Queue-Aware Optimal Resource Allocation for the LTE Downlink

A Project Report

submitted by

HUSSAM AHMED P

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

May 2013

THESIS CERTIFICATE

This is to certify that the thesis titled **Queue-Aware Optimal Resource Allocation for the LTE Downlink**, submitted by **Hussam Ahmed P**, to the Indian Institute of Technology, Madras, in partial fullfilment of the requirements for the award of the degree of **Master of Technology**, is a bona fide record of the project work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Srikrishna Bhashyam
Project Guide
Associate Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Dr. Krishna Jagannathan
Project Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 21st May 2013

ACKNOWLEDGEMENTS

I have immense pleasure to acknowledge my sincere gratitude to my project guides, Dr. Srikrishna Bhashyam and Dr. Krishna Jagannathan for their help and guidance during the project. Their valuable suggestions and encouragement helped me a lot in carrying out the work as well as bringing the thesis to this form.

I would like to thank my dear friends Azif, Aneesh, Aseem, Asit, Renuka, Maria, Kruthika, Prathyusha and Vinay for their support and encouragement in the successful completion of this project work. I acknowledge my gratitude to Henry Scheffé who literally rescued me with his lemma while I was struggling with a proof.

Above all, I am forever indebted to the Almighty and to my parents, for their cheerful encouragement, unfailing patience and consistent support.

ABSTRACT

KEYWORDS: Resource Allocation; OFDM; Lyapunov stability; Limited channel information; LTE

We address the problem of optimal downlink resource allocation in an OFDMA system, in a scenario where very limited channel quality information (CQI) is available at the base-station. Our work is particularly applicable in the context of the LTE downlink, since the feedback mechanism we consider closely resembles one of the CQI reporting modes in LTE. Specifically, the users only report the indices of their best M sub-bands and an effective CQI corresponding to these best M bands. Our policy simultaneously performs optimal sub-band assignment and rate allocation, by taking into account channel quality as well as the queue backlogs of each user. The technical novelty of our work lies in exploiting a limit theorem on the best SNRs reported by the users, and combining it within a Lyapunov stability framework. We show that our policy is throughput maximizing among all policies which are constrained to the CQI mechanism considered. Numerical results indicate that in terms of throughput and average delay, our policy compares favorably to existing resource allocation policies such as proportional fair.

We also address the problem of resource allocation in a generic single hop network with a constraint on the deterministic delay of every packet. We tackle this problem by using Lyapunov optimization technique together with the maximization of a concavely extended utility function. The policy tracks the delay of the head-of-line data in each queue and deterministically bound this number. We adopt a stopping rule based approach to analyse the evolution of the delay of the head-of-line data. We show that our policy ensures deterministic delay guarantees and yield a thoughput utility that differs from the optimal value by not more than an amount that is inversely proportional to the delay bound. Our results hold for any generic single hop network with any random arrival process and time varying channel rates.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

LTE	Long Term Evolution
OFDM	Orthogonal Frequency Division Multiplexing
CQI	Channel Quality Information
SNR	Signal to Noise Ratio
BS	Base Station
OFDMA	Orthogonal Frequency Division Multiple Access
UE	User Equipment
RB	Resource Block
EESM	Effective Exponential SNR Mapping
QoS	Quality of Service
KKT	Karush-Kuhn-Tucker
i.i.d.	independent and identically distributed

NOTATION

M	Number of sub-bands reported by a user
K	Number of UEs
i	i^{th} UE/queue
j	j^{th} sub-band
t	t^{th} slot
$A_i(t)$	Number of arrivals to i^{th} queue during slot t
$Q_i(t)$	Queue length of i^{th} UE during slot t
λ_i	Mean arrival rate to i^{th} queue
$\gamma_{ave,i}$	Average SNR for the i^{th} UE
N	Number of sub-bands
$\gamma_i^j(t)$	SNR on the j^{th} sub-band for the i^{th} UE in slot t
$\gamma_i^{(j)}(t)$	SNR on the j^{th} best sub-band for the i^{th} UE in slot t
$\gamma_i^{eff}(t)$	EESM of the best M sub-bands for the i^{th} UE in slot t
I_i	Set of indices of best M sub-bands in descending order of the SNRs
i_j	Index of the j^{th} best sub-band for the i^{th} UE
$\underline{\gamma}^{eff}(t)$	Set of all EESMs reported by UEs
$\underline{I}$	Set of all index matrices reported by UEs
$\underline{Q}$	Set of all queue lengths
$[i, j]$	i^{th} UE - j^{th} sub-band pair
$\hat{I}$	Set of distinct sub-bands reported by at least one UE
M'	Number of distinct sub-bands reported by at least one UE
P	Total power budget at BS during each slot
$C_{i,j}$	Instantaneous capacity of $[i, j]$
$r_{i,j}$	Rate of transmission assigned to $[i, j]$
$P_{i,j}(r_{i,j})$	Outage probability for $[i, j]$ when the assigned rate is $r_{i,j}$
$G_{i,j}(r_{i,j})$	Goodput for $[i, j]$ when the assigned rate is $r_{i,j}$
$\mu_i(t)$	Amount of service for i^{th} queue during slot t
$a_{i,j}$	Fraction of time for which the j^{th} sub-band is assigned to i^{th} UE

$H_{i,j}(t)$	Indicates whether the transmission through $[i, j]$ is successful or not
$\underline{\lambda}$	Arrival rate vector
Λ	Stability region of the network
$\mathcal{P}$	Family of all resource allocation with uniform power with best M feedback
U_j	Set of all UEs which reported j as one of their best M sub-bands
$\hat{P}_{i,j}(r)$	Outage probability for $[i, j]$ calculated using weak limit approximation
$i(j)$	UE allocated to the j^{th} sub-band by the policy
$r^*_{i(j),j}$	Rate assigned to $[i(j), j]$ by the policy
$L()$	Lyapunov function
$\Delta()$	Condtional Lyapunov drift
$\{\alpha_j\}, \{\beta_{i,j}\}, \{\delta_{i,j}\}$	Non-negative Lagrange multipliers
A_{max}	Maximum number of arrivals to one queue in one slot
$D_i(t)$	Amount of data dropped from i^{th} queue during slot t
$\underline{D}(t)$	Drop vector during slot t
$\underline{y}$	Difference between the rate of arrivals and rate of dropping
$g()$	Concave utility function
$\hat{g}()$	Concavely extended utility function
$\underline{\phi}(t)$	Auxiliary vector during slot t
$Z_i(t)$	Virtual queue length for i^{th} UE during slot t
$N_i(t)$	Amount of head-of-line data in i^{th} queue during slot t
$W_i(t)$	Waiting time of head-of-line data in i^{th} queue during slot t
$1_{Q_i}(t)$	Indicates whether i^{th} queue is empty during slot t
$1_{D_i}(t)$	Indicates whether data are dropped from i^{th} queue during slot t
$1_{N_i}(t)$	Indicates whether $N_i(t) \geq \mu_i(t)$
$1_{A_i}(t)$	Indicates whether i^{th} queue has arrivals during slot t

CHAPTER 1

Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is employed in most of the emerging high data-rate wireless cellular standards such as Long Term Evolution (LTE) [8] and IEEE 802.16 (WiMAX). In this thesis, we tackle the problem of optimal resource allocation in downlink of an OFDMA system, in a scenario where very limited channel quality information (CQI) is available at the base-station (BS). Our work is particularly applicable in the context of LTE downlink, since the feedback mechanism we consider closely resembles one of the CQI reporting modes in LTE.

In an OFDM system such as the LTE downlink, the available bandwidth (say 20 MHz) is divided into several hundred sub-carriers (e.g., 512, 1024, or 2048). These sub-carriers need to be allocated to multiple user equipments (UEs). In practice, a *resource block* (RB) pair consisting of 12 contiguous sub-carriers and 14 OFDM symbols in time is the smallest resource allocation unit [9]. After accounting for unusable tones, this leaves us with about 50 to 100 RBs to allocate to the UEs.

In order to schedule the UEs opportunistically, the base-station, in principle, needs to obtain channel quality information from each UE, on each of the resource blocks. This is highly impractical, since it leads to an enormous amount of control overheads on the uplink. To overcome this, UEs in an LTE system report CQI to the base-station in a very sparse manner.

1.1 Related Work

Various reduced feedback mechanisms have been studied in the literature in the context of resource allocation in OFDM downlink. In [14], the CQI of each UE is fed back only for those sub-bands[1] whose quality is better than a certain threshold. The feedback overhead is even more reduced in [26], where the UEs report one-bit per sub-band

[1] A sub-band typically consists of one to three resource blocks.

whenever the channel quality exceeds the threshold. In [7], an opportunistic feedback strategy is considered wherein only the channel gains of pre-specified number of best sub-bands are reported. A variation of this policy has been considered in [10, 11]. In [10], the UEs feedback the average gain of the best M sub-bands and the corresponding indices while in [11], each UE reports an Effective Exponential Signal-to-noise ratio Mapping (EESM) of the best M sub-bands and their respective indices. In effect, EESM translates the different SNRs on parallel channels into a single effective flat-fading SNR [32].

In this thesis, we assume a CQI feedback mechanism similar to [10, 11], since it closely resembles one of the CQI reporting modes – namely, the UE-selected sub-band feedback mode – defined in the LTE standards [1]. Specifically, the UEs only report the indices of their best M sub-bands, where M is a small number (say 2 to 5), and the EESM corresponding to these best M bands.

Downlink resource allocation for OFDM systems has been studied from various perspectives in recent years. In [15], resource allocation in downlink OFDM is posed as a utility maximization problem, which includes proportionally fair resource allocation [18, 31] as a special case. The optimal power and sub-carrier allocation are then determined using convex duality techniques. While [15] assumed full CQI availability except for an estimation error term, [27] takes imperfect CQI into account by factoring for outages due to erroneous CQI at the base-station. In [34], the authors consider opportunistic resource allocation in OFDM under various fairness constraints, and propose a Hungarian algorithm based solution. It is worth noting that [15, 27, 34] assume fully backlogged buffers (i.e., that the base-station always has data to send to the UEs), and do not consider any queuing dynamics.

There is a vast literature on optimal server allocation to constrained queueing systems with time-varying connectivities. Most of the literature in this area based on the landmark papers [30, 29] which introduced Lyapunov techniques for resource allocation. Subsequently, these Lyapunov methods, which explicitly take queue lengths into account for making resource allocation decisions, have been applied in various contexts including high-speed switches [20], satellites [21], wireless [22], and optical networks [6]. In addition to being inherently throughput maximizing, Lyapunov based resource allocation policies can also been used to ensure Quality of Service (QoS) metrics such as delay guarantees [17, 16] and fairness [24].

In the above Lyapunov based resource allocation policies, the resource allocation decision is based on the UE's channel quality as well as queue backlogs, and these are typically assumed to be available perfectly and instantaneously at the base station. In contrast, [33] proposes a throughput optimal resource allocation algorithm under *delayed* channel information; their policy utilizes the conditional expectation of the channel quality, given the delayed measurements. In [19], a cross layer resource allocation policy which maximizes the throughput under delayed CQI and takes into account the channel *outage* event is proposed.

Another important point to note is that the above Lyapunov based resource allocation policies are optimal with respect to the network throughput but does not guarantee any QoS requirements such as bounded delay. In [23], a delay based utility maximization algorithm for a single hop network is proposed which guarantees a bound on the deterministic delay of the packets. The policy also ensures that the throughput-utility is very close to the optimal value. But the author restricts the class of arrival process to bernauli arrivals which simplifies the analysis substantially. There has also been recent work on low-complexity dynamic resource allocation for OFDM [3, 4] to ensure low delay, but these papers do not consider sparse CQI feedback.

1.2 Our Contributions

In the first part of the thesis, we propose a queue-aware resource allocation policy for the OFDM downlink that is optimized for the specific form of the CQI available at the base-station. As described earlier, we assume that the UEs only report the indices of their best M sub-bands, where M is a small number, and the EESM corresponding to these best M sub-bands. We develop a sub-band assignment and rate allocation algorithm which is throughput maximizing under this CQI scenario, when the total number of sub-bands is large. In other words, our algorithm is guaranteed to keep the queueing system stable for all traffic rates that can be stabilized by any resource allocation policy which is constrained to this CQI scenario.

One of the technical contributions of the paper lies in obtaining an explicit characterization of the outage probability on each of the M reported sub-bands. In order to obtain the outage probability expression, we exploit a 'Gumbel' limit theorem on the joint distribution of the best M sub-bands, which subsequently leads to an explicit expression for the conditional density, given the EESM. It is worth commenting that the Gumbel weak limit is an attractor for the extremal values of a fairly large family of distributions [2], so that our work does not crucially depend on the assumption that the sub-band gains are i.i.d. Rayleigh distributed. Another distinguishing feature of our resource allocation policy is that it naturally decouples for each sub-band, and does not entail solving any computationally intensive matching problems [3, 34].

In the second part, we propose a delay based resource allocation algorithm based on Lyapunov optimization technique which ensures deterministic delay guarantees for every packet in any generic single hop network. We adopt a method which is similar to [23] where the author uses the objective function as a concavely extended utility function. The delay bound is guaranteed by tracking the waiting time of the head-of-line data in all the queue during every slot and making the scheduling decisions based on those waiting times. We model the evolution of the waiting time of the head-of-line data using the stopping rule framework. Specifically, for any large enough integer D, we can construct an algorithm that ensures all non- dropped packets have delay less than or equal to D slots, with total throughput-utility that differs from optimal by O(1/D).

1.3 Organization of Thesis

The thesis is organized as follows. The Chapter 2 proposes a throughput optimal resource allocation algorithm for the LTE downlink. A delay based resource allocation policy is proposed in Chapter 3 for any generic single hop network which is near throughput optimal but guarantees an upper bound for the delay. The expression for outage probabilities are obtained in Chapter 4 under the weak limit approximation. The Chapter 5 presents the simulation results and finally the Chapter 6 gives the conclusion of the work.

CHAPTER 2

Throughput Optimal Resource Allocation Policy

In this chapter, we propose an algorithm for the throughput optimal resource allocation in the context of the CQI feed back mechanism to that in LTE. Using Lyapunov stabilty frame work, we prove that the policy is asymptotically throughput optimal by showing that the policy can stabilize set of all arrival rates that can be stabilized by any other policy.

2.1 System Model

Consider a downlink system with one BS and K UEs. The BS maintains a separate queue corresponding to each UE. Time is slotted, and the queue corresponding to i^{th} UE receives exogenous arrivals according to a random process. We denote the amount of data that enters queue i during time slot t by $A_i(t)$, and the queue length corresponding to the i^{th} UE during slot t by $Q_i(t)$. We assume that the arrival process $A_i(t)$ is i.i.d. from slot to slot, with mean λ_i and a finite second moment.

We assume that the channel between the BS and i^{th} UE is a frequency selective Rayleigh fading channel. We remark that this Rayleigh fading assumption is not crucial to our work, but it makes exposition easier. OFDM transmission with N_c sub-carriers is used. The SNR for the i^{th} UE on the j^{th} sub-carrier follows an exponential distribution. The average SNR for the i^{th} UE is denoted as $\gamma_{ave,i}$.

We assume that the downlink channel gains of the UEs are not known to the BS unless the UEs feedback their CQI to the BS. This corresponds to a scenario where the uplink and the downlink channels are not reciprocal, or a scenario where the UEs are not transmitting any data on the uplink, so that reciprocity (even if present) cannot be exploited. In order to reduce feedback overhead, we assume that the sub-carriers are grouped into N sub-bands in such a way that the channel can be approximated as flat-fading in each sub-band. Further, we consider the 'best M' feedback mechanism similar to [11], where each UE reports (i) the EESM corresponding to its best $M(\ll N)$ sub-bands, and (ii) the indices of those sub-bands.

Let $\gamma_i^j(t)$ be the SNR on the j^{th} sub-band for the i^{th} UE in slot t and $\gamma_i^{(1)}(t), \ldots, \gamma_i^{(N)}(t)$ be the ordered sub-band SNRs for i^{th} UE in descending order. The EESM for the best M sub-bands corresponding to the i^{th} UE in slot t, denoted γ_i^{eff}, is defined by

$$\gamma_i^{eff}(t) = -\eta \ln \left(\frac{1}{M} \sum_{j=1}^{M} e^{-\frac{\gamma_i^{(j)}(t)}{\eta}} \right), \tag{2.1}$$

where η is a parameter that depends on the modulation and coding scheme (MCS). Hence, the i^{th} UE reports the following two quantities to the BS during each slot.

(i) The EESM γ_i^{eff},

(ii) The index set $I_i = \{i_1, i_2, \ldots i_M\}$,

where i_j is the index of the j^{th} best sub-band of the i^{th} UE.

Since we are considering a downlink problem, the BS is assumed to know the instantaneous queue lengths $Q_i(t)$ for all the UEs.

2.2 Problem Formulation

In this section, we develop a mathematical formulation of the optimal resource allocation problem. As mentioned earlier, the following information is assumed to be available with the BS during time-slot t (for simplified notation, we omit t):

(i) The EESMs $\underline{\gamma}^{eff} = [\gamma_1^{eff}, \gamma_2^{eff}, \ldots, \gamma_K^{eff}]$

(ii) The index sets $\underline{I} = [I_1, I_2, \ldots, I_K]$

(iii) The queue length vector $\underline{Q} = [Q_1, Q_2, \ldots, Q_K]$.

Given this information, our aim is to come up with a resource allocation policy which can maximize throughput while keeping all queues at the BS stable. In order to make this statement precise, we develop some terminology and notation.

A resource allocation policy performs the following two operations in each slot.

- Sub-band assignment: For each sub-band j that is reported by at least one UE, the policy determines a unique UE to assign the sub-band. (Recall that a sub-band can be allocated to at most one UE due to interference considerations, whereas a UE can be allocated multiple sub-bands).

- Rate allocation: Given that j^{th} sub-band is assigned to i^{th} UE, determine the rate $r_{i,j}$ at which data transmission will take place on j^{th} sub-band.

From now on, we use the notation $[i,j]$ for the i^{th} UE - j^{th} sub-band pair. In the interest of simplicity, we restrict our attention to policies which allocate equal power to all scheduled sub-bands, although our framework can be modified to include optimal power allocation for different sub-bands. To be precise, define $\hat{I} = \cup_{i=1}^{K} I_i$ as the set of all distinct sub-bands reported by at least one UE, and let $M' = |\hat{I}|$ denote the number of such distinct sub-bands. Assume that the BS has a power budget of P for transmissions during each slot. Then, the base station allocates power P/M' to each sub-band. Let $C_{i,j}$ be the instantaneous capacity of $[i,j]$. Under the above assumptions, we have

$$C_{i,j} = \log_2\left(1 + \frac{P}{M'}\gamma_i^j\right). \tag{2.2}$$

For a reliable communication over a sub-band, the rate assigned to $[i,j]$, $r_{i,j}$, should not exceed $C_{i,j}$. Given $\underline{\gamma}^{eff}$ and the index sets $\underline{I}$, we say $[i,j]$ is in outage if the rate allocated to $[i,j]$ is greater than $C_{i,j}$. The outage probability for $[i,j]$ when the assigned rate is $r_{i,j}$ is defined as follows:

$$P_{i,j}(r_{i,j}) = \mathbb{P}\{C_{i,j} < r_{i,j} | \gamma_i^{eff}, I_i\} \tag{2.3}$$

We define a natural metric, namely *goodput*, as the average successfully transmitted rate over a sub-band [28]. The goodput for $[i,j]$ when the assigned rate is $r_{i,j}$ is defined as follows:

$$G_{i,j}(r_{i,j}) = r_{i,j}(1 - P_{i,j}(r_{i,j})). \tag{2.4}$$

Next, we briefly review the queueing dynamics and stability considerations of the queueing system at the BS.

2.2.1 Stability considerations

The queue evolution equation for the i^{th} UE can be written as

$$Q_i(t+1) = \max\{Q_i(t) - \mu_i(t), 0\} + A_i(t), \tag{2.5}$$

where $A_i(t)$ and $\mu_i(t)$ are arrival and service processes of the i^{th} UE queue. Here, $\mu_i(t)$ is the amount of data served from the i^{th} UE queue during slot t, and can be written as

$$\mu_i(t) = \sum_{j=1}^{N} a_{i,j} r_{i,j} H_{i,j}(t).$$

In the above expression, $a_{i,j}$ denotes the fraction of time the j^{th} sub-band is allocated to the i^{th} UE during slot t. Clearly

$$\sum_{i=1}^{K} a_{i,j} \leq 1. \tag{2.6}$$

Later, we will show that our optimal policy allocates a sub-band to at most one UE during each time-slot. Next, $H_{i,j}(t)$ is an indicator random variable which takes a value 1 whenever the transmission through $[i, j]$ during slot t is successful and 0 otherwise. Thus, $\mathbb{P}\{H_{i,j}(t) = 0\} = P_{i,j}^{N}(r_{i,j})$.

In the spirit of [13], we say that the queueing system at the BS is *strongly stable* if for each UE i,

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} E[Q_i(t)] < \infty. \tag{2.7}$$

Denote by $\mathcal{P}$ the family of all resource allocation which allocate equal power to all scheduled sub-bands, and have access only to the parameters $\underline{\gamma}^{eff}$, $\underline{I}$, and $\underline{Q}$ in order to make the resource allocation decisions during each slot. Let Λ be the *stability region* of the network, which is defined as (the closure of) the set of all arrival rates $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_K)$ for which there exists some policy $\Pi \in \mathcal{P}$ under which the queueing system is strongly stable.

Our goal is to find a resource allocation policy in $\mathcal{P}$ which is *throughput optimal*, in the sense that it keeps the queuing system stable for *all* arrival rates in the interior of Λ. We propose the throughput optimal resource allocation algorithm in the next section.

2.3 Throughput Optimal Resource Allocation Policy

During each time slot, the scheduler at the BS observes $\underline{\gamma}^{eff}$, $\underline{I}$, and $\underline{Q}$, and implements the following steps :

1. Determine $\hat{I} = \cup_{i=1}^{K} I_i$ and $M' = |\hat{I}|$.
2. for $j = 1$ to M' do
3. Determine $U_j = \{i|j \in I_i\}$.
4. Calculate an estimate of the outage probability $\hat{P}_{i,j}(r)$ as a function of r for each $i \in U_j$. (See Chapter 4)
5. Calculate
$$r^*_{i,j} = \arg\max_r \{r(1 - \hat{P}_{i,j}(r))\} \; \forall \, i \in U_j.$$
6. Calculate
$$i(j) = \arg\max_{i \in U_j} \{Q_i(t) r^*_{i,j}(1 - \hat{P}_{i,j}(r^*_{i,j}))\}.$$
7. Assign j^{th} sub-band to $i(j)^{th}$ UE, and transmit at rate $r^*_{i(j),j}$.
8. end for

2.3.1 Discussion

In the first step, the scheduler determines the set of all distinct sub-bands reported by the UEs. Then, for each such sub-band j, the scheduler determines the set U_j of all UEs who report that sub-band as being one of their best M sub-bands. In step 4, the outage probability on $[i, j]$ is computed, as explained in Section 4. In step 5, the scheduler computes the rate that ensures the best goodput for each UE $i \in U_j$. Finally, in steps 6 and 7, the scheduler assigns j^{th} sub-band to the i^{th} UE that has the maximum queue-length goodput product.

Notice that the above algorithm assigns every reported sub-band to a unique UE. Also, no power is assigned to sub-bands that are not reported by any UE.

2.3.2 Lyapunov Analysis

In this section, we derive the optimal resource allocation policy as a Lyapunov drift minimizing policy, and prove that it is throughput optimal. We define the quadratic

Lyapunov function

$$L(\underline{Q}(t)) = \sum_{i=1}^{K} (Q_i(t))^2,$$

and consider the *conditional Lyapunov drift*

$$\Delta(\underline{Q}(t)) = \mathbb{E}\left\{L(\underline{Q}(t+1)) - L(\underline{Q}(t)) | \underline{Q}(t)\right\}.$$

We obtain the following inequality by squaring the both sides of (2.5).

$$\begin{aligned}(Q_i(t+1))^2 \leq &(Q_i(t))^2 + \left(\sum_{j=1}^{N} a_{i,j} r_{i,j} H_{i,j}(t)\right)^2 + (A_i(t))^2 \\ &- 2Q_i(t)\left(\sum_{j=1}^{N} a_{i,j} r_{i,j} H_{i,j}(t) - A_i(t)\right).\end{aligned}$$

Taking the sum over all the UEs and using the fact that the sum of squares of non-negative variables is less than or equal to the square of the sum, we get the following inequality.

$$\begin{aligned}L(\underline{Q}(t+1)) - L(\underline{Q}(t)) \leq \sum_{i=1}^{K} (A_i(t))^2 + 2\sum_{i=1}^{K} A_i(t) Q_i(t) \left(\sum_{j=1}^{N}\sum_{i=1}^{K} a_{i,j} r_{i,j} H_{i,j}(t)\right)^2 \\ - 2\sum_{i=1}^{K} Q_i(t) \sum_{j=1}^{N} a_{i,j} r_{i,j} H_{i,j}(t). \quad (2.8)\end{aligned}$$

Using (2.6), we get the following upper bound

$$\sum_{i=1}^{K} a_{i,j} r_{i,j} H_{i,j}(t) \leq \max_i \{C_{i,j}\} < \infty, \ \forall\, j.$$

Thus, taking conditional expectations and exploiting the independence of $A_i(t)$ and $Q_i(t)$, we get

$$\Delta(\underline{Q}(t)) \leq B + 2\sum_{i=1}^{K} Q_i(t)\lambda_i - 2\sum_{i=1}^{K}\sum_{j=1}^{N} Q_i(t) a_{i,j} r_{i,j} (1 - P_{i,j}(r_{i,j})), \quad (2.9)$$

where

$$B = \left(\sum_{j=1}^{N} \max_i \{C_{i,j}\}\right)^2 + \sum_{i=1}^{K} \mathbb{E}\left[A_i(t)^2\right] < \infty.$$

We know from [13, Lemma 4.1] that the Lyapunov drift becoming negative for large queue backlogs is a sufficient condition for the strong stability of the queueing system. With this in mind, we seek the policy that maximizes the negative term on the right hand side of (2.9). We therefore formulate the optimal resource allocation problem as follows.

$$\max_{\{a_{i,j}\},\{r_{i,j}\}} \sum_{i=1}^{K} \sum_{j=1}^{N} Q_i(t) a_{i,j} G_{i,j}(r_{i,j}), \tag{2.10}$$

subject to

$$\begin{aligned} \sum_{i=1}^{K} a_{i,j} &\leq 1,\ \forall\, j, && (C1) \\ a_{i,j} &\geq 0,\ \forall\, i, j, && (C2) \\ r_{i,j} &\geq 0,\ \forall\, i, j. && (C3) \end{aligned}$$

We assume that it is possible to come up with modulation and coding schemes for any desired rate $r_{i,j}$. Then, the above problem is a convex optimization problem and the solution can be obtained easily by using Karush-Kuhn-Tucker (KKT) conditions [5]. The solution is discussed next.

2.3.3 Minimizing the Lyapunov Drift

We now solve the convex optimization problem (2.10) and arrive at our resource allocation policy. Introducing the non-negative Lagrange multipliers $\{\alpha_j\}$,$\{\beta_{i,j}\}$,$\{\delta_{i,j}\}$ for constraints $(C1)$-$(C3)$ respectively, the following conditions also must be satisfied at the optimal solution (superscript $(\cdot)^*$ denotes optimal values).

$$Q_i(t) G_{i,j}(r_{i,j}^*) + \beta_{i,j}^* - \alpha_j^* = 0,\ \forall\, i, j. \tag{2.11}$$

$$Q_i(t) a_{i,j}^* \frac{\partial G_{i,j}(r_{i,j}^*)}{\partial r_{i,j}^*} + \delta_{i,j}^* = 0,\ \forall\, i, j. \tag{2.12}$$

$$\alpha_j^* \left(\sum_{i=1}^{K} a_{i,j}^* - 1 \right) = 0, \; \forall \, j. \quad (2.13)$$

$$\beta_{i,j}^* a_{i,j}^* = 0, \; \forall \, i, j. \quad (2.14)$$

$$\delta_{i,j}^* r_{i,j}^* = 0, \; \forall \, i, j. \quad (2.15)$$

Proposition 1. *The optimal sub-band allocation for problem (2.10) assigns a sub-band exclusively to the UE with the largest corresponding queue-length goodput product.*

Proof. It follows from (2.15) and (2.12) that if $r_{i,j}^* > 0$, then $\delta_{i,j}^* = 0$, i.e., $\frac{\partial G_{i,j}(r_{i,j}^*)}{\partial r_{i,j}^*} = 0$. Thus, $r_{i,j}^*$ is obtained by maximizing goodput of i^{th} UE on j^{th} sub-band (Step 5 of our policy in Section **??**). Similarly, it follows from (2.14) and (2.11) that if $a_{i,j}^* > 0$, then $\beta_{i,j}^* = 0$, i.e., $Q_i(t)G_{i,j}(r_{i,j}^*) = \alpha_j^*$. If $a_{i,j}^* = 0$, then $\beta_{i,j}^* \geq 0$, i.e., $Q_i(t)G_{i,j}(r_{i,j}^*) \leq \alpha_j^*$. Hence, the j^{th} sub-band is assigned to the UE with largest queue-length goodput product $Q_i(t)G_{i,j}(r_{i,j}^*)$. If multiple UEs have the same queue-length goodput product for the same sub-band j, the sub-band can be shared in any arbitrary manner among these users without affecting optimality in terms of the objective function in (2.10). □

Proposition 1 shows that the optimal sub-band allocation assigns each reported sub-band j to the UE which has the maximum queue-length goodput product on the j^{th} sub-band. This establishes that the policy in the Section 2.3 minimizes the Lyapunov drift.

Since the proposed policy ensures the "most negative" Lyapunov drift among the class $\mathcal{P}$, it seems plausible that our policy should be able to stabilize the queueing system, whenever some policy in $\mathcal{P}$ can do so. The following theorem asserts that this is indeed true.

Theorem 1. *The resource allocation policy proposed in Section 2.3 is asymptotically throughput optimal, when the number of sub-bands is large.*

Proof. If the arrival rate vector $\underline{\lambda}$ is stabilizable by some policy $\Pi \in \mathcal{P}$ then $\exists \underline{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_K)$ with $\epsilon_i > 0 \; \forall \; i$ such that

$$\lambda_i \leq \sum_{j=1}^{N} b_{i,j} r_{i,j} (1 - P_{i,j}(r_{i,j})) - \epsilon_i, \quad \forall \; i,$$

where $b_{i,j}$ is the fraction of j^{th} sub-band allocated to i^{th} UE and $r_{i,j}$ is the rate assigned to $[i, j]$ by policy Π. Next, we invoke Scheffé's Lemma [25], which asserts the uniform convergence of $|P_{i,j}(r) - \hat{P}_{i,j}(r)|$ to zero. Thus, for large N,

$$|P_{i,j}(r) - \hat{P}_{i,j}(r)| \leq \delta_{i,j}^{N}, \; \forall \; r, i, j,$$

where $\delta_{i,j}^{N}$ is a small positive number independent of r.. Hence, for large N,

$$1 - P_{i,j}(r) - \delta_{i,j}^{N} \leq 1 - \hat{P}_{i,j}(r) \leq 1 - P_{i,j}(r) + \delta_{i,j}^{N} \tag{2.16}$$

Since for every sub-band, our policy assigns $a_{i,j}^{*}$ and $r_{i,j}^{*}$ such that $\sum_{i=1}^{K} Q_i(t) a_{i,j}^{*} r_{i,j}^{*} (1 - \hat{P}_{i,j}(r_{i,j}^{*}))$ is maximized, the following inequality holds good $\forall \; j, \{b_{i,j}\}, \{r_{i,j}\}$.

$$\sum_{i=1}^{K} Q_i(t) b_{i,j} r_{i,j} (1 - \hat{P}_{i,j}(r_{i,j})) \leq \sum_{i=1}^{K} Q_i(t) a_{i,j}^{*} r_{i,j}^{*} (1 - \hat{P}_{i,j}(r_{i,j}^{*})).$$

Using (2.16) we get,

$$\sum_{i=1}^{K} Q_i(t) b_{i,j} r_{i,j} (1 - P_{i,j}(r_{i,j}) - \delta_{i,j}^{N}) \leq \sum_{i=1}^{K} Q_i(t) a_{i,j}^{*} r_{i,j}^{*} (1 - P_{i,j}(r_{i,j}^{*}) + \delta_{i,j}^{N}).$$

Therefore, for our policy, the Lyapunov drift can be upper bounded as

$$\Delta(\underline{Q}(t)) \leq B - \sum_{i=1}^{K} Q_i(t) \left(\epsilon_i - \sum_{j=1}^{N} (a_{i,j}^{*} r_{i,j}^{*} + b_{i,j} r_{i,j}) \delta_{i,j}^{N} \right).$$

Note that at most M of the $a_{i,j}^{*}$ and $b_{i,j}$ are non-zero for each user i which ensures that the summation is finite even if N is large. Thus, for any $\underline{\epsilon}$, there exists a large enough

N for which

$$\epsilon_i - \sum_{j=1}^{N} (a^*_{i,j} r^*_{i,j} + b_{i,j} r_{i,j}) \delta^N_{i,j} > 0, \; \forall \, i,$$

which ensures that the Lyapunov drift becomes negative as queues grow. i.e., the proposed policy stabilizes all the arrival rates which can be stabilized by any other policy for large enough N. Hence it is asymptotically throughput optimal. □

CHAPTER 3

Delay Based Resource Allocation Policy

In this chapter, we develop a resource allocation policy for a generic single hop network based on the Lyapunov optimization technique described in [13]. Since we impose an additional constraint that the delay of the data should be deterministically bounded it is intuitive to expect that certain arrivals have to be dropped from the buffer inorder to ensure stability for all arrival rates inside the stablizable region of the system. We adopt a method which is similar to [23] where the author uses the objective function as a concavely extended utility function. The delay bound is guaranteed by tracking the waiting time of the head-of-line data in all the queues in every slot and by making the scheduling decisions based on those waiting times.

3.1 Network model

Consider a time slotted generic single hop network with K links numbered as $\{1, 2, \ldots, K\}$. The data arrive randomly every slot and are put into separate queues for different links. Let $A_i(t)$ be the amount of data arrived into queue i during slot. The arrivals to the network are assumed to be i.i.d. over slots and independent over different queues. We assume that the amount of data that can be arrived to a queue in one slot is bounded above by A_{max}. Let $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots \lambda_K)$ be the arrival rate vector. Let $\underline{Q}(t) = (Q_1(t), Q_2(t), \ldots, Q_K(t))$ be the queue-length vector for the network in the beginning of slot t. The data in the queue are marked with their integer arrival times which is used to determine the waiting time in the system. The queue evolution is given by the following equation.

$$Q_i(t+1) = \max\{Q_i(t) - \mu_i(t) - D_i(t), 0\} + A_i(t)$$

where $\mu_i(t)$ is the service given to queue i during slot t and $D_i(t)$ is the amount of data dropped from the i^{th} queue during slot t. The link capacities are assumed to be time varying and are denoted by $\underline{C}(t) = (C_1(t), C_2(t), \ldots C_K(t))$ for the K links during slot t. We assume that the capacity of the link is upper bounded by C_{max}. Let $I_c(t)$ denote the link quality information available at the controller during slot t. Due to the limited link state information at the controller, the transmission can become failure whenever the rate exceeds the capacity of the link. Let $\underline{x}(t) = (x_1(t), x_2(t), \ldots x_K(t))$ be the transmission rate vetor during slot t. Let $P_i^{out}(x_i)$ be the probability that the trasmission rate x_i exceeds the capacity of the link i given the link state information $I_c(t)$ at the controller. The service rate $\mu_i(t)$ is given as follows.

$$\mu_i(t) = x_i(t) 1_i(t)$$

where $1_i(t)$ is an indicator random variable that takes 1 when the transmission through link i is successful. Thus

$$1_i(t) = \begin{cases} 1 & \text{with probability} \quad 1 - P_i^{out}(x_i) \\ 0 & \text{with probability} \quad P_i^{out}(x_i) \end{cases} \tag{3.1}$$

For each queue i, define $Y_i(t) = \lambda_i - D_i(t)$. Let $\underline{y}$ be the time average expectation of $Y_i(t)$

$$\underline{y} = \underline{\lambda} - \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t} \mathbb{E}\{\underline{D}(\tau)\}.$$

The vector $\underline{y}$ is the difference between the rate of arrivals and the rate of droppings, and hence it represents the throughtput vector provided the queues are stable.

3.2 The Optimization Objective

The goal is to propose a delay-based transmission scheme with data dropping that the solves the following problem :

$$\max \quad g(\underline{y}), \tag{3.2}$$

subject to

$$\begin{aligned} \underline{y} &\in \Lambda, \\ 0 \le \; y_i &\le \lambda_i, \; \forall\, i, \end{aligned}$$

where $g(\underline{y})$ is a continuous and concave utility function of the K-dimensional vector. The function is assumed to be defined over the hyper-cube $\mathcal{Y} = \{\underline{y} \mid 0 \le y_i \le A_{max} \forall i\}$. We make the following additional assumption.

Assumption 1 : For each queue i, for any vectors $\underline{y}$ and $\underline{w}$ such that $\underline{y} \in \mathcal{Y}$, $\underline{w} \in \mathcal{Y}$ and $\underline{y} + \underline{w} \in \mathcal{Y}$ we have :

$$g(\underline{y} + \underline{w}) \le g(\underline{y}) + \sum_{i=1}^{K} \nu_i w_i, \tag{3.3}$$

where $\nu_i \ge 0 \; \forall\, i$. Note that the assumption is equivalent to say that for each i, the i^{th} partial derivative of $g()$ is bounded above by a finite constant ν_i.

Let g^* be the maximum value of the objective in the problem (3.2). Our aim is to find a policy which achieve a utility close to g^* and simultaneously guarantees a deterministic delay bound for the non dropped data. In the following sections, we define the required machinery to obtain the solution easily.

3.3 Concave Extension of Utility Function

Suppose that $g(\underline{y})$ satisfies Assumption 1, and define the concave extension of $g(\underline{y})$ as the function $\hat{g}(\underline{y})$ defined over the extended hyper-cube $\hat{\mathcal{Y}} = \{\underline{y} \mid -A_{max} \le y_i \le A_{max} \;\forall\, i\}$ given by,

$$\hat{g}(\underline{y}) = g(\max\{\underline{y}, 0\}) + \sum_{i=1}^{K} \nu_i \min\{y_i, 0\},$$

where $\max\{\underline{y}, 0\} = (\max\{y_1, 0\}, \max\{y_2, 0\}, \ldots, \max\{y_K, 0\})$. Since (3.3) holds, we have the following inequality.

$$\hat{g}(\underline{y}) \le \hat{g}(\underline{y}_i) + \nu_i (y_i + A_{max}), \tag{3.4}$$

where $\underline{y}_i$ is formed from $\underline{y}$ by replacing y_i by $-A_{max}$. This method of concavely extending the utility function is crucial to control the delay to be bounded.

3.4 Equivalent Problem with Virtual Queues

The optimization problem (3.2) can be easily transformed to the following problem using an auxillary vector $\underline{\phi}(t)$.

$$\max \ \hat{g}(\underline{\phi}), \tag{3.5}$$

subject to

$$\begin{aligned} y_i &\geq \phi_i \,\forall\, i, \\ -A_{max} \leq \phi_i &\leq A_{max}, \,\forall\, i, \\ \overline{\underline{Q}} &< \infty, \end{aligned}$$

where $\overline{\underline{Q}}$ and $\underline{y}$ are achievable on the network and $\overline{\underline{Q}} = \lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t} \mathbb{E}\{\underline{Q}(\tau)\}$.

We solve the above problem by using the Lyapunov optimization technique for stabilizing a set of virtual queues $Z(t) = (Z_1(t), Z_2(t), \ldots, Z_K(t))$ with the update equation as follows. For the i^{th} queue,

$$Z_i(t+1) = \max\{Z_i(t) - \lambda_i + D_i(t) + \phi_i(t), 0\}. \tag{3.6}$$

Stabilizing this virtual queues ensures that the first constraint in problem (3.5) is satisfied. Note that the virtual queue updation equation needs the knowledge of the actual arrival rate vector $\underline{\lambda}$. In-order to incorporate the delay into the resource allocation policy, we define $W_i(t)$ as the waiting time of the head-of-line data in the i^{th} queue on slot t. Let $W_i(t) = 0$ if the queue i is empty. Let $N_i(t)$ be the amount of head-of-line data in the i^{th} queue. Define $1_{Q_i}(t)$ as an indicator variable that is 1 if $Q_i(t) > 0$, and is zero if the queue is empty. Define $1_{D_i}(t)$ as an indicator random variable which takes a value 1 whenever $D_i(t) > 0$ and zero otherwise. Let $1_{N_i}(t)$ be an indicator variable which takes a value 1 if $N_i(t) \geq \mu_i(t)$ and zero otherwise. We observe that $W_i(t)$ satisfies the following update equation.

$$W_i(t+1) = 1_{Q_i}(t)\max\{W_i(t) + 1 - \rho(1_{D_i}(t), J_i(t), 1_{N_i}(t)), 0\} + (1 - 1_{Q_i}(t))1_{A_i(t)} \tag{3.7}$$

where $\rho(1_{D_i}(t), J_i(t), 1_{N_i}(t))$ represents the time after which the data at the head-of-line at the end of slot t (after the service during slot t) have arrived with respect to the arrival epoch of the head-of-line data in the beginning of the slot and $1_{A_i(t)}$ is an indicator variable which takes 1 if $A_i(t) > 0$ and zero otherwise. We have,

$$\rho(1_{D_i}(t), J_i(t), 1_{N_i}(t)) = \begin{cases} 0, & 1_{D_i}(t) = 0 \ \&\ 1_{N_i}(t) = 1 \\ T_i(t), & 1_{D_i}(t) = 1 \ \&\ 1_{N_i}(t) = 1 \\ J_i(t), & 1_{N_i}(t) = 0 \end{cases}$$

where

$$J_i(t) = \min\{\omega \mid \sum_{\tau=1}^{\omega} A_i(t - W_i(t) + \tau) > \mu_i(t) - N_i(t)\}. \tag{3.8}$$

and

$$T_i(t) = \min\{\omega \mid A_i(t - W_i(t) + \omega) > 0, \omega > 0\}.$$

The update equation of the waiting time of the head-of-line data given by (3.7) can be understood as follows: If the queue is empty, the value of $W_i(t+1)$ is 1 if and only if there is a new arrival in the current slot t. Alternatively if the queue is non empty, $\mu_i(t)$ amount of data is served from the queue. Now if $\mu_i(t)$ is less than the amount of head-of-line data in the queue (i.e., $1_{N_i}(t) = 1$), then the remaining packets are either dropped or retained in the queue. Hence $W_i(t+1)$ can become $W_i(t) + 1 - T_i(t)$ or $W_i(t) + 1$ depending on the value of $D_i(t)$. Here, $T_i(t)$ represent the interarrival time between the head-of-line data and the subsequent arrivals. On the other hand, if $1_{N_i}(t) = 0$, the data which arrived $J_i(t)$ slots after the head-of-line data will become the new head-of-line in the next slot which makes $W_i(t+1)$ to be $W_i(t) + 1 - J_i(t)$.

Claim 1. *For any queue i in every slot t,*

$$\rho(1_{D_i}(t), J_i(t), 1_{N_i}(t)) \geq J_i(t) - 1 \tag{3.9}$$

Proof. Note that $J_i(t)$ becomes 1 whenever $1_{N_i}(t) = 1$. Hence, whenever $1_{N_i}(t) = 1$, irrespective of whether data are dropped or not, $\rho(1_{D_i}(t), J_i(t), 1) \geq J_i(t) - 1 = 0$.

Now if $1_{N_i}(t) = 0$, then $\rho(1_{D_i}(t), J_i(t), 1) = J_i(t) \geq J_i(t) - 1$.

$\square$

Without loss of generality, assume that $\lambda_i > 0$ for all the queues (else, just remove the queues that have no arrivals). Define $\Theta(t) = [\underline{Z}(t); \underline{W}(t); \underline{N}(t)]$ where $\underline{W}(t) = (W_1(t), W_2(t), \ldots W_K(t))$ and $\underline{N}(t) = (N_1(t), N_2(t), \ldots N_K(t))$. We use the following non-negative Lyapunov function:

$$L(\Theta(t)) = \frac{1}{2}\sum_{i=1}^{K} Z_i(t)^2 + \frac{1}{2}\sum_{i=1}^{K} \lambda_i W_i(t)^2. \tag{3.10}$$

3.5 Minimizing the Drift-Minus-Utility

Define $\Delta(\Theta(t))$ as the conditinal Lyapunov drift given by,

$$\Delta(\Theta(t)) = \mathbb{E}\left\{L(\Theta(t+1)) - L(\Theta(t))|\Theta(t)\right\}.$$

Now using the Lyapunov optimization frame work in [13], our strategy is to make transmission and dropping decisions to minimize a bound on the following drift-minus-utility expression every slot:

$$\zeta = \Delta(\Theta(t)) - V\mathbb{E}\{\hat{g}(\underline{\phi}(t))|\Theta(t)\},$$

where V is a non-negative control parameter. We have the following preliminary lemmas.

Lemma 1. *During every slot t, for queue i*

$$\mathbb{E}\{J_i(t)^2|\Theta(t)\} < \infty$$

Proof. We have,

$$\{J_i(t) = n\} = \{A^{n-1} < \mu_i(t) - N_i(t)\} \cap \{A^n > \mu_i(t) - N_i(t)\}$$

where $A^n = \sum_{\tau=1}^{n} A_i(t - W_i(t) + \tau)$. Define $P_n = \mathbb{P}(J_i(t) = n|\Theta(t), \mu_i(t) = \mu)$.

Now, Conditioning on $\mu_i(t) = \mu$, we have,

$$\begin{aligned} P_n &\leq \mathbb{P}\{A^{n-1} < \mu - N_i(t)\} \\ &\leq \mathbb{P}\{A^{n-1} < C_{max}\} \end{aligned}$$

Let m be the number of slots between $t - W_i(t)$ and $t - W_i(t) + n$ (excluding the boundaries) in which there are no arrivals. Now, if $n \geq \lfloor C_{max} \rfloor + 1$, then

$$m \geq n - 1 - \lfloor C_{max} \rfloor$$

Hence, for $n \geq \lfloor C_{max} \rfloor + 1$

$$\begin{aligned} P_n &\leq k_1 a_0^{n-1-\lfloor C_{max} \rfloor} \\ &\leq k_2 a_0^n \end{aligned}$$

where $a_0 = \mathcal{P}(A_i(t) = 0)$ and k_1 and k_2 are finite constants. Now,

$$\begin{aligned} \mathbb{E}\{J_i(t)^2 | \Theta(t), \mu_i(t)\} &= \sum_{n=1}^{\infty} n^2 P_n \\ &\leq k_3 + k_2 \sum_{n=\lfloor C_{max} \rfloor + 1}^{\infty} n^2 a_0^n < \infty \end{aligned}$$

where k_3 is a finite positive constant. Hence $\mathbb{E}\{J_i(t)^2 | \Theta(t)\} < \infty$. $\square$

Lemma 2. *Every slot t, the Lyapunov drift satisfies:*

$$\Delta(\Theta(t)) \leq B - \sum_{i=1}^{K} Z_i(t)\mathbb{E}\{Y_i(t) - \phi_i(t)|\Theta(t)\} - \sum_{i=1}^{K} \lambda_i W_i(t)\mathbb{E}\{J_i(t) - 2|\Theta(t)\},$$

where B is a finite positive constant.

Proof. We get the following inequality by squaring (3.6).

$$Z_i(t+1)^2 \leq Z_i(t)^2 + \phi_i(t)^2 + Y_i(t)^2 - 2Z_i(t)(Y_i(t) - \phi_i(t)) - 2Y_i(t)\phi_i(t)$$

Summing over all the queues and taking conditional expection we get,

$$\mathbb{E}\{\sum_{i=1}^{K}(Z_i(t+1)^2 - Z_i(t)^2)|\Theta(t)\} \leq B_1 - 2\sum_{i=1}^{K}\mathbb{E}\{Z_i(t)(Y_i(t) - \phi_i(t))|\Theta(t)\}$$

where B_1 is a finite positive constant. Now using (3.7) and Claim 1, we get the following bound.

$$W_i(t+1)^2 \leq W_i(t)^2 + 4 + J_i(t)^2 - 2W_i(t)(J_i(t) - 2)$$

Summing over all the queues, taking conditional expectation and using Lemma 1,

$$\mathbb{E}\{\sum_{i=1}^{K}\lambda_i(W_i(t+1)^2 - W_i(t)^2)|\Theta(t)\} \leq B_2 - 2\sum_{i=1}^{K}\lambda_i\mathbb{E}\{W_i(t)(J_i(t) - 2)|\Theta(t)\}$$

where B_2 is a finite positive constant. Adding the above bounds, the Lemma follows directly. □

Lemma 3. *Every slot t, the drift-minus-utility satisfies :*

$$\zeta \leq B_3 + \sum_{i=1}^{K}(2\lambda_i + N_i(t))W_i(t) - \mathbb{E}\{V\hat{g}(\underline{\phi}(t)) - \sum_{i=1}^{K} Z_i(t)\phi_i(t)|\Theta(t)\}$$
$$- \sum_{i=1}^{K} Z_i(t)\mathbb{E}\{Y_i(t)|\Theta(t)\} - \sum_{i=1}^{K} W_i(t)\mathbb{E}\{\mu_i(t)|\Theta(t)\},$$

where B_3 is a finite constant.

Proof. Every slot t, conditioning on the service $\mu_i(t)$, the random variable $J_i(t)$ can be viewed as a *stopping rule* which depends on the set of independent arrivals to the i^{th} queue. Applying Wald's equality we get,

$$\mathbb{E}\{A^{J_i(t)}|\mu_i(t), \Theta(t)\} = \lambda_i \mathbb{E}\{J_i(t)|\mu_i(t), \Theta(t)\}. \tag{3.11}$$

where $A^{J_i(t)} = \sum_{\tau=1}^{J_i(t)} A_i(t - W_i(t) + \tau)$. Clearly, by definition of $J_i(t)$, $A^{J_i(t)} > \mu_i(t) - N_i(t)$. Thus,

$$\mathbb{E}\{J_i(t)|\mu_i(t), \Theta(t)\} > \frac{\mu_i(t) - N_i(t)}{\lambda_i}. \tag{3.12}$$

Now taking expectation over $\mu_i(t)$ we get,

$$\mathbb{E}\{J_i(t)|\Theta(t)\} > \frac{\mathbb{E}\{\mu_i(t)|\Theta(t)\} - N_i(t)}{\lambda_i}. \tag{3.13}$$

From Lemma 2 and (3.13), the lemma follows directly. □

Our delay based resource allocation policy in the next section makes control decision for $\underline{\phi}(t), \underline{D}(t)$ and the also the transmission rate allocation.

3.6 Delay Based Resource Allocation Policy

During slot t, the BS observes $\Theta(t)$, $I_c(t)$, and perform the following operations.

1. *Congestion control step*:

   Choose $\underline{\phi}(t)$ as the solution to the following problem:

$$\max \quad V\hat{g}(\underline{\phi}(t)) - \sum_{i=1}^{K} Z_i(t)\phi_i(t) \tag{3.14}$$

   subject to

$$-A_{max} \leq \quad \phi_i(t) \quad \leq A_{max},\ \forall\, i,$$

2. *Transmission rate allocation*:

   Choose $\{x_i\}$ as the solution to the following:

$$\max \quad \sum_{i=1}^{K}(W_i(t) + m_i(t)Z_i(t))x_i(1 - P_i^{out}(x_i)) \tag{3.15}$$

   where $m_i(t) = 1$ if $W_i(t) \geq Z_i(t)$ and zero otherwise. Transmit $\mu_i(t) = x_i(t)1_i(t)$ data from the i^{th} queue.

3. *Dropping decision*:

   If $m_i(t) = 1$ and $N_i(t) > \mu_i(t)$, then drop the remaining head-of-line data in the i^{th} queue.

The following lemma is useful to prove that the delay is deterministically bounded.

Lemma 4. *If $Z_i(t) > V\nu_i$ for a particular slot and queue i, then the congestion control step in the above policy chooses $\phi_i(t) = -A_{max}$ for that slot.*

Proof. The value of $\phi_i(t)$ is determined by maximizing $V\hat{g}(\underline{\phi}(t))|\Theta(t) - \sum_{i=1}^{K} Z_i(t)\phi_i(t)$ over $-A_{max} \leq \phi_i(t) \leq A_{max}$. By (3.4) we know that for any vector $\underline{\phi}(t)$ such that $-A_{max} \leq \phi_i(t) \leq A_{max}$,

$$V\hat{g}(\underline{\phi}(t)) - \sum_{m=1}^{K} Z_m(t)\phi_m(t) \leq V\hat{g}(\underline{\phi}_i(t)) + V\nu_i(\phi_i(t) + A_{max}) - \sum_{m=1}^{K} Z_m(t)\phi_m(t)$$

Because $V\nu_i < Z_i(t)$, the upper bound is maximized at $\phi_i(t) = -A_{max}$ and equality holds if and only if $\phi_i(t) = -A_{max}$. Hence the lemma. □

Theorem 2. *The delay based resource allocation policy given above minimizes the upper bound for the drift-minus-utiltiy given in Lemma 3. i.e., the following quantity is maximized:*

$$\mathbb{E}\{V\hat{g}(\underline{\phi}(t)) - \sum_{i=1}^{K} Z_i(t)\phi_i(t)|\Theta(t)\} + \sum_{i=1}^{K} Z_i(t)\mathbb{E}\{Y_i(t)|\Theta(t)\} + \sum_{i=1}^{K} W_i(t)\mathbb{E}\{\mu_i(t)|\Theta(t)\}$$

Proof. Note that the $\phi_i(t)$ terms appear separably in the expression, and hence they can be optimally chosen by maximizing the following quantity:

$$V\hat{g}(\underline{\phi}(t)) - \sum_{i=1}^{K} Z_i(t)\phi_i(t)$$

subject to $-A_{max} \leq \phi_i(t) \leq A_{max}$ for all queues i. This is precisely the congestion control step of the policy.

Next, we have to maximize the following quantity by optimally allocating the resources to the users:

$$\sum_{i=1}^{K} W_i(t)\mathbb{E}\{\mu_i(t)|\Theta(t)\} - \sum_{i=1}^{K} Z_i(t)\mathbb{E}\{D_i(t)|\Theta(t)\}$$

The amount of data dropped from the queue i in slot t, $D_i(t)$ is given by,

$$D_i(t) = m_i(t)\max\{N_i(t) - \mu_i(t), 0\}$$

The above equation can be understood as follows. When $m_i(t) = 1$ i.e., $W_i(t) \geq Z_i(t)$, and also if there are more data in the head-of-line of queue i than the service in slot t, then the remaining data are dropped. In any other case, no data are dropped from the queue. This is exactly the dropping strategy given in the policy. Now,

$$\mathbb{E}\{D_i(t)|\Theta(t)\} = m_i(t)\mathbb{P}(\mu_i(t) < N_i(t))(N_i(t) - \mathbb{E}\{\mu_i(t)|\Theta(t)\}),$$
$$\leq m_i(t)(N_i(t) - \mathbb{E}\{\mu_i(t)|\Theta(t)\})$$

Now, maximizing (3.6) is equivalent to maximize the following:

$$\sum_{i=1}^{K}(W_i(t) + m_i(t)Z_i(t))\mathbb{E}\{\mu_i(t)|\Theta(t)\}.$$

This is precisely the transmission rate allocation step in the proposed policy. □

Theorem 3. *The proposed resource allocation policy ensures a deterministic bound on both the virtual queue size and the delay of the head-of-line data in every queue in every slot. More precisely, for any queue i,*

$$Z_i(t) \leq \lceil V\nu_i \rceil + 2A_{max} \text{ and } W_i(t) \leq \lceil V\nu_i \rceil + 2A_{max}. \tag{3.16}$$

Proof. Consider the i^{th} queue. At t = 0, when the queues are empty, clearly, $Z_i(0) \leq \lceil V\nu_i \rceil + 2A_{max}$ and $W_i(0) \leq \lceil V\nu_i \rceil + 2A_{max}$. Hence the theorem holds for $t = 0$. We use this as the basis for the following proof by induction. We assume that (3.16) holds for given t. From (3.6) we have,

$$Z_i(t+1) \leq Z_i(t) + D_i(t) + \phi_i(t) \leq Z_i(t) + 2A_{max}.$$

If $Z_i(t) \leq \lceil V\nu_i \rceil$, then $Z_i(t+1) \leq \lceil V\nu_i \rceil + 2A_{max}$. Now if $Z_i(t) > \lceil V\nu_i \rceil$, by Lemma 4, $\phi_i(t) = -A_{max}$ and hence $Z_i(t+1) \leq Z_i(t) \leq \lceil V\nu_i \rceil + 2A_{max}$. Therefore, by induction, $Z_i(t)$ is bounded above as in (3.16). Similarly from (3.7) we have,

$$W_i(t+1) \leq W_i(t) + 1.$$

If $W_i(t) \leq \lceil V\nu_i \rceil + 2A_{max} - 1$, then $W_i(t+1) \leq \lceil V\nu_i \rceil + 2A_{max}$. Now if $W_i(t) > \lceil V\nu_i \rceil + 2A_{max} - 1$, then $W_i(t) = \lceil V\nu_i \rceil + 2A_{max}$ (since $W_i(t)$ must be an integer) which makes $W_i(t) \geq Z_i(t)$ and hence the head-of-line data will be completely drained off (either by successful transmission or if not by dropping). Therefore, $W_i(t+1) \leq W_i(t) \leq \lceil V\nu_i \rceil + 2A_{max}$ and hence by induction, the theorem holds for all t. □

Theorem 4. *The delay based resource allocation policy ensures the delay of all non-dropped data to be less than or equal to* D *slots, with total throughput-utility that differs from optimal by* $\mathcal{O}(1/D)$.

Proof. Clearly Theorem 3 asserts that the delay of any non-dropped data in the i^{th} queue is bounded by $\lceil V\nu_i \rceil + 2A_{max}$ slots. Hence by choosing V such that, $D = \lceil V\nu_{max} \rceil + 2A_{max}$ where $\nu_{max} = \max_i \nu_i$, the delay of any non-dropped packet can be deterministically bounded by D slots.

Now by Theorem 2, the proposed policy minimizes the drift-minus-utility. The Lyapunov Optimization given in [13, Theorem 5.4] says that the policy which minimizes the drift-minus-utility gives a lower bound for the throughput-utility as follows.

$$\liminf_{t\to\infty} g(y(t)) \geq g^* - B/V, \tag{3.17}$$

where $y(t) = \underline{\lambda} - \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\{\underline{D}(\tau)\}$. Hence the throughput differs from the optimal by $\mathcal{O}(1/D)$. □

CHAPTER 4

Derivation of Outage Probability

In this section, we describe how the BS estimates the outage probability on $[i, j]$ in Step 4 of our algorithm, using only the parameters $\underline{\gamma}^{eff}$ and $\underline{I}$. We utilize a limit theorem on the order statistics of the SNRs to derive an expression for the conditional joint distribution of the SNRs on the best M sub-bands for each UE, given the EESM and the sub-band indices. For ease of exposition, we assume that the SNRs on the sub-bands of a given UE are i.i.d. exponentially distributed. This assumption will hold well in the case of Rayleigh fading in a rich multi-path environment, with number of paths comparable to the number of sub-bands. However, we remark that the limit theorem we are about to exploit holds for a fairly large class of distributions – namely, those which lie within the Gumbel domain of attraction [2]. Therefore, our policy remains asymptotically throughput optimal for this class of sub-band SNR distributions.

We first state a result which follows from [12, Theorem 15] regarding the order statistics of M extremal values, drawn from N i.i.d. exponential random variables .

Theorem 5. *Let $Z_1, Z_2, \ldots, Z_N$ be a sequence of i.i.d. unit exponential random variables, and $Z_{(1)}, Z_{(2)}, \ldots, Z_{(N)}$ be the corresponding order statistics in descending order. Then*

$$(e^{-\tilde{Z}_{(1)}}, e^{-\tilde{Z}_{(2)}}, \ldots, e^{-\tilde{Z}_{(M)}}) \xrightarrow{\mathfrak{D}} (Y_1, Y_2, \ldots Y_M),$$

as $N \to \infty$, where $\tilde{Z}_{(i)} = Z_{(i)} - \ln N$, $Y_i = \sum_{j=1}^{i} X_j$ and X_js are i.i.d. unit exponential random variables.

Proof. Note that $1-e^{-Z_1}, 1-e^{-Z_2}, \ldots, 1-e^{-Z_N}$ is a sequence of i.i.d. standard uniform random variables. Now directly applying the [12, Theorem 15], the result follows. □

Lemma 5. *Let $\underline{Y}^{(n)} = (Y_1, Y_2, \ldots, Y_n)$ with the entries $Y_i = \sum_{j=1}^{i} X_j, \forall\, i = 1, \ldots, n$, where X_js are i.i.d. unit exponential random variables. The joint pdf of $\underline{Y}^{(n)}$ is given by*

$$f_{\underline{Y}^{(n)}}(y_1, y_2, \ldots, y_n) = e^{-y_n}, 0 \leq y_1 \leq y_2 \leq \ldots \leq y_n. \quad (4.1)$$

Proof. Let $f_{X_i}(x_i)$ and $f_{Y_i}(y_i)$ denote the pdf of X_i and Y_i respectively. Thus,

$$f_{X_i}(x) = e^{-x}, x \geq 0, \forall i.$$

Consider $n = 2$.

$$\begin{aligned} f_{\underline{Y}^{(2)}}(y_1, y_2) &= f_{Y_1}(y_1) f_{Y_2|Y_1=y_1}(y_2), \quad 0 \leq y_1 \leq y_2, \\ &= f_{X_1}(y_1) f_{X_2}(y_2 - y_1), \\ &= e^{-y_2}, \qquad\qquad 0 \leq y_1 \leq y_2. \end{aligned}$$

Hence the lemma holds for $n = 2$. We use this as the basis for the following proof by induction. We assume that the (4.1) holds for given n. Then,

$$\begin{aligned} f_{\underline{Y}^{(n+1)}}(y_1, y_2, \dots, y_{n+1}) &= f_{\underline{Y}^{(n)}}(y_1, y_2, \dots, y_n) f_{Y_{n+1}|\underline{Y}^{(n)}=(y_1, y_2, \dots, y_n)}(y_{n+1}), \\ &= e^{-y_n} f_{X_{n+1}}(y_{n+1} - y_n), \\ &= e^{-y_{n+1}}, \qquad 0 \leq y_1 \leq y_2 \leq \dots \leq y_{n+1}. \end{aligned}$$

So by induction, the lemma holds for all $n \geq 2$. □

Consider the i^{th} UE. Let $S_i^j = e^{-(\gamma_i^{i_j} - \ln N)}$ for $j = 1, \dots, M$. Assuming that the number of sub-bands N is large, we can use Theorem 5 and Lemma 5 , to approximate the unconditional joint pdf of $\underline{S}_i^{(M)} = (S_i^1, S_i^2, \dots, S_i^M)$ as

$$f_{\underline{S}_i^{(M)}}(s_1, s_2, \dots, s_M) = e^{-s_M}, \quad 0 \leq s_1 \leq s_2 \leq \dots \leq s_M.$$

Numerical results indicate that this approximation is good even for moderate values of N. Next, define $S_i^{eff} = e^{-(\gamma_i^{eff} - \ln N)}$. Choosing the parameter η in (2.1) as unity for simplicity, we have

$$S_i^{eff} = \frac{1}{M} \sum_{j=1}^{M} S_i^j.$$

Note that S_i^{eff} is known to the BS. Next, conditioned on $S_i^{eff} = s$ and $I_i = I$, $\underline{S}_i^{(M)}$ takes values only on the hyperplane $\frac{1}{M} \sum_{j=1}^{M} S_i^j = s$. Hence, we ignore the M^{th} best

SNR and calculate the joint CDF [1] of $(S_i^1, \dots S_i^{M-1}, S_i^{eff})$ as follows.

$$\begin{aligned}
F_{S_i^1,\dots S_i^{M-1},S_i^{eff}}(s_1,\dots s_{M-1},s) &= \mathbb{P}(S_i^1 \le s_1, \dots, S_i^{M-1} \le s_{M-1}, S_i^{eff} \le s), \\
&= \mathbb{P}(S_i^1 \le s_1, \dots, S_i^{M-1} \le s_{M-1}, S_i^{M} \le Ms - \sum_{j=1}^{M-1} s_j), \\
&= F_{\underline{S}_i^{(M)}}(s_1,\dots s_{M-1}, Ms - \sum_{j=1}^{M-1} s_j).
\end{aligned}$$

Now taking the partial derivatives with respect to $s_1, \dots, s_{M-1}, s$, we get,

$$\begin{aligned}
f_{S_i^1,\dots S_i^{M-1},S_i^{eff}}(s_1,\dots s_{M-1},s) &= M f_{\underline{S}_i^{(M)}}(s_1,\dots s_{M-1}, Ms - \sum_{j=1}^{M-1} s_j), \\
&= M e^{-(Ms - \sum_{j=1}^{M-1} s_j)}, \;\; 0 \le s_1 \le s_2 \le \dots \le Ms - \sum_{j=1}^{M-1} s_j.
\end{aligned}$$

Next we calculate the conditional joint pdf of the best $M-1$ sub-bands.

$$\begin{aligned}
f_{\underline{S}_i^{(M-1)}|S_i^{eff}=s, I_i=I}(s_1, s_2, \dots, s_{M-1}) &= \frac{f_{S_i^1,\dots S_i^{M-1},S_i^{eff}}(s_1,\dots s_{M-1},s)}{f_{S_i^{eff}}(s)}, \\
&= \frac{M e^{-(Ms - \sum_{j=1}^{M-1} s_j)}}{f_{S_i^{eff}}(s)}, 0 \le s_1 \le s_2 \le \dots \le Ms - \sum_{j=1}^{M-1} s_j.
\end{aligned}$$

Now, we can find the conditional marginal density $f_{S_i^j|S_i^{eff}=s,I_i=I}(s_j)$ for $j = 1, \dots, M-1$ by integrating out the other variables in the above expression. Similarly, the $f_{S_i^M j|S_i^{eff}=s,I_i=I}(s_M)$ can be obtained by ignoring the best sub-band SNR. The outage probability can thus be determined as

$$\begin{aligned}
\hat{P}_{i,i_j}(r) &= \mathbb{P}(C_{i,i_j} \le r | \gamma_i^{eff} = \gamma, I_i = I), \\
&= \mathbb{P}(\log_2(1 + \frac{P}{M'}\gamma_i^{i_j}) \le r | \gamma_i^{eff} = \gamma, I_i = I), \\
&= \mathbb{P}(S_i^j \ge N e^{-\frac{2^r-1}{\frac{P}{M'}}} | S_i^{eff} = N e^{-\gamma}, I_i = I), \\
&= \int_y^N f_{S_i^j|S_i^{eff}=Ne^{-\gamma},I_i=I}(s_j) ds_j.
\end{aligned} \tag{4.2}$$

where $y = Ne^{-\left(\frac{2^r-1}{\frac{P}{M'}}\right)}$. Closed form expressions for the outage probabilities (4.2) can be obtained through some tedious computations.We provide explicit expressions for the

[1] F(.) denotes Cumulative Distribution Function (CDF).

cases $M = 3$ and $M = 4$ in Appendix A. The outage probabilities computed in (4.2) are used in Step 4 of the resource allocation algorithm.

CHAPTER 5

Simulation Results

In this section, we present simulation results that demonstrate the throughput gains achieved by the proposed policy over other existing policies. We also demonstrate that the limiting approximation we use to obtain closed form outage probability expressions is a good approximation.

The proposed policy (labeled as "optimal" in the plots) is throughput optimal among all policies that use the limited channel feedback scheme described in Section **??**. Three important components of our policy are: (1) evaluation of the *conditional expected CQI* for each sub-band from the EESM, (2) evaluation of goodput while *accounting for outage probability*, and (3) optimal utilization of *queue length* information. To illustrate the importance of each component of our proposed policy, we compare the proposed policy with the following policies (each of the heuristic policies ignores *at least* one component of our proposed policy): (1) a throughput optimal policy with perfect CQI (labelled "Perfect CQI"), (2) a policy that uses queue length information but assumes that the reported EESM is the CQI for all the best M reported sub-bands (labelled "Heuristic 1"), (3) a policy that uses queue length information and evaluates the conditional expected CQI for the best M reported sub-bands *without* accounting for outage probability (labelled "Heuristic 2"), and (4) a proportionally fair rate allocation policy that uses the conditional expected CQI and goodput evaluation without using queue length information (labelled "PF").

A single-cell OFDM downlink with $K = 100$ UEs is simulated. The number of subcarriers is 512 and there are 12 subcarriers in each sub-band. Two channel models are considered: (1) IID sub-bands, and (2) Correlated sub-bands resulting from a 6-path channel with an uniform power-delay profile where each path is Rayleigh fading. The arrival traffic for the i^{th} UE is assumed to be Poisson with with parameter λ_i. The channel feedback from each UE is assumed to be the best M sub-bands and EESM for these sub-bands.

Figures 5.1 and 5.2 show the average queue length (averaged across UEs and time

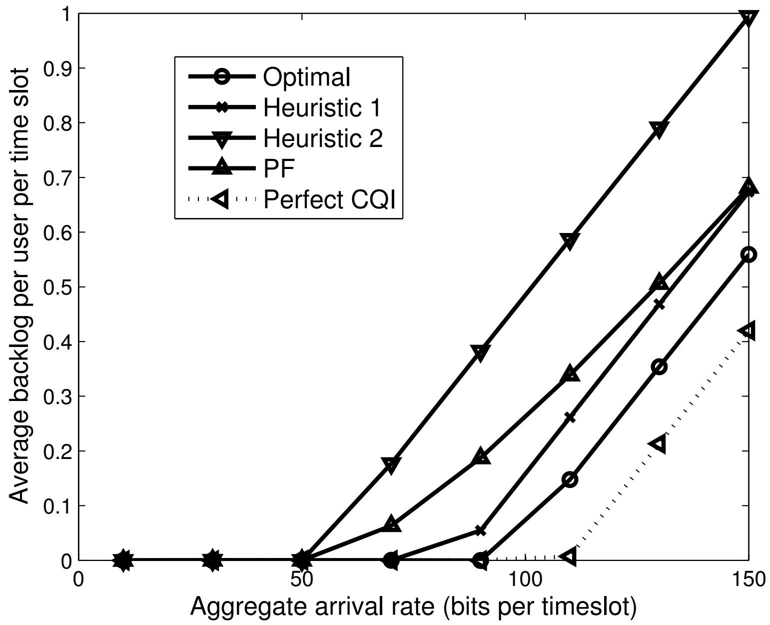

Figure 5.1: IID sub-bands case: $M = 3$, $N = 43$, $\eta = 1$, $\gamma_{ave,i} = 1 \, \forall \, i$.

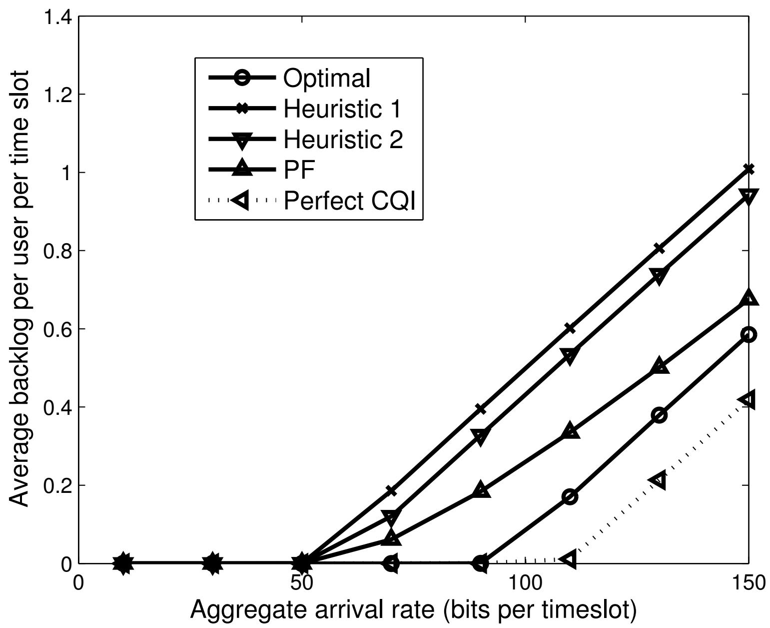

Figure 5.2: Correlated sub-bands case: $M = 3$, $N = 43$, $\eta = 1$, $\gamma_{ave,i} = 1 \, \forall \, i$.

slots) versus the aggregate arrival rate (i.e., sum of λ_i's) for the IID and correlated sub-band cases respectively. λ_i is chosen as $i\lambda$, i.e., each UE has a different arrival traffic rate, and λ is varied to change the arrival traffic load. Also, $M = 3$ and the number of sub-bands $N = 43$. It is clear that the proposed policy can support significantly higher arrival traffic for the same average queue length than the heuristic policies. The Perfect CQI policy is also shown to quantify the loss due to limited feedback. It is also

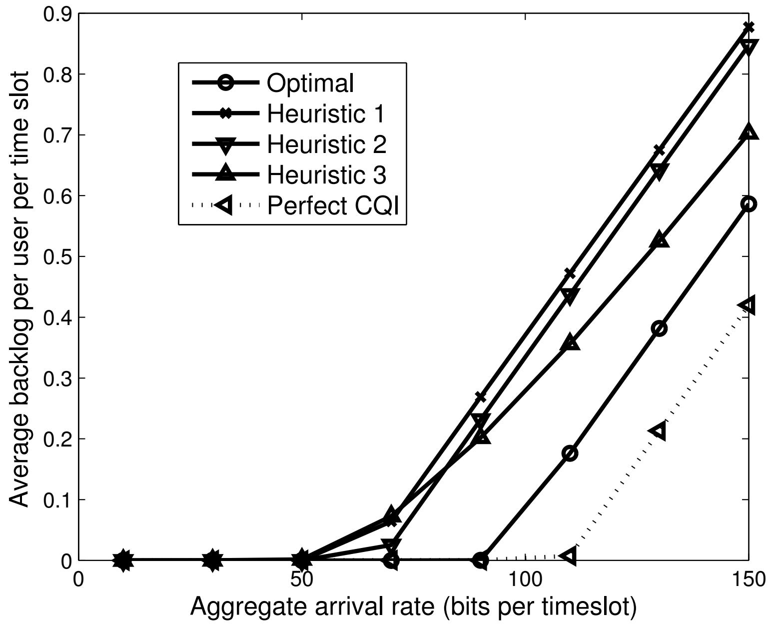

Figure 5.3: IID sub-bands case: $M = 4$, $N = 43$, $\eta = 1$, $\gamma_{ave,i} = 1 \, \forall \, i$.

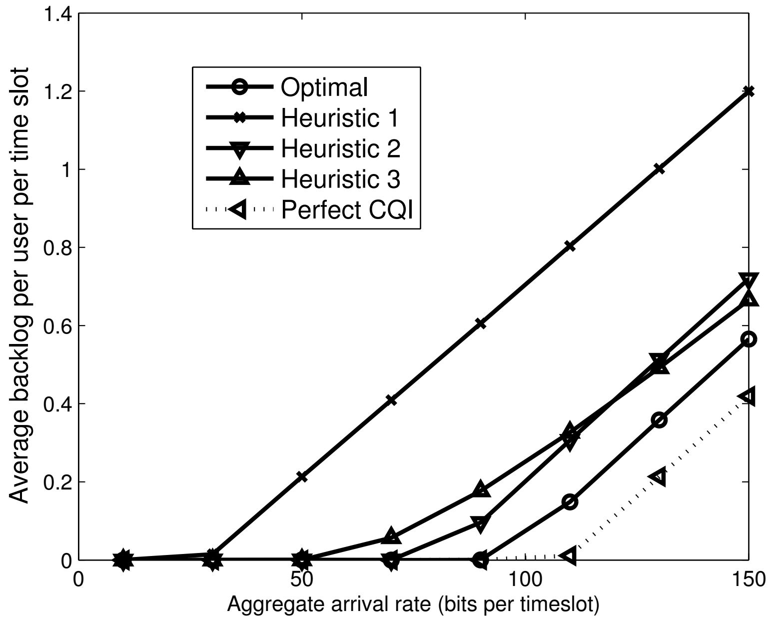

Figure 5.4: Correlated sub-bands case: $M = 4$, $N = 43$, $\eta = 1$, $\gamma_{ave,i} = 1 \, \forall \, i$.

clear that the proposed policy provides similar performance gains even in the correlated sub-band case. Similar results have been observed for $M = 4$ (figures 5.3 and 5.4).

Figure 5.5 shows the conditional CDF of the CQI of the best sub-band given a particular EESM for the best M sub-bands. Four cases of N (the total number of sub-bands) are shown. Note that the number of subcarriers is $12N$. It can be observed that: (1) the weak limit approximation is very good for $N = 22$ and $N = 43$, and (2) the IID

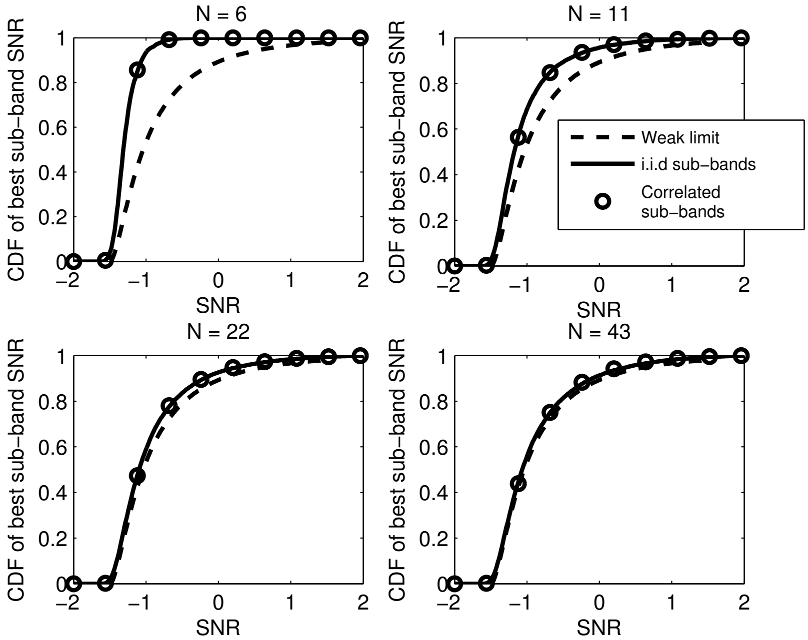

Figure 5.5: Conditional CDF of best sub-band given $\gamma_{eff} = -1.5 + \log N$, $M = 3$, $\eta = 1$.

and correlated sub-band cases are not very different.

CHAPTER 6

Conclusion

We proposed two queue-aware policies for allocating sub-bands in the LTE downlink when each UE reports the best M sub-band indices and a single effective CQI for these bands. The throughput optimality of the first policy was shown using the Lyapunov stability framework. The policy assigns each sub-band to the UE with the best queue-length goodput product for that sub-band. The goodput was obtained by deriving analytical expressions for the conditional outage probability of each sub-band given the effective CQI. The conditional outage probability was derived by exploiting a limit theorem on the joint distribution of the SNR of the best sub-bands. The proposed policy supports signifcantly higher arrival traffic than existing policies like: (1) proportional fair allocation based on CQI that does not consider queue information, (2) queue-aware policies that use the effective CQI as the CQI of each sub-band, and (3) queue-aware policies that do not account for *outage* in the estimation of goodput. The second policy ensures a bound on deterministc delay, say D slots, for all the UEs provided the throughput is not less than the optimal value by $\mathcal{O}(1/D)$. The policy is obtained using Lyapunov optimization technique which involves minimization of drift-minus-utility. The concavely extended utility function is used in the analysis which is crucial in bounding the delay.

APPENDIX A

Expressions for Outage Probability for special cases

Case 1. *No of sub-bands reported,* $M = 3$.

The region for which the conditional joint pdf of S_i^1 and S_i^2 is non-zero is shown by the shaded area in Figure A.1

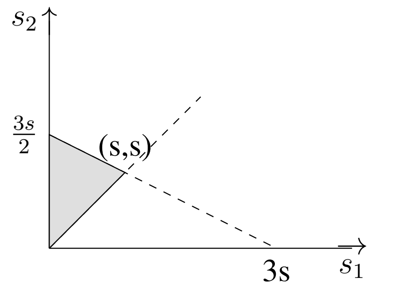

Figure A.1: Region for which the conditional joint pdf of S_i^1 and S_i^2 given $S_i^{eff} = s$ is non-zero

We can find the marginal density as follows. Note that conditional pdfs are non-zero only for the specified region.

(i) Pdf of S_i^{eff}.

– For $0 \leq s \leq \infty$,

$$f_{S_i^{eff}}(s) = \int_0^s \int_{s_1}^{\frac{3s-s_1}{2}} 3e^{-(3s-s_1-s_2)} ds_2 ds_1,$$
$$= \frac{9}{2}e^{-s} + \frac{3}{2}e^{-3s} - 6e^{-\frac{3s}{2}}.$$

(ii) Best sub-band SNR.

– For $0 \leq s_1 \leq s$,

$$f_{S_i^1|S_i^{eff}=s}(s_1) = \int_{s_1}^{\frac{3s-s_1}{2}} \frac{3e^{-(3s-s_1-s_2)}}{f_{S_i^{eff}}(s)} ds_2,$$
$$= \frac{6e^{\left(-\frac{3s}{2}+\frac{s_1}{2}\right)} - 6e^{2s_1}}{3 - 12e^{\frac{3s}{2}} + 9e^{2s}}.$$

(iii) Second best sub-band SNR.

– For $0 \leq s_2 \leq s$,

$$f_{S_i^2|S_i^{eff}=s}(s_2) = \int_0^{s_2} \frac{3e^{-(3s-s_1-s_2)}}{f_{S_i^{eff}}(s)} ds_1,$$
$$= \frac{6e^{s_2}(-1+e^{s_2})}{3-12e^{\frac{3s}{2}}+9e^{2s}}.$$

– For $s \leq s_2 \leq \frac{3s}{2}$,

$$f_{S_i^2|S_i^{eff}=s}(s_2) = \int_0^{3s-2s_2} \frac{3e^{-(3s-s_1-s_2)}}{f_{S_i^{eff}}(s)} ds_1,$$
$$= \frac{6e^{3s}(e^{-s_2}-e^{-3s+s_2})}{3-12e^{\frac{3s}{2}}+9e^{2s}}.$$

(iv) Third best sub-band SNR.

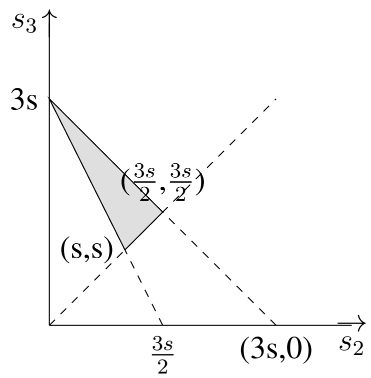

Figure A.2: Region for which the conditional joint pdf of S_i^2 and S_i^3 given $S_i^{eff} = s$ is non-zero

In-order to find the expression for $f_{S_i^3|S_i^{eff}=s}(s_3)$, we follow a similar procedure but by ignoring the best sub-band SNR. The joint pdf of $(S_i^{eff}, S_i^2, \ldots, S_i^3)$ is given by

$$f_{S_i^{eff},S_i^2,\ldots,S_i^3}(s, s_1, \ldots, s_M) = M f_{\underline{S}_i^{(3)}}(3s-s_2-s_3, s_2, s_3),$$
$$= 3e^{-s_3}, \;\; 0 \leq 3s-s_2-s_3 \leq s_2 \leq s_3.$$

The conditional joint pdf can is given by

$$f_{S_i^2,S_i^3|S_i^{eff}=s,I_i=I}(s_2, s_3) = \frac{3e^{-s_3}}{f_{S_i^{eff}}(s)}, \;\; 0 \leq 3s-s_2-s_3 \leq s_2 \leq s_3. \quad \text{(A.1)}$$

The region specified in (A.1) is shown in Figure A.2. The conditional marginal pdf is obtained as follows.

- For $s \leq s_3 \leq \frac{3s}{2}$,

$$f_{S_i^3|S_i^{eff}=s}(s_3) = \int_{\frac{3s-s_3}{2}}^{s_3} \frac{3e^{-s_3}}{f_{S_i^{eff}}(s)} ds_2,$$
$$= \frac{9e^{-s_3}(s_3 - s)}{3e^{-3s} - 12e^{-\frac{3s}{2}} + 9e^{-s}}.$$

- For $\frac{3s}{2} \leq s_3 \leq 3s$,

$$f_{S_i^3|S_i^{eff}=s}(s_3) = \int_{\frac{3s-s_3}{2}}^{3s-s_3} \frac{3e^{-s_3}}{f_{S_i^{eff}}(s)} ds_2,$$
$$= \frac{3e^{-s_3}(3s - s_3)}{3e^{-3s} - 12e^{-\frac{3s}{2}} + 9e^{-s}}.$$

The outage probabilities $\hat{P}_{i,i_j}(r)$ for $j = 1, \ldots, 3$ can be obtained as follows. let $y = Ne^{-\left(\frac{2^r-1}{\frac{P}{M'}}\right)}$.

(i) Best sub-band.

- For $0 \leq y \leq s$,

$$\hat{P}_{i,i_1}(r) = \frac{3e^{2s} + e^{2y} - 4e^{\left(\frac{3s}{2}+\frac{y}{2}\right)}}{1 - 4e^{\frac{3s}{2}} + 3e^{2s}}.$$

(ii) Second best sub-band.

- For $0 \leq y \leq s$,

$$\hat{P}_{i,i_2}(r) = \frac{2e^s}{1 + 2e^{\frac{s}{2}} + 3e^s} + \frac{(e^s - e^y)(e^s + e^y - 2)}{1 - 4e^{\frac{3s}{2}} + 3e^{2s}}.$$

- For $s \leq y \leq \frac{3s}{2}$,

$$\hat{P}_{i,i_2}(r) = \frac{2e^{-y}(e^{\frac{3s}{2}} - e^y)^2}{1 - 4e^{\frac{3s}{2}} + 3e^{2s}}.$$

(iii) Third best sub-band.

- For $s \leq y \leq \frac{3s}{2}$,

$$\hat{P}_{i,i_3}(r) = \frac{3e^{-3s} - 12e^{-\frac{3s}{2}} - 9se^{-y} + 9ye^{-y} + 9e^{-y}}{3e^{-3s} - 12e^{-\frac{3s}{2}} + 9e^{-s}}.$$

- For $\frac{3s}{2} \leq y \leq 3s$,

$$\hat{P}_{i,i_3}(r) = \frac{3e^{-3s} - 3e^{-y} + 9se^{-y} - 3ye^{-y}}{3e^{-3s} - 12e^{-\frac{3s}{2}} + 9e^{-s}}.$$

Case 2. *Number of sub-bands reported, M = 4.*

We can find the marginal densities as follows. Let $s_4 = 4s - s_1 - s_2 - s_3$.

(i) Pdf of S_i^{eff}.

– For $0 \leq s \leq \infty$,

$$f_{S_i^{eff}}(s) = \int_0^s \int_{s_1}^{\frac{4s-s_1}{3}} \int_{s_2}^{\frac{4s-s_1-s_2}{2}} 4e^{-s4} ds_3 ds_2 ds_1,$$
$$= \frac{32}{3}e^{-s} + 8e^{-2s} - \frac{2}{3}e^{-4s} - 18e^{-\frac{4s}{3}}.$$

(ii) Best sub-band SNR.

– For $0 \leq s_1 \leq s$,

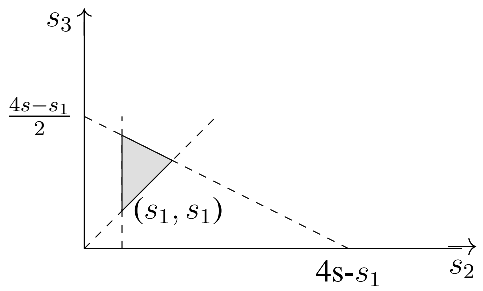

Figure A.3: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $0 \leq s_1 \leq s$.

$$f_{S_i^1|S_i^{eff}=s}(s_1) = \int_{s_1}^{\frac{4s-s_1}{3}} \int_{s_2}^{\frac{4s-s_1-s_2}{2}} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_3 ds_2,$$
$$= \frac{e^{\frac{s_1}{2}-4s}\left(2e^{\frac{5s_1}{2}} - 8e^{2s+\frac{s_1}{2}} + 6e^{\frac{16s-s_1}{6}}\right)}{f_{S_i^{eff}}(s)}.$$

(iii) Second best sub-band SNR.

- For $0 \leq s_2 \leq s$,

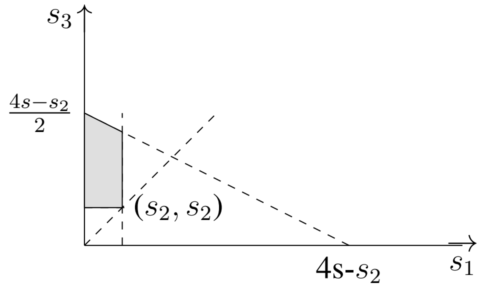

Figure A.4: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $0 \leq s_2 \leq s$.

$$f_{S_i^2|S_i^{eff}=s}(s_2) = \int_0^{s_2} \int_{s_2}^{\frac{4s-s_1-s_2}{2}} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_3 ds_1,$$
$$= \frac{4e^{\frac{s_2}{2}-2s}\left(2e^{\frac{s_2}{2}}-2\right) - 4e^{2s_2-4s}\left(e^{s_2}-1\right)}{f_{S_i^{eff}}(s)}.$$

- For $s \leq s_2 \leq \frac{4s}{3}$,

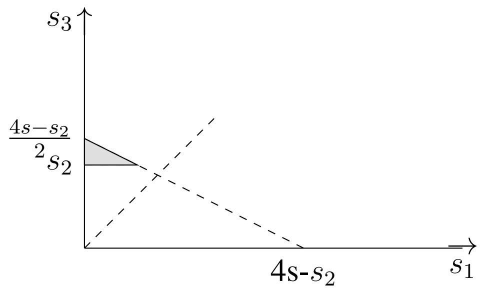

Figure A.5: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $s \leq s_2 \leq \frac{4s}{3}$.

$$f_{S_i^2|S_i^{eff}=s}(s_2) = \int_0^{4s-3s_2} \int_{s_2}^{\frac{4s-s_1-s_2}{2}} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_3 ds_1,$$
$$= \frac{4e^{-s_2} - 8e^{\frac{s_2}{2}-2s} + 4e^{2s_2-4s}}{f_{S_i^{eff}}(s)}.$$

(iv) Third best sub-band SNR.

– For $0 \leq s_3 \leq s$,

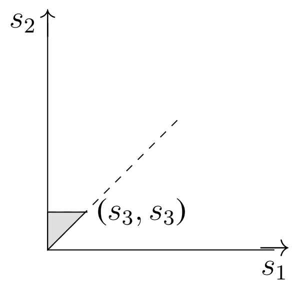

Figure A.6: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $0 \leq s_3 \leq s$.

$$f_{S_i^3|S_i^{eff}=s}(s_3) = \int_0^{s_3} \int_0^{s_2} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_1 ds_2,$$
$$= \frac{2e^{s_3}(e^{s_3}-1)^2 e^{-4s}}{f_{S_i^{eff}}(s)}.$$

– For $s \leq s_3 \leq \frac{4s}{3}$,

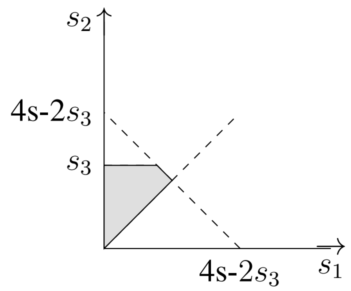

Figure A.7: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $s \leq s_3 \leq \frac{4s}{3}$.

$$f_{S_i^3|S_i^{eff}=s}(s_3) = \int_0^{2s-s_3} \int_0^{s_2} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_1 ds_2 + \int_{2s-s_3}^{s_3} \int_0^{4s-2s_3-s_2} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_1 ds_2,$$
$$= \frac{4e^{-2s} - 4e^{2s_3-4s} - 8se^{-s_3} + 8s_3e^{-s_3} + 2(e^{2s}-e^{s_3})^2 e^{-4s-s_3}}{f_{S_i^{eff}}(s)}.$$

– For $\frac{4s}{3} \leq s_3 \leq 2s$,

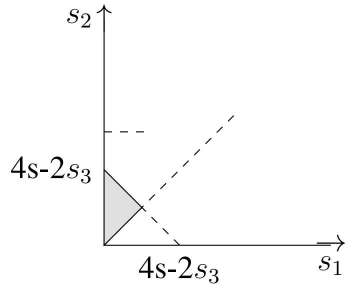

Figure A.8: Region for which the conditional joint pdf of $\underline{S}_i^{(3)}$ given $S_i^{eff} = s$ is non-zero for $\frac{4s}{3} \leq s_3 \leq 2s$.

$$f_{S_i^3|S_i^{eff}=s}(s_3) = \int_0^{2s-s_3}\int_0^{s_2} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_1 ds_2 + \int_{2s-s_3}^{4s-2s_3}\int_0^{4s-2s_3-s_2} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_1 ds_2,$$
$$= \frac{4e^{-2s} - 4e^{-s_3} + 8se^{-s_3} - 4s_3e^{-s_3} + 2(e^{2s} - e^{s_3})^2 e^{-4s-s_3}}{f_{S_i^{eff}}(s)}.$$

(iv) Fourth best sub-band SNR.

– For $s \leq s_4 \leq \frac{4s}{3}$,

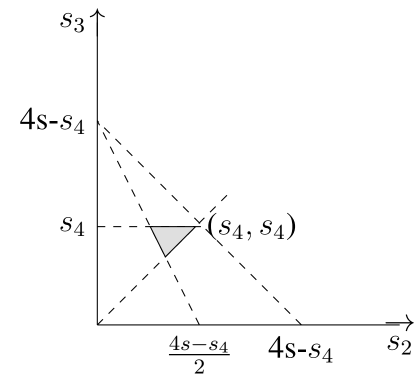

Figure A.9: Region for which the conditional joint pdf of (S_i^2, S_i^3, S_i^4) given $S_i^{eff} = s$ is non-zero for $s \leq s_4 \leq \frac{4s}{3}$.

$$f_{S_i^4|S_i^{eff}=s}(s_4) = \int_{\frac{4s-s_4}{3}}^{s_4}\int_{\frac{4s-s_4-s_3}{2}}^{s_3} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_2 ds_3,$$
$$= \frac{16e^{-s_4}(s-s_4)^2}{3f_{S_i^{eff}}(s)}.$$

– For $\frac{4s}{3} \leq s_4 \leq 2s$,

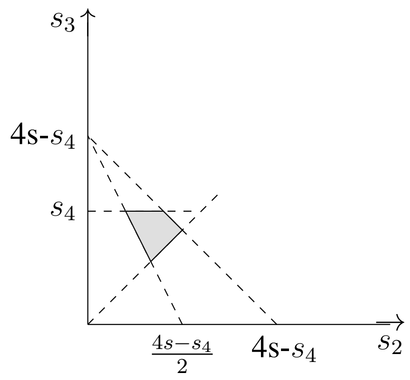

Figure A.10: Region for which the conditional joint pdf of (S_i^2, S_i^3, S_i^4) given $S_i^{eff} = s$ is non-zero for $\frac{4s}{3} \leq s_4 \leq 2s$.

$$\begin{aligned} f_{S_i^4|S_i^{eff}=s}(s_4) &= \int_{\frac{4s-s_4}{3}}^{\frac{4s-s_4}{2}} \int_{\frac{4s-s_4-s_3}{2}}^{s_3} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_2 ds_3 + \int_{\frac{4s-s_4}{2}}^{s_4} \int_{\frac{4s-s_4-s_3}{2}}^{4s-s_4-s_3} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_2 ds_3, \\ &= \frac{e^{-s_4}(4s-s_4)^2 - 3e^{-s_4}(4s-3s_4)(12s-5s_4)}{12 f_{S_i^{eff}}(s)}. \end{aligned}$$

– For $2s \leq s_4 \leq 4s$,

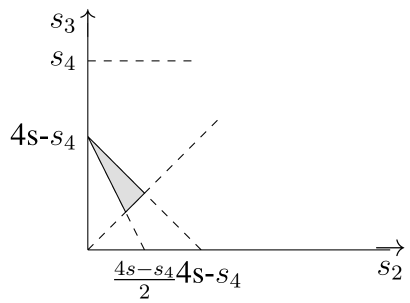

Figure A.11: Region for which the conditional joint pdf of (S_i^2, S_i^3, S_i^4) given $S_i^{eff} = s$ is non-zero for $2s \leq s_4 \leq 4s$.

$$\begin{aligned} f_{S_i^4|S_i^{eff}=s}(s_4) &= \int_{\frac{4s-s_4}{3}}^{\frac{4s-s_4}{2}} \int_{\frac{4s-s_4-s_3}{2}}^{s_3} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_2 ds_3 + \int_{\frac{4s-s_4}{2}}^{4s-s_4} \int_{\frac{4s-s_4-s_3}{2}}^{4s-s_4-s_3} \frac{4e^{-s_4}}{f_{S_i^{eff}}(s)} ds_2 ds_3, \\ &= \frac{e^{-s_4}(4s-s_4)^2}{3 f_{S_i^{eff}}(s)}. \end{aligned}$$

The expressions for outage probabilities are evaluated in Mathematica and are omitted here.

REFERENCES

[1] (2009). Evolved universal terrestrial radio access (E-UTRA); physical layer procedures (release 8). Technical report, TS 36.213 (v8.8.0), 3rd Generation Partnership Project (3GPP).

[2] **Bertin, E.** and **M. Clusel** (2006). Generalised extreme value statistics and sum of correlated variables. *Journal of Physics A*.

[3] **Bodas, S.**, **S. Shakkottai**, **L. Ying**, and **R. Srikant**, Low-complexity scheduling algorithms for multi-channel downlink wireless networks. *In INFOCOM*. 2010.

[4] **Bodas, S.**, **S. Shakkottai**, **L. Ying**, and **R. Srikant**, Scheduling for small delay in multi-rate multi-channel wireless networks. *In INFOCOM*. 2011.

[5] **Boyd, S.** and **L. Vandenberghe**, *Convex optimization*. Cambridge university press, 2004.

[6] **Brzezinski, A.** and **E. Modiano** (2005). Dynamic reconfiguration and routing algorithms for IP-over-WDM networks with stochastic traffic. *Journal of Lightwave Technology*, **23**(10), 3188.

[7] **Choi, Y.-J.** and **S. Bahk**, Selective channel feedback mechanisms for wireless multichannel scheduling. *In Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*. IEEE Computer Society, 2006.

[8] **Cox, C.**, *An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications*. Wiley, 2012.

[9] **Dahlman, E.**, **S. Parkvall**, **J. Skold**, and **P. Beming**, *3G Evolution: HSPA and LTE for Mobile Broadband*. Academic Press, 2008, second edition.

[10] **Donthi, S.** and **N. Mehta** (2011). An accurate model for eesm and its application to analysis of cqi feedback schemes and scheduling in lte. *Wireless Communications, IEEE Transactions on*, **10**(10), 3436 –3448. ISSN 1536-1276.

[11] **Donthi, S.** and **N. Mehta** (2011). Joint performance analysis of channel quality indicator feedback schemes and frequency-domain scheduling for lte. *Vehicular Technology, IEEE Transactions on*, **60**(7), 3096 –3109. ISSN 0018-9545.

[12] **Ferguson, T.**, *A Course in Large Sample Theory: Texts in Statistical Science*, volume 38. Chapman & Hall/CRC, 1996.

[13] **Georgiadis, L.**, **M. Neely**, and **L. Tassiulas**, *Resource allocation and cross-layer control in wireless networks*. Now Pub, 2006.

[14] **Gesbert, D.** and **M. S. Alouini**, How much feedback is multi-user diversity really worth? *In Proc. IEEE ICC*, volume 4. 2004.

[15] **Huang, J.**, **V. G. Subramanian**, **R. Agrawal**, and **R. A. Berry** (2009). Downlink scheduling and resource allocation for ofdm systems. *Wireless Communications, IEEE Transactions on*, **8**(1), 288–296.

[16] **Jaramillo, J. J.**, **R. Srikant**, and **L. Ying** (2011). Scheduling for optimal rate allocation in ad hoc networks with heterogeneous delay constraints. *Selected Areas in Communications, IEEE Journal on*, **29**(5), 979–987.

[17] **Kittipiyakul, S.** and **T. Javidi** (2009). Delay-optimal server allocation in multi-queue multiserver systems with time-varying connectivities. *Information Theory, IEEE Transactions on*, **55**(5), 2319 –2333. ISSN 0018-9448.

[18] **Kushner, H. J.** and **P. A. Whiting** (2004). Convergence of proportional-fair sharing algorithms under general conditions. *Wireless Communications, IEEE Transactions on*, **3**(4), 1250–1259.

[19] **Manikandan, C.**, **S. Bhashyam**, and **R. Sundaresan** (2009). Cross-layer scheduling with infrequent channel and queue measurements. *Wireless Communications, IEEE Transactions on*, **8**(12), 5737 –5742. ISSN 1536-1276.

[20] **McKeown, N.**, **A. Mekkittikul**, **V. Anantharam**, and **J. Walrand** (1999). Achieving 100% throughput in an input-queued switch. *IEEE Transactions on Communications*, **47**(8), 1260–1267.

[21] **Neely, M.**, **E. Modiano**, and **C. Rohrs**, Power and server allocation in a multi-beam satellite with time varying channels. *In IEEE INFOCOM*, volume 3. 2002.

[22] **Neely, M.**, **E. Modiano**, and **C. Rohrs**, Dynamic power allocation and routing for time varying wireless networks. *In IEEE INFOCOM*, volume 1. 2003.

[23] **Neely, M. J.**, Delay-based network utility maximization. *In INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010.

[24] **Neely, M. J.**, **E. Modiano**, and **C.-P. Li** (2008). Fairness and optimal stochastic control for heterogeneous networks. *Networking, IEEE/ACM Transactions on*, **16**(2), 396–409.

[25] **Pollard, D.**, *A user's guide to measure theoretic probability*, volume 8. Cambridge University Press, 2001, 57.

[26] **Sanayei, S.** and **A. Nosratinia** (2007). Opportunistic downlink transmission with limited feedback. *Information Theory, IEEE Transactions on*, **53**(11), 4363 –4372. ISSN 0018-9448.

[27] **Stefanatos, S.** and **N. Dimitriou**, Downlink ofdma resource allocation under partial channel state information. *In Communications, 2009. ICC'09. IEEE International Conference on*. IEEE, 2009.

[28] **Stefanatos, S.** and **N. Dimitriou**, Downlink ofdma resource allocation under partial channel state information. *In Communications, 2009. ICC '09. IEEE International Conference on*. 2009. ISSN 1938-1883.

[29] **Tassiulas, L.** and **A. Ephremides** (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *Automatic Control, IEEE Transactions on*, **37**(12), 1936 –1948. ISSN 0018-9286.

[30] **Tassiulas, L.** and **A. Ephremides** (1993). Dynamic server allocation to parallel queues with randomly varying connectivity. *Information Theory, IEEE Transactions on*, **39**(2), 466 –478. ISSN 0018-9448.

[31] **Viswanath, P.**, **D. N. C. Tse**, and **R. Laroia** (2002). Opportunistic beamforming using dumb antennas. *Information Theory, IEEE Transactions on*, **48**(6), 1277–1294.

[32] **Westman, E.** (2006). Calibration and evaluation of the exponential effective sinr mapping (eesm) in 802.16. *The Royal Institute of Technology (KTH) , Master's thesis*. ISSN 1536-1276.

[33] **Ying, L.** and **S. Shakkottai** (2011). On throughput optimality with delayed network-state information. *Information Theory, IEEE Transactions on*, **57**(8), 5116–5132.

[34] **Zhang, Z.**, **Y. He**, and **E. Chong**, Opportunistic downlink scheduling for multiuser ofdm systems. *In Wireless Communications and Networking Conference, 2005 IEEE*, volume 2. 2005. ISSN 1525-3511.