lambda_V/\mu_V(\overline{\gamma}))$ are both valid or either of them is valid.

As pointed out in [53, 79], meta-stability is a highly undesirable property because the

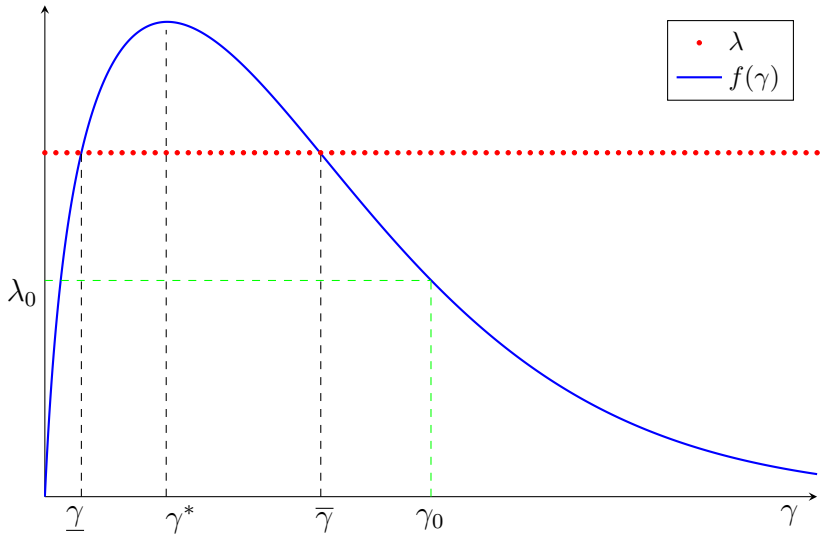

Figure 4.4: The unimodal function $f(\gamma)$ and its intersection with total input arrival rate.

state of a network fluctuates – over long periods of time – between different stable states. Such long oscillations make it impossible to predict the average network performance in terms of throughput and delay. As a consequence, a proper level of quality-of-service cannot be guaranteed.

Along the same line as in slotted ALOHA, we provide a solution for inhomogeneous CSMA systems to avoid meta-stability. In particular, if the MPR technique is carefully designed, then the network is proven to be globally stable, and so meta-stability can be completely eliminated. This leads to an important design criterion, especially when quality-of-service guarantee is in great need.

Theorem 4.3.4 In the inhomogeneous persistent CSMA system with symmetric MPR model, meta-stability can be avoided if

$$\sum_{v \in \mathcal{V}} N_v p_v^{(N)} < \gamma^*$$

Algorithm 1 A procedure to determine $\boldsymbol{\rho}^*$ and the state of a V-class system.

Input: $\boldsymbol{\lambda}^V = (\lambda_1, \cdots, \lambda_V)$
Output: $\boldsymbol{\rho}^*$ and the state of system.

if $\lambda < \lambda_0$ **then**
  Find $\underline{\gamma}$ as the root of $f(\gamma) - \lambda = 0$
  return $\boldsymbol{\rho}^* = \left(\lambda_1/\mu_1(\underline{\gamma}), \cdots, \lambda_V/\mu_V(\underline{\gamma})\right)$
else
  if $\lambda > \lambda_0$ & $\gamma_0 > \gamma^*$ **then**
    Set $\underline{\gamma}$ to be the root of $f(\gamma) - \lambda = 0$ in $[0, \gamma^*]$
    Set $\overline{\gamma}$ to be the root of $f(\gamma) - \lambda = 0$ in $(\gamma^*, \gamma_0]$
    Set $\boldsymbol{\rho}_1 = \left(\lambda_1/\mu_1(\underline{\gamma}), \cdots, \lambda_V/\mu_V(\underline{\gamma})\right)$
    Set $\boldsymbol{\rho}_2 = (\lambda_1/\mu_1(\overline{\gamma}), \cdots, \lambda_V/\mu_V(\overline{\gamma}))$
    if $\boldsymbol{\rho}_1 < \mathbf{1}$ & $\boldsymbol{\rho}_2 < \mathbf{1}$ **then**
      state = BISTABLE
      return $\boldsymbol{\rho}_1^* = \boldsymbol{\rho}_1, \boldsymbol{\rho}_2^* = \boldsymbol{\rho}_2$ and state
    else
      if $\boldsymbol{\rho}_1 < \mathbf{1}$ **then** $\boldsymbol{\rho}^* = \boldsymbol{\rho}_1$ **end if**
      if $\boldsymbol{\rho}_2 < \mathbf{1}$ **then** $\boldsymbol{\rho}^* = \boldsymbol{\rho}_2$ **end if**
      state = STABLE
      return $\boldsymbol{\rho}^*$ and state
    end if
  else
    state = UNSTABLE
    return state
  end if
end if

and

$$q_1 \leq 2q_2 \leq \cdots \leq Kq_K \quad \text{or} \quad q_1 \geq 2q_2 \geq \cdots \geq Kq_K \tag{4.12}$$

Proof. The proof follows from Lemma 4.3.2 and the fact that the arrival rates corresponding to bistable and unstable regions should be avoided, hence $\gamma_0 < \gamma^*$. □

As an interesting consequence of Theorem 4.3.4, meta-stability is automatically avoided when $K = 2$.

4.3.4 Throughput and Delay Analysis

Here, we characterize the throughput and delay performance based on the mean-field approximation, especially the limiting stability region. Our results apply to a network with an arbitrary number of classes.

In this section we only focus on stable and bistable regions because if an arrival rate vector is in the unstable region then a rate control protocol like TCP Vegas can be used to reduce the arrival rate.

Recall that, under the mean-field approximation, the average throughput of a class-v user is given by $R_v\left(\boldsymbol{\rho}^{(N)}\right)$, which is asymptotically exact. This leads to the following throughput result:

Proposition 4.3.5 The aggregate throughput of the inhomogeneous persistent CSMA with all-or-nothing symmetric MPR in a V-class network is given by

$$R_{\text{MPR-CSMA}} = \sum_{v=1}^{V} N_v^{(N)} R_v\left(\boldsymbol{\rho}^{(N)}\right) + o(1)\,, \tag{4.13}$$

where the $o(1)$ term is understood as $N \to \infty$. In particular, the saturated throughput is given by $\sum_v N_v R_v(\mathbf{1})$, where $\mathbf{1}$ is an all-one vector of length V.

In order to apply Proposition 4.3.5, we need to obtain the value of the utilization probabilities $\boldsymbol{\rho}^{(N)}$ for any given $\{\lambda_v\}$ and $\{p_v\}$. Our strategy is to approximate $\boldsymbol{\rho}^{(N)}$ with the limiting utilization probabilities $\boldsymbol{\rho}^{(\infty)}$. Specifically, we set $\tilde{\lambda}_v = \lambda_v N$ and $\tilde{p}_v = p_v N$ and then solve the system of equations given in (4.9) to obtain a solution $\boldsymbol{\rho}^*$. As we discussed previously, Algorithm 1 can be used for this purpose.

Let us now look at the accuracy of (4.13) in a 2-class network through simulations. Consider a network with N users where half of the users belong to class 1. We assume that $q_1 = 0.96$, $q_2 = 0.89$ and fix the transmission probability of class-2 users at $p_2 = 0.2$. We then compute the network aggregate throughput while varying the transmission probability of class-1 users.

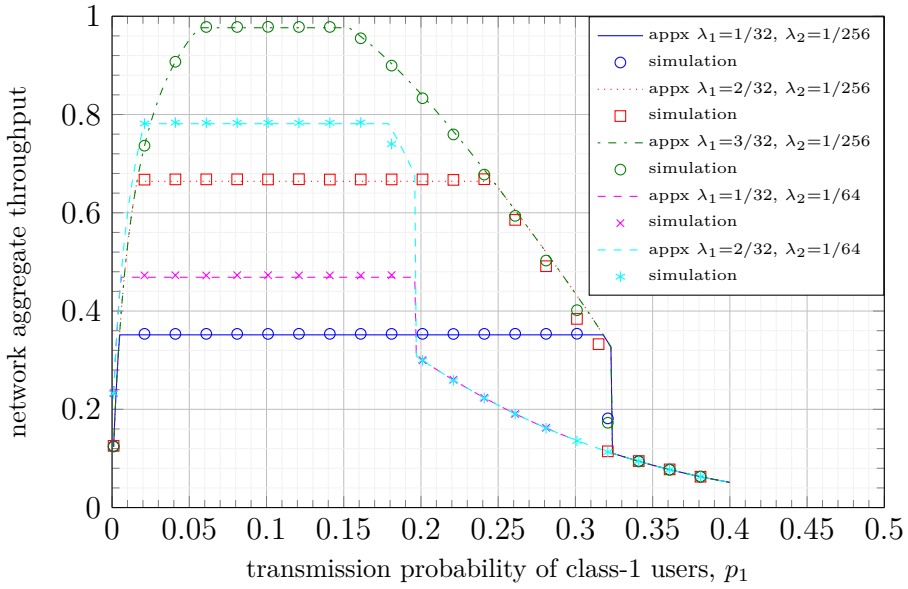

Figure 4.5: Aggregate throughput for a 2-class network with $\kappa = 10$, $p_2 = 0.2$, $q_1 = 0.96$, $q_2 = 0.89$ and $N_1 = N_2 = 10$.

Fig. 6.2 shows the network aggregate throughput versus the transmission probability of class-1 users when $N = 20$. It can be seen that the aggregate throughput closely matches the analytical result for a variety of packet arrival rates.

Fig. 4.6 shows the maximum aggregate throughput, i.e., the maximum total system throughput while all the queues are stable, versus the number of users N when $p_1 = \frac{5}{6N}$ and $p_2 = \frac{7}{6N}$. First of all, the benefit of symmetric MPR based CSMA can be observed compared with conventional CSMA. Moreover, it can be seen that the aggregate throughput approximation matches the simulations considerably well.

We now turn our attention to study the delay performance. Recall that there are two types of delays, namely, service delay and the total delay (packet delay). Let $D_{v,\text{CSMA}}$ denote the service delay of a HOL packet of a class-v user. Now consider a V-class network and assume that the system is stable. We have the following result for the service delay.

Proposition 4.3.6 The service delay of a class-v user in the inhomogeneous persistent

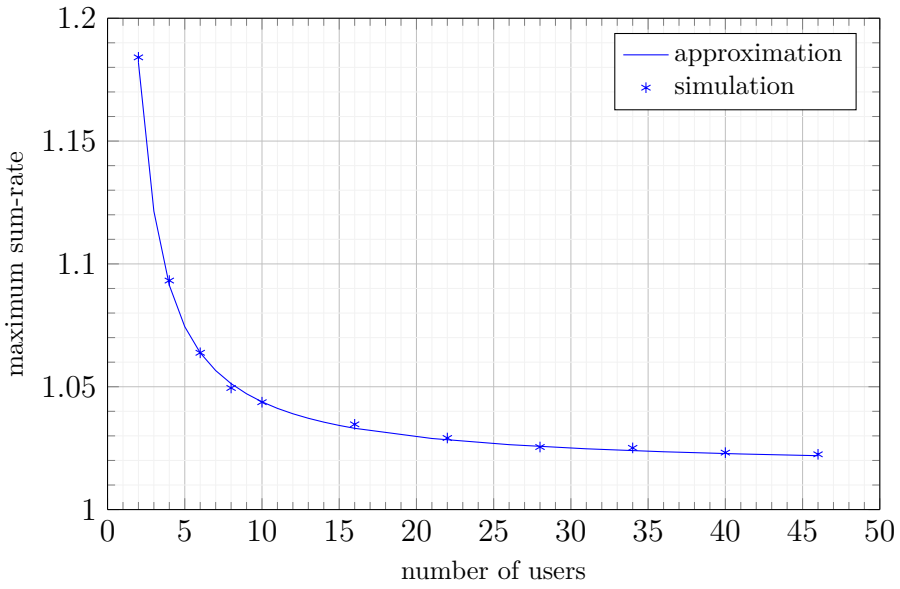

Figure 4.6: Maximum aggregate throughput of CSMA with symmetric MPR with $K = 2$ and $\kappa = 10$.

CSMA system with symmetric MPR is

$$D_{v,\mathrm{CSMA}} = \frac{\rho_v^{(N)}}{\lambda_v} + o(1) \quad \text{for} \ \ v \in \mathcal{V}\,. \tag{4.14}$$

Proof. Suppose that $\rho_v^{(N)} \in (0,1)$ for all $v \in \mathcal{V}$ is the unique solution to $\lambda_v = R_v(\boldsymbol{\rho}^{(N)})$. A class-$v$ user is then stable with the arrival rate λ_v and we can apply Little's law to obtain

$$\lambda_v D_{v,\mathrm{CSMA}} = \rho_v^{(N)} + o(1)\,.$$

This proves (4.14).

We next provide an alternative proof for (4.14) which might give some additional insight. Note that the success probability $P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})$ for a class-v user can be obtained from the

throughput function as follows

$$P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}) = \frac{P_v(\boldsymbol{\rho}^{(N)})}{\rho_v^{(N)}}\,. \tag{4.15}$$

Hence, the service delay for a class-v user can be written in a recursive fashion as follows

$$\begin{aligned} D_{v,\mathrm{CSMA}} = \tau P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}) + (1 + D_{v,\mathrm{CSMA}}) P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) \\ + (\tau + D_{v,\mathrm{CSMA}}) \left(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) - P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})\right) + o(1)\,. \end{aligned}$$

Therefore

$$D_{v,\mathrm{CSMA}} = \frac{P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) + \tau(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}))}{P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})} + o(1)\,. \tag{4.16}$$

In fact the system is either in idle, collision or successful transmission states. If the system is in idle state, then the HOL packet transmission is delayed by one time slot. If the class-v user is in the process of transmitting a packet successfully, the HOL packet transmission is delayed by τ time slots. Lastly, if the system is in collision state or there is any other successful transmission in progress, it takes τ more time slots for the HOL packet before it is successfully transmitted. □

Denote by $T_{v,\mathrm{CSMA}}$ the total delay experienced by a packet in the inhomogeneous persistent CSMA system with symmetric MPR. We characterize the total delay in the following theorem.

Theorem 4.3.7 Suppose that for all v, $\rho_v^{(N)} < 1$, i.e., the system is globally stable. The total delay of a class-v packet in the inhomogeneous persistent CSMA system with symmetric MPR is

$$T_{v,\mathrm{CSMA}} = \frac{\rho_v^{(N)}\left(\frac{1}{\lambda_v} - \frac{1}{\tau}\right) + \frac{\tau-1}{2}\left(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)})\right)}{1 - \rho_v^{(N)}} + o(1)\,. \tag{4.17}$$

Proof. See Appendix B.3. □

4.3.5 Adjusting Protocol Parameters

The simple expressions derived before can be used to adjust the transmission probabilities to meet certain throughput or delay requirements. For instance, assume a two-class network with given arrival rates. From the stability region in Theorem 4.3.3, we can immediately exclude the set of transmission probabilities for the two classes that destabilize the system or are too small that results is low throughput. Alternatively, one can exploit the delay expression in Theorem 4.3.7 to obtain the transmission probabilities such that a certain delay constraint is satisfied for a class. As a result, all the users in such a class will be stable. To further demonstrate the benefit of our delay analysis in system design, we present the following example.

Example 12 Consider a two-class network where N_1 users belong to class 1 and N_2 users are in class 2 with fixed arrival rates λ_1 and λ_2 respectively. Assume that a packet of a class-v user is required to experience an average finite total delay no more than T_v.

Since the delays are finite therefore all the queues are stable. Let $x_v = \rho_v p_v$ for $v \in \{1, 2\}$. For the given λ_v's, we can then compute x_1^* and x_2^* as the unique solution to $\lambda_v = R_v(x_1, x_2)$. In addition, from the delay constraint for each class and (4.17) it follows that

$$p_v \geq \frac{\left(\frac{1}{\lambda_v} - \frac{1}{\tau} + T_v\right) x_v^*}{T_v - \frac{\tau - 1}{2}\left(1 - P^{\text{IDLE}}(x_1^*, x_2^*)\right)} \quad \text{for } \; v \in \{1, 2\}\,.$$

The above gives the design criteria for transmission probabilities of every class given the delay constraints. A similar approach can be applied to the general case of $V > 2$ classes.

Lastly, if in a real scenario there are more than two classes of high-end and low-end users, as long as the requirement for the arrival rates of some users is such that they are below the arrival rates of high/low-end users, by a dominant system argument we see that these users enjoy a delay and throughput performance no worse than that experienced by high/low-end users. Obviously if new users with rate requirement higher than high-end users are required

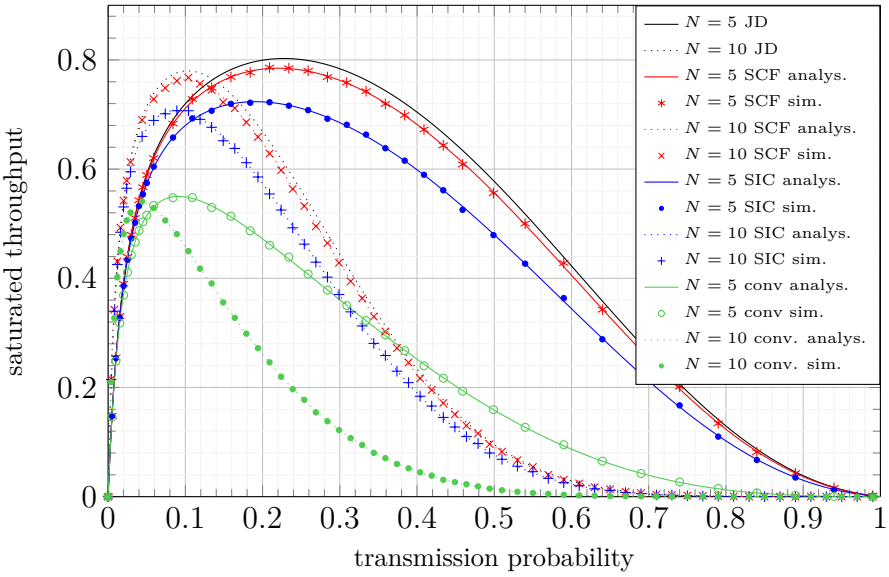

Figure 4.7: Saturated throughput benefit of SCF-based CSMA compared to CSMA with SIC and conventional CSMA in low $\mathsf{SNR} = 6\,\mathrm{dB}$, with $p_2 = 0.8p_1$, message rate $= 1$, $K = 1$, $\kappa = 10$.

to join the network, we have to add another class of users, but the throughput and delay guarantee can be fulfilled easily based on the result we have obtained.

4.4 Numerical Results

To validate our theoretical results and to further demonstrate the benefits of CSMA with SCF, we conduct simulations in various scenarios. Fig. 4.7 depicts saturated throughput, i.e., $R_v(\mathbf{1})$ for CSMA with SCF, CSMA with SIC and conventional CSMA as a function of transmission probabilities in a two-class network when $N = 5, 10$, $N_1 = 3/5N$ and $\mathsf{SNR} = 6\,\mathrm{dB}$. It is evident that SCF-based CSMA performs close to CSMA with optimal joint decoding (JD) and significantly improves the saturated throughput compared to conventional CSMA. Further, CSMA with SCF universally outperforms CSMA with SIC offering high throughput for a wider range of transmission probabilities.

Fig. 4.8 shows the network aggregate throughput performance of CSMA with SCF, CSMA

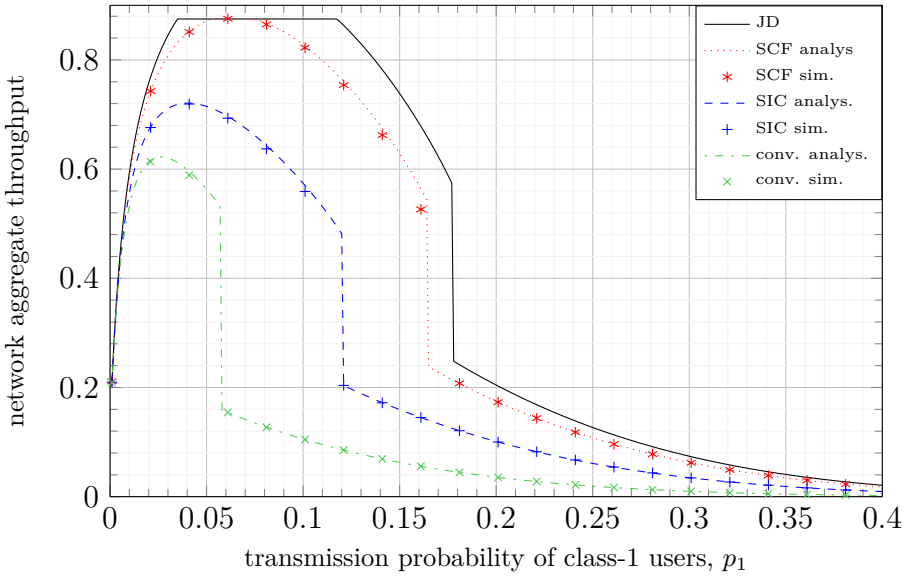

Figure 4.8: Throughput benefit of SCF-based CSMA compared to CSMA with SIC and conventional CSMA at SNR = 15 dB with message rate = 2, $N_1 = 12$, $N_2 = 8$, $\lambda_1 = 1/16$, $\lambda_2 = 1/64$, $p_2 = 1/4$, $K = 1$, $\kappa = 10$.

with SIC, conventional CSMA and CSMA with optimal JD for a network with $N_1 = 12$ class-1 users, $N_2 = 8$ class-2 users and SNR = 15 dB. CSMA with SCF clearly performs better than the other techniques and relatively close to optimal joint decoding. As the transmission probability of class-1 users p_1 varies from 0 to 1, eventually all the users become saturated (unstable queues) and the throughput of each user equals the rate function $R_v(\mathbf{1})$ for all $v = \{1, 2\}$.

We next look at the service delay and packet delay of symmetric MPR techniques in low and high SNR regimes. Figures 4.9, 6.3 demonstrates the service and packet delays as the arrival rate changes in a 2-class network with $N = 30$ ($N_1 = 20$ users in class-1) at SNR = 6 dB. We see that the analytical delay very well agrees with the simulation. Note that, in contrast to the existing work, the delay performance is presented for each class separately. In addition, it can be seen that CSMA with SCF is stable for a larger range of arrival rates. Similarly, the same behavior in terms of better performance of CSMA with

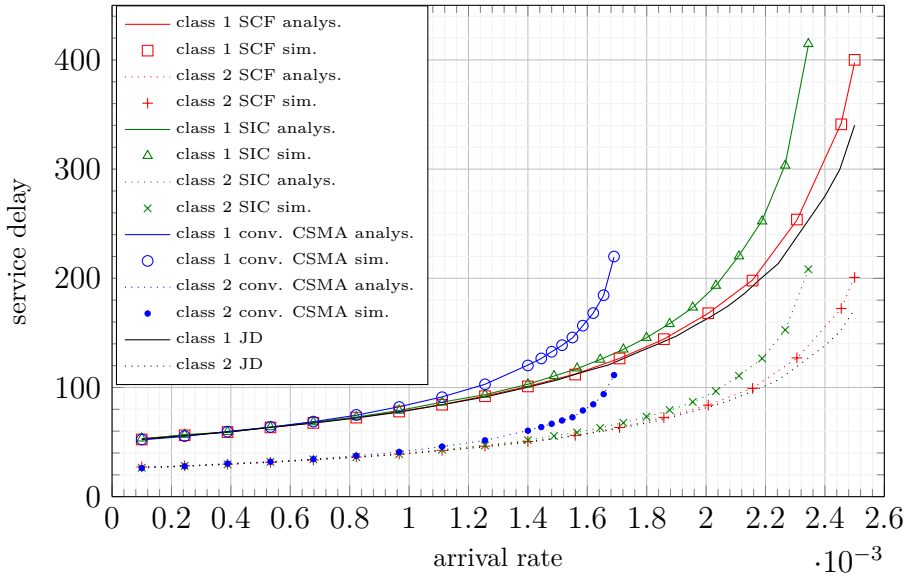

Figure 4.9: Service delay performance of inhomogeneous persistent CSMA schemes for a 2-class network at SNR $= 6\,\text{dB}$ with message rate $= 1$, $N_1 = 20$, $N_2 = 10$, $p_1 = 1/40$, $p_2 = 1/20$, $K = 1$, $\kappa = 10$.

SCF can be seen in Figures 4.11, 4.12 for the same network at SNR $= 15\,\text{dB}$.

4.5 Summary

In this chapter, we studied the inhomogeneous CSMA protocol with all-or-nothing symmetric MPR. In particular, we have derived throughput and delay expressions, which are asymptotically exact as the number of users grows. Based on these expressions, we have provided some theoretical guidelines for network design, and evaluated the performance of various MPR techniques in terms of throughput, packet delay and service delay. The results shed some light on the system design and performance analysis of MPR based CSMA systems, highlighting some interesting properties of MPR-based CSMA.

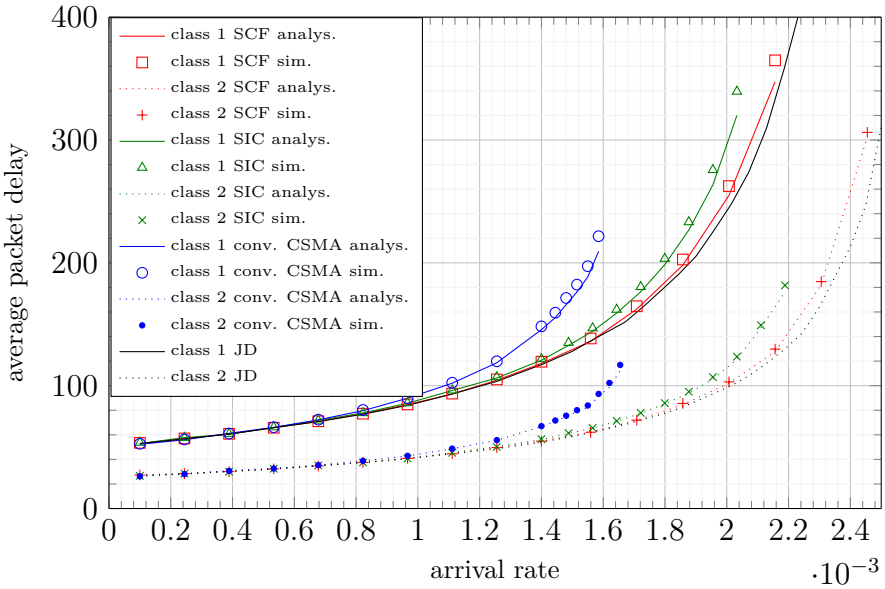

Figure 4.10: Average total delay (packet delay) performance of inhomogeneous persistent CSMA schemes for a 2-class network. The simulation setup is the same as that of Fig. 4.9.

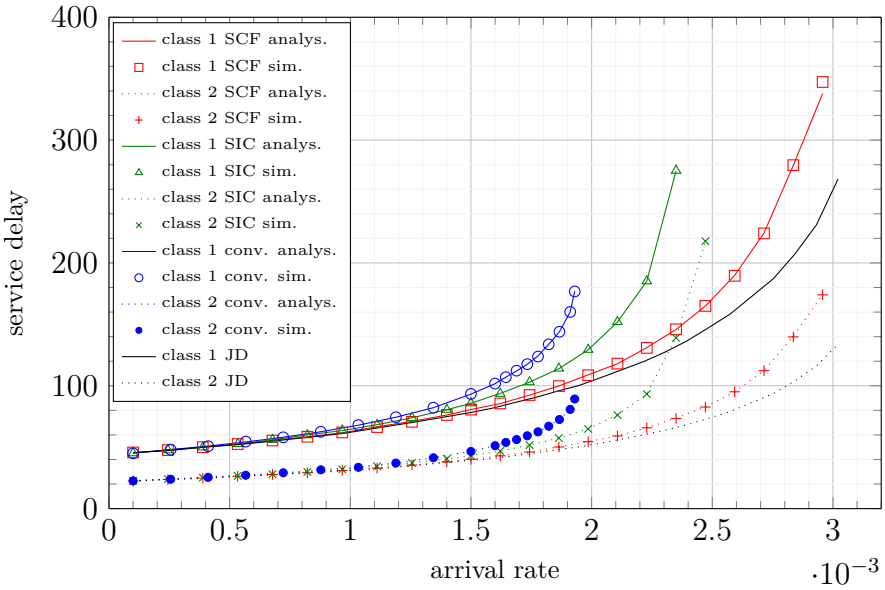

Figure 4.11: Service delay performance of inhomogeneous persistent CSMA schemes for a 2-class network at $\mathsf{SNR} = 15\,\mathrm{dB}$ with message rate $= 2$, $N_1 = 20$, $N_2 = 10$, $p_1 = 1/40$, $p_2 = 1/20$, $K = 1$, $\kappa = 10$.

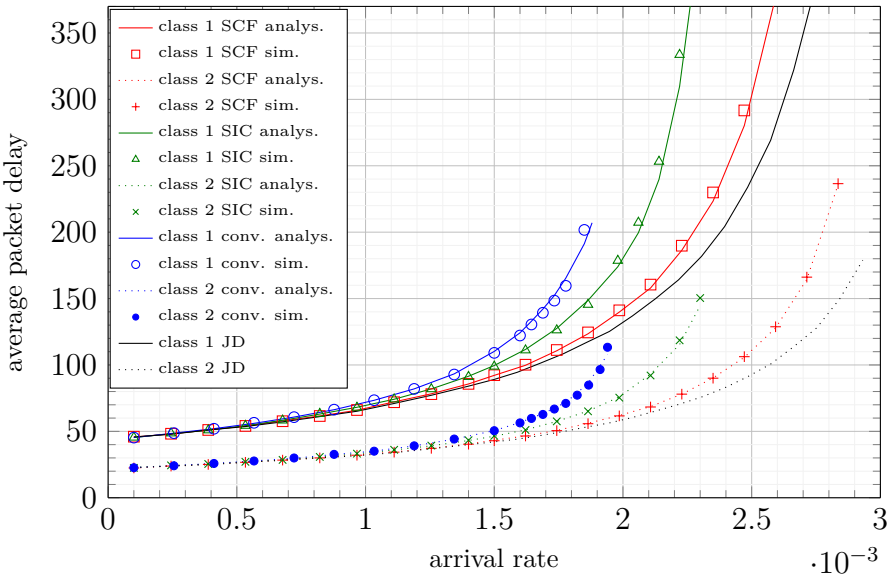

Figure 4.12: Average total delay (packet delay) performance of inhomogeneous persistent CSMA schemes for a 2-class network. The simulation setup is the same as that of Fig. 4.11.

Chapter 5

EXTENSION I: CSMA WITH RANDOM BACK-OFF

In this chapter, we extend the CSMA system model of Chapter 4 to a CSMA system with back-off mechanism in which a user transmission probability is determined by its back-off stage. The number of back-off stages can be finite or infinite. All the users are assume to be saturated, i.e., every user has a packet to transmit at all times. By changing their back-off stage, users in a CSMA system can reduce the risk of collision. The system is studied under the all-or-nothing symmetric MPR channel model assumption. We exploit a mean-field approximation to find the stationary (invariant) probability of the back-off stage of a user and show the accuracy of the approximation through simulations and numerical results.

5.1 System Model

Consider a CSMA random-access network of N users which contend to transmit packets to an access point (AP) equipped with M antennas. Assume that time is slotted and a packet can only be transmitted at the start of a slot. Packets are of constant length whose successful transmission (including a SIFS, DIFS and an ACK) takes exactly τ time slots. We study a slight variant of the persistent CSMA with a back-off mechanism.

Each user has a packet to transmit at all times, i.e., users are saturated. Suppose that there are L back-off stages $\mathcal{L} = \{0, 1, \cdots, L-1\}$ corresponding to L possible user transmission probabilities such that $1 > p_0^{(N)} > p_1^{(N)} > \cdots > p_{L-1}^{(N)} \geq 0$. At the beginning of each time slot, user i in the back-off stage ℓ_i transmits a packet with probability $p_{\ell_i}^{(N)}$ picked from the set of allowed probabilities $\{p_0^{(N)}, p_1^{(N)}, \cdots, p_{L-1}^{(N)}\}$ based on whether a collision occurred in the last time slot. If a collision occurs while user i in the back-off stage ℓ is transmitting a packet, it then changes its back-off stage to $\ell + 1$ hence reducing its transmission probability. This

is a mechanism to prevent a situation where transmissions from different users repeatedly collide with each other. If user i is already in stage $L-1$ and collision happens then its back-off stage becomes 0 in the following time slot. We assume that the back-off stage ℓ and its transmission probability are related as $p_\ell^{(N)}(t) = p_0^{(N)} c^{-\ell}$ where $c \in \mathbb{R}_+$ is called the back-off factor and $c \geq 2$.

At the end of each slot, the AP tries to recover the transmitted packets via a symmetric MPR technique such as compute-and-forward (C&F) and successive compute-and-forward (SCF). Specifically, we assume a all-or-noting MPR model where n concurrent transmissions can be recovered by the AP with probability q_n and the MPR capability is K.

5.2 A Mean-Field Approximation of CSMA with Back-off

Denote by $X_i^{(N)}(m)$ the back-off stage of user i at time slot m. Let $\boldsymbol{X}^{(N)}(m)$ be a vector whose i-th entry is $X_i^{(N)}(m)$. The system can be described by the Markov process $\left\{\boldsymbol{X}^{(N)}(m) : m = 0, 1, \cdots\right\}$. Since the evolution of system is symmetric with respect to back-off stage, equivalently we can use a new Markov chain $\left\{\boldsymbol{Q}^{(N)}(m) : m = 0, 1, \cdots\right\}$ which is the empirical measure of $\left\{X_i^{(N)}(m)\right\}_{i=1}^N$ defined as

$$Q_\ell^{(N)}(m) = \frac{1}{N} \sum_{j=1}^{N} 1_{\left\{X_j^{(N)}(m)=\ell\right\}}.$$

Although the Markov chain $\boldsymbol{Q}_\ell^{(N)}(m)$ is easy to describe, it is in general very difficult to analyze. In fact, the dynamics of the CSMA with back-off depends on the complex interaction amongst N queues, which makes it quite challenging to find the conditions that guarantee the existence of a steady state distribution for $\{\boldsymbol{Q}^{(N)}(m)\}$. As such, we study the behavior of the system asymptotically when N tends to infinity. To this end, we exploit a mean-field technique as we did in the previous chapters for slotted ALOHA and persistent CSMA. Mean-field approximation has been successfully applied in the context of communication and queueing networks [29, 53, 75, 77] where the mean-field decoupling assumption is justified. Assuming a mean-field approximation, the properly scaled process is convergent in

the limit when N grows large to a dynamical system described by a set of ordinary differential equations (ODEs). When the convergence holds, the states of any finite number of users become decoupled in the sense that the steady state distribution of $\{\boldsymbol{Q}^{(N)}(m)\}$ is equivalent to the product measure. Thus we have

$$\lim_{N\to\infty} Q_\ell^N(\lfloor Nt \rfloor) \longrightarrow Q_\ell(t)\,,$$

where $Q_\ell(t)$ is the solution to a dynamical system which describes the CSMA system asymptotically when N approaches infinity. Under certain conditions such a dynamical system is globally stable and its solution $Q_\ell(t)$ converges to an invariant point denoted by $Q_\ell^{\rm st}$.

5.2.1 Mean-Field Analysis

Consider a sequence of systems index by the number of users N. For any fixed N, the transmission probability of user i in back-off stage ℓ_i at time t is $p_{\ell_i}^{(N)}(t)$. In order to guarantee a fair share of throughput for each user we re-scale the transmission probabilities by $1/N$. So we assume that

$$Np_{\ell_i}^{(N)}(t) \to p_0 c^{-\ell_i},\ \text{as } N\to\infty \text{ and } \forall\, t\in\mathbb{Z}_+\,,$$

where $p_0 = \lim_{N\to\infty} Np_0^{(N)}$. That is, the transmission probability $p_i^{(N)}(t)$ scales like $p_0 c^{-\ell_i}/N$. Note that p_0 can be larger than 1, since it is no longer a probability. Next, we compute the transition probabilities for a typical user, say user i in back-off stage ℓ_i. As previously mentioned, after a successful transmission, the back-off stages of the users involved in packet transmission become zero, hence $p_{\ell_i}^{(N)}(t+1) = p_0^{(N)}$. When a collision occurs we have $p_{\ell_i}^{(N)}(t+1) = p_0^{(N)} c^{-(\ell_i+1)}$ if $\ell_i < L-1$ and $p_{\ell_i}^{(N)}(t+1) = p_0^{(N)}$ otherwise. Let $p_\ell = \lim_{N\to\infty} Np_\ell^{(N)}$ for all $\ell\in\mathcal{L}$. Let $\boldsymbol{\nu} = \boldsymbol{Q}^{(N)}(t)$ be the state of the system at time slot $\lfloor Nt \rfloor$. User i in back-off stage ℓ_i will have a successful transmission with probability $p_{\rm succ}^{(N)}(\ell_i,\boldsymbol{\nu})/N$ or experiences a collision with probability $p_{\rm col}^{(N)}(\ell_i,\boldsymbol{\nu})/N$. Let $\{i, j_1, \cdots, j_{K-1}\}$ be the indices of K users that

transmit their packets simultaneously. We can then write the transition probabilities for user i as follows.

$$p_{\text{succ}}^{(N)}(\ell_i, \boldsymbol{\nu})/N = p_{\ell_i}/N \Bigg(q_1 \prod_{j \neq i} (1 - p_{\ell_j}/N) + q_2 \sum_{j_1 \neq i} p_{\ell_{j_1}}/N \prod_{j \neq i, j_1} (1 - p_{\ell_j}/N) + \cdots$$

$$+ q_K \sum_{j_1, \cdots, j_K \neq i} \prod_{k=1}^{K-1} p_{\ell_{j_k}}/N \prod_{j \neq j_1, \cdots, j_{K-1}} (1 - p_{\ell_j}/N) \Bigg)$$

$$= p_{\ell_i}/N \prod_{j \neq i} (1 - p_{\ell_j}/N) \Bigg(q_1 + q_2 \sum_{j_1 \neq i} \frac{p_{\ell_{j_1}}/N}{1 - p_{\ell_{j_1}}/N} + \cdots$$
$$+ q_K \sum_{j_1, \cdots, j_K \neq i} \prod_{k=1}^{K-1} \frac{p_{\ell_{j_k}}/N}{1 - p_{\ell_{j_k}}/N} \Bigg)$$

$$= \frac{p_{\ell_i}/N}{1 - p_{\ell_i}/N} \prod_{\ell=0}^{L-1} (1 - p_\ell/N)^{N\nu_\ell} \Bigg(q_1 + q_2 \sum_{\ell=0}^{L-1} \binom{N\nu_\ell - 1_{\{\ell=\ell_i\}}}{1} \frac{p_\ell/N}{1 - p_\ell/N} + \cdots$$
$$+ q_K \sum_{\substack{i_0 + \cdots + i_{L-1} = K-1 \\ 0 \leq i_j \leq K-1 \, \forall j \in \mathcal{L}}} \prod_{\ell=0}^{L-1} \binom{N\nu_\ell - 1_{\{\ell=\ell_i\}}}{i_\ell} \left(\frac{p_\ell/N}{1 - p_\ell/N} \right)^{i_\ell} + o(1/N) \Bigg) . \tag{5.1}$$

Note that $\binom{N\nu_\ell}{i_\ell}$ tends to $(N\nu_\ell)^{i_\ell}/i_\ell!$ when N approaches infinity. From (5.1) and using multinomial theorem for large enough N we have

$$\overline{p}_{\text{succ}}(\ell_i, \boldsymbol{Q}(t)) \triangleq \lim_{N \to \infty} p_{\text{succ}}^{(N)}(\ell_i, \boldsymbol{\nu}) = p_{\ell_i} \Big(q_1 + \frac{q_2}{1!} \zeta(t) + \cdots + \frac{q_K}{(K-1)!} \zeta(t)^{K-1} \Big) e^{-\zeta(t)} , \tag{5.2}$$

where

$$\zeta(t) \triangleq \sum_{\ell=0}^{L-1} p_0 c^{-\ell} Q_\ell(t) ,$$

because the empirical measure converges to $\boldsymbol{Q}(t) = [Q_0(t), Q_1(t), \cdots]$ in the limit when $N \to \infty$. $\zeta(t)$ can be interpreted as the average attempt probability of a user at time t. When a packet transmission is not successful it can be due to a) packet loss or decoding failure at the receiver or b) collision. In the first case there is no collision and the packet is lost or dropped with the following probability

$$p_{\ell_i}\Big(1 - q_1 + \frac{1-q_2}{1!}\zeta(t) + \cdots + \frac{1-q_K}{(K-1)!}\zeta(t)^{K-1}\Big)e^{-\zeta(t)}\,.$$

Hence, collision probability is

$$\overline{p}_{\text{col}}(\ell_i, \boldsymbol{Q}(t)) = p_{\ell_i}\left(1 - \left(1 + \frac{\zeta(t)}{1!} + \cdots + \frac{\zeta(t)^{K-1}}{(K-1)!}\right)e^{-\zeta(t)}\right)\,. \tag{5.3}$$

By applying mean-field analysis we can show that the scaled process $Q^{(N)}(\lfloor Nt \rfloor)$ evolves according to the following dynamical system.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}Q_\ell(t) &= \overline{p}_{\text{col}}\left(\ell - 1, \boldsymbol{Q}(t)\right) Q_{\ell-1}(t) \\ &\qquad - \Big(\overline{p}_{\text{succ}}\left(\ell, \boldsymbol{Q}(t)\right) + \overline{p}_{\text{col}}\left(\ell, \boldsymbol{Q}(t)\right)\Big)Q_\ell(t)\,,\end{aligned} \tag{5.4}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}Q_0(t) &= \sum_{\ell=0}^{L-1}\overline{p}_{\text{succ}}\left(\ell, \boldsymbol{Q}(t)\right) Q_\ell(t) \\ &\qquad - \Big(\overline{p}_{\text{succ}}\left(0, \boldsymbol{Q}(t)\right) + \overline{p}_{\text{col}}\left(0, \boldsymbol{Q}(t)\right)\Big)Q_0(t) \\ &\qquad\quad + \overline{p}_{\text{col}}(L-1, \boldsymbol{Q}(t))Q_{L-1}(t)\,.\end{aligned} \tag{5.5}$$

The above dynamical system can be interpreted as follows. Suppose that a user is in some back-off stage, say $\ell \neq 0$. The first term $\overline{p}_{\text{col}}\left(\ell - 1, \boldsymbol{Q}(t)\right) Q_{\ell-1}(t)$ in (5.4) is the probability that a user in state (back-off stage) $\ell - 1$ transits to stage ℓ. On the other hand, the term $\Big(\overline{p}_{\text{succ}}\left(\ell, \boldsymbol{Q}(t)\right) + \overline{p}_{\text{col}}\left(\ell, \boldsymbol{Q}(t)\right)\Big)Q_\ell$ represents the probability that the back-off stage becomes 0 in case of a successful transmission or changes to $\ell + 1 \mod L$ due to a collision.

5.2.2 Stationary Distribution of a Back-off Stage

Define $f_{\boldsymbol{a}}(x) = \left(a_1 + \frac{a_2}{1!}x + \cdots + \frac{a_K}{(K-1)!}x^{K-1}\right)e^{-x}$ for some $x \in \mathbb{R}_+$ where $\boldsymbol{a} = [a_1, \cdots, a_K]$, $a_i \in [0,1]$ for all $1 \leq i \leq K$. Let $\boldsymbol{q} = [q_1, \cdots, q_K]$ an let $\mathbf{1}$ be the all-one vector of size K. In the following sections, we look at two cases to analyze the dynamical system (5.4, 5.5), namely 1) infinite back-off stage 2) finite back-off stage, and provide the conditions under which a stationary (invariant) distribution for the back-off stage exists.

5.2.3 Infinite Number of Back-off Stages

It can be confirmed that

$$\begin{aligned}
&\overline{p}_{\text{succ}}\left(0, \boldsymbol{Q}(t)\right) + \overline{p}_{\text{col}}\left(0, \boldsymbol{Q}(t)\right) \\
&\qquad = p_{\ell_i}\left(1 - \left(1 - q_1 + \frac{1-q_2}{1!}\zeta(t) + \cdots + \frac{1-q_K}{(K-1)!}\zeta(t)^{K-1}\right)e^{-\zeta(t)}\right) \\
&\qquad = p_{\ell_i}\Big(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\Big)\,.
\end{aligned}$$

Therefore, when there are infinite number of back-off stages, the system's behavior can be described by the following dynamical system

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_\ell(t) = p_{\ell-1}\Big(1 - f_{\mathbf{1}}(\zeta(t))\Big)Q_{\ell-1}(t) - p_\ell\Big(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\Big)Q_\ell(t)\,, \qquad \ell \in \{1, 2, \cdots\}\,; \quad (5.6)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_0(t) = \zeta(t) f_{\boldsymbol{q}}(\zeta(t)) - p_0\Big(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\Big)Q_0(t) \qquad \ell = 0\,. \quad (5.7)$$

Let $\boldsymbol{Q}^{\text{st}} = [Q_0^{\text{st}}, Q_1^{\text{st}}, \cdots]$ denote the invariant solution (fixed point) of the above dynamical system. Define $\zeta_s \triangleq \lim_{t\to\infty}\zeta(t)$. We find the fixed points as follows.

$$\left(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\right)p_\ell Q_\ell^{\text{st}} = \left(1 - f_{\mathbf{1}}(\zeta_s)\right)p_{\ell-1}Q_{\ell-1}^{\text{st}}\,, \qquad \ell \in \{1, 2, \cdots\}\,; \quad (5.8)$$

$$\left(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\right)p_0 Q_0^{\text{st}} = \zeta_s f_{\boldsymbol{q}}(\zeta_s)\,. \quad (5.9)$$

From (5.8) we get that

$$Q_\ell^{\rm st} = \left(\frac{c - cf_{\mathbf{1}}(\zeta_s)}{1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell Q_0^{\rm st}\,.$$

In order for $\boldsymbol{Q}^{\rm st}$ to be a probability distribution we must have, $\sum_{\ell\geq 0} Q_\ell^{\rm st} = 1$. Thus

$$Q_0^{\rm st} = \frac{1}{\sum_{\ell\geq 0}\left(\frac{c-cf_{\mathbf{1}}(\zeta_s)}{1-f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell}\,. \tag{5.10}$$

Assuming that the invariant solution exists, we have

$$\begin{aligned}\zeta_s = \sum_{\ell\geq 0} p_0 c^{-\ell} Q_\ell^{\rm st} &= \frac{\sum_{\ell\geq 0} p_0 c^{-\ell}\left(\frac{c-cf_{\mathbf{1}}(\zeta_s)}{1-f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell}{\sum_{\ell\geq 0}\left(\frac{c-cf_{\mathbf{1}}(\zeta_s)}{1-f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell} \\ &= p_0\frac{1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) - c(1 - f_{\mathbf{1}}(\zeta_s))}{f_{\mathbf{1}}(\zeta_s) - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\,.\end{aligned} \tag{5.11}$$

So, the fixed point of (5.8) and (5.9) is the solution to the following equation

$$\frac{\zeta_s}{p_0} = \frac{1 - c + f_{\boldsymbol{c}-\mathbf{1}+\boldsymbol{q}}(\zeta_s)}{f_{\boldsymbol{q}}(\zeta_s)}\,. \tag{5.12}$$

It is easy to verify that the first derivative of the right-hand side (RHS) of (5.12) is non-positive which implies that the RHS is a non-increasing function in ζ_s starting from $(0, 1)$. On the other hand, the left-hand side (LHS) of (5.12) is a linear function with intercept point at the origin. Thus, the RHS and the LHS functions always coincide at exactly one point.

We now turn to the case of finite number of back-off stages.

5.2.4 Finite Number of Back-off Stages

The ordinary differential equations for this case is as follows

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_\ell(t) = p_{\ell-1}\Big(1 - f_{\mathbf{1}}(\zeta(t))\Big)Q_{\ell-1}(t) - p_\ell\left(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\right)Q_\ell(t)\,, \quad 1 \leq \ell \leq L-1\,; \tag{5.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_0(t) = \zeta(t)f_{\boldsymbol{q}}(\zeta(t)) + p_{L-1}(1 - f_{\mathbf{1}}(\zeta(t))Q_{L-1}(t) - p_0\Big(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\Big)Q_0(t)\,. \tag{5.14}$$

The fixed point of the above dynamical system is obtained by solving the following equations

$$p_\ell \left(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\right) Q_\ell^{\text{st}} = \left(1 - f_{\mathbf{1}}(\zeta_s)\right) p_{\ell-1} Q_{\ell-1}^{\text{st}}\,, \qquad 1 \le \ell \le L-1\,; \tag{5.15}$$

$$\zeta_s f_{\boldsymbol{q}}(\zeta_s) = p_0\Big(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t))\Big) Q_0^{\text{st}} - (1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta(t)) p_{L-1} Q_{L-1}^{\text{st}}(t). \tag{5.16}$$

After some algebraic manipulation it can be see that a fixed point of the above dynamical system is the solution to the following equation

$$\zeta_s = \frac{\sum_\ell \left(\frac{1-f_{\mathbf{1}}(\zeta_s)}{1-f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell}{\sum_\ell \frac{1}{p_\ell}\left(\frac{1-f_{\mathbf{1}}(\zeta_s)}{1-f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)}\right)^\ell}\,. \tag{5.17}$$

Observe that in general, there may be multiple solutions to (5.17) in which case the system becomes meta-stable. This, for instance, depends on the choice of transmission probabilities p_ℓ.

The above discussions yields the following results.

5.3 Main Results

Focusing on the dynamical system representation of CSMA with back-off, here we present the main result.

Theorem 5.3.1 For a CSMA system with infinite number of back-off stages, and under the symmetric MPR model assumption, the stationary(invariant) distribution exists if

$$f_{\boldsymbol{c}-\mathbf{1}+\boldsymbol{q}}\left(\frac{c}{c-1}p_0\right) > c - 1\,.$$

Proof. See Appendix C.1. □

For the finite number of back-off stages a similar result follows:

Corollary 5.3.2 For a CSMA with finite number of back-off stages where $p_\ell = p_0 c^{-\ell}$ for $\ell \ge 0$ and under symmetric MPR model assumption, the stationary distribution exists.

Proof. See Appendix C.2. □

Now that we have the stationary distribution of the back-off stages, the system is similar to an inhomogeneous persistent CSMA which we discussed in chapters 4. We further have the assumption that the queues are saturated. Therefore, the throughput can be obtained from the results obtained in chapter 4.

Theorem 5.3.3 For large enough N, the average saturated throughput of CSMA with back-off under the symmetric MPR model is

$$R_{\text{B\_CSMA}} = \frac{\tau\zeta_s f(\zeta_s)}{\tau(1-e^{-\zeta_s})+e^{-\zeta_s}}\,. \tag{5.18}$$

Proof. Since the network is saturated by taking $\rho_i = 1$ for all i's. We have

$$P_{\text{idle}} = \lim_{N\to\infty} \mathbb{E}_{\boldsymbol{Q}^{\text{st}}}\left[\prod_{i=1}^{N}\left(1-\frac{p_i}{N}\right)\right] = e^{-\zeta_s}\,. \tag{5.19}$$

Furthermore, from (5.1) the average success probability of a user is

$$P_{\text{succ}} = \mathbb{E}_{Q^{\text{st}}}\left[\overline{p}_{\text{succ}}(\ell_i, \boldsymbol{Q}(t))\right] = \zeta_s f(\zeta_s)\,. \tag{5.20}$$

Thus

$$R_{\text{B\_CSMA}} = \frac{\tau P_{\text{succ}}}{\tau(1-P_{\text{idle}})+P_{\text{idle}}}\,. \tag{5.21}$$

The throughput then follows by substituting (5.19) and (5.20) into (5.21). □

5.3.1 Global Stability of the Dynamical System (5.13, 5.14)

In practice, the case of finite number of back-off stages is of more interest. For this reason we only focus on the stability of the dynamical system corresponding to this case. The

dynamical system (5.13, 5.14) is a nonlinear ODE, so we linearize it around its unique fixed point obtained as the solution to (5.12). If the linearized system is stable (all eigenvalues with negative real part) then the corresponding non-linear system will be asymptotically stable and since the fixed point is unique, the dynamical system will be globally stable.

Theorem 5.3.4 The dynamical system (5.13, 5.14) is globally stable and

$$\lim_{t\to\infty} Q_\ell(t) = Q_\ell^{\text{st}}, \qquad \forall\, \ell \in \{0, \ldots, L-1\} \tag{5.22}$$

Proof. See Appendix C.3. □

5.4 Simulations

To demonstrate the accuracy of the mean-field approximation we conduct simulations in this section. In the first scenario we assume that the network size is $N = 100$, the back-off factor $c = 2$ and the number of back-off stages is set to $L = 15$. The AP with $M = 2$ antennas decodes received packets using three MPR techniques, namely successive interference cancellation (SIC), compute-and-forward (C&F), successive compute-and-forward (SCF) an joint decoding (JD) with decoding probabilities $q_1 = 0.91$, $q_2 = 0.31$ and $q_1 = 0.91$, $q_2 = 0.61$ and $q_1 = 0.91$, $q_2 = 0.66$, and $q_1 = 0.91$, $q_2 = 0.8$, respectively. We then find the stationary distribution of each back-off stage and compute the ratio of the probabilities of being in stages 1, 2 and 3 to that of stage 0 as shown in Fig. 5.1. In addition, we compute the fixed point of the system from the equation (5.15) as shown in Fig. 5.1 by yellow markers. It can be seen that the mean-field approximation concentrates near the average of the probability of a back-off stage.

Table 5.1 provides the throughput results for SIC, C&F, SCF and JD. We find a close match between the mean-field approximation of the throughput and the actual throughput resulted from the simulations.

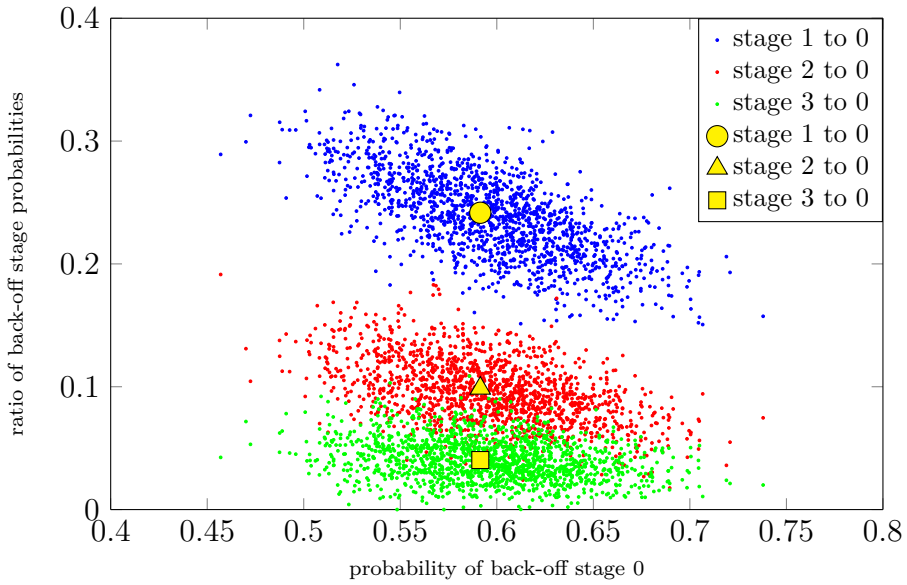

Figure 5.1: The ratio of the back-off stages 1, 2, 3 probabilities to that of stage 0. The yellow markers indicate the mean-field approximations.

Table 5.1: Throughput results.

throughput	SIC	C&F	SCF	JD
simulation	0.7076	0.8443	0.8676	0.9296
approximation	0.7074	0.8423	0.8655	0.9311

5.5 Summary

In this chapter, we study a CSMA system with back-off mechanism under all-or-nothing symmetric MPR channel model. Each user has a transmission probability corresponding to its back-off stage. The number of back-off stages can be finite or infinite. Furthermore, all the users are assumed to be saturated, i.e., every user has a packet to transmit at all times. Compared to a unsaturated system, the chance of collision is clearly higher in a saturated system. Back-off mechanism is aimed at reducing collision by way of adjusting transmission probabilities. Mean-field approximation is then used to find the stationary probability of the

back-off stages for both finite and infinite cases of back-off stages. Finally, the accuracy of the mean-field approximation is shown through simulations and numerical results.

Chapter 6

EXTENSION II: THE MULTIPLE AP CASE

The application of compute-and-forward (C&F) to random access scenarios – as we did in the previous chapters – indicates its potential as a powerful physical-layer network coding technique. To further bridge the gap between the theoretical advances of C&F and its real-world use, in this chapter, we extend the use of C&F from the case of single access point (AP) to the case of multiple APs, which is common in reality. We take into account several important practical constraints, such as channel variations, decentralized operations, and random data traffic. Two major challenges are identified, and novel solutions are proposed to address these challenges. In particular, we first introduce the AP cooperation problem to fully recover packets from packet collisions, and develop a distributed algorithm which is computationally efficient. We then introduce the joint channel estimation and active user recovery problem, and propose an algorithm based on spare recovery techniques. In addition, we provide accurate throughput and delay expressions for C&F-based CSMA protocols. These expressions, together with our trace-driven simulations, demonstrate the significant advantages of C&F-based CSMA over conventional CSMA.

6.1 Motivation

In chapters 2 and 3 we explored the use of C&F as a special case of the symmetric MPR model in random-access protocols, including slotted-ALOHA and CSMA, with a focus on the scenario with a single access point (AP). In this chapter, we aim to go a step further towards realizing the theoretical benefits of C&F in random-access networks, with a particular focus of the scenario with *multiple* APs.

The multi-AP scenario is motivated by the fact that today's random-access networks

are comprised of densely deployed APs. With existing random-access protocols, a dense deployment of APs does not scale well with the throughput demands. When one user (client) is transmitting to an AP, the nearby users have to remain silent to avoid packet collisions. The use of C&F in a multi-AP system helps to scale the network throughput. For instance, even if two users are transmitting their packets at the same time as illustrated in Fig. 6.1, with C&F each AP can decode a linear combination of the transmitted packets, and then forwards it to a central controller. As long as these two combinations are linearly independent, the controller is able to recover the original packets by solving a system of linear equations. In this way, the cooperation among APs increases the throughput, which is particularly useful in highly populated locations where packet collisions often occur.

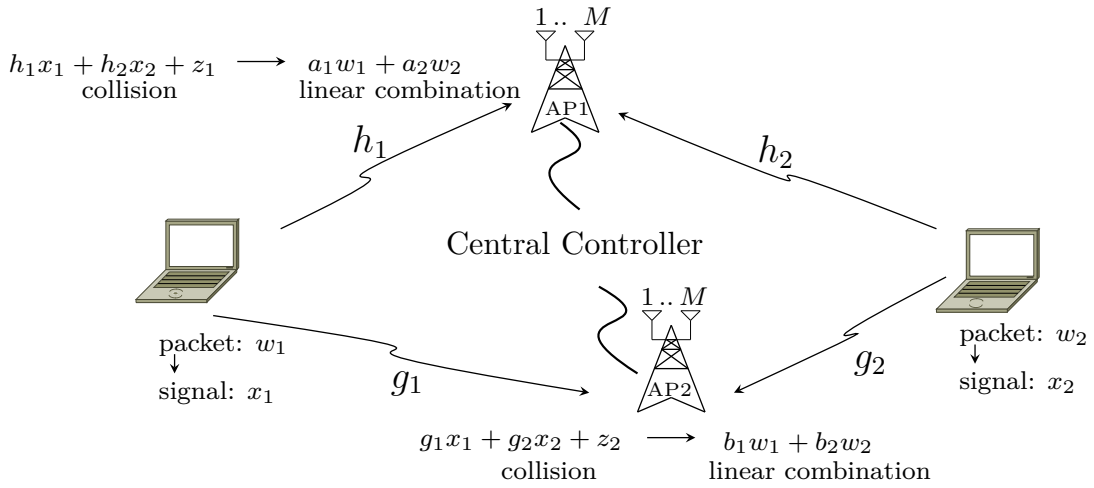

Figure 6.1: The use of C&F in a two-AP scenario.

Although the basic idea seems simple at first glance, in the multi-AP setting two major technical challenges remain open towards designing a practical random-access protocol with C&F. First, unlike the single-AP case (where the AP itself resolves packet collisions), each AP in a multi-AP system only decodes one linear combination. This poses a new question: *Which linear combination shall each AP decode?* Intuitively, if each AP in Fig. 6.1 simply

decodes the most reliable linear combination, then the linear combinations collected at the central controller could be linearly dependent, making it impossible to recover the original packets. In other words, the APs should collaborate with each other to ensure that each decoded linear combination is as reliable as possible, and at the same time, the collection of combinations is full rank.

In Section 6.2, we formally define this AP cooperation problem and develop a distributed solution that achieves the best possible reliability under the full-rank constraint. Our distributed solution is computationally efficient, and remarkably fast when the number of colliding packets is less than 5. Also, we show that our AP cooperation problem contains the network-coding design problem [14, 56] as a special case, and our new distributed solution has lower computational cost than existing solutions. Therefore, as a by-product of this work, we provide a more efficient solution to the network-coding design problem.

Second, the APs need to identify colliding users and estimate their channel conditions before decoding linear combinations. Conventional channel-estimation methods only work for the single-transmitter case, because they assign an *identical* training sequence to every user in the network. With multiple concurrent transmitters, one needs to assign *different* training sequences to different users. Thus the question is: *How shall we design such training sequences?* Our key observation is that, though the number of users can be very large in a dense network with multiple APs, the number of colliding users is often small under random-access protocols. The *sparsity* allows us to draw upon sparse recovery techniques developed in compressed sensing to design training sequences.

In Sec. 6.3, we formally define the channel estimation and active user recovery problem. We then propose to use an algorithm based on sparse recovery techniques. Our algorithm makes use of Reed-Muller signatures and has computational complexity *sub-linear* of the number of users.

Furthermore, we show that, once the above two challenges have been addressed, there is a unified perspective on the single-AP case and the multi-AP case. Based on this unified perspective, we are able to provide accurate throughput and delay expressions for C&F-based

CSMA protocols. These expressions can be used to tune system parameters (such as the transmission probabilities) and to characterize the advantages of C&F-based CSMA over conventional CSMA.

Finally, we conduct extensive simulations driven by a real-world channel condition trace collected from a 44-node wireless network [83]. Our simulation results show that C&F-based CSMA achieves a multi-fold increase in network throughput over a wide range of system parameters, and can greatly reduce the average delays.

6.2 C&F for the Multi-AP Case: AP Cooperation

In this section, we first define the AP cooperation problem and then develop a distributed solution. We also discuss the connection between our AP cooperation problem and the network-coding design problem [14, 56] (for multi-source multi-relay networks), and show that our distributed solution overcomes major shortcomings of existing methods for the network-coding design problem.

6.2.1 AP Cooperation Problem

Consider a network with N wireless users and K APs. Suppose that there are n active users involved in a packet collision, and the received signal at the kth AP is given by

$$\mathbf{y}_k = \sum_{\ell=1}^{n} h_{k\ell}\mathbf{x}_\ell + \mathbf{z}_k\,. \tag{6.1}$$

With C&F, when a collision occurs, the AP can decode multiple linear combinations. If these decoded linear combinations are full rank, the AP can recover the original packets by solving a system of linear equations. More specifically, suppose that a packet collision involves n active users and the AP tries to decode n (linearly independent) linear combinations with coefficient vectors $\{\mathbf{a}_i\}$. Then, the AP can recover the original packets as long as

$$\min_{i=1,\ldots,n} R_{\mathsf{comp}}(\mathbf{a}_i;\mathbf{h}) > R\,.$$

So, these coefficient vectors $\{\mathbf{a}_i\}$ should be chosen such that the value of $\min_i R_{\mathsf{comp}}(\mathbf{a}_i; \mathbf{h})$ is maximized. That is, the AP needs to solve the following optimization problem

$$\begin{aligned} \text{maximize} \quad & \min_{i=1,\ldots,n} R_{\mathsf{comp}}(\mathbf{a}_i; \mathbf{h}) \\ \text{subject to} \quad & \mathsf{rank}\,(\mathbf{A}) = n\,, \end{aligned} \tag{6.2}$$

where $\mathbf{A}$ is a matrix with $\mathbf{a}_i$ as its ith column.

This problem has been solved in Chapter 4 (also see [84,85]). Let $\{\mathbf{a}_i^*\}$ denote an optimal solution. Then the probability that the corresponding n linear combinations are successfully decoded is

$$q_n = \mathbb{P}\left[\min_i R_{\mathsf{comp}}(\mathbf{a}_i^*; \mathbf{h}) > R\right], \tag{6.3}$$

where the probability is taken over the choice of $\mathbf{h}$.

Unlike the single-AP case, the K APs in the system need to cooperate with each other. Recall that the kth AP can decode a linear combination with coefficient vector $\mathbf{a}_k$ as long as $R_{\mathsf{comp}}(\mathbf{a}_k; \mathbf{h}_k) > R$, where $\mathbf{h}_k = (h_{k1}, \ldots, h_{kn})^T$ is the channel-gain vector. Let $\mathbf{A}$ be a matrix with $\mathbf{a}_k$ at its kth column. Clearly, if $\mathbf{A}$ is of rank n, then the central controller can recover all the original packets by solving a system of linear equations. Therefore, all the APs need to solve the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \min_{k=1,\ldots,K} R_{\mathsf{comp}}(\mathbf{a}_k; \mathbf{h}_k) \\ \text{subject to} \quad & \mathsf{rank}\,(\mathbf{A}) = n\,. \end{aligned} \tag{6.4}$$

6.2.2 Our Distributed Solution

Here, we develop a distributed algorithm for the AP cooperation problem (6.4), which consists of two steps. In the first step, the kth AP computes a so-called dominant solution as defined in [18]:

$$\Omega_k = \left\{\mathbf{a}_k^{(1)}, \mathbf{a}_k^{(2)}, \ldots, \mathbf{a}_k^{(n)}\right\}, \tag{6.5}$$

where $\mathbf{a}_k^{(1)}, \ldots, \mathbf{a}_k^{(n)}$ have the following property

$$\begin{aligned}\mathbf{a}_k^{(1)} &= \arg\max\{R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h}_k) \mid \mathbf{a} \text{ is nonzero}\}, \\ \mathbf{a}_k^{(2)} &= \arg\max\{R_{\mathsf{comp}}(\mathbf{a}; \mathbf{h}_k) \mid \mathbf{a} \text{ and } \mathbf{a}_k^{(1)} \text{ are linearly indep.}\},\end{aligned}$$

and so on. That is, $\mathbf{a}_k^{(1)}$ has the largest computation rate among all non-zero vectors, and $\mathbf{a}_k^{(2)}$ has the largest computation rate among all vectors linearly independent of $\mathbf{a}_k^{(1)}$, and so on.

When $n \leq 4$, this step can be implemented by using the lattice-reduction method proposed in [86], which is remarkably fast compared to alternative methods. When $n > 4$, this step can be implemented by using the greedy search algorithm presented in [18].

In the second step, the destination collects all the solution sets $\{\Omega_k\}_{k=1}^K$, and then tries to pick up n vectors from n distinct sets such that these n vectors are linearly independent and the minimum value of their computation rates is maximized. This step can be easily done. For instance, consider a special case of $n = 2$ and $K = 3$. Suppose that $\mathbf{a}_1^{(1)} = (1, 1)$, $\mathbf{a}_1^{(2)} = (1, 0)$, $\mathbf{a}_2^{(1)} = (1, 1)$, $\mathbf{a}_2^{(2)} = (0, 1)$, $\mathbf{a}_3^{(1)} = (1, 0)$, $\mathbf{a}_3^{(2)} = (2, 1)$, and they have the following order based on their computation rates

$$\mathbf{a}_1^{(1)} \succ \mathbf{a}_2^{(1)} \succ \mathbf{a}_1^{(2)} \succ \mathbf{a}_3^{(1)} \succ \mathbf{a}_2^{(2)} \succ \mathbf{a}_3^{(2)},$$

where $\mathbf{a}_1^{(1)} \succ \mathbf{a}_2^{(1)}$ means $R_{\mathsf{comp}}(\mathbf{a}_1^{(1)}; \mathbf{h}_1) \geq R_{\mathsf{comp}}(\mathbf{a}_2^{(1)}; \mathbf{h}_2)$. Then, the second step will output $\{\mathbf{a}_1^{(2)}, \mathbf{a}_2^{(1)}\}$, because they are linearly independent and we cannot do better (since $\mathbf{a}_1^{(1)}$ and $\mathbf{a}_2^{(1)}$ are linearly dependent).

6.2.3 Proof of Correctness

We only present the proof for the case $n = 2$. Other cases can be proved similarly, which require a more involved characterization of linear dependency.

Let $\{\mathbf{a}_k^*\}_{k=1}^K$ be an optimal solution to Problem (6.4). Then, we have $\mathsf{rank}(\mathbf{A}^*) = 2$.

That is, there exist two linearly independent vectors in $\{\mathbf{a}_k^*\}_{k=1}^K$. Without loss of generality, we assume that $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ are linearly independent and $R_{\mathsf{comp}}(\mathbf{a}_1^*;\mathbf{h}_1) \geq R_{\mathsf{comp}}(\mathbf{a}_2^*;\mathbf{h}_2)$. Furthermore, we can show that $R_{\mathsf{comp}}(\mathbf{a}_2^*;\mathbf{h}_2) = R_{\mathsf{comp}}(\mathbf{A}^*)$.

To establish the optimality of our algorithm, we only need to show that our algorithm will produce a solution with minimum computation rate no worse than $R_{\mathsf{comp}}(\mathbf{a}_2^*;\mathbf{h}_2)$. In particular, we will show that we can pick up $\mathbf{a}_i \in \Omega_i$ $(i = 1, 2)$ such that $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent and the minimum computation rate $\min\{R_{\mathsf{comp}}(\mathbf{a}_1;\mathbf{h}_1), R_{\mathsf{comp}}(\mathbf{a}_2;\mathbf{h}_2)\}$ is no smaller than $R_{\mathsf{comp}}(\mathbf{a}_2^*;\mathbf{h}_2)$.

To this end, we consider the following four cases based on whether $\mathbf{a}_i^*$ is linearly independent of $\mathbf{a}_i^{(1)}$:

1. $\mathbf{a}_i^*$ is linearly dependent of $\mathbf{a}_i^{(1)}$ for $i = 1, 2$.

2. $\mathbf{a}_i^*$ is linearly independent of $\mathbf{a}_i^{(1)}$ for $i = 1, 2$.

3. Only $\mathbf{a}_1^*$ is linearly dependent of $\mathbf{a}_1^{(1)}$.

4. Only $\mathbf{a}_2^*$ is linearly dependent of $\mathbf{a}_2^{(1)}$.

For Case 1), we choose $\mathbf{a}_i = \mathbf{a}_i^{(1)}$ $(i = 1, 2)$. Since $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ are linearly independent, so are $\mathbf{a}_1$ and $\mathbf{a}_2$. Moreover, by the property of $\mathbf{a}_i^{(1)}$, we have $R_{\mathsf{comp}}(\mathbf{a}_i;\mathbf{h}_i) \geq R_{\mathsf{comp}}(\mathbf{a}_i^*;\mathbf{h}_i)$. Hence, our solution $\{\mathbf{a}_1, \mathbf{a}_2\}$ is no worse than $\{\mathbf{a}_1^*, \mathbf{a}_2^*\}$.

For Case 2), we choose $\mathbf{a}_1 = \mathbf{a}_1^{(1)}$, and $\mathbf{a}_2 = \mathbf{a}_2^{(1)}$ if $\mathbf{a}_2^{(1)}$ is linearly independent of $\mathbf{a}_1$, otherwise $\mathbf{a}_2 = \mathbf{a}_2^{(2)}$. Clearly, the selected $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent. Moreover, by the property of $\mathbf{a}_1^{(1)}$, $R_{\mathsf{comp}}(\mathbf{a}_1;\mathbf{h}_1) \geq R_{\mathsf{comp}}(\mathbf{a}_1^*;\mathbf{h}_1)$; similarly, $R_{\mathsf{comp}}(\mathbf{a}_2;\mathbf{h}_2) \geq R_{\mathsf{comp}}(\mathbf{a}_2^*;\mathbf{h}_2)$. Once again, our solution $\{\mathbf{a}_1, \mathbf{a}_2\}$ is no worse than $\{\mathbf{a}_1^*, \mathbf{a}_2^*\}$.

For Case 3), we use the same choice as for Case 2), and again, $\{\mathbf{a}_1, \mathbf{a}_2\}$ is no worse than $\{\mathbf{a}_1^*, \mathbf{a}_2^*\}$. Finally, Case 4) is very similar to Case 3), and the proof is essentially the same.

6.2.4 A Related Problem

We note that the AP cooperation problem generalizes the network-coding design problem for multi-source multi-relay networks with C&F studied in [14, 15, 56]. Although their setup looks similar to ours, their focus is on the information-theoretic analysis, whereas our focus is on the networking performance under practical constraints. The network coding design problem for multi-source multi-relay networks with C&F is defined as follows.

Consider a system where n sources are transmitting packets to a single destination through n relays. Assume that the final destination has (wired) backhaul links to the relays. The kth relay ($k = 1, \ldots, n$) decodes a linear combination with coefficient vector $\mathbf{a}_k$ and then forwards it to the destination. The network-coding design problem is to choose the coefficient vectors $\{\mathbf{a}_k\}$ such that they are linearly independent and minimum value of their computation rates is maximized.

Clearly, when $n = K$, the AP cooperation problem (6.4) reduces to the network-coding design problem. However, existing solutions in the literature are not satisfactory. For instance, [56] noticed that the network-coding design problem is an instance of integer programming (IP), and suggested some standard methods for solving IP problems, including the cutting-plane method and the branch-and-bound method. These methods, however, are generally of high computational complexity, because they do not take into account any special structure of the network-coding design problem.

[14] proposed a distributed algorithm for the network-coding design problem, which also consists of two steps. In the first step, each relay k attempts to find a candidate set $\Omega_k = \left\{\mathbf{a}_k^{(1)}, \mathbf{a}_k^{(2)}, \ldots, \mathbf{a}_k^{(T)}\right\}$, where $\mathbf{a}_k^{(1)}$ has the largest computation rate, $\mathbf{a}_k^{(2)}$ has the second largest computation rate, and so on. Here, T is a parameter to control the length of the candidate set, which is set by experience. (Note that T is typically much larger than n as explained in [14].) Once T is set, this step can be implemented by using the Fincke-Pohst method [14]. In the second step, the destination collects all the candidate sets $\{\Omega_k\}_{k=1}^n$, and then tries to pick up n linearly independent vectors (one from each candidate set) such that

the minimum value of their computation rates is maximized.

However, compared to our distributed algorithm, there are two major shortcomings of the above distributed algorithm :

1. It is unclear how to set the parameter T.

2. The complexity of the Fincke-Pohst method is high.

In other words, our distributed algorithm provides a more efficient solution to the network-coding design problem.

6.3 Joint Channel Estimation and Active User Recovery

Previously, we assume that the channel coefficients and the set of active users are known to the APs. (This assumption is also made in [84] and [85] for the single-AP case.) In this section, we show how to relax this assumption by making use of sparse recovery techniques.

6.3.1 Joint Channel Estimation and Active User Recovery

Let N be the number of clients in the network. Let $\mathbf{s}_\ell \in \mathbb{C}^T$ be a unique codeword (signature) of length T symbols, assigned to the ℓ-th client. This signature is used as the header for every packet of the ℓ-th client. Clearly, the smaller T is, the lower the overhead. In particular, we assume that $T \ll N$ in a dense network. The header of the received signal at the kth AP is

$$\begin{aligned} \mathbf{y}_k &= \sum_{\ell=1}^{N} h_{k\ell} z_\ell \mathbf{s}_\ell + \mathbf{n} \\ &= \mathbf{S}\mathbf{H}_k \mathbf{z} + \mathbf{n}\,, \end{aligned} \tag{6.6}$$

where $h_{k\ell}$ is the channel coefficient from the ℓth client to the kth AP, $\mathbf{S} = [\mathbf{s}_1|\cdots|\mathbf{s}_N]$ is a matrix whose columns are signatures, $\mathbf{H}_k \in \mathbb{C}^{N\times N}$ is a diagonal matrix representing channel coefficients from users to the kth AP, $\mathbf{z} \in \{0,1\}^N$ represents client activity state, i.e., $z_i = 1$

if client i is active, and $z_i = 0$ if it is inactive, and lastly $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is additive white Gaussian noise (AWGN).

The objective of the kth AP is to estimate the set $\mathbf{z}$ of active users as well as their corresponding channel coefficients. Hence, we call it the joint channel coefficient estimation and active client recovery problem.

6.3.2 Connection with Sparse Recovery

We note that Eq. (6.6) can be rewritten as

$$\mathbf{y}_k = \mathbf{S}\mathbf{v}_k + \mathbf{n}\,, \tag{6.7}$$

where $\mathbf{v}_k = \mathbf{H}_k\mathbf{z}$. The kth AP tries to estimate $\mathbf{v}_k$, which contains the information of active users and their channel coefficients. Clearly, this is a sparse recovery problem as we need to reconstruct vector $\mathbf{v}_k$ which a $N \times 1$ high dimensional vector and the columns of matrix $\mathbf{S}$ are $T \times 1$ low dimensional vectors ($T \ll N$) representing the signatures.

Compressed sensing (CS) and sparse recovery techniques aim at reconstruction of a high dimensional sparse vector using a small number of random linear measurements [87]. CS-based Multi-User Detection (MUD) in the random-access channels have been investigated in a number of works [88,89]. In [88] the high dimensional vector space is the message space and the measurements are collided signatures taken from an i.i.d. Gaussian codebook. Capacity bounds and minimum number of measurements are then characterized for the on-off random multiple access channel.

Fletcher *et. al.* [88] used an i.i.d random Gaussian codebook with sequential orthogonal matching pursuit (SeqOMP) which is analogous to successive interference cancellation decoding. For practical applications, however, random codebook matrix incurs large memory size. Furthermore, the active power shaping proposed in [88] is not a realistic assumption as users are not aware of the channel coefficients. Alternatively, deterministic codebook matrices have a much more efficient storage requirement and fast reconstruction/decoding

time. Reed-Muller codes with a very fast reconstruction algorithm are proposed in [90]. The complexity of Reed-Muller decoding depends on the number of measurements T and not on the message space dimension N [90]. So, we propose to use deterministic signatures constructed from second-order Reed-Muller codes where the decoding complexity is sub-linear in the dimension of the message space N.

6.3.3 MUD with Reed-Muller Signatures

The rth-order Reed-Muller code $\mathrm{RM}(r, m)$ over $\mathbb{F}_2$ for $r \geq 2$ is a linear error-correcting code whose message space consists of degree r polynomials over $\mathbb{F}_2$ in m variables [91]. Thus, each codeword is of length $T = 2^m$ symbols over $\mathbb{F}_2$. For the details of Reed-Muller code construction see [92].

Howard *et. al.* [90] used $\mathrm{RM}(2, m)$ to construct a deterministic sensing matrix and proposed a fast reconstruction algorithm. Based on their construction, we assume that each client is assigned a unique signature of length n bits consisting of two binary vectors $\mathbf{b} \in \mathbb{Z}^{n_1}$ and $\mathbf{c} \in \mathbb{Z}^{n_2}$. $\mathbf{c}$ is then mapped to a $m \times m$ symmetric matrix $\mathbf{P}(\mathbf{c})$ as follows

$$\mathbf{P}(\mathbf{c}) = \sum_{i=1}^{n_2} c_i \mathbf{B}(i) \mod 2\,, \tag{6.8}$$

where c_i is the ith entry of vector $\mathbf{c}$ and $\mathbf{B}(i)$ is the ith element of the ordered set $\mathbf{B}$ of basis for the space of $m \times m$ binary symmetric matrices [93].

Let $\phi_{\mathbf{b},\mathbf{c}}$ be a $\mathrm{RM}(2, m)$ codeword, we then have

$$\phi_{\mathbf{b},\mathbf{c}}(\mathbf{a}) = \exp\left[j\pi\left(\frac{1}{2}\mathbf{d}^{\mathsf{T}}\mathbf{P}(\mathbf{c})\mathbf{d} + \tilde{\mathbf{b}}^{\mathsf{T}}\mathbf{d}\right)\right], \tag{6.9}$$

where $\tilde{\mathbf{b}} \in \mathbb{Z}^m$ is formed by padding $\mathbf{b}$ with $m - n_1$ zeros, and codewords are indexed by $\mathbf{d} \in \mathbb{Z}^m$.

For the codebook $\mathrm{RM}(2, m)$, the set of basis $\mathbf{B}$ of size $m(m+1)/2$ can be calculated in advance and stored at each client where it is used to compute $\mathbf{P}(\mathbf{c})$ requiring $n_2 m^2$ operations.

For decoding RM$(2, m)$ we modify the scheme described in Algorithm 4 in [93]. The main steps of the decoding algorithm are as follows:

1. The first 2^m symbols of a client's packet constitutes its signature which is input to the Reed-Muller decoder where $\hat{\mathbf{b}}_j$ and $\hat{\mathbf{c}}_j$ are then estimated and used to detect an active client. The detected active client that appears more than any other at the output of the K APs (majority vote) is then added to the set of current detected active users whose corresponding codewords form the columns of the matrix $\mathbf{S}_{\hat{\mathcal{D}}}$.

2. Channel coefficients $\mathbf{h}_k$ for all $k \in \{1, \cdots, K\}$ are estimated using the set of current detected active users. The estimation is based on minimization of $\|\mathbf{y}_k - \mathbf{S}_{\hat{\mathcal{D}}}\mathbf{h}_k\|_2$. Update the residual signal $\mathbf{r}_k = \mathbf{y}_k - \mathbf{S}_{\hat{\mathcal{D}}}\mathbf{h}_k$.

3. If $\|\mathbf{r}_k\| < \epsilon$ for some pre-set small $\epsilon > 0$ then $\hat{\mathcal{D}}_k$ is the final estimate of the set of active users observed by the kth AP. Otherwise the residual signal is used as the input to the RM decoder. This procedure is repeated for no more than a pre-determined maximum number of iterations $i_{\max}$ or until the energy of the residual signal is below ϵ; whichever is satisfied earlier.

4. Finally the APs exchange their set of active users with each other and their intersection will be the set of current active users based on which the channel coefficients are re-estimated.

The proposed channel coefficients estimation and active client set detection scheme is summarized in Algorithm 2.

6.4 A Unified Analysis of Single-AP and Multi-AP Cases

In this section, we will show that, once the AP cooperation problem and the channel estimation problem have been solved, there is a unified perspective on the performance analysis of the single-AP case and the multi-AP case.

Algorithm 2 Channel coefficient estimation and active client set detection algorithm

Input: $\mathbf{y}_1, \ldots, \mathbf{y}_K$

Output: Channel coefficients $\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K$, active client set $\hat{\mathcal{D}}$.

Initialize: $\mathbf{r}_k = \mathbf{y}_k$, $\hat{\mathcal{D}}_k = \emptyset$, $\hat{\mathcal{D}} = \emptyset$
for all $k \in \{1, \ldots, K\}$ **do**
  while $i \leq i_{\max}$ and $\|\mathbf{r}_k\|_2 > \epsilon$ **do**
    Run RM decoder with $\mathbf{r}_k$ as input to detect the index of a client d.
    $\hat{\mathcal{D}}_k \leftarrow \hat{\mathcal{D}}_k \cup \{d\}$
    Form the matrix $\mathbf{S}_{\hat{\mathcal{D}}_k}$ by placing the codewords corresponding to $\hat{\mathcal{D}}_k$ as the columns.
    Find the channel coefficients:

$$\hat{\mathbf{h}}_k \leftarrow \arg\max_{\mathbf{h}_k} \|\mathbf{y}_k - \mathbf{S}_{\hat{\mathcal{D}}_k}\mathbf{h}_k\|_2$$

    Update the residual $\mathbf{r}_k \leftarrow \mathbf{r}_k - \mathbf{y}_k$
  end while
end for
APs' pass $\hat{\mathcal{D}}_k$ among themselves over a backhaul link
Each AP computes the final set of active users as $\hat{\mathcal{D}} = \cap_k \hat{\mathcal{D}}_k$
for all $k \in \{1, \ldots, K\}$ **do**
  Refine channel coeffs. $\hat{\mathbf{h}}_k \leftarrow \arg\max_{\mathbf{h}_k} \|\mathbf{y}_k - \mathbf{S}_{\hat{\mathcal{D}}}\mathbf{h}_k\|_2$
end for
return $\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K$ and $\hat{\mathcal{D}}$

Let $\{\mathbf{a}_k^*\}$ be an optimal solution to the AP cooperation problem (6.4). Suppose that there are n active users. Then, the destination is able to recover n original packets as long as $\min_k R_{\mathsf{comp}}(\mathbf{a}_k^*; \mathbf{h}_k) > R$. Let q_n denote the probability that these n original packets are recoverable. Then, we have

$$q_n = \mathbb{P}\left[\min_k R_{\mathsf{comp}}(\mathbf{a}_k^*; \mathbf{h}_k) > R\right], \tag{6.10}$$

where the probability is taken over the choice of $\{\mathbf{h}_k\}$.

Clearly, these success probabilities $\{q_n\}$ abstract the decoding operations at multiple APs, and thus determine the throughput and delay performance. In this sense, a multi-AP system can be viewed as a single-AP system with success probabilities $\{q_n\}$ computed in

(6.10) (rather than calculated in (6.3)). With this perspective, we can apply Proposition 4.3.5 and Proposition 4.3.7 to our multi-AP scenario, obtaining accurate throughput and delay expressions. These expressions can then be used as a theoretical guideline to tune the transmission probabilities $\{p_v\}$ and to characterize the advantages of C&F-based CSMA protocols over conventional CSMA. We will illustrate the first point through the following example. The second point is discussed in the next section.

Example 13 Consider a 2-class network where N_1 users belong to class 1 and N_2 users are in class 2 with fixed arrival rates λ_1 and λ_2 respectively. Suppose that we want the clients in both classes to experience the same total delay, i.e., $T_1 = T_2$. To achieve this, we first compute x_1^* and x_2^* as we did in example one. From (4.17) for $v \in \{1, 2\}$ we then have

$$\rho_v^* = \frac{T_v - \frac{\tau-1}{2}\left(1 - P^{\text{IDLE}}(x_1^*, x_2^*)\right)}{T_v + \frac{1}{\lambda_v} - \frac{1}{\tau}},$$

Finally $p_v^* = \frac{x_v^*}{\rho_v^*}$ gives the proper transmission probabilities such that all clients experience identical total delay.

6.5 Trace-Driven Simulations

To understand the performance of C&F-based CSMA protocols under realistic channel variations, we conduct simulations using a real-world channel condition trace [83]. The trace, collected from a 44-node wireless network in an office area, contains over 9300 measured channel impulse responses (CIR) for a total of $44 \times 43 = 1892$ pair-wise links.

We strategically pick 2 nodes as 2 APs, and the remaining 42 nodes as wireless users. Furthermore, we assume that 2/3 of the users are in class 1 ($N_1 = 28$) and the rest are class-2 users ($N_2 = 14$). The transmission of a packet is assumed to take $\tau = 10$ time slots. During the subsequent simulations, we generate channel coefficients for these sender-receiver (client-AP) pairs based on the statistics obtained from the trace. Specifically, for each client-AP pair, we first collect all the measured CIRs from the trace, based on which we estimate the variance of its channel gains. The estimated variance captures the large-scale fading

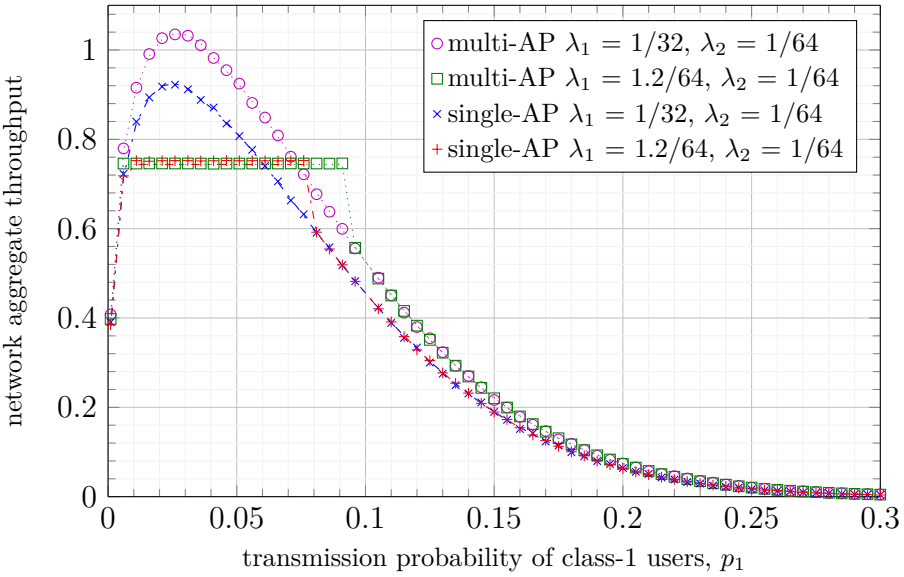

Figure 6.2: Single AP v.s. multi-AP aggregate throughput for the 2-class network with $\tau = 10$, $p_2 = 1/42$, $N_1 = 28$ and $N_2 = 14$.

effect (i.e., the path loss) [94]. To model the small-scale fading effect, we use the Rayleigh fading model [94]. Based on the traces we compute the conditional probabilities for the 2-AP scenario as $q_1 = 0.9946$, $q_2 = 0.8800$.

We first look at the network aggregate throughput for single AP and multi-AP scenarios. Fig. 6.2 illustrates the network aggregate throughput for class-1 users while varying the transmission probability of class-1 users with the transmission probability of class-2 users fixed at $p_2 = 1/42$. It can be seen that depending on the traffic load in class-1 and class-2, either class or both can become stable or unstable, and the aggregate throughput closely matches our analytical result. In addition, the multi-AP scheme offers a better aggregate throughput compared to single-AP case as a result of AP cooperation.

We then compare the delay performance for single AP and multi-AP scenarios. Fig. 6.3 depicts the total delay experienced by a packet in the two-class network. It is evident that the delay analysis matches the simulations for both classes quite well. The multi-AP scenario also supports higher arrival rates for a given finite delay. Note that the delay performance is

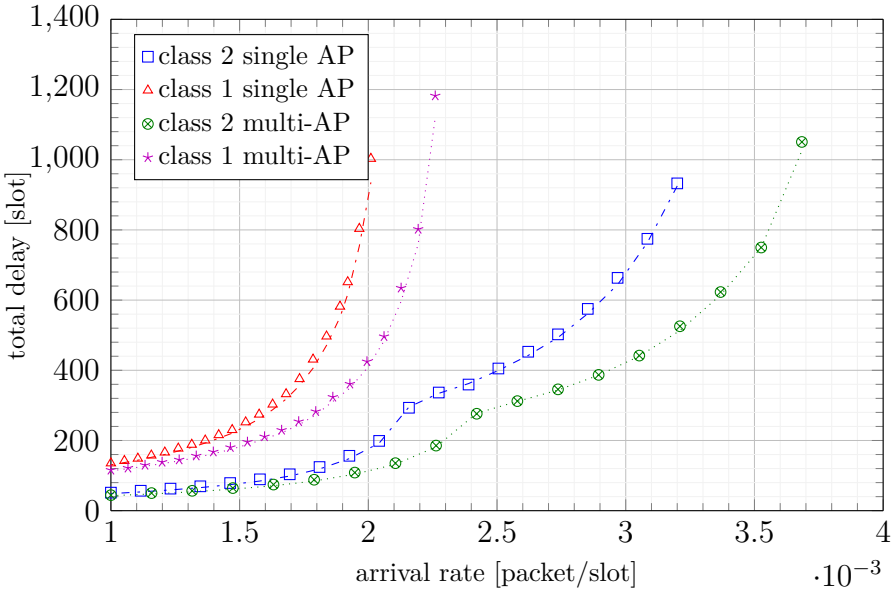

Figure 6.3: Total delay of a packet for the 2-class network. The simulation setup is the same as that of Fig. 6.2.

given for each class separately. In other words, the simple expression provided for the total delay enables us to compute the delay for a user in any class. This is very helpful for system designers—as we have seen in the previous two examples presented in Section 6.4—to tune system parameters like transmission probabilities.

Next, we compare the aggregate throughput performance of C&F-based CSMA with that of conventional CSMA in Fig. 6.4. The performance gain we obtain by using C&F is considerably good. For instance, in $\lambda_1 = 0.015$ and $\lambda_2 = 0.015$, the CSMA with C&F gives a much higher aggregate throughput for a larger range of transmission probabilities of class-1 users. As an example, at $p_1 = 0.1$, aggregate throughput offered by the multi-AP scenario with C&F-based CSMA is 4.7 times greater than that of conventional CSMA, indicating that the throughput scales well with our C&F-based scheme.

Finally, we look into the benefit of AP cooperation as opposed to a non-cooperative multi-AP case where each AP decodes the most reliable linear combination *independently* and then forwards it to the central controller. As can be seen in Fig. 6.5, the cooperation between

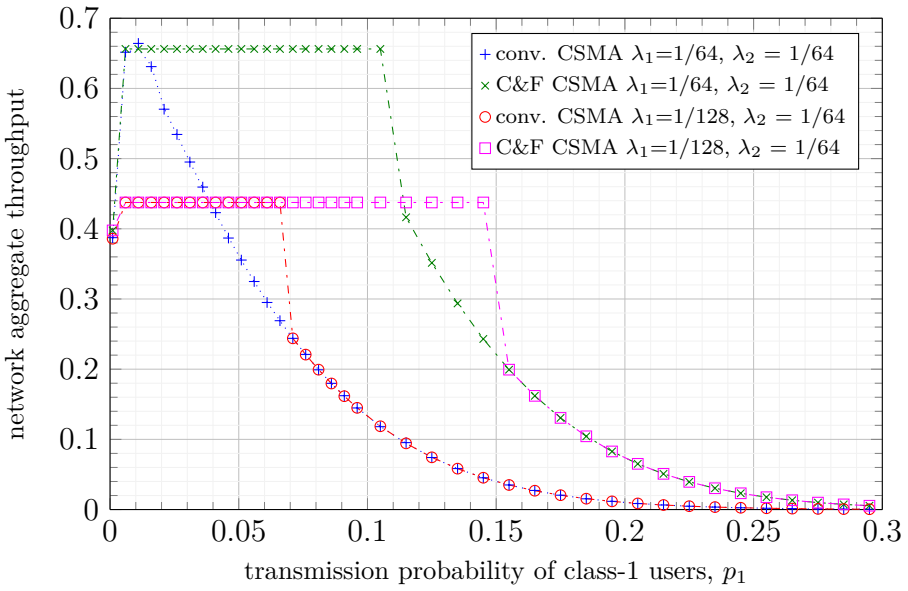

Figure 6.4: C&F-based CSMA compared with conventional CSMA in network aggregate throughput. The simulation setup is the same as that of Fig. 6.2.

the two AP's improves the network aggregate throughput significantly. The reason is that when AP's cooperate with each other their decoded linear combinations are more likely to be linearly independent and thus collision can be resolved by the central controller. However, in the non-cooperative scheme there is a higher chance of linear dependence between the decoded linear combinations. With a small cooperation overhead, therefore, the throughput scales well thanks to the C&F decoding technique.

6.6 Alternative Physical-Layer Techniques

Reasonable alternatives to C&F may include physical-layer network coding (PNC) [11], multiuser detection via successive interference cancellation (SIC) [95], and ZigZag decoding [96]. However, they are not particularly suitable for multi-AP scenarios. Conventional PNC decodes an XOR-combination from a collision [11]. In a two-AP setup, even if both APs are able to decode the XOR of the colliding packets, the controller cannot recover the individual packets due to the linear dependency.

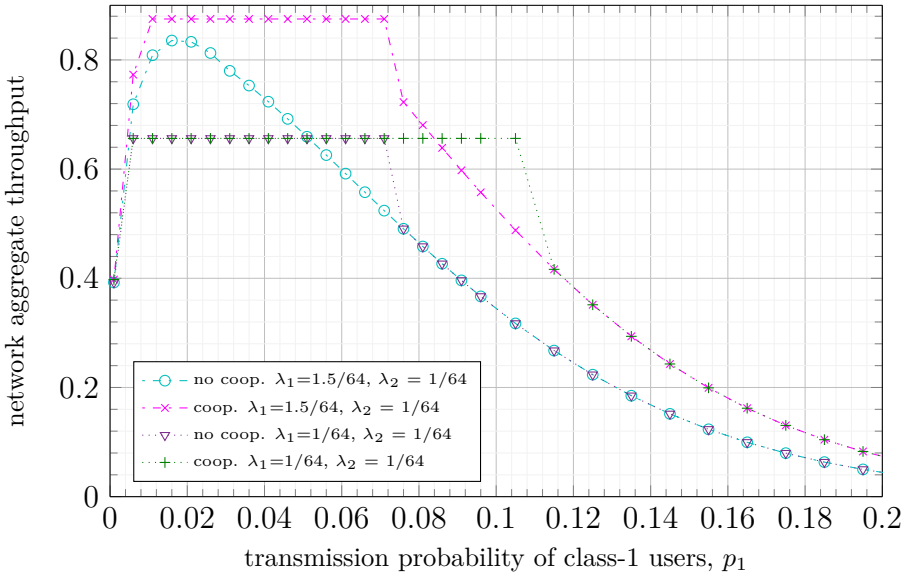

Figure 6.5: Aggregate throughput comparison between the cooperation scheme and non-cooperation in a 2-AP scenario. The simulation setup is the same as that of Fig. 6.2.

An SIC receiver attempts to recover all packets from a single collision. In many instances, this may not be possible, even though it might still be possible in such instances to decode one linear combination of the colliding packets (clearly an "easier" problem, and the one solved by C&F).

ZigZag decoding makes use of *multiple* collisions by exploiting distinct temporal offsets in different collisions. If ZigZag decoding is used in a two-AP setup, each AP must wait for at least two collisions in order to recover the colliding packets. In sharp contrast, C&F exploits spatial diversity and can reconstruct the colliding packets from a single collision.

6.7 Summary

C&F is a promising physical-layer technique that has a potential to greatly improve network throughput. Unlike most prior work that focuses on information-theoretic analysis and encoding-decoding methods of C&F, we have explored the use of C&F in random-access networks, taking into account several important practical constraints. In this chapter, we

extend the single-AP random access scenario to the multi-AP case where APs are allowed to cooperate with each other. In particular, we have identified two major challenges and developed novel solutions. Furthermore, we have provided accurate throughput and delay expressions for C&F-based CSMA systems. Using these expressions, together with our trace-driven simulations, the significant advantages of C&F-based CSMA over conventional CSMA are demonstrated.

Chapter 7

CONCLUDING REMARKS

Understanding the behavior of random access networks — even under collision assumption — is a difficult problem that has resisted a full understanding over the years. This dissertation proposes a mean-field approach for the analysis of stability and performance of random access systems over multi-packet reception (MPR) channels. We solve the problem for the special case of all-or-nothing MPR channel as well as symmetric MPR channel model which is promising due to its connection to symmetric MPR techniques such as compute-and-forward, successive compute-and-forward and successive interference cancellation. It appears that mean-field technique is a key tool to better understand the behavior of random access protocols and provide actual guidelines for system design.

We conclude by outlining the main contributions of this thesis.

7.1 Contributions

I. *Slotted ALOHA with All-or-Nothing Multi-packet Reception* Most of the analysis of stability conditions of random access schemes — for both MPR and non-MPR channels — have focused on homogeneous networks. We apply the mean-field approach to solve the special case of all-or-nothing and symmetric MPR channel models for inhomogeneous networks with V classes. We analyze the slotted ALOHA protocol as the simplest random access scheme and develop a sample-path-based mean-field analysis. It turns out that the behavior of slotted ALOHA over all-or-nothing channels can be described by a deterministic dynamical system composed of a set of V coupled ordinary differential equations representing $M/M/1$ queues. The solution to the dynamical system yields a solution for the trajectory of the underlying stochastic process in the large users

limit and provides a very good approximation for the finite system case even when the number of users is small.

Exploring the meta-stability of MPR-based slotted ALOHA was another key objective. This is undesired in practical systems as it impedes provisioning of quality of service guarantees. Yet, the mean-field approach enables identifying constraints on the physical-layer that avoids meta-stability. In other words, our results provide design criteria for lattice code design such that a slotted ALOHA system with symmetric MPR model becomes globally stable. In addition, the stability results provide the means for system performance analysis in terms of throughput and delay.

II. *CSMA with All-or-Nothing Multi-packet Reception*

The mean-field approach applied to MPR-based slotted ALOHA provides a starting point for the analysis of more advanced random access schemes like Carrier Sense Multiple Access (CSMA). The problem of persistent CSMA with all-or-nothing MPR is studied. Again, our focus is on the inhomogeneous case (as opposed to the homogeneous case where all the mobile users are assumed to have identical packet arrival rates and transmission probabilities) as it remains largely open in the literature. In this dissertation, we take a first step towards this open problem by deriving an approximation for the stability region as well as throughput and delay expressions for inhomogeneous CSMA, focusing on the all-or-nothing symmetric MPR. Interestingly, this family of MPR models is general enough to include a number of useful MPR scenarios —such as successive interference cancellation (SIC), compute-and-forward (C&F), and successive compute-and-forward (SCF)—as special cases. Based on these throughput and delay expressions, we provide theoretical guidelines for meeting quality-of-service requirements and for achieving global stability; we also evaluate the performances of various MPR techniques, highlighting the clear advantages offered by SCF.

III. *CSMA with Random Back-off*

We then extend the CSMA system model to a scenario where users are saturated, i.e., every user has a packet to transmit at all times, and use a back-off mechanism to reduce the probability of collision. In the CSMA with back-off model, the transmission probability of a user is determined by its back-off stage. With mean-field technique we analyze the two cases of finite and infinite number of back-off stages and find the stationary (invariant) probability of a user's back-off stage. In addition, it is shown that the mean-field approximation for these cases is quite accurate.

IV. *The Multiple AP Case*

The final extension studied is the multi-AP scenario where multiple APs cooperate with each other to resolve collision in a CSMA system over all-or-nothing MPR channel model. This is an interesting problem given the growth in the number of wireless subscribers and the consequent dense deployments of WLANs. The extension from single AP to multi-AP scenario posed two new challenges. In particular, a new efficient distributed scheme is developed for collision resolution. In addition, an algorithm based on spare recovery technique is proposed for joint channel estimation and active user recovery. The AP cooperation problem is closely related to the network-coding design problem. As we have shown in Chapter 6 (Section 6.2.4), our distributed solution to AP cooperation makes a particular use of the structure of the problem, and so is more computational efficient than existing solutions to the network-coding design problem.

From a theoretical perspective, this study dealt with major challenges in analyzing slotted ALOHA and persistent CSMA random access schemes which remained open problems for years. The seminal work by Bordenave *et. al.* [29] on slotted ALOHA with collision model paved the way for further analysis of such schemes. We exploit a mean-field approach – as used by [29] – to tackle the problem of random access under MPR assumption and provide solution to special cases. This problem is still challenging enough that we have to focus on all-or-nothing and symmetric MPR channel models.

On the practical side, the adoption of Multi-user Multiple-Input Multiple-Output (MIMO) techniques for the uplink in recent releases of IEEE 802.11 standard indicates great interest in incorporating advanced signal processing techniques to mitigate interference in receivers. Along this line, some studies [97–100] on slotted ALOHA with interference cancellation have been proposed.

7.2 What is Next?

We expect more work both on the theory and practice of random access protocols and specially on more advanced versions of such protocols. In spite of decades of research on random access protocols and their performance, a thorough understanding of even the simplest protocols of this kind is lacking. We believe that our key proposal in this thesis is the use of mean-field technique to analyze MPR-based random access schemes. We have studied the slotted ALOHA and persistent CSMA as well as a couple extensions. It would be interesting to see if the mean-field framework presented here can be extended to the case of CSMA with back-off in presence of random arrival under the all-or-nothing MPR assumption. This is a case which has the most similarity to the IEEE 802.11 CSMA protocol.

All the random access schemes in this thesis assume an all-or-nothing or symmetric MPR channel model. Although these models are still general enough to include a number of MPR techniques, we hope to see results for the general case of asymmetric MPR.

The meta-stability phenomenon studied in this thesis can be further studied by computing the average time interval between fluctuations due to the existence multiple stable states in random access systems. For example, it would be interesting to see how MPR model affects the average time between consecutive oscillations in a bi-stable CSMA system. Perhaps large deviation theory is the proper tool for developing such results.

Finally, our multi-AP system model studied in Chapter 6 assumes a single-cell scenario where collisions occur among the users within a single cell. In practical scenarios however, due to dense deployment of multi-cell WLANs, there are overlapping cells and hence there exist collisions between users associated with different APs. It is worth relaxing this assumption

and study cooperation schemes to cope up with such scenarios and yet achieve throughput and delay benefits offered by MPR-based protocols.

BIBLIOGRAPHY

[1] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6463–6486, Oct. 2011.

[2] Y. Cheng, J. Bellardo, P. Benkö, A. C. Snoeren, G. M. Voelker, and S. Savage, "Jigsaw: Solving the puzzle of enterprise 802.11 analysis," in *SIGCOMM*, pp. 39–50, 2006.

[3] W. L. Huang, K. Letaief, and Y. J. Zhang, "Cross-layer multi-packet reception based medium access control and resource allocation for space-time coded MIMO/OFDM," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3372–3384, Sept. 2008.

[4] P. X. Zheng, Y. Zhang, and S. C. Liew, "Multipacket reception in wireless local area networks," in *Proc. of IEEE Int. Conf. Commun.*, vol. 8, pp. 3670–3675, Jun. 2006.

[5] D. Chan, T. Berger, and L. Tong, "Carrier sense multiple access communications on multipacket reception channels: Theory and applications to IEEE 802.11 wireless networks," *IEEE Trans. Commun.*, vol. 61, pp. 266–278, Jan. 2013.

[6] D. Chan and T. Berger, "Performance and cross–layer design of CSMA for wireless networks with multipacket reception," in *Proc. of the 38th Asilomar Conf. Signals, Systems, and Computers*, vol. 2, pp. 1917–1921, Nov. 2004.

[7] K. Tan, H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker, "SAM: Enabling practical spatial multiple access in wireless LAN," in *Proc. of the 15th Annual Int. Conf. on Mobile Computing and Networking*, pp. 49–60, Sept. 2009.

[8] Y. J. Zhang, S. C. Liew, and D. R. Chen, "Sustainable throughput of wireless LANs with multipacket reception capability under bounded delay-moment requirements," *IEEE Trans. Mobile Comput.*, vol. 9, pp. 1226–1241, Sept. 2010.

[9] S. Ghez, S. Verdú, and S. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Trans. Autom. Control*, vol. 33, pp. 640–649, Jul. 1988.

[10] B. Nazer and M. Gastpar, "Computing over multiple-access channels with connections to wireless network coding," in *Proc. of IEEE Int. Symp. on Inf. Theory*, pp. 1354–1358, Jul. 2006.

[11] S. Zhang, S.-C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. of ACM Int. Conf. on Mobile Computing and Networking*, pp. 358–365, ACM, 2006.

[12] P. Popovski and H. Yomo, "The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *Proc. of IEEE Int. Conf. on Commun.*, vol. 9, pp. 3885–3890, Jun. 2006.

[13] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proc. IEEE*, vol. 99, pp. 438–460, Mar. 2011.

[14] L. Wei and W. Chen, "Compute-and-forward network coding design over multi-source multi-relay channels," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3348–3357, Sept. 2012.

[15] M. E. Soussi, A. Zaidi, and L. Vandendorpe, "Compute-and-forward on a multi-user multi-relay channel," *IEEE Wireless Commun. Lett.*, vol. 3, pp. 589–592, Dec. 2014.

[16] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, pp. 5641–5654, Nov. 2010.

[17] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, pp. 3450–3482, Jun. 2014.

[18] C. Feng, D. Silva, and F. R. Kschischang, "An algebraic approach to physical-layer network coding," *IEEE Trans. Inf. Theory*, vol. 59, pp. 7576–7596, Nov. 2013.

[19] U. Niesen, B. Nazer, and P. Whiting, "Computation alignment: Capacity approximation without noise accumulation," *IEEE Trans. Inf. Theory*, vol. 59, pp. 3811–3832, Jun. 2013.

[20] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, Third ed., 1999.

[21] B. S. Tsybakov and V. A. Mikhailov, "Ergodicity of a slotted ALOHA system," *Probl. Peredachi Inf.*, vol. 15, pp. 73–87, 1979.

[22] R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Trans. Inf. Theory*, vol. 34, pp. 918–930, Sept. 1988.

[23] W. Szpankowski, "Stability conditions for multidimensional queueing systems with computer applications," *Oper. Res.*, vol. 36, pp. 944–957, Nov. 1988.

[24] V. Anantharam, "The stability region of the finite-user slotted ALOHA protocol," *IEEE Trans. Inf. Theory*, vol. 37, pp. 535–540, May 1991.

[25] S. Kompalli and R. Mazumdar, "On the stability of finite queue slotted ALOHA protocol," *IEEE Trans. Inf. Theory*, vol. 59, pp. 6357–6366, Oct. 2013.

[26] Z. Rosberg, "A positive recurrence criterion associated with multidimensional queueing processes," *J. Appl. Probab.*, vol. 17, no. 3, pp. pp. 790–801, 1980.

[27] M. Jonckheere and S. Borst, "Stability of multi-class queueing systems with state-dependent service rates," in *Proc. of Intl. Conf. on Performance Evaluation Methodologies and Tools*, ACM, 2006.

[28] S. Borst, M. Jonckheere, and L. Leskela, "Stability of parallel queueing systems with coupled service rates," *Discrete Event Dynamic Systems*, vol. 18, no. 4, pp. 447–472, 2008.

[29] C. Bordenave, D. McDonald, and A. Proutiere, "Asymptotic stability region of slotted ALOHA," *IEEE Trans. Inf. Theory*, vol. 58, pp. 5841–5855, Sept. 2012.

[30] C. Bordenave, D. McDonald, and A. Proutiere, "A particle system in interaction with a rapidly varying environment: Mean field limits and applications," *ArXiv Mathematics e-prints*, Jan. 2007.

[31] S. Ghez, S. Verdú, and S. Schwartz, "Optimal decentralized control in the random access multipacket channel," *IEEE Trans. Autom. Control*, vol. 34, pp. 1153–1163, Nov. 1989.

[32] J. Sant and V. Sharma, "Performance analysis of a slotted ALOHA protocol on a capture channel with fading," *Queueing Systems*, vol. 34, no. 1-4, pp. 1–35, 2000.

[33] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2636–2656, Jul. 2005.

[34] D. Chan, T. Berger, and L. Tong, "On the stability and optimal decentralized throughput of CSMA with multipacket reception capability," in *Proc. of Allerton Conf. on Commun., Control, and Comput.*, Sept.–Oct. 2004.

[35] H. Jin, J. B. Seo, and D. K. Sung, "Stability analysis of p-persistent slotted CSMA systems with finite population," *IEEE Trans. Commun.*, vol. 62, pp. 4373–4386, Dec. 2014.

[36] S. Wu, W. Mao, and X. Wang, "Performance study on a CSMA/CA-based MAC protocol for multi-user MIMO wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 3153–3166, Jun. 2014.

[37] Y. Gai, S. Ganesan, and B. Krishnamachari, "The saturation throughput region of p-persistent CSMA.," in *IEEE Inf. Theory Workshop*, pp. 116–119, 2011.

[38] Y. H. Bae, B. D. Choi, and A. Alfa, "Achieving maximum throughput in random access protocols with multipacket reception," *IEEE Trans. Mobile Comput.*, vol. 13, pp. 497–511, Mar. 2014.

[39] Y. J. Zhang, P. X. Zheng, and S. C. Liew, "How does multiple-packet reception capability scale the performance of wireless local area networks?," *IEEE Trans. Mobile Comput.*, vol. 8, pp. 923–935, Jul. 2009.

[40] R. H. Gau, "Modeling the slotted nonpersistent CSMA protocol for wireless access networks with multiple packet reception," *IEEE Commun. Lett.*, vol. 13, pp. 797–799, Oct. 2009.

[41] F. Babich and M. Comisso, "Theoretical analysis of asynchronous multi-packet reception in 802.11 networks," *IEEE Trans. Commun.*, vol. 58, pp. 1782–1794, Jun. 2010.

[42] L. Tong, Q. Zhao, and G. Mergen, "Multipacket reception in random access wireless networks: From signal processing to optimal medium access control," *IEEE Commu. Mag.*, vol. 39, pp. 108–112, Nov. 2001.

[43] S. Wang, Q. Song, X. Wang, and A. Jamalipour, "Distributed MAC protocol supporting physical-layer network coding," *IEEE Trans. Mobile Comput.*, vol. 12, no. 5, pp. 1023–1036, 2013.

[44] Z. Chen, P. Fan, and K. Ben Letaief, "Throughput optimized multi-source cooperative networks with compute-and-forward," in *Proc. 23rd Conf. Wireless and Optical Commun.*, pp. 1–5, May 2014.

[45] M. E. Soussi, A. Zaidi, and L. Vandendorpe, "Compute-and-forward on a multiaccess relay channel: Coding and symmetric-rate optimization," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1932–1947, Apr. 2014.

[46] L. Wei and W. Chen, "Compute-and-forward network coding design over multi-source multi-relay channels," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3348–3357, Sept. 2012.

[47] Z. Chen, P. Fan, and K. Ben Letaief, "Compute-and-forward: Optimization over multisource-multirelay networks," *IEEE Trans. Veh. Technol.*, vol. 64, pp. 1806–1818, May 2015.

[48] J. Goseling, M. Gastpar, and J. Weber, "Random access with physical-layer network coding," in *IEEE Inf. Theory Workshop*, pp. 1–7, Feb. 2013.

[49] J. Goseling, M. Gastpar, and J. Weber, "Physical-layer network coding on the random-access channel," in *Proc. of IEEE Int. Symp. on Inf. Theory*, pp. 2339–2343, Jul. 2013.

[50] J. Goseling, C. Stefanovic, and P. Popovski, "Sign compute resolve for random access," in *Proc. of Allerton Conf. on Commun., Control, and Comput.*, pp. 675–682, 2014.

[51] J. Goseling, M. Gastpar, and J. Weber, "Random access with physical-layer network coding," *IEEE Trans. Inf. Theory*, vol. 61, pp. 3670–3681, Jul. 2015.

[52] A. A. Borovkov, "Asymptotic stabilization in the decentralized ALOHA algorithm. diffusion approximation," *Prob. of Inf. Trans.*, vol. 25, pp. 42–49, 1989.

[53] N. Vvedenskaya and Y. Suhov, "Multi-access system with many users: Stability and metastability," *Prob. of Inf. Trans.*, vol. 43, pp. 263–269, Sept. 2007.

[54] S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. 23, pp. 891–904, Sept. 1975.

[55] R. Rom and M. Sidi, *Multiple Access Protocols: Performance and Analysis*. Springer-Verlag New York, Inc., 1990.

[56] M. E. Soussi, A. Zaidi, and L. Vandendorpe, "Compute-and-forward on a multiaccess relay channel: Coding and symmetric-rate optimization," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1932–1947, Apr. 2014.

[57] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commu. Mag.*, vol. 48, pp. 164–174, Nov. 2010.

[58] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1859–1874, Sept. 2004.

[59] M. F. Guo, X. Wang, and M.-Y. Wu, "On the capacity of k-MPR wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 3878–3886, Jul. 2009.

[60] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Sel. Areas Commun.*, vol. 12, pp. 796–807, Jun. 1994.

[61] B. Nazer, V. R. Cadambe, V. Ntranos, and G. Caire, "Expanding the compute-and-forward framework: Unequal powers, signal levels, and multiple linear combinations," *CoRR*, vol. abs/1504.01690, 2015.

[62] O. Ordentlich, U. Erez, and B. Nazer, "Successive integer-forcing and its sum-rate optimality," in *Proc. of Allerton Conf. on Commun., Control, and Comput.*, pp. 282–292, Oct. 2013.

[63] J. Zhu and M. Gastpar, "Multiple access via compute-and-forward," *Computing Research Repository (CoRR)*, Jul. 2014.

[64] M. P. Wilson, K. R. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, pp. 5641–5654, Nov. 2010.

[65] P. Q. Nguyen and D. Stehlé, "Low-dimensional lattice basis reduction revisited," *ACM Trans. Algorithms*, vol. 5, pp. 46:1–46:48, Nov. 2009.

[66] N. di Pietro, J. J. Boutros, G. Zemor, and L. Brunel, "Integer low-density lattices based on construction A," in *IEEE Inf. Theory Workshop*, pp. 422–426, Sept. 2012.

[67] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.

[68] O. Abari, H. Rahul, D. Katabi, and M. Pant, "Airshare: Distributed coherent transmission made seamless," in *Proc. of IEEE Int. Conf. on Computer Commun.*, pp. 1742–1750, Apr. 2015.

[69] P. C. Wang, Y. C. Huang, and K. R. Narayanan, "Asynchronous physical-layer network coding with quasi-cyclic codes," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 309–322, Feb. 2015.

[70] C. Feng, D. Silva, and F. Kschischang, "Blind compute-and-forward," in *Proc. of IEEE Int. Symp. on Inf. Theory*, pp. 403–407, Jul. 2012.

[71] L. You, S. C. Liew, and L. Lu, "Network-coded multiple access II: Toward real-time operation with improved performance," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 264–280, Feb. 2015.

[72] A. Mejri and G. R. B. Othman, "Practical physical layer network coding in multi-sources relay channels via the compute-and-forward," in *Wireless Commun. and Netw. Conf. Workshops*, pp. 166–171, Apr. 2013.

[73] G. Cocco, N. Alagha, C. Ibars, and S. Cioni, "Practical issues in multi-user physical layer network coding," in *Advanced Satellite Multimedia Syst. Conf. and the 12th Signal Process. for Space Commun. Workshop*, pp. 205–211, Sept. 2012.

[74] F. Rossetto and M. Zorzi, "On the design of practical asynchronous physical layer network coding," in *IEEE 10th Workshop on Signal Process. Advances in Wireless Commun.*, pp. 469–473, Jun. 2009.

[75] C. Graham, "Chaoticity on path space for a queueing network with selection of the shortest queue among several," *J. Appl. Probab.*, vol. 37, pp. 198–211, Mar. 2000.

[76] M. Mitzenmacher, "The power of two choices in randomized load balancing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, pp. 1094–1104, Oct. 2001.

[77] G. Sharma, A. Ganesh, and P. Key, "Performance analysis of contention based medium access control protocols," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1665–1682, Apr. 2009.

[78] M. Mitzenmacher, B. Prabhakar, and D. Shah, "Load balancing with memory," in *FOCS*, 2002.

[79] N. Antunes, C. Fricker, P. Robert, and D. Tibi, "Metastability of CDMA cellular systems," in *Proc. of ACM Int. Conf. on Mobile Comput. and Netw.*, pp. 206–214, 2006.

[80] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part I - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. 23, pp. 1400–1416, Dec. 1975.

[81] D. Bertsekas and R. Gallager, *Data Networks (2nd Ed.)*. Prentice-Hall, Inc., 1992.

[82] W. Szpankowski, "Stability conditions for some distributed systems: Buffered random access systems," *J. of Adv. Appl. Prob.*, vol. 26, pp. 498–515, Jun. 1994.

[83] N. Patwari, "Measured channel impulse response data set." `http://span.ece.utah.edu/pmwiki/pmwiki.php?n=Main.MeasuredCIRDataSet`, Sept. 2007.

[84] S. Ashrafi, C. Feng, S. Roy, and F. Kschischang, "Slotted ALOHA with compute-and-forward," in *Proc. of IEEE Int. Symp. on Inf. Theory*, Jun. 2015.

[85] S. Ashrafi, C. Feng, and S. Roy, "Performance analysis of CSMA with multi-packet reception: The inhomogeneous case," *Submitted to IEEE Trans. Commun.*, 2015.

[86] P. Q. Nguyen and D. Stehle, "Low-dimensional lattice basis reduction revisited," *ACM Trans. Algorithms*, vol. 5, pp. 1–48, Oct. 2009.

[87] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[88] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," *CoRR*, vol. abs/0903.1022, 2009.

[89] X. Li, A. Rueetschi, A. Scaglione, and Y. Eldar, "Compressive link acquisition in multiuser communications," *IEEE Trans. Signal Process.*, vol. 61, pp. 3229–3245, Jun. 2013.

[90] S. Howard, A. Calderbank, and S. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes," in *42nd Annual Conf. on Inf. Sci. and Syst.*, pp. 11–15, Mar. 2008.

[91] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley-Interscience, 2005.

[92] A. R. Calderbank, A. C. Gilbert, and M. J. Strauss, "List decoding of noisy reed-muller-like codes," 2006.

[93] D. Guo, J. Luo, L. Zhang, and K. Shen, "Compressed neighbor discovery for wireless networks," *CoRR*, vol. abs/1012.1007, 2010.

[94] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[95] D. Halperin, T. Anderson, and D. Wetherall, "Taking the sting out of carrier sense: Interference cancellation for wireless LANs," in *Proc. of ACM Mobicom*, pp. 339–350, Sept. 2008.

[96] S. Gollakota and D. Katabi, "ZigZag decoding: Combating hidden terminals in wireless networks," in *Proc. of ACM SIGCOMM*, pp. 159–170, Aug. 17-22, 2008.

[97] K. R. Narayanan and H. D. Pfister, "Iterative collision resolution for slotted ALOHA: An optimal uncoordinated transmission policy," in *Int. Symp. Turbo Codes and Iter. Inf. Processing*, pp. 136–139, Aug. 2012.

[98] L. Toni and P. Frossard, "Prioritized random mac optimization via graph-based analysis," *IEEE Trans. Commun.*, vol. 63, pp. 5002–5013, Dec. 2015.

[99] C. Stefanovic and P. Popovski, "ALOHA random access that operates as a rateless code," *IEEE Trans. Commun.*, vol. 61, pp. 4653–4662, Nov. 2013.

[100] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, pp. 6815–6832, Dec. 2015.

[101] D. R. Curtiss, "Recent extentions of Descartes' rule of signs," *Annals of Mathematics*, vol. 19, no. 4, pp. 251–278, 1918.

Appendix A

PROOFS FOR CHAPTER 3

A.1 Proof of Theorem 3.4.5 (Stability of Finite System)

In Theorem 3.4.2 we provided the necessary and sufficient conditions for the stability of ODEs corresponding to the scaled process in the limit. We now show that the stability of the original finite system can be inferred from exactly the same conditions. To this end, we follow a proof technique similar to that in [29] which is based on a dominant system argument.

A.1.1 Sufficiency:

First we note that an alternative representation of the stability region of the ODE is possible by using simpler sets. Let

$$\Lambda^{\infty}(\boldsymbol{\alpha}, V) = \left\{\boldsymbol{\lambda} \in \mathbb{R}_{+}^{V} : \ \exists \boldsymbol{\rho} \in [0,1]^{V}, \ \lambda_v = \rho_v p_v \chi_{\boldsymbol{\alpha}}(\sigma) e^{-\sigma}\right\}, \tag{A.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$, $\chi_{\boldsymbol{\alpha}}(\sigma) = \sum_{i=1}^{K} \alpha_i \frac{q_i}{(i-1)!}\sigma^{i-1}$ and $\sigma = \sum_v \beta_v \rho_v p_v$. $\Lambda^{\infty}(\boldsymbol{\alpha}, V)$ is the stability region of the limiting system assuming that with the probability α_i the AP decoding threshold is $i \in \{1, \cdots, K\}$. The alternative representation of the stability region of limiting system $\Lambda^{\infty}(\boldsymbol{\alpha}, V)$ is presented the following proposition. Define

$$\partial_v \Lambda(\boldsymbol{\alpha}, V) \triangleq \left\{\boldsymbol{\lambda} \in \mathbb{R}_{+}^{V} : \exists \boldsymbol{\rho} \in \partial_v [0,1]^{V}, \lambda_v = \rho_v p_v \chi_{\boldsymbol{\alpha}}(\sigma) e^{-\sigma}\right\}, \tag{A.2}$$

and let $\Lambda(\boldsymbol{\alpha}, V)$ be the region below one of the V boundaries defined above.

Proposition A.1.1 If

$$\sum_{v\in\mathcal{V}} \beta_v p_v < \gamma^*, \tag{A.3}$$

then

$$\Lambda(\boldsymbol{\alpha}, V) = \Lambda^{\infty}(\boldsymbol{\alpha}, V). \tag{A.4}$$

Proof. Let $z_v = \rho_v p_v$ for all $v \in \mathcal{V}$. Define $\langle \beta, z\rangle \triangleq \sum_u \beta_u z_u$ where $\beta^T = (\beta_1, \cdots, \beta_V)$ and $z^T = (z_1, \cdots, z_V)$. Define the V-dimensional mapping $g : [0, p_1] \times \cdots \times [0, p_V] \to \Lambda^{\infty}(\boldsymbol{\alpha}, V)$ with the coordinatewise functions as

$$g_v(z) = z_v \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) e^{-\langle \beta, z\rangle} \qquad \forall\, v \in \mathcal{V}\,. \tag{A.5}$$

The proof relies on Inverse Function Theorem according to which the determinant of the Jacobian matrix of the above mapping has to be non-zero. Let $\nabla_z^T f = (\partial f/\partial z_1, \cdots, \partial f/\partial z_V)$. The Jacobian is

$$J = \left(I_V \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) - z\left(\beta^T \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) - \nabla_z^T \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle)\right)\right) e^{-\langle \beta, z\rangle}\,,$$

The matrix $z\left(\beta^T \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) - \nabla_z^T \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle)\right)$ is a rank one matrix with an eigenvalue equal to $\langle \beta, z\rangle \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) - \langle \nabla_z \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle), z\rangle$ and $V-1$ eigenvalues equal to one. Hence the determinant of J is

$$\det(J) = \left((1 - \langle \beta, z\rangle)\, \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle) + \langle \nabla_z \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle), z\rangle\right) e^{-\langle \beta, z\rangle}\,.$$

Furthermore

$$\frac{\partial \chi_{\boldsymbol{\alpha}}(\langle \beta, z\rangle)}{\partial z_v} = \left(\alpha_2 q_2 + \frac{\alpha_3 q_3}{1!}\langle \beta, z\rangle + \cdots + \frac{\alpha_K q_K}{(K-2)!}\langle \beta, z\rangle^{K-2}\right)\beta_v \qquad \forall\, v \in \mathcal{V}\,.$$

Hence

$$\det(J) = \left((1-\sigma)\chi_{\boldsymbol{\alpha}}(\sigma) + \sigma\frac{\partial\chi_{\boldsymbol{\alpha}}(\sigma)}{\partial\sigma}\right)e^{-\langle\beta,z\rangle},$$

and

$$\psi(\sigma) \triangleq (1-\sigma)\chi_{\boldsymbol{\alpha}}(\sigma) + \sigma\frac{\partial\chi_{\boldsymbol{\alpha}}(\sigma)}{\partial\sigma} = \sum_{i=0}^{K}\frac{(i+1)\alpha_{i+1}q_{i+1} - i\alpha_i q_i}{i!}\sigma^i \tag{A.6}$$

were $q_0 = q_{K+1} = 0$. According to Theorem 3.5.1 for $\boldsymbol{\alpha} = \mathbf{1}$, $\psi(\sigma)$ has exactly one positive root which is γ^*. For certain $\boldsymbol{\alpha}$ (arising from the induction argument following this proposition), we claim that (A.6) can be written as $\psi(\sigma + \beta_v p_v)$. As a result, $\gamma^* - \beta_v p_v$ is the root of $\psi(\sigma + \beta_v p_v)$, hence as long as $\sigma < \gamma^* - \beta_v p_v$ the Jacobian is always non-zero. So, we must have

$$\begin{aligned}\sigma &= \sum_v \beta_v\rho_v p_v \\ &\leq \sum_v \beta_v p_v \\ &\leq \gamma^*,\end{aligned}$$

to ensure bijectivity and continuity of the mapping. □

Proposition A.1.1 basically gives a sufficient condition for function g to be a one-to-one mapping between the boundary of the stability region of the limiting system and the union of the surfaces defined in (A.2). Let $\boldsymbol{\lambda}^V = (\lambda_1, \cdots, \lambda_V)$. Define $\hat{\Lambda}(\boldsymbol{\alpha}, V)$ as the region below the following boundaries

$$\begin{aligned}\partial_v\hat{\Lambda}^N(\boldsymbol{\alpha}, V) = \Bigg\{\boldsymbol{\lambda} \in \mathbb{R}_+^V :\ \exists\boldsymbol{\rho} \in \partial_v[0,1]^V,\ \lambda_u = \frac{\rho_u p_u}{1-\rho_u p_u/N} \\ \times\prod_w (1-\rho_w p_w/N)^{\beta_w^N N}\,\chi_{\boldsymbol{\alpha}}\left(\sum_{w\in\mathcal{V}}\frac{\rho_w p_w/N}{1-\rho_w p_w/N}\beta_w^N N\right)\ \forall u \in \mathcal{V}\Bigg\}.\end{aligned}$$

It is clear that $\boldsymbol{\lambda}^N + \epsilon\mathbf{1}^N \in \hat{\Lambda}^N$ iff $\boldsymbol{\lambda}^V + \epsilon\mathbf{1}^V \in \hat{\Lambda}^N(\mathbf{1}^K, V)$. Then by induction on V

the stability is shown. For $V = 1$ the result is already known as we have a homogeneous network [33]. We assume that the result holds for $|\mathcal{V}| \leq V$ classes. For the $V + 1$ case, we note that $\boldsymbol{\lambda}^{V+1} + \epsilon \mathbf{1}^{V+1} \in \hat{\Lambda}^N(\mathbf{1}, V + 1) = \Lambda(\mathbf{1}, V + 1)$ for large enough N. Further, there exists $\boldsymbol{\rho} \in \partial_v[0, 1]^{V+1}$ and a rate vector such that

$$\lambda_u = \rho_u p_u \chi_{\boldsymbol{\alpha}_v}(\sigma_v) e^{-\sigma_v} \qquad \text{for } u \neq v \tag{A.7}$$

$$\lambda_v + \epsilon \leq p_v \chi_{\boldsymbol{\alpha}_v}(\sigma_v) e^{-\sigma_v}, \tag{A.8}$$

where $\sigma_v = \sigma - \beta_v p_v$ and $\boldsymbol{\alpha}_v = (\alpha_{1,v}, \cdots, \alpha_{K,v})$ whose elements $\alpha_{i,v}$'s can be computed as a function of β_v, p_v and ρ_v. The Jacobian associated with $\boldsymbol{\alpha}_v$ results in $\psi(\sigma_v + \beta_v p_v)$. Now let us look at the queues not in class v. There exists a small $\epsilon > 0$ and a large enough N such that

$$\boldsymbol{\lambda}_{-v}^{V+1} + \epsilon \mathbf{1}^{V+1} \in \hat{\Lambda}^N \left(\boldsymbol{\alpha}_v^N, V \right), \tag{A.9}$$

where $\boldsymbol{\lambda}_{-v}^{V+1}$ is the arrival rate vector with λ_v removed, $\boldsymbol{\alpha}_v^N$ is a vector similar to $\boldsymbol{\alpha}_v$ but for the finite system. In other words, we switch from the stability region of the dynamical system to that of the finite system by properly choosing ϵ and N. Now we consider a dominant system where all queues in class v are saturated. By the induction hypothesis the original subsystem (system without class-v queues) is stable and the corresponding steady state distribution is tight. Then according to Theorem 3.4.2 we see that the limiting subsystem corresponding to the network of queues not in class v is stable.

So we have two systems (with class-v queues removed): one is the original system, another one is the dominant system. Both share the same mapping from $\partial_v[0, 1]^V$ to two regions. If this mapping is bijective then the non-empty probability ρ_u, and σ's of the two systems will be equal. Therefore, we assume that the mapping is bijective (Proposition A.1.1) and this enables us to relate the stability region of the dominant system to that of the dynamical system. The case of class-v queues is identical to the stability analysis of a homogeneous

network ($V = 1$) with redefined number of users and AP decoding threshold probabilities.

A.1.2 Necessity:

Observe that when $\boldsymbol{\lambda}^V - \epsilon \mathbf{1}^V \notin \Lambda(\mathbf{1}^K, V)$ then there exists a constant $0 < c < 1$ and $\epsilon > 0$ such that $c\boldsymbol{\lambda}^V_{-v} + \epsilon \mathbf{1}^{V-1} \in \partial_v \Lambda(\boldsymbol{\alpha}_v, V-1)$ and

$$\lambda_v > p_v \chi_{\boldsymbol{\alpha}_v}(\sigma_v) e^{-\sigma_v} + \epsilon \,.$$

We pick an arrival vector $\boldsymbol{\lambda}^V - \epsilon \mathbf{1}^V \notin \Lambda(\mathbf{1}^K, V)$. There exists some $\epsilon' > 0$ and $N_{\epsilon'}$ such that

$$\boldsymbol{\lambda}^V_{-v} + \epsilon' \mathbf{1}^V \in \hat{\Lambda}^N \left(\boldsymbol{\alpha}^N_v, V-1 \right), \tag{A.10}$$

$$\lambda_v > p_v \chi_{\boldsymbol{\alpha}_v}(\sigma_v) e^{-\sigma_v} + \epsilon', \tag{A.11}$$

with $\alpha_i = 1$ for all $i \in \{1, \cdots, K\}$.

Pick a class-v queue, say user i. Consider a dominant system where all the queues from class v except user i are saturated. Define a stopping time for user i in the dominant system and look at the evolution of queue-length for user i. Let $A_i(t)$ and $D_i(t)$ represent user i's arrivals and departures up to time t. From (A.11) and the law of large numbers we have

$$\lim_{t \to \infty} \left(A_i(t)/t - D_i(t)/t \right) > \epsilon'/N \,. \tag{A.12}$$

Therefore, there exists a t_0 such that for all $t > t_0$

$$A_i(t) - D_i(t) > \epsilon' t/N \,. \tag{A.13}$$

Define a stopping time for user i as follows

$$T_i^N = \inf \left\{ t \in \mathbb{R}_+ : X_i^N(t) < \epsilon' t/N \right\} .$$

The stopping time event occurs if the queue-length of user i falls below $\epsilon' t/N$ threshold (not saturated). In fact we look at the queue-length evolution and compare it to the line $\epsilon' t/N$ as time goes on. If $\mathrm{P}(T_i^N = \infty) > 1 - \delta$ then the queue-length of user i becomes unbounded. Now consider a dominant system where class-v queues are saturated. We note that this system is coupled to the original system only in $[0, t]$. Then the queues not in class v are stable based on the argument we did in the sufficiency part. It then suffices to pick a large enough initial queue-length for the class-v queues such that the queue-length of user i does not drop below the threshold. For a sufficiently large queue-length for class-v users at $t = 0$ there exists a $\delta > 0$ such that

$$\mathbb{P}(T_i^N = \infty) > 1 - \delta\,,$$

which implies that the system is transient.

A.2 Proof of Theorem 3.4.2 (Stability of Fluid Model: Limiting System)

The dynamical system (3.12) can be interpreted as Kolmogorov equation of an $M/M/1$ queue with state-dependent arrival rate and service rate. Let $X = (X_1, X_2, \cdots, X_V)$ and $Y = (Y_1, Y_2, \cdots, Y_V)$ be random variables in $\mathbb{R}_+^V$ whose entries are buffer (queue) lengths for two systems. Let $X(t)$ and $Y(t)$ be the random vectors at time t. Let $Q^X(t)$ and $Q^Y(t)$ be the tail distributions of the queue-lengths at $t \geq 0$ corresponding to system X and system Y, respectively. The following lemma shows that queue-length evolution has a monotone property.

Lemma A.2.1(Monotonicity of $Q(t)$) If $Q^X(0) \preccurlyeq_{\mathrm{st}} Q^Y(0)$ then $Q^X(t) \preccurlyeq_{\mathrm{st}} Q^Y(t)$ for any $t > 0$ and all $1 \leq i \leq N$. Moreover, $Q^X(t_1) \preccurlyeq_{\mathrm{st}} Q^X(t_2)$ for any $t_2 \geq t_1$.

Proof. Let $\tilde{X}(t) = X(t)$ and define $\tilde{Y}(t)$ such that $\tilde{Y}(t)$ and $Y(t)$ are identically distributed. We establish a coupling between $\tilde{X}(t)$ and $\tilde{Y}(t)$.

We know that $Q^X(0) \preccurlyeq_{\mathrm{st}} Q^Y(0)$. Denote by F_1 and F_2 the cumulative distribution function corresponding to $Q^X(0)$ and $Q^Y(0)$, respectively. Let $U \sim \mathcal{U}[0, 1]$ be a uniformly distributed random variable. Then $\tilde{X}_i(0) = F_1^{-1}(U) \leq F_2^{-1}(U) = F_2^{-1}(F_2(\tilde{Y}_i(0))) = \tilde{Y}_i(0)$ for

all $1 \leq i \leq N$.

Suppose that arrivals occur at the beginning of each time slot and departures occur at an arbitrary small time ϵ, before the start of a time slot. Let $0 < t_1 < t_2 < \cdots < t_{n-1} < t_n < \cdots$ be the values of arrival times and potential departure times. We now construct a coupling between the two systems $(\tilde{X}(t), \tilde{Y}(t))$ such that for any sample path (sequence of arrivals and departures) the claim holds. We already have $\tilde{X}(0) \leq \tilde{Y}(0)$. Suppose that $\tilde{X}(t) \leq \tilde{Y}(t)$ for all $t_{n-1} < t < t_n$ for some $n \geq 1$. We show that $\tilde{X}(t_{n+1}) \preccurlyeq_{\text{st}} \tilde{Y}(t_{n+1})$. First, suppose that t_n is an arrival time and that the arrival occurs on the jth longest queue of $\tilde{X}(t)$. We couple $\tilde{Y}$ to $\tilde{X}$ by assuming that the arrival occurs on the jth longest queue of $\tilde{Y}(t)$. Clearly $\tilde{X}_j(t_{n+1}) = \tilde{X}_j(t_n) + 1 \leq \tilde{Y}_j(t_n) + 1 = \tilde{Y}_j(t_{n+1})$. Next suppose that t_n is a departure time. Further suppose that we have k nonempty queues of $\tilde{Y}(t)$, thus there are at most k nonempty queues of $\tilde{X}(t)$. Number $\tilde{X}$ and $\tilde{Y}$ systems in a descending order of their queue-lengths, so, the first k queues of $\tilde{Y}$ are non-empty and at most the first k queues of $\tilde{X}$. We assume if a departure occurs at the jth queue of $\tilde{X}(t)$ the same holds for $\tilde{Y}(t)$. Then

$$\begin{aligned}
\tilde{X}(t_{n+1}) &= \tilde{X}(t_n) - 1_{\{\tilde{X}(t_n)>0\}} \\
&\leq \tilde{Y}(t_n) - 1_{\{\tilde{Y}(t_n)>0\}} \\
&= \tilde{Y}(t_{n+1})\,.
\end{aligned}$$

The same argument holds when multiple departures occur.

Now suppose that at time t_1, system Y starts with empty queues and system X starts with $Q^X(t_1) = Q^Y(t_2)$ for $t_2 \geq t_1$. By monotonicity property as proved above, $Q^X(t_2) \succcurlyeq_{\text{st}} Q^Y(t_2) = Q^X(t_1)$. □

Now we prove Theorem 3.4.2.

A.2.1 Necessary Conditions:

From Proposition 3.4.1, the workload drift $\Delta(\gamma) = 0$ admits invariant states only if $\sum_v \beta_v \lambda_v < \zeta(\gamma^*)$. This gives us the first necessary condition as in (3.17). Now assume the following two cases:

1. $\sum_v \beta_v p_v < \overline{\gamma}$: From (3.10), $\gamma(t) < \overline{\gamma}(\lambda, q)$ for all t. In a stable system $\gamma(t)$ can only converge to either $\underline{\gamma}(\lambda, q)$ or $\overline{\gamma}(\lambda, q)$. Since $\gamma(t) < \overline{\gamma}(\lambda, q)$ then γ can potentially converge to $\underline{\gamma}(\lambda, q)$ (The proof of convergence is provided in the next section) in which case each queue behaves as an $M/M/1$ queue with service rate $p_v \chi(\underline{\gamma}) e^{-\underline{\gamma}}$ whose stability necessitates (3.19). Otherwise if $\lambda_v > p_v \chi(\underline{\gamma}) e^{-\underline{\gamma}}$ then some queues will be unstable. Let us assume that all the queues are stable, we show this results in contradiction. From the proof for the sufficiency (see below) we know that $\gamma \to C$ as $t \to \infty$ if the system is initialized with empty queues and is stable.

$$C = \sum_v \beta_v p_v \left(1 - \frac{\lambda_v}{p_v \chi(C) e^{-C}}\right).$$

   We should have $\lambda_v < p_v \chi(C) e^{-C}$ for all $v \in \mathcal{V}$. But $\lambda_v > p_v \chi(\underline{\gamma}) e^{-\underline{\gamma}}$ for some v which implies that $C < \underline{\gamma}$. Thus $\mathrm{d}W(t)/\mathrm{d}t > 0$ and the workload diverges which contradicts the stability assumption. So some queues are indeed unstable.

2. $\sum_v \beta_v p_v \geq \overline{\gamma}$: Depending on the initial condition of the queues, γ may converge to either $\underline{\gamma}$ or $\overline{\gamma}$. The former occurs if the system begins with empty queues and the latter can be achieved if the initial condition of the queues is stationary distribution of an $M/M/1$ queue with service rate $p_v \chi(\underline{\gamma}) e^{-\underline{\gamma}}$. Thus the system does not have a unique stable point, i.e., is not globally stable.

A.2.2 Sufficient Conditions:

We use superscript μ as in $\gamma^\mu(t)$ to show that the system starts with queue-length values generated by the measure μ. When the system is initially empty we use $\mu = 0$. Suppose

that queues are all empty at $t = 0$. From (3.10) we have $\gamma^0(0) = 0$. From Lemma A.2.1 we have $Q_v^0(t_1) \preccurlyeq_{\text{st}} Q_v^0(t_2)$ for $t_2 \geq t_1$ and all $v \in \mathcal{V}$. Thus $\gamma^0(t)$ is non-decreasing. $W^0(t)$ is also a non-decreasing function as $W^0(t) = \sum_{k\geq 0} Q_{v,k}^0(t)$, hence from (3.14) for all $t \geq 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t} W(t) \geq 0 \Rightarrow \sum_v \beta_v \lambda_v \geq \chi(\gamma^0(t)) e^{-\gamma^0(t)} \,. \tag{A.14}$$

Furthermore, $\gamma^0(t)$ is a bounded function, $\gamma^0(t) \leq \sum_v \beta_v p_v$ for all $t \geq 0$, thus it converges to some C from below $\gamma^0(t) \uparrow C$ as $t \to +\infty$ and $C \leq \underline{\gamma}(\lambda, q)$ otherwise from (A.14), $\frac{\mathrm{d}}{\mathrm{d}t} W(t) < 0$ which is contradiction. Therefore, for all $v \in \mathcal{V}$:

$$\begin{aligned} \lambda_v &< p_v \chi\left(\underline{\gamma}(\lambda, q)\right) e^{-\underline{\gamma}(\lambda,q)} \\ &\leq p_v \chi(C) e^{-C} \,, \end{aligned}$$

which implies that a single $M/M/1$ queue of class v is stable ($W^0(\infty) < \infty$) and the probability that the queue is non-empty (utilization) is

$$\rho_v = \frac{\lambda_v}{p_v \chi(C) e^{-C}} \,. \tag{A.15}$$

Finally,

$$\begin{aligned} C &= \gamma^{(0)}(\infty) \\ &= \sum_n \beta_v p_v \rho_v \\ &= \sum_v \beta_v p_v \frac{\lambda_v}{p_v \chi(C) e^{-C}} \,, \end{aligned}$$

thus $C\chi(C)e^{-C} = \sum_v \beta_v \lambda_v$ and therefore $C = \underline{\gamma}(\lambda, q)$.

Now we turn to the case where initial queue-length distribution is not the zero measure. Suppose

$$\overline{Q}_v(0) \succ_{\text{st}} Q_v^u(0) \succcurlyeq_{\text{st}} \underline{Q}_v(0) \succcurlyeq_{\text{st}} Q_v^l(0)$$

which implies that

$$\overline{\gamma} > \gamma^u(t) \geq \underline{\gamma} \geq \gamma^l(t) \geq \gamma^0(t)\,.$$

Starting from (3.14) we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}W^u(t) &= \sum_v \beta_v\lambda_v - \gamma^u(t)\chi(\gamma^u(t))e^{-\gamma^u(t)} \\ &= \lambda - \lambda + \underline{\gamma}(t)\chi(\underline{\gamma}(t))e^{-\underline{\gamma}(t)} - \gamma^u(t)\chi(\gamma^u(t))e^{-\gamma^u(t)} \\ &= \zeta(\underline{\gamma}) - \zeta(\gamma^u(t))\,.\end{aligned}$$

ζ is a Lipschitz function so there exists an $M > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}W^u(t) = \zeta(\underline{\gamma}) - \zeta(\gamma^u(t)) \tag{A.16}$$

$$\leq M(\underline{\gamma} - \gamma^u(t))\,. \tag{A.17}$$

(A.16) implies that $\lim_{t\to\infty}\gamma^u(t) = \underline{\gamma}$. To see why this is true we write

$$\begin{aligned}M\int_0^\infty (\gamma^u(t) - \underline{\gamma})\mathrm{d}t &\leq W^u(0) - W^u(\infty) \\ &\overset{\text{(a)}}{\leq} W^u(0) - W^0(\infty) \\ &< \infty\,,\end{aligned}$$

where in (a) we used the fact that $Q_v^u(t) \succcurlyeq_{\mathrm{st}} \underline{Q}_v^0(t)$ and thus $W^u(t) \geq W^0(t)$ for all $v \in \mathcal{V}$ and $t \geq 0$. Since $\gamma^u(t) \geq \underline{\gamma}$ we must have

$$\lim_{t\to\infty}\gamma^u(t) = \underline{\gamma}\,.$$

By a similar argument we have

$$\lim_{t\to\infty}\gamma^l(t) = \underline{\gamma}\,.$$

The convergence of $\gamma(t)$ to $\underline{\gamma}$ implies again that the system is indeed a network of $M/M/1$

queues and the rest of the argument is the same as the case where queues are initially empty. This completes the proof.

A.3 Proof of Theorem 3.5.1 (Meta-Stability)

We note that $\zeta(\gamma)$ has exactly one positive extremum if the first order derivative of the function has only one positive root. Thus,

$$p(\gamma) \triangleq q_1 + \sum_{k=2}^{K+1} \frac{kq_k - (k-1)q_{k-1}}{(k-1)!}\gamma^{k-1} = 0 \tag{A.18}$$

Equation (A.18) is a degree K polynomial in γ. The first and last coefficients of this polynomial are positive and negative respectively. According to Descartes' rule of sign the number of positive roots of a polynomial with real coefficients ordered by descending variable exponent is either equal to the number of sign differences between consecutive nonzero coefficients, or is less than it by an even number. For (A.18) to have one positive root, it is sufficient to have one sign difference. Equivalently, all the coefficients $(kq_k - (k-1)q_{k-1})/(k-1)!$ for all $2 \leq k \leq K$ should be either all positive or all negative which yields the lemma. We assume that $kq_k - (k-1)q_{k-1} \geq 0$. Moreover, $p(0) > 0$ and for any large enough γ we have $p(\gamma) < 0$. So $p(\gamma)$ has one positive root. Finally, if $\gamma < \gamma^*$ then the system cannot be bi-stable (see Fig. 3.4).

A.4 Proof of Theorem 3.7.3 (Total Delay of All-or-Nothing Slotted ALOHA)

Suppose that there are n packets in the buffer of a class-v queue upon the arrival of a new packet. Denote by $\pi_{v,n}$ the stationary distribution of the state (queue-length) of a class-v queue immediately before an arrival. Let $\pi'_{v,n}$ be the stationary distribution at the end of the following time slot. The the following relationships hold

$$\pi'_{v,n} = \begin{cases} \pi_{v,n} P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^*) + (1 - P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}))\pi_{v,n-1} & \text{if} \quad n > 0 \\ \pi_{v,0} P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^*) & \text{if} \quad n = 0 \end{cases}$$

Let $T_v(n)$ represent the packet delay of a class-v user conditioned on its queue-length being n. Furthermore, let $X_v^{(N)}$ be the queue-length of a class-v user. Clearly $T_v(n) = (n+1)D_{v,\text{ALOHA}}$. Let $\boldsymbol{\rho}^*$ be the vector of non-empty probabilities when the system is in steady state. The average total delay $T_{v,\text{ALOHA}}$ is

$$\begin{aligned}
T_{v,\text{ALOHA}} &= \sum_{n\geq 0} T_v(n)\pi'_{v,n} \\
&= D_{v,\text{ALOHA}}\pi_{v,0} + \sum_{n>0}(n+1)D_{v,\text{ALOHA}}\left(\pi_{v,n}P_v^{\text{SUCC}}(\boldsymbol{\rho}^*) + (1-P_v^{\text{SUCC}}(\boldsymbol{\rho}^*))\pi_{v,n-1}\right) \\
&= D_{v,\text{ALOHA}}\left(\pi_{v,0}P_v^{\text{SUCC}}(\boldsymbol{\rho}^*) + \mathbb{E}\left[X_v^{(N)}\right]P_v^{\text{SUCC}}(\boldsymbol{\rho}^*) + (1-P_v^{\text{SUCC}}(\boldsymbol{\rho}^*))\left(1+\mathbb{E}\left[X_v^{(N)}\right]\right)\right) \\
&= D_{v,\text{ALOHA}}\mathbb{E}\left[X_v^{(N)}\right] + \pi_{v,0} + D_{v,\text{ALOHA}}\left(1-\frac{\lambda_v}{\rho_v^*}\right) \\
&\overset{\text{(a)}}{=} D_{v,\text{ALOHA}}\lambda_v T_{v,\text{ALOHA}} + \frac{\rho_v^*}{\lambda_v} - \rho_v^*\,.
\end{aligned}$$

Note that in (a) we used the Little's law $\mathbb{E}\left[X_v^{(N)}\right] = \lambda_v T_{v,\text{ALOHA}}$. This completes the proof.

A.5 Proof of Theorem 3.4.3 (Convergence in the Transient Regime)

For a process $\{X(n)\}_{n\in\mathbb{Z}_+}$, we use $X^{(N)}(t)$ to denote the scaled process at time $t \in \mathbb{R}_+$ where we look at the state of the system at time slot $n = \lfloor Nt \rfloor$.

$$X^{(N)}(t) = X(\lfloor Nt \rfloor)$$

We use a sample-path based approach to prove the convergence. We show that over any finite interval, any sub-sequence of $W_v^{(N)}(t)$ for all $v \in \mathcal{V}$ admits a further sub-sequence that converges coordinate-wise to a Lipschitz-continuous function.

Let $A^{(N)}(t)$ and $D^{(N)}(t)$ be the cumulative arrival and departure processes at time t. Upon every arrival or departure $A^{(N)}(t)$ and $D^{(N)}(t)$ will change by $\frac{1}{N}$. We therefore have

$$W_{v,\ell}^{(N)}(t) = W_{v,\ell}^{(N)}(0) + A_{v,\ell}^{(N)}(t) - D_{v,\ell}^{(N)}(t)$$

Proposition A.5.1 Fix $T > 0$ and $v \in \mathcal{V}$. There exist Lipschitz functions $w_v(t)$, $a_v(t)$, $d_v(t)$, and an increasing sequence $\{N_k\}_{k\in\mathbb{Z}_+} \subset \mathbb{N}$ such that

$$\lim_{N_k\to\infty} \left\| W_v^{(N_k)}(t) - w_v(t) \right\|_T = 0$$
$$\lim_{N_k\to\infty} \left\| A_v^{(N_k)}(t) - a_v(t) \right\|_T = 0$$
$$\lim_{N_k\to\infty} \left\| D_v^{(N_k)}(t) - d_v(t) \right\|_T = 0$$

Proof. We have

$$A_{v,\ell}(m) = A_{v,\ell}(m-1) + \frac{1}{N} \sum_{j=1}^{NQ_{v,\ell-1}^{(N)}(m)} 1_{\{u_j(m)<\lambda_v/N\}}$$
$$D_{v,\ell}(m) = D_{v,\ell}(m-1) + \sum_{j=1}^{NQ_{v,\ell+1}^{(N)}(m)} 1_{\left\{u_j(m)<\mu_{v,\ell}^{(N)}\left(k,Q_{v,\ell+1}^{(N)}(m)\right)/N\right\}}$$

where $u_j(m) \sim \mathcal{U}[0,1]$. We then write

$$\begin{aligned}
\left|W_{v,\ell}^{(N)}(t_2) - W_{v,\ell}^{(N)}(t_1)\right| &\le \left|A_{v,\ell}^{(N)}(t_2) - A_{v,\ell}^{(N)}(t_1)\right| + \left|D_{v,\ell}^{(N)}(t_2) - D_{v,\ell}^{(N)}(t_1)\right| \\
&= \frac{1}{N}\left(\sum_{m=\lfloor Nt_1\rfloor+1}^{\lfloor Nt_2\rfloor} \sum_{j=1}^{NQ_{v,\ell-1}^{(N)}(m)} 1_{\{u_j(m)<\lambda_v/N\}} + \sum_{m=\lfloor Nt_1\rfloor+1}^{\lfloor Nt_2\rfloor} \sum_{j=1}^{NQ_{v,\ell-1}^{(N)}(m)} 1_{\left\{u_j(m)<\mu_{v,\ell}^{(N)}\left(Q_{v,\ell-1}^{(N)}(m)\right)/N\right\}} \right) \\
&\le \frac{1}{N}\left(\sum_{j=1}^{N} \sum_{m=\lfloor Nt_1\rfloor+1}^{\lfloor Nt_2\rfloor} 1_{\{u_j(m)<\lambda_v/N\}} + \sum_{j=1}^{N} \sum_{m=\lfloor Nt_1\rfloor+1}^{\lfloor Nt_2\rfloor} 1_{\left\{u_j(m)<\mu_{v,\ell}^{(N)}\left(Q_{v,\ell-1}^{(N)}(m)\right)/N\right\}} \right) \\
&\overset{(a)}{\le} (\lambda_v + \bar{\mu}_{v,\ell})(t_2 - t_1) + \delta_N \,, \qquad \text{(A.19)}
\end{aligned}$$

where (a) follows from the functional law of large numbers for the cumulative arrival and service processes and that there exist a positive sequence $\{\delta_N\}_{N\in\mathbb{N}}$ with $\delta_N \to 0$ as $N \to \infty$ such that for all $t_1, t_2 \in [0,T]$ and $t_2 > t_1$, (A.19) holds.

Now consider any subsequence $\{W_{v,\ell}^{(N_j)}(t)\}_{j\in\mathbb{N}}$. Using $\lim_{N\to\infty} \left\| W_v^{(N)}(0) - w_v(0) \right\| = 0$, and (A.19) it can then be shown that there exists a further subsequence convergent to a

Lipschitz function. That is

$$\lim_{s\to\infty} \left\| W_v^{(N_{j_s})}(t) - w_v(t) \right\|_T = 0\,.$$

So far we showed that each coordinate of $W_v^{(N_j)}(t)$ admits a limit point. Next we construct the limit point of $W_v^{(N_j)}(t)$. This can be achieved by using a family of nested sequences $\{N_j^i\}_{i,j\in\mathbb{N}}$ such that $\{N_j^i\}_{j\in\mathbb{N}}$ is a subsequence of $\{N_j^{i-1}\}_{j\in\mathbb{N}}$. Let $w_{v,1}(t)$ be the Lipschitz continuous limit point of $\{W^{(N_{v,1})}(t)\}$. So there exits a subsequence $\{W_{v,1}^{(N_j^1)}(t)\}_{j\in\mathbb{N}}$ convergent to $w_{v,1}(t)$. Define the limit point $w_v(t) = [w_{v,1}(t), w_{v,2}(t), \cdots]$ where $w_{v,i}(t)$ is the limit point of $\{W_{v,i}^{(N_j^i)}(t)\}_{j\in\mathbb{N}}$. By definition we have

$$0 \le W_{v,\ell}^{(N)}(t) \le W_{v,1}^{(N)}(t) \qquad \forall\, \ell \ge 1,\ v \in \mathcal{V},\ t \in [0,T]$$

which implies that $0 \le w_{v,\ell}(t) \le w_{v,1}(t)$ for all $v \in \mathcal{V}$ and $t \in [0,T]$. Define $c_1 \triangleq \max_v \lambda_v + \bar{\mu}_{v,\ell}$. We then have

$$\sup_{t\in[0,T]} w_{v,\ell}(t) \le \sup_{t\in[0,T]} w_{v,1}(t) \le w_{v,1}(0) + c_1 T\,.$$

Let $N_1 = 1$. Define

$$N_\ell = \min\left\{ N : N \ge N_{\ell-1},\ \sup_{0\le l\le \ell} \left\| W_{v,l}^{(N)}(t) - w_{v,l}(t) \right\|_T \le \frac{1}{\ell},\ \ \forall v \in \mathcal{V} \right\}.$$

For $\ell \ge 2$ we then have

$$\begin{aligned}
\left\| W_v^{(N_\ell)}(t) - w_v(t) \right\|_T &= \sup_{t\in[0,T]} \sqrt{\sum_{l=1}^{\infty} \left| W_{v,l}^{(N_\ell)}(t) - w_{v,l}(t) \right|^2 2^{-l}} \\
&\le \frac{1}{\ell}\sqrt{\sum_{l=0}^{\ell} 2^{-l}} + \sqrt{\sum_{l=\ell+1}^{\infty} 2^{-l}(w_{v,1}(0) + c_1 T)^2} \\
&\le \frac{\sqrt{2 - 2^{-\ell}}}{\ell} + 2^{-\ell/2}\left(w_{v,1}(0) + c_1 T\right)\,.
\end{aligned}$$

Thus

$$\lim_{\ell\to\infty} \left\| W_v^{(N_\ell)}(t) - w_v(t) \right\|_T = 0\,.$$

Similarly we can argue for the convergence of $A_v^{(N)}(t)$ and $D_v^{(N)}(t)$ to $a_v(t)$ and $d_v(t)$, respectively. □

Proposition A.5.1 shows that for all $v \in \mathcal{V}$, any subsequence $\{W_v^{(N_j)}(t)\}_{j\in\mathbb{N}}$ admits a further convergent subsequence. It remains to show that the limit point $w_v(t)$ is indeed the solution of the dynamical system.

Proposition A.5.2 Fix $T > 0$. Let $w_v(t)$ be a limit point of $W_v^{(N)}(t)$ and coordinatewise differentiable at t. Then for all $\ell \geq 1$ and $v \in \mathcal{V}$

$$\begin{aligned} &\text{(i)} \quad \frac{\mathrm{d}}{\mathrm{d}t} a_{v,\ell}(t) = \lambda_v Q_{v,\ell-1}(t)\,, \\ &\text{(ii)} \quad \frac{\mathrm{d}}{\mathrm{d}t} d_{v,\ell}(t) = \mu_v(\gamma(t)) Q_{v,\ell}(t)\,. \end{aligned}$$

Proof. From Proposition A.5.1 we have that $\{A_v^{(N_j)}(t)\}_{j\in\mathbb{N}}$ is a subsequence convergent to a Lipschitz function $a_v(t)$. Thus

$$\lim_{N\to\infty} \left(A_{v,\ell}^{(N)}(t) - A_{v,\ell}^{(N)}(t-\epsilon) \right) = a_{v,\ell}(t) - a_{v,\ell}(t-\epsilon)\,,$$

and

$$\lim_{N\to\infty} \left(W_{v,\ell-1}^{(N)}(t) - W_{v,\ell}^{(N)}(t) \right) = w_{v,\ell-1}(t) - w_{v,\ell}(t)\,.$$

Furthermore, $W_{v,\ell}^{(N)}(z)$ is a Lipschitz function and there exists a positive sequence $\{\delta_N\}_{N\in\mathbb{N}}$ with $\delta_N \to 0$ as $N \to \infty$ such that for all $z \in [t-\epsilon, t]$

$$w_{v,\ell}(t) - c_1(z-t) - \delta_N \leq W_{v,\ell}^{(N)}(z) \leq w_{v,\ell}(t) + c_1(z-t) + \delta_N\,.$$

Thus

$$\begin{aligned} Q_{v,\ell}^{(N)}(z) &= W_{v,\ell}^{(N)}(z) - W_{v,\ell+1}^{(N)}(z) \\ &\leq w_{v,\ell}(t) - w_{v,\ell+1}(t) + 2c_1(z-t) + 2\delta_N \\ &= Q_{v,\ell}(t) + 2c_1(z-t) + 2\delta_N \,. \end{aligned}$$

Similarly $Q_{v,\ell}^{(N)}(z) \geq Q_{v,\ell}(t) - 2c_1(z-t) - 2\delta_N$. Now we write

$$\begin{aligned} A_{v,\ell}^{(N)}(t) - A_{v,\ell}^{(N)}(t-\epsilon) &= \frac{1}{N} \sum_{m=\lfloor N(t-\epsilon)\rfloor+1}^{\lfloor Nt\rfloor} \sum_{j=1}^{NQ_{v,\ell-1}^{(N)}(t)} 1_{\{u_j(m)<\lambda_v/N\}} \\ &\leq \epsilon \sum_{j=1}^{N(Q_{v,\ell-1}(t)+2c_1\epsilon+2\delta_N)} 1_{\{u_j(m)<\lambda_v/N\}} \,. \end{aligned}$$

Hence

$$\begin{aligned} a_{v,\ell}(t) - a_{v,\ell}(t-\epsilon) &= \lim_{N\to\infty} \left(A_{v,\ell}^{(N)}(t) - A_{v,\ell}^{(N)}(t-\epsilon) \right) \\ &\leq \lim_{N\to\infty} \epsilon \sum_{j=1}^{N(Q_{v,\ell-1}(t)+2c_1\epsilon+2\delta_N)} 1_{\{u_j(m)<\lambda_v/N\}} \\ &= \lambda_v \epsilon \, Q_{v,\ell-1}(t) + 2c_1\lambda_v\epsilon^2 \,. \end{aligned}$$

So $|a_{v,\ell}(t) - a_{v,\ell}(t-\epsilon) - \lambda_v \epsilon Q_{v,\ell-1}(t)| \leq 2c_1\epsilon^2$, and we get the result as

$$\frac{\mathrm{d}}{\mathrm{d}t} a_{v,\ell}(t) = \lim_{\epsilon \searrow 0} \frac{a_{v,\ell}(t) - a_{v,\ell}(t-\epsilon)}{\epsilon} = \lambda_v Q_{v,\ell-1}(t) \,.$$

The argument for case (ii) is along the same line and omitted for the sake of brevity. □

Theorem 3.4.2 provides the sufficient and necessary conditions for the solution of dynamical systems to be unique in which case $\{Q_v^{(N)}(t)\}_{N\in\mathbb{N}}$ for all $v \in \mathcal{V}$ converges to the solution of dynamical system over the compact set $[0, T]$.

A.6 Proof of Theorem 3.4.4 (Convergence in the Stationary Regime)

Let $C_b(\mathcal{Q})$ be the space of bounded continuous functions from $\mathcal{Q}$ to $\mathbb{R}$. Pick a $g \in C_b(\mathcal{Q})$. For any $\boldsymbol{Q}_v^0 \in \mathcal{Q}$ and $t \geq 0$, since $\boldsymbol{\pi}_v^{(N)}$ is the invariant measure, we have

$$\int_{\boldsymbol{Q}_v^0} \mathbb{E}_{\boldsymbol{Q}_v^0}\left[g\left(\boldsymbol{Q}_v^{(N)}(t)\right)\right] d\boldsymbol{\pi}_v^{(N)} = \int_{\boldsymbol{Q}_v^0} g\left(\boldsymbol{Q}_v^0\right) d\boldsymbol{\pi}_v^{(N)}. \tag{A.20}$$

In addition, $\boldsymbol{\pi}_v^{(N)}$ is tight. Then by Prokhorov's theorem $\boldsymbol{\pi}_v^{(N)}$ is relatively compact, i.e., $\boldsymbol{\pi}_v^{(N)}$ has a subsequence which is convergent to a limit point $\boldsymbol{\pi}$. Furthermore, $g(x) \in C_b(\mathcal{Q})$ thus for any $\epsilon > 0$ such that $\|x - y\| < \epsilon$ there exists a $\delta > 0$ that $\|g(x) - g(y)\| < \delta$. Thus for any $t \geq 0$

$$\begin{aligned}
\left\|\mathbb{E}_{\boldsymbol{Q}_v^0}\left[g(\boldsymbol{Q}_v^{(N)}(t)) - g(\boldsymbol{Q}_v(t))\right]\right\|_T = \Bigg\| & \mathbb{E}_{\boldsymbol{Q}_v^0}\left[\left(g(\boldsymbol{Q}_v^{(N)}(t)) - g(\boldsymbol{Q}_v(t))\right) 1_{\left\{\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T \leq \epsilon\right\}}\right] \\
& \cdot \mathbb{P}\left(\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T \leq \epsilon\right) \\
& + \mathbb{E}_{\boldsymbol{Q}_v^0}\left[\left(g(\boldsymbol{Q}_v^{(N)}(t)) - g(\boldsymbol{Q}_v(t))\right) 1_{\left\{\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T > \epsilon\right\}}\right] \\
& \cdot \mathbb{P}\left(\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T > \epsilon\right) \Bigg\|_T \\
\leq \delta + 2\|g\| & \, \mathbb{P}\left(\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T > \epsilon\right).
\end{aligned}$$

Thus

$$\begin{aligned}
& \left\|\int_{\boldsymbol{Q}_v^0} \mathbb{E}_{\boldsymbol{Q}_v^0}\left[g\left(\boldsymbol{Q}_v^{(N)}(t)\right)\right] d\boldsymbol{\pi}_v^{(N)} - \int_{\boldsymbol{Q}_v^0} g\left(\boldsymbol{Q}_v(t)\right) d\boldsymbol{\pi}_v^{(N)}\right\|_T \\
& \qquad \leq \int_{\boldsymbol{Q}_v^0} \left\|\mathbb{E}_{\boldsymbol{Q}_v^0}\left[g\left(\boldsymbol{Q}_v^{(N)}(t)\right)\right] - g\left(\boldsymbol{Q}_v(t)\right)\right\|_T d\boldsymbol{\pi}_v^{(N)} \\
& \qquad \leq \delta + 2\|g\| \int_{\boldsymbol{Q}_v^0} \mathbb{P}\left(\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T > \epsilon\right) d\boldsymbol{\pi}_v^{(N)}.
\end{aligned}$$

By theorem 3.4.3, $\mathbb{P}\left(\left\|\boldsymbol{Q}_v^{(N)}(t) - \boldsymbol{Q}_v(t)\right\|_T > \epsilon\right) \to 0$ as $N \to \infty$. Further, from A.20 and the fact that $\boldsymbol{\pi}_v^{(N)} \to \boldsymbol{\pi}$ as $N \to \infty$ we obtain

$$\left\| \int_{\boldsymbol{Q}_v^0} g\left(\boldsymbol{Q}_v(t)\right) d\boldsymbol{\pi} - \int_{\boldsymbol{Q}_v^0} g\left(\boldsymbol{Q}_v^0\right) d\boldsymbol{\pi} \right\|_T \longrightarrow 0\,.$$

So, $\boldsymbol{\pi}$ is an invariant measure of $\boldsymbol{Q}_v(t)$. Assuming g to be the indicator function the result follows, however, the indicator function is not a continuous function. Instead, we can approximate it by a sequence of continuous functions that monotonically converge to the indicator function. We then apply monotone convergence theorem and the result follows.

Finally, from Theorem 3.4.2 we have that $\lim_{t\to\infty} \boldsymbol{Q}_v(t) = \boldsymbol{Q}_v^I$ for all initial conditions in $\mathcal{Q}$. In addition, under the conditions of Theorem 3.4.2 the system has the unique fixed point $\boldsymbol{Q}_v^I$.

Now consider a ball $B_\epsilon\left(\boldsymbol{Q}_v^I\right)$ centered at $\boldsymbol{Q}_v^I$ with radius ϵ. Since the dynamical system converges to $\boldsymbol{Q}_v^I$ so there exits a $t_0 > 0$ such that for any $t > t_0$ the solution of the dynamical system $\boldsymbol{Q}_v(t) \subset B_\epsilon\left(\boldsymbol{Q}_v^I\right)$ for any subset of $\mathcal{Q}$ as the initial conditions. Since $\boldsymbol{\pi}$ is the invariant measure of $\boldsymbol{Q}_v(t)$ therefore $\boldsymbol{\pi}\left(B_\epsilon\left(\boldsymbol{Q}_v^I\right)\right) = 1$, so

$$\boldsymbol{\pi} = \delta_{\boldsymbol{Q}_v^I}\,,$$

where δ is the Dirac delta measure.

Appendix B

PROOFS FOR CHAPTER 4

B.1 Proof of Lemma 4.3.2

Define $h(\gamma) \triangleq \gamma\chi(\gamma)e^{-\gamma}$. Let $\tilde{\gamma}$ be the maximizer of h. We assume that h is unimodal. Note that $f(\gamma) = \frac{h(\gamma)}{e^{-\gamma}+\tau(1-e^{-\gamma})}$ is a continuous function over $[0, \gamma_0]$ as its denominator is always positive. In order for f to be unimodal it should have exactly one maximum, thus its first order derivative should have only one positive root. Setting the first order derivative of f equal to zero we get

$$\tau\left(\chi(\gamma) + \gamma\frac{\partial\chi(\gamma)}{\partial\gamma} - \gamma\chi(\gamma)\right)e^{\gamma} - (\tau-1)\left(\chi(\gamma) + \gamma\frac{\partial\chi(\gamma)}{\partial\gamma}\right) = 0\,. \tag{B.1}$$

Equivalently

$$g(\gamma) \triangleq \dot{h}(\gamma) - (\tau-1)f(\gamma)e^{-\gamma} = 0\,,$$

where $\dot{h}(\gamma)$ is the first order derivative of h. Observe that $g(0) > 0$ and since $\dot{h}(\gamma_0) < 0$ for any $\gamma_0 > \tilde{\gamma}$ (as h is unimodal) then $g(\gamma_0) < 0$. Hence g has at least one positive root in $[0, \gamma_0]$.

By substituting Taylor series expansion of exponential function in (B.1) we have

$$p(\gamma) \triangleq \tau\sum_{i=0}^{K} a_i\gamma^i \sum_{n\geq 0}\frac{\gamma^n}{n!} - (\tau-1)\sum_{i=0}^{K-1} b_i\gamma^i = 0\,, \tag{B.2}$$

where

$$a_i = \frac{i+1}{i!}q_{i+1} - \frac{1}{(i-1)!}q_i \quad \text{and} \quad b_i = \frac{i+1}{i!}q_{i+1}\,. \tag{B.3}$$

Equation (B.2) is a degree ∞ polynomial in γ. We showed that $p(\gamma)$ has at least one positive root. To prove that (B.2) has exactly one positive root we apply generalized Descartes' rule of signs [101] to impose one sign difference to the coefficients of $p(\gamma)$.

Let c_m represent the coefficients of $p(\gamma)$. Then

$$c_k = \begin{cases} \tau \sum_{n=0}^{k} \frac{a_n}{(k-n)!} - (\tau - 1) b_k & 0 \leq k \leq K-1 \\ \tau \sum_{n=0}^{K} \frac{a_n}{(k-n)!} & k > K-1 \end{cases} \tag{B.4}$$

We observe from (B.3) that $a_K < 0$ as $q_{K+1} = 0$. Suppose that $a_i \geq 0$ for all $0 \leq i < K$. In this case $c_k \geq 0$ for all $0 \leq i < K$. This can be shown by induction on k. For $k = 0$, clearly $c_0 > 0$. Suppose that for some $k - 1 > 0$, $c_{k-1} > 0$. Then

$$c_k = \tau \sum_{n=0}^{k} \frac{a_n}{(k-n)!} - (\tau - 1) b_k \tag{B.5}$$

$$= c_{k-1} + \tau \frac{(k+1)q_{k+1} - kq_k}{k!} - (\tau - 1)\frac{k+1}{k!} q_{k+1} + (\tau - 1)\frac{k}{(k-1)!} q_k \tag{B.6}$$

$$= c_{k-1} + \tau \frac{(k+1)q_{k+1} - kq_k}{k!} + (\tau - 1)\frac{k^2 q_k - (k+1)q_{k+1}}{k!} \tag{B.7}$$

$$\geq c_{k-1} + \tau \frac{(k+1)q_{k+1} - kq_k}{k!} + (\tau - 1)\frac{kq_k - (k+1)q_{k+1}}{k!} \tag{B.8}$$

$$= c_{k-1} + \frac{(k+1)q_{k+1} - kq_k}{k!} \tag{B.9}$$

$$> 0 \tag{B.10}$$

The last step follows if $(k+1)q_{k+1} - kq_k \geq 0$ or equivalently $a_k \geq 0$. Now if the rest of the coefficients be negative then according to the Descartes' rule of sign the number of positive roots of a polynomial with real coefficients ordered by descending variable exponent is either equal to the number of sign differences between consecutive nonzero coefficients, or is less than it by an even number. Therefore in order to have one positive root, it is sufficient to have one sign difference. Equivalently, all the coefficients c_k for $k \geq K$ should be negative. We observe that the assumption $a_i \geq 0$ ensures that $c_k < 0$ for all $k \geq K_0$ for some K_0.

Because

$$\begin{aligned}
c_k &= \tau \sum_{n=0}^{K} \frac{a_n}{(k-n)!} \\
&= \tau \left(\frac{a_0}{k!} + \frac{a_1}{(k-1)!} + \cdots + \frac{a_K}{(k-K)!} \right) \\
&= \tau \left(\frac{q_1}{k!0!} + \frac{2q_2 - q_1}{1!(k-1)!} + \cdots + \frac{(K+1)q_{K+1} - Kq_K}{K!(k-K)!} \right) \\
&= \frac{\tau}{k!} \left(\binom{k}{0} q_1 + \binom{k}{1} (2q_2 - q_1) + \cdots + \binom{k}{K} ((K+1)q_{K+1} - Kq_K) \right) \\
&\leq \frac{\tau}{k!} \left(\binom{k}{K} q_1 + \binom{k}{K} (2q_2 - q_1) + \cdots + \binom{k}{K} ((K+1)q_{K+1} - Kq_K) \right) \\
&< 0 \,.
\end{aligned}$$

And $c_k > 0$ for $K \leq k \leq K_0$ which follows by induction starting from $k = K$ and $c_K > 0$. So if $a_i \geq 0$ then f is unimodal. Equivalently we have to have $((i+1)q_{i+1} - iq_i)/i! \geq 0$ for all $1 \leq i \leq K-1$ which yields the result.

B.2 Proof of Theorem 4.3.3

The stability region of the limiting system is the set of vectors $\tilde{\boldsymbol{\lambda}}$ such that the limiting system is stable. The proof is analogous to slotted ALOHA case which is discussed in Chapter 3. Here, we only discuss the global stability of CSMA system. Suppose that a rate vector $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \cdots, \tilde{\lambda}_V)$ is given. From (4.10) and the fact that $f(\gamma)$ is a unimodal function we have that if $\lambda < f(\gamma^*)$ then the equation $f(\gamma) = \lambda$ has two distinct roots $\underline{\gamma}$ and $\overline{\gamma}$. In addition, by the definition of γ, we have $0 \leq \gamma \leq \gamma_0$ so if $\underline{\gamma} \leq \gamma_0 < \overline{\gamma}$ then $\underline{\gamma}$ is the only valid root. This implies that there exists a unique set of non-empty probabilities $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_V)$ and the system is globally stable.

B.3 Proof of Theorem 4.3.7

Suppose that there are n packets in the buffer of a class-v queue upon the arrival of a new packet. An arrival can occur at an idle, collision or a success time slot. So we compute the delay conditioned on the state of the system. Let $T_v(n)$ be the delay being experienced by a newly arrived packet at a class-v queue until it is successfully transmitted. Conditioned on the state of a time slot we have the followings cases:

- *Idle*: At the time of a new arrival, there is a chance that one of the n packets is already in the process of a successful transmission. Therefore, the delay being experienced by the newly arrived packet can be written as follows
  - if $n = 0$ then $T_v(n) = D_{v,\text{CSMA}}$
  - if $n > 0$

$$T_v(n) = \begin{cases} nD_{v,\text{CSMA}} & \frac{1/\tau P_v^{\text{SUCC}}(\boldsymbol{\rho}^{(N)})}{P^{\text{IDLE}}(\boldsymbol{\rho}^{(N)})+1/\tau(1-P^{\text{IDLE}}(\boldsymbol{\rho}^{(N)}))} \\ (n+1)D_{v,\text{CSMA}} & \text{otherwise} \end{cases}$$

- *Collision or Success*: If the new arrival occurs in a transmission period (whether a collision or successful transmission), it is either at the first slot of a transmission period or any other time slot.
  - with probability $1/\tau(P_v^{\text{COL}}(\boldsymbol{\rho}^{(N)})+P_v^{\text{SUCC}}(\boldsymbol{\rho}^{(N)}))$ the arrival occurs in the first time slot and $T(n)$ can be computed as we discussed above for the idle case.
  - with probability $(\tau-1)/\tau(P_v^{\text{COL}}(\boldsymbol{\rho}^{(N)}) + P_v^{\text{SUCC}}(\boldsymbol{\rho}^{(N)}))$ the arrival occurs in any time slot but the first time slot. For all $n \geq 0$ we then have

$$T_v(n) = R + (n+1)D_{v,\text{CSMA}}\,,$$

where R is a random variable representing the residual time until a transmission period ends.

Let $Q_{v,n}$ be the probability that a class-v user has n packets in its buffer at the steady state. Further let $T_{v,\mathrm{CSMA}} = \sum_{n\geq 0} T_v(n) Q_{v,n} + o(1)$ be the total delay of a class-v packet. We write (B.11).

$$T_{v,\mathrm{CSMA}} = \sum_{n>0} \Big(P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) + 1/\tau \Big(P_v^{\mathrm{COL}}(\boldsymbol{\rho}^{(N)}) + P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}) \Big) \Big) \tag{B.11}$$
$$\Bigg[n D_{v,\mathrm{CSMA}} \left(\frac{1/\tau P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})}{P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) + 1/\tau(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}))} \right)$$

$$+ (n+1) D_{v,\mathrm{CSMA}} \left(1 - \frac{1/\tau P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})}{P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) + 1/\tau(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}))} \right) \Bigg]$$
$$+ Q_{v,0} \Big(D_{v,\mathrm{CSMA}} P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) + D_{v,\mathrm{CSMA}} 1/\tau \Big(P_v^{\mathrm{COL}}(\boldsymbol{\rho}^{(N)}) + P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}) \Big) \Big)$$

$$+ \frac{\tau - 1}{\tau} (P_v^{\mathrm{COL}}(\boldsymbol{\rho}^{(N)}) + P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)})) \sum_{n\geq 0} (R + (n+1) D_{v,\mathrm{CSMA}}) Q_{v,n} + o(1)\,.$$

After some manipulation of (B.11) and since $(1 - Q_{v,0}) P_v^{\mathrm{SUCC}}(\boldsymbol{\rho}^{(N)}) = \lambda_v$ we obtain the following

$$T_{v,\mathrm{CSMA}} = \frac{D_{v,\mathrm{CSMA}}(1 - \lambda_v) + d_v}{1 - \lambda_v D_{v,\mathrm{CSMA}}} + o(1), \tag{B.12}$$

where

$$d_v = (\tau - 1) \left(\frac{1}{2} \left(1 - P^{\mathrm{IDLE}}(\boldsymbol{\rho}^{(N)}) \right) + \frac{1}{\tau} \lambda_v D_{v,\mathrm{CSMA}} \right),$$

which results in Theorem 4.3.7.

Appendix C

PROOFS FOR CHAPTER 5

C.1 Proof of Theorem 5.3.1

The invariant measure exists only if the geometric summation in (5.10) is convergent. Suppose that the stationary distribution exists and let ζ_s be the fixed point of the equation (5.12). Thus, the condition

$$0 < \frac{c - c f_{\mathbf{1}}(\zeta_s)}{1 - f_{\mathbf{1-q}}(\zeta_s)} < 1\,, \tag{C.1}$$

should hold. This gives us the following condition.

$$f_{\mathbf{c-1+q}}(\zeta_s) > c - 1\,. \tag{C.2}$$

Furthermore,

$$\zeta_s = \sum_{\ell \geq 0} p_0 c^{-\ell} Q_\ell^{\text{st}} \leq \frac{c}{c-1} p_0\,.$$

And

$$f_{\mathbf{c-1+q}}\left(\frac{c}{c-1} p_0\right) > c - 1\,.$$

C.2 Proof of Corollary 5.3.2

Clearly the left-hand-side (LHS) of (5.17) is an increasing function in ζ_s. We now show that the right-hand-side (RHS) of (5.17) is a non-increasing function in ζ_s and thus a fixed point always exists. Define

$$g(\zeta_s) \triangleq \frac{1 - f_{\mathbf{1}}(\zeta_s)}{1 - f_{\mathbf{1-q}}(\zeta_s)}\,.$$

Observe that $g(\zeta_s)$ is an increasing function in ζ_s because the sign of the first derivative of $g(\zeta_s)$ is determined by the sign of the following term which is positive.

$$\left(\frac{\partial f_{\mathbf{1}-\boldsymbol{q}}}{\partial \zeta_s}(\zeta_s) - \frac{\partial f_{\mathbf{1}}}{\partial \zeta_s}(\zeta_s)\right)(1 - f_{\mathbf{1}}) - \frac{\partial f_{\mathbf{1}}}{\partial \zeta_s}(\zeta_s) f_{\boldsymbol{q}} > 0\,. \tag{C.3}$$

Furthermore, the sign of the first order derivative of (5.17) is determined by the sign of the following term

$$\sum_{i=1}^{L-1} i g^{i-1}(\zeta_s) \sum_{j=0}^{L-1} c^j g^j(\zeta_s) - \sum_{i=1}^{L-1} i c^i g^{i-1}(\zeta_s) \sum_{j=0}^{L-1} g^j(\zeta_s)\,. \tag{C.4}$$

It is not difficult to see that for some $n \in \{0, 1, \cdots, 2L-2\}$, a typical term of (C.4) is of the following form

$$g^{n-1}(\zeta_s) \sum_{i,j:i+j=n} i(c^j - c^i) \tag{C.5}$$

$$= g^{n-1}(\zeta_s) \sum_{i<j:i+j=n} -(i-j)(c^i - c^j) \tag{C.6}$$

$$< 0\,. \tag{C.7}$$

Therefore, the RHS of (5.17) is a non-increasing function in ζ_s.

C.3 Proof of Theorem 5.3.4

To linearize the system we derive the Jacobian matrix computed at the fixed point. Let $\boldsymbol{x}(t) = \boldsymbol{Q}(t) - \boldsymbol{Q}^{\mathrm{st}}$. Assume that $\boldsymbol{x}(t) = (x_0(t), \ldots, x_{L-1}(t))^{\intercal}$. We show that $\boldsymbol{x}(t)$ converges to zero vector exponentially fast as t goes to infinity. After linearization of the original system, we obtain the following set of ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = A\boldsymbol{x}(t). \tag{C.8}$$

$A = [a_{i,j}]$ is a $L \times L$ matrix where $0 \leq i, j \leq L-1$, and

$$a_{0,0} = p_0(f_{\mathbf{1}}(\zeta_s) - 1) + \zeta_s f^{(0)}_{\boldsymbol{q}}(\zeta_s) - p_{L-1}Q^{\mathrm{st}}_{L-1} f^{(0)}_{\mathbf{1}}(\zeta_s) + p_0 Q^{\mathrm{st}}_0 f^{(0)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\,, \tag{C.9}$$

$$a_{0,j} = p_j f_{\boldsymbol{q}}(\zeta_s) + \zeta_s f^{(j)}_{\boldsymbol{q}}(\zeta_s) - p_{L-1}Q^{\mathrm{st}}_{L-1} f^{(j)}_{\mathbf{1}}(\zeta_s) + p_0 Q^{\mathrm{st}}_0 f^{(j)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) \quad 1 \leq j \leq L-2\,, \tag{C.10}$$

$$a_{0,L-1} = p_{L-1} f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) + \zeta_s f^{(L-1)}_{\boldsymbol{q}}(\zeta_s) - p_{L-1}Q^{\mathrm{st}}_{L-1} f^{(L-1)}_{\mathbf{1}}(\zeta_s) + p_0 Q^{\mathrm{st}}_0 f^{(L-1)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\,, \tag{C.11}$$

where $f^{(j)}_{\boldsymbol{q}}(\zeta_s) \triangleq \frac{\partial f_{\boldsymbol{q}}(\zeta(t))}{\partial Q_j(t)}$ is evaluated at $\zeta(t) = \zeta_s$. Furthermore, for $1 \leq i \leq L-1$

$$\begin{aligned} a_{i,j} = &-p_{i-1}Q^{\mathrm{st}}_{i-1} f^{(j)}_{\mathbf{1}}(\zeta_s) + p_i Q^{\mathrm{st}}_i f^{(j)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) \\ &+ (1 - f_{\mathbf{1}}(\zeta_s)) p_{i-1} 1_{\{j=i-1\}} - p_i \left(1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)\right) 1_{\{j=i\}}\,. \end{aligned} \tag{C.12}$$

Observe that $\sum_i x_i(t) = 0$. Further, it can be seen that $a_{i,i}$'s are all negative. Let s_i be the sign of $x_i(t)$ such that $s_i = 1$ if $x_i(t) \geq 0$ and $s_i = -1$ otherwise.

$$\frac{\mathrm{d}}{\mathrm{d}t}|x_i(t)| = -|a_{i,i}||x_i(t)| + \sum_{j \neq i} a_{i,j} s_i s_j |x_j(t)|\,. \tag{C.13}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_i |x_i(t)| = -\sum_i |a_{i,i}||x_i(t)| + \sum_i s_i \sum_{j \neq i} a_{i,j} s_j |x_j(t)| \tag{C.14}$$

$$< \sum_i \sum_{j \neq i} a_{i,j} s_i s_j |x_j(t)| \tag{C.15}$$

$$= \sum_i \sum_{\substack{j \neq i \\ s_i s_j = 1}} a_{i,j}|x_j(t)| - \sum_i \sum_{\substack{j \neq i \\ s_i s_j = -1}} a_{i,j}|x_j(t)| \tag{C.16}$$

$$< \sum_i \max_{\substack{j \neq i \\ s_i s_j = 1}} a_{i,j} \sum_{\substack{j \neq i \\ s_i s_j = 1}} |x_j(t)| - \sum_i \min_{\substack{j \neq i \\ s_i s_j = -1}} a_{i,j} \sum_{\substack{j \neq i \\ s_i s_j = -1}} |x_j(t)| \tag{C.17}$$

$$= -\sum_i \max_{\substack{j \neq i \\ s_i s_j = 1}} a_{i,j}|x_i(t)| - \sum_i \min_{\substack{j \neq i \\ s_i s_j = -1}} a_{i,j}|x_i(t)| \tag{C.18}$$

$$< -2 \min_{i \neq j} a_{i,j} \sum_i |x_i(t)|\,. \tag{C.19}$$

Now we show that $a_{i,j} > 0$ for all $i \neq j$. First, observe that $f^{(j)}_{\boldsymbol{q}}(\cdot) = -p_j f_{\Delta \boldsymbol{q}}(\cdot)$ where $\Delta \boldsymbol{q} = [q_1 - q_2, q_2 - q_3, \cdots, q_K]$ for any real vector $\boldsymbol{q}$. We write

$$
\begin{aligned}
a_{0,j} &= p_j f_{\boldsymbol{q}}(\zeta_s) + \zeta_s f^{(j)}_{\boldsymbol{q}}(\zeta_s) - p_{L-1} Q^{\mathrm{st}}_{L-1} f^{(j)}_{\mathbf{1}}(\zeta_s) + p_0 Q^{\mathrm{st}}_0 f^{(j)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) && \text{(C.20)}\\
&= p_j f_{\boldsymbol{q}}(\zeta_s) - p_j \zeta_s f_{\Delta \boldsymbol{q}}(\zeta_s) + p_{L-1} Q^{\mathrm{st}}_{L-1} p_j f_{\Delta \mathbf{1}}(\zeta_s) - p_0 Q^{\mathrm{st}}_0 p_j f_{\Delta \mathbf{1}-\boldsymbol{q}}(\zeta_s) && \text{(C.21)}\\
&> p_j (1-\zeta_s) f_{\boldsymbol{q}}(\zeta_s) + p_j \zeta_s \left(q_2 + \frac{q_3}{1!}\zeta_s + \cdots + \frac{q_K}{(K-2)!}\zeta_s^{K-2} \right) e^{-\zeta_s} \\
&\qquad + p_0 Q^{\mathrm{st}}_0 p_j \left(f_{\Delta \mathbf{1}}(\zeta_s) - f_{\Delta \mathbf{1}-\boldsymbol{q}}(\zeta_s) \right) && \text{(C.22)}\\
&\geq 0\,. && \text{(C.23)}
\end{aligned}
$$

Moreover, for $i \neq j$

$$
\begin{aligned}
a_{i,j} &= -p_{i-1} Q^{\mathrm{st}}_{i-1} f^{(j)}_{\mathbf{1}}(\zeta_s) + p_i Q^{\mathrm{st}}_i f^{(j)}_{\mathbf{1}-\boldsymbol{q}}(\zeta_s) + (1 - f_{\mathbf{1}}(\zeta_s)) p_{i-1} && \text{(C.24)}\\
&> p_{i-1} Q^{\mathrm{st}}_{i-1} p_i f_{\Delta \mathbf{1}}(\zeta_s) - \frac{1 - f_{\mathbf{1}}(\zeta_s)}{1 - f_{\mathbf{1}-\boldsymbol{q}}(\zeta_s)} p_{i-1} Q^{\mathrm{st}}_{i-1} p_i f_{\Delta \mathbf{1}-\boldsymbol{q}}(\zeta_s) && \text{(C.25)}\\
&\geq p_{i-1} Q^{\mathrm{st}}_{i-1} p_i \left(f_{\Delta \mathbf{1}}(\zeta_s) - f_{\Delta \mathbf{1}-\boldsymbol{q}}(\zeta_s) \right) && \text{(C.26)}\\
&= p_{i-1} Q^{\mathrm{st}}_{i-1} p_i f_{\Delta \boldsymbol{q}}(\zeta_s) && \text{(C.27)}\\
&\geq 0\,. && \text{(C.28)}
\end{aligned}
$$

We then let $\epsilon = 2 \min_{i \neq j} a_{i,j} > 0$ and finally,

$$
\frac{\mathrm{d}}{\mathrm{d}t} \sum_i |x_i(t)| < -\epsilon \sum_i |x_i(t)|, \tag{C.29}
$$

which shows that $x_i(t)$'s converge to zero exponentially fast, thus $\lim_{t \to \infty} Q_i(t) = Q^{\mathrm{st}}_i$.