Performance Analysis of Retrial Access Schemes for Optical Communication Networks with Slow and Fast Tuning

by

WALID M. HAMDY

S.B. Massachusetts Institute of Technology (1986)
Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering and Computer Science at the
Massachusetts Institute of Technology
May 1988

© Walid M. Hamdy 1988

The author hereby grants to MIT permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author

Department of Electrical Engineering and Computer Science
May 5, 1988

Certified by

Robert S. Kennedy
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Departmental Committee on Graduate Students
Performance Analysis of Retrial Access Schemes for Optical Communication Networks with Slow and Fast Tuning

by

Walid M. Hamdy

Submitted to the
Department of Electrical Engineering and Computer Science
on May 5, 1988 in partial fulfillment of the requirements for the Degree of Master of Science

Abstract

Usual laser tuning techniques, such as electronic tuning, offer high speed tuning but with a tuning range that is only a small fraction of the theoretical bandwidth of single mode fiber. On the other hand, other tuning techniques, such as thermal or mechanical tuning, offer tuning ranges that are close to the theoretical fiber bandwidth, but that are too slow for modulation.

We propose a frequency division approach that uses both tuning techniques, thereby improving system performance. Specifically, slow tuning tunes the fast tuning bandwidth range over the entire fiber bandwidth thus allowing complete accessibility of all the frequency channels created by the frequency division. Channels have bandwidths that are much larger than that needed by any one user, and hence are further divided into \( N \) subchannels.

We model the network as a blocking network, with blocked users allowed to retry in other channels by using slow tuning. Two forms of retrial are considered: Random and Nearest Neighbor. The performance measures used are the channel blocking probabilities and the subchannel access delay.

Thesis Supervisor: Robert S. Kennedy

Title: Professor of Electrical Engineering,
Department of Electrical Engineering and Computer Science
Contents

1 Introduction .................................................. 8
  1.1 Background ................................................. 8
  1.2 Motivation and Goals of this Thesis ......................... 9
  1.3 Brief Overview of Current Laser Technology ................. 10
    1.3.1 Laser Stability ...................................... 11
    1.3.2 Laser Linewidth ...................................... 11
    1.3.3 Laser Tunability ...................................... 12
  1.4 Summary of Previous Related Works ........................ 13
  1.5 Thesis Overview ........................................... 14

2 System Model .................................................. 15
  2.1 Definitions ............................................... 15
    2.1.1 Characterization of User Traffic ...................... 17
    2.1.2 Coordination of Transmitters and Receivers ............. 18
    2.1.3 Intra-channel Multiple Access ........................ 19
  2.2 Limits of Slow Tuning Performance ........................ 22
  2.3 Discussion ................................................. 24
  2.4 Retrial Schemes ........................................... 25

3 The Random Retrial Scheme .................................... 27
  3.1 System Model .............................................. 27
    3.1.1 Average Tuning Time .................................. 27
    3.1.2 Number of Retrials .................................... 29
    3.1.3 Constraints Imposed by the Random Retrial Scheme .... 30
  3.2 Overflow Traffic and the Equivalent Random Method ....... 31
## CONTENTS

3.2.1 Overflow Streams .......................... 31
3.2.2 The Equivalent Random Method .......... 34
3.2.3 Random Switch Effect ...................... 35
3.3 Random Retrial with Poisson Inputs .......... 38
  3.3.1 Subchannel Access Delay with Random Retrials / Poisson Inputs .......... 41
3.4 Random Retrial with Non-Poisson Inputs .... 43
3.A Proof that $z = 1$ iff the Overflow Process is Poisson .................. 48

4 The Nearest Neighbor Retrial Scheme .......... 49
  4.1 System Model .............................. 49
    4.1.1 Average Tuning Time .................. 51
    4.1.2 Number of Retrials ...................... 52
    4.1.3 Constraints Imposed by Nearest Neighbor Retrial .................. 53
  4.2 Nearest Neighbor Retrial with Poisson Inputs .......... 53
  4.3 Subchannel Access Delay with Nearest Neighbor Retrial ............. 55

5 Conclusions ................................ 59
  5.1 Summary .................................. 59
  5.2 Directions For Further Research .......... 60
List of Figures

2.1 General System Model from the Frequency Viewpoint .................................. 16
2.2 Slow Tuning moving the Fast Tuning Bandwidth ........................................... 16
2.3 Overlapping Channels .................................................................................. 20
2.4 Blocking Probability with and without Slow Tuning ..................................... 23

3.1 Random Retrial Method .............................................................................. 28
3.2 Feedback due to a user retrying in the same channel on the second retrial .... 29
3.3 Alternative Description of Overflow Process ............................................. 34
3.4 Notation of the Equivalent Random Method .............................................. 35
3.5 Random Switch Effect .............................................................................. 37
3.6 Channel Input/Output Processes .................................................................. 38
3.7 Channel Input/Output Processes for Large $L$ and Poisson Input .................. 40
3.8 Blocking Probability for Random Retrial Scheme with Poisson Inputs I ...... 41
3.9 Blocking Probability for Random Retrial Scheme with Poisson Inputs II .... 42
3.10 Blocking Probability for Random Retrial Scheme with Poisson Inputs III ... 42
3.11 Mean Subchannel Access Delay With Random Retrial / Poisson Input I ..... 44
3.12 Mean Subchannel Access Delay With Random Retrial / Poisson Input II .... 44
3.13 Channel Input/Output Processes for Large $L$ and Non-Poisson Input ........ 45
3.14 Approximate System for Random Retrial Non-Poisson Inputs ................. 46

4.1 Nearest Neighbor Retrial with One Way Tuning ......................................... 50
4.2 Direction Probabilities for Nearest Neighbor Tuning ................................. 51
4.3 Composition of channel $i$ retrial stream ................................................. 52
4.4 Normalized Expected Subchannel Delay - System I .................................. 56
4.5 Normalized Subchannel Delay Variance - System I .................................... 56
4.6 Normalized Expected Subchannel Delay - System II .................................. 57
LIST OF FIGURES

4.7 Normalized Subchannel Delay Variance - System II ................. 57
Acknowledgments

First and foremost, I deeply thank Professor Kennedy for his guidance, patience and constant cheerfulness, without which this research would have been a true hardship. His guidance, both as a supervisor and as a teacher, were (and still are) invaluable in my course of study.

Many thanks also go to Professor Humblet for his advice in matters both related to this research and in other academic matters. His unending energy and enthusiasm, which were almost contagious at times, provided a strong driving force for me to do my best.

The LCN group was (and still is) a very enjoyable group to work among. I thank all of its members, past and present, for creating such a friendly environment. From outside the group, two names deserve mentioning: Mohamed Akra and Emre Telatar: Thank you both for all your help and advice.

To my old buddies, Serge, Riaz and Chris: Thanks for the several outings we made. I really needed them.

My parents and my younger sister deserve much more than a passing acknowledgment here. Their constant and unending support, as well their reassuring presence at times of difficulties, is impossible to repay. Nevertheless, I will try.

This research was supported by a fellowship from the Ministry of Higher Education of the Kingdom of Saudi Arabia. I gratefully acknowledge its support, as well as the helpful attitude of the staff of the Saudi Arabian Educational Mission, and in particular Dr. Amal Hashimi.
Chapter 1

Introduction

1.1 Background

Laser tunability is one of the major factors limiting the accessibility of the tremendous bandwidth (on the order of 30,000 Ghz) of single mode optical fiber. While major strides have been made in the development of very widely tunable lasers [Mur 87] [Hei 87], the accessible bandwidth using only electronic tuning is still only a fraction of the theoretical fiber bandwidth.

However, there exist alternative tuning techniques (mechanical and thermal tuning) which have very large tuning bandwidths. Unfortunately, they are not fast enough for modulation. Hence, the question posed is what improvement will be achieved and at what cost, if one of these techniques is used along with electronic tuning.

Several multiaccess techniques have been proposed to allow sharing of the large bandwidth capacity of single mode fibers. Frequency division multiple access is currently the most attractive method. This is mainly due to the promise of frequency selectivity techniques such as optical filtering and coherent optics. These methods allow for the multiplexing of a very large number of frequency channels on the same fiber with little cross-channel interference ([Gi12 87],[Kim 87],[ShE 87],[StHS 87]).

Time division multiple access techniques, on the other hand, require lasers with the capability of producing extremely short pulses ($\sim 10^{-15}$ sec), which is currently difficult to achieve (but admittedly may not be so in the future).

A third multiaccess method which has attracted attention is code division multiple access (CDMA) which, loosely speaking, has the advantage of eliminating network control at the cost of wasting bandwidth. In most practical cases, however, some form of control
must be retained for CDMA to achieve acceptable performance. CDMA will be described briefly in chapter 2, but for a comprehensive discussion of the application of CDMA to single mode optical fiber networks (using direct detection), refer to [Esc 88].

In this thesis, we adopt the frequency division approach, but as will be seen in chapter 2, due to the special characteristics of the system under consideration, TDMA and CDMA can also be applied in this system.

The notions of fast and slow tuning are not widespread and therefore deserve further clarification. Fast tuning refers to tuning methods in which the tuning rate (in Hz/sec) is very large, and hence can be used for modulation. Electronic tuning is considered such a technique.

Slow tuning refers to tuning methods in which the tuning rate is very small, and cannot be used for modulation. Mechanical and thermal tuning are considered of this type. The former relies on mechanically adjusting certain elements in the laser so as to modify the frequency output of the laser. The latter changes the frequency output by adjusting the temperature of the laser.

The other difference between fast and slow tuning is the accessible bandwidth of each. The former has a relatively narrow tuning bandwidth, whereas the latter has a very large tuning bandwidth. For example, in [WyD 83] a tuning bandwidth of 55 nm (~7000 GHz) is achieved using mechanical tuning, whereas typical values for electronic tuning bandwidths are 10 GHz or less.

In this thesis, we investigate the performance improvement that can be achieved by utilizing slow tuning in lasers for optical fiber communication systems. The approach is mainly mathematical, and will concentrate on system issues, rather than devices or technology. It is nonetheless true that technology, and specifically laser technology, plays a vital role in the motivation for this work. We therefore briefly cover the current state of laser technology in Section 1.3, enabling the justification of the assumptions made in the next chapter.

1.2 Motivation and Goals of this Thesis

In most previous works, the assumptions made regarding laser tunability have typically been one of two extremes. At one end is the assumption that all users have full access to the entire accessible bandwidth, with no delay resulting from tuning. The other extreme is where no tuning is assumed (i.e. all transmitters and receivers are “frozen” at some
CHAPTER 1. INTRODUCTION

predefined frequencies). One recent work which assumes neither of these two extremes is [Won 88] where the assumption made is that each user can access only a fixed portion of the entire bandwidth.

It appears then that a new model of tunability that allows for the possibility of using slow tuning to improve system performance is desirable.

Based on the characteristics of both tuning methods given above, a natural division of tasks between both tuning techniques would thus be as follows:

1. Electronic tuning for data modulation.

2. Slow tuning for shifting the electronic tuning range in the entire bandwidth.

Heuristically, the performance of such a system should fall between that of a system using only electronic tuning (i.e. one with no slow tuning) and and that of a system using electronic tuning that has full bandwidth accessibility (i.e. "instantaneous" slow tuning). A more detailed comparison between these two systems is given in chapter 2.

Intuitively speaking, having two tuning methods should always give better performance than one alone (except for the two extreme cases above). The main cost will largely be in system complexity (e.g. the cost of lasers with the capability of using the two tuning methods). However, it is possible that other tradeoffs may exist as well. Thus the main goal of this thesis is to investigate the merits as well as costs of using two tuning methods in single mode optical fiber communication systems.

Another goal is to analyze different methods of using the added capability of slow tunability. Specifically, how and when should slow tunability be used? Implicit in these questions is the question of whether slow tuning should be used at all.

1.3 Brief Overview of Current Laser Technology

The area of laser technology is quite broad, and a further narrowing of the subject is necessary here. Specifically, we concentrate only on the aspects that are of concern to applying lasers in single mode fiber communication systems.

The main areas of laser technology that are of concern are laser frequency stability, linewidth and tunability.

---

1System performance will be defined in chapter 2.
CHAPTER 1. INTRODUCTION

1.3.1 Laser Stability

The inherent frequency stability of lasers is still rather poor (around Hundreds of Ghz), and thus some method of frequency stabilization is almost always needed. This is particularly so in frequency division systems where channel frequency separation is usually required to be very small.

Several methods of achieving laser frequency stability have been reported [Soll 87][Tai 85][Gl2 87]. The basic idea used in all of these methods is to use feedback to lock the laser to some extremely accurate reference. Feedback is achieved by converting the frequency deviation into a photocurrent intensity variation, comparing the photocurrent with some reference, and using the difference signal as a feedback signal.

Three common reference types used are:

1. Fabry-Perot Interferometer [Gl1 87]
2. Fiber Optic Ring Resonator [Tai 85].
3. Atomic-Molecular Absorption Line [Tsu 82].

As an example, the first reference type results in a frequency stability that allows frequency spacings between channels to be as low as 500 Mhz [Gl1 87].

1.3.2 Laser Linewidth

The linewidth $\Delta \nu$ of a laser specifies roughly how broad the emitted spectrum is. It is given by the following [Hen 86]

$$\Delta \nu = \frac{R}{4\pi I}(1 + \alpha^2)$$

where

$R =$ spontaneous emission rate.
$I =$ intensity.
$\alpha =$ linewidth parameter = ratio of changes in real and imaginary parts of the refractive index with respect to the change in the carrier number.
Naturally, one would prefer a linewidth as narrow as possible so as to allow a small frequency channel separation, and thus a large number of channels. Linewidth narrowing can be achieved in several ways:

1. Increase output power $P$, since $\Delta \nu \propto 1/P$.

2. Increase data rate, since the minimum acceptable linewidth increases with data rate.

3. Improve laser design and construction.

The first method is not always feasible, since lasers usually have an output power constraint that always must be met for reliable operation. In addition, increasing output power can result in pronounced nonlinear effects, such as Raman scattering and four wave mixing [Abe 87].

Increasing the data rate is not always feasible either, since usually the electronics used in the system cannot operate at extremely high rates (or what is otherwise known as the electronic bottleneck).

New laser design, on the other hand, can result in enormous linewidth reduction. For example, by adding a passive Section, AR coating one end and adding an external reflector can result in $\Delta \nu P = 10$ KHz·mW. However, such designs are bulky and not very practical.

Large linewidth reduction can also be achieved by external feedback. With weak feedback (e.g. DFB laser), linewidths of approximately 10 Mhz can be achieved. Moderate feedback (SLM laser in a cavity) can result in a 1 Mhz linewidth. With strong feedback, linewidths of 10 KHz or less can be achieved. Strong feedback suffers from “coherence collapse”, in which a laser can become unstable and lose its coherence [Hen 86].

1.3.3 Laser Tunability

Already briefly mentioned above, laser tuning can be accomplished in several ways: By varying the injection current (usually performed by electronics), by varying the temperature (thermal or temperature tuning), and by mechanical means as in the case of large optical cavity lasers with diffraction gratings that are mechanically rotatable and translatable.

Intense research efforts in this area have produced excellent results. Using external cavity lasers and mechanical tuning, ranges of up to 18,000 Ghz were obtained [WaCK 87]. However, electronic tuning is preferred over mechanical tuning, and grated external cavity lasers are quite bulky. One group [Hei 87] reported obtaining a tuning range of
approximately 1000 GHz using electrooptic tuning of an InGaAsP-Ti:LiNbO extended cavity laser. A second group [Mur 87] obtained a frequency range of 720 GHz using electronic tuning of a 1.55 μm DBR laser. These values, however, do not exactly correspond to our definition of fast bandwidth, since the modulation bandwidth (which is what concerns us) attainable by these lasers is only a fraction of the above values.

Assumptions on Laser Technology:

In this thesis we assume that all lasers (both transmitters and local oscillators) have the capability of slow and fast tuning. Whether both will be used in transmitters and receivers will be discussed in chapter 2.

We also assume that all lasers are absolutely stable, and have zero linewidth, thus resulting in no cross channel interference. These assumptions allow us to concentrate on the aspect of tuning, and should not be understood as belittling the importance of non-zero linewidth and non-perfect stability effects.

1.4 Summary of Previous Related Works

Most of the papers related to this work are found in the fields of traffic flow, service systems and congestion theory as applied to optical communication networks. Several recent works in this area are [Abb 88], [Lie 88] and [Won 88]. In [Abb 88], the emphasis is mainly on investigating the performance of an optical communication network with heterogeneous users, and analyzing the performance of several sharing schemes. In this system all users have full access to the entire bandwidth.

[Won 88] analyzes the problem of channel scheduling in optical fiber networks with frequency concurrency (frequency division). Each user has access to a limited and fixed subset of the channels, thereby allowing only limited tunability. The emphasis is on systems with a small number of channels. This allows for a Markovian analysis of the system.

In [Lie 88], the capacity assignment problem is analyzed in the context of an optical communication network, but with the assumption that all transmitters and receivers are not tunable, and hence each user would generally need more than one transmitter and one receiver.
1.5 Thesis Overview

In the following chapter we discuss in detail the system model, and the main assumptions of the thesis. A brief description of retrial methods is then given. The comparison of perfect slow tuning against no slow tuning is examined in Section 2.2, and is followed by a discussion of the consequences of some of the assumptions, as well as a discussion on retrial methods.

Chapter 3 discusses the Random Retrial scheme. After describing the special characteristics of this method, the constraints imposed on the system by this retrial scheme is then given. This is followed by a performance analysis of this system with Poisson inputs, and non-Poisson inputs.

The Nearest Neighbor Retrial scheme is then investigated in Chapter 4. As in Chapter 3, the special characteristics and constraints of a system with this retrial scheme are first given. The performance of this system with Poisson inputs is then analyzed, and a lower bound on the blocking probability, together with a lower bound on the expected channel access delay and its variance are given.

Finally Chapter 5 gives a summary of the work and some directions for further research.
Chapter 2

System Model

In this chapter we discuss in detail the system model as well as the underlying assumptions used throughout this thesis. The description of the user traffic and the problem of transmitter-receiver coordination are also discussed. Section 2.2 investigates the two extreme cases of this model, namely one with no slow tuning, and one with "instantaneous" slow tuning. In Section 2.3, we discuss several constraints imposed on the system by using retrial schemes.

2.1 Definitions

Let the total fiber system bandwidth be $W_T$ Hz. We assume that slow tuning has a bandwidth equal to $W_T$ Hz (i.e. slow tuning spans the entire fiber bandwidth). Let $W_C$ Hz be the fast tuning bandwidth, such that

$$W_R \leq W_C \leq W_{C_{\text{max}}},$$

(2.1)

Where $W_{C_{\text{max}}}$ is the maximum fast tuning range, and $W_R$ is the modulation bandwidth required by each user. The upper limit simply states that $W_C$ does not necessarily have to be the maximum fast tuning bandwidth. In fact, as we shall see in Chapter 4, there exist situations in which $W_C$ should equal its smallest possible value, $W_R$. The lower limit states that the fast tuning bandwidth must be at least as large as $W_R$, the data modulation bandwidth required by each user.

Typically, we assume that $W_R \ll W_{C_{\text{max}}} \ll W_T$. Typical values may be $W_T = 50$ THz, $W_{C_{\text{max}}} = 10$ GHz, $W_R = 100$ Mhz.
Figure 2.1: General System Model from the Frequency Viewpoint

Figure 2.2: Slow Tuning moving the Fast Tuning Bandwidth
Since $W_C$ can support at least one user simultaneously, we may envision $W_T$ being frequency divided into $L$ disjoint\footnote{we will see later that disjointness is not necessary; it is however, analytically convenient} channels, each of bandwidth $W_C$, where we assume $L$ to be an integer. See Fig. 2.1.

Each channel can then be considered as a separate communication channel with a capacity of $N$ users where $N$ is limited by

$$1 \leq N \leq \frac{W_{C_{\text{max}}}}{W_R}$$

(2.2)

Slow tuning delay refers to inter-channel tuning delay (as opposed to intra-channel tuning delay which we assume to be zero). We assume that slow tuning has a constant tuning rate over $W_T$. That is, if we define $T \triangleq$ Tuning period from channel 1 to channel $L$, (see Fig. 2.2), then the tuning period between neighboring channels is just $T/(L - 1)$. In general, the tuning period from channel $i$ to channel $j$ is

$$T_{ij} = \frac{|i - j|T}{L} \quad 1 \leq i, j \leq L$$

(2.3)

In general, a constant slow tuning rate need not be the case since tuning over only one channel may take longer than $T/L - 1$ due to an overhead delay that must be added for each channel “stop”. We will return to this issue in Chapter 4.

In summary then, the system constants which are assumed given are:

- $W_T$: The total system bandwidth = The slow tuning bandwidth.
- $W_{C_{\text{max}}}$: The maximum fast tuning bandwidth.
- $W_R$: The user data modulation bandwidth.
- $T$: The end-to-end system bandwidth slow tuning delay.

### 2.1.1 Characterization of User Traffic

For analytical purposes, we assume an infinite subscriber population. Subscribers currently using or attempting to use the system are named users. Subscribers are evenly divided among the $L$ channels. That is, each channel has an infinite subscriber population, and the processes of users arriving at their assigned channels is a Poisson process with mean
interarrival time $1/\lambda$. In addition, we assume that the probability distribution function of the call holding times is negative exponentially distributed with mean $1/\mu$. That is,

$$G(x) = \begin{cases} 
1 - e^{-\mu x} & x \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

Subscribers who attempt to use their own channel but find $N$ users in the channel at that time will get blocked. Usually, loss systems assume that blocked users are lost (cleared) ($M/M/N$-loss or $M/M/N/N$ system). Queueing (or delay) systems on the other hand allow users to wait until served. A third less common model is the retrial model, where blocked users form a new population (the blocked population) of which each member independently retries. Interretrial times are usually taken to be negative exponentially distributed with mean $= 1/\nu$. Furthermore, some analyses of retrial systems do not allow an infinite retrial period. Instead, it is assumed that the retrial periods (the length of time during which retrials are allowed) are negative exponentially distributed with mean $= 1/\gamma$ [Sys 86].

Strictly speaking, our system uses none of these models. The novel idea here is to modify the retrial model above so that blocked users immediately employ slow tuning to tune to another channel where they can retry.

### 2.1.2 Coordination of Transmitters and Receivers

The capability of slow tuning need not be available to both transmitters and receivers. In fact, we may have the following four combinations:

(a) No slow tuning for either transmitters or receivers: In this case, we do not have full transmitter-receiver accessibility, since only transmitters and receivers on the same channel may communicate. The $L$ channels in this case are completely independent.

(b) Only transmitters have slow tuning capability: Complete transmitter-receiver accessibility exists. The channels are no longer independent since transmitters may need to tune to channels other than their own to communicate with the intended receivers.

(c) Only receivers have slow tuning capability: This case is identical to the previous one except that the problem of transmitter collision (two or more transmitters trying to communicate with the same receiver simultaneously) is eliminated.

---

2Thus we assume only homogeneous or one type services
CHAPTER 2. SYSTEM MODEL

(d) Both transmitters and receivers have slow tuning capability: Clearly, full transmitter-receiver accessibility exists. This case has the added flexibility of allowing any transmitter-receiver pair to communicate in any of the L channels.

In this thesis, we assume we do not have a central controller which has full knowledge of the network state. This major assumption is made for two main reasons:

- With a network theoretically capable of serving thousands, possibly tens of thousands of users, a great deal of burden is placed on such a controller. Furthermore, with the expected increase in the robustness of lasers (see Chapter 1), it does not seem too absurd to envision an optical network with distributed control.

- If a central controller is used, the issue of slow-fast-tuning is transformed into the problem of finding an algorithm by which a central controller can assign subchannels to transmitters and receivers such that some performance measure (e.g. circuit access delay) is optimized. While this may be an interesting problem by itself, it is tangential to this work.

The problem of coordination is concerned with devising a method or rule by which transmitters and receivers can be synchronized in frequency and time (in the case of CDMA, on the chip level).

Coordination may be achieved if the task of finding a free subchannel is left entirely to the transmitter. In that case, once the transmitter has access to a free subchannel, it may inform the receiver (through a set-up channel, for example) of the location of this subchannel (i.e. the channel number, and the subchannel number).

2.1.3 Intra-channel Multiple Access

In this subsection we discuss three common methods of dividing each channel bandwidth among the N users.

1. Frequency Division Multiple Access (FDMA)

In FDMA, \( W_C \) is simply divided up into \( N \) equal sized subchannels, each of bandwidth \( W_C / N = W_R \). Note that by using FDMA as an intra-channel multiaccess method, we have a two hierarchy FDMA system, since the entire bandwidth \( W_T \) is partitioned into
Chapter 2. System Model

Figure 2.3: Overlapping Channels

$L$ channels of bandwidth $W_C$, each of which is further subdivided into $W_R$ bandwidth subchannels. Thus the maximum number of users at any time $= \frac{W_T}{W_C} \frac{W_C}{W_R} = \frac{W_T}{W_R} = LN$.

It should be noted as well that channel disjointness is not strictly necessary here. In fact, neighboring channels may have a maximum overlap of $(N-1)W_R$ (i.e. neighboring channels share $N-1$ subchannels) See fig. 2.3. However, allowing channels to overlap does not appear to have any benefits, since only one new subchannel appears when tuning between neighboring channels.

FDMA is usually either fixed or dynamic. In the former, each user (transmitter or receiver) is assigned a single channel (channels are not subdivided), and a transmitter that wishes to communicate with a specific receiver simply tunes to that receiver's respective channel. In dynamic FDMA, channel allocation is performed on a dynamic basis, usually by a central controller, and thus results in a more efficient use of bandwidth at the cost of increased complexity.

Strictly speaking, our system is a dynamic FDMA system, since all transmitters and receivers have tuning capability. However, it differs from usual dynamic FDMA in two respects. First, no central controller is assumed, and hence allocating subchannels must be done in a distributed fashion (e.g. by contention). Second, allocating subscribers to channels provides each user with a "home" channel which it must try first, thus making the first trial deterministic.

2. Time Division Multiple Access (TDMA)
CHAPTER 2. SYSTEM MODEL

The use of TDMA is feasible here since the bandwidth of each channel \(W_C\) is much less than \(W_T\), and thus the pulse widths needed would be much larger than those needed with no division of \(W_T\) (a minimum of 10 ps vs 10 fs). Nonetheless, the constraint on \(W_C\) should be evident. As in FDMA, \(N = \frac{w_c}{w_r} = \frac{T_A}{T_C}\) and \(L = \frac{w_T}{w_C}\).

3. Code Division Multiple Access (CDMA)

In CDMA, each communicating transmitter-receiver pair is given a separate codeword which the transmitter uses to "spread" the signal. Different codewords are chosen to have minimum cross-correlation, thus permitting each receiver to "despread" the transmitted signal with minimum interference from other transmitter-receiver pairs. CDMA has the additional property of trading bandwidth efficiency with network control. This appears to be a very attractive feature that fits well with optical fiber network requirements. However, in reality, some form of network control is still needed (e.g. for codeword distribution).

Determining the maximum number of users that can be supported in one channel is not as straightforward in this case as in FDMA or TDMA due to the nonorthogonality of users. However, two methods that may be used are:

(a) Specifying a maximum BER suffered by any user: In this method the most recent user entering a channel transmits a test sequence and measures the resulting BER. If it exceeds a specified threshold, this user "disconnects" from this channel and retries in another channel.

(b) Specifying a maximum PSD in the channel bandwidth: In this method the entering user measures the power spectral density in the channel bandwidth. The threshold PSD may either be defined by a maximum PSD peak or a maximum PSD average. If the threshold is exceeded, the user retries in another channel.

For a comprehensive discussion of CDMA see [SiOSL 85].

Regardless of which multiaccess method is used, we assume that a maximum number of users \(N\) are allowed in each channel simultaneously. In fact, it is not even necessary to assume that all channels use the same multiaccess method. However, for illustrative purposes, we will assume henceforth that disjoint FDMA is used as the channel multiaccess method in all the channels, since it is easiest to visualize, as well as being relatively straightforward to implement.
CHAPTER 2. SYSTEM MODEL

2.2 Limits of Slow Tuning Performance

In this Section we look at the performance of this system in the limit of zero tuning time \((T = 0)\), and infinite tuning time \((T = \infty)\). These two extremes should at least give us a rough idea of what to expect with finite but nonzero \(T\).

- \(T = \infty\):

  No thermal tuning is used, and each user sees only its own channel. If no retrials are allowed (or equivalently, if the interretrial time distribution is identical to the interarrival distribution of the original users), the system performance is identical to that of the classical \(M/M/N/N\) system. That is,

  \[
  Pr(\text{blocking}) = B(N, \rho) = \frac{\rho^N / N!}{\sum_{i=0}^{\infty} \rho^i / i!}
  \]

  where \(\rho \triangleq \lambda / \mu\)

- \(T = 0\):

  In this case, an arriving user sees the whole accessible bandwidth \(W_T\) available. Hence, again if one assumes that no retrials are allowed, the system performance in this case is just that of the \(M/M/LN/LN\) system with load \(L\rho\), or simply

  \[
  Pr(\text{blocking}) = B(LN, L\rho) = \frac{(L\rho)^{LN} /LN!}{\sum_{i=0}^{LN} (L\rho)^i / i!}
  \]

  For a comparison of these two systems, see Fig. 2.4, where the blocking probability \(P_b\) is plotted as a function of the normalized load \(\bar{\rho} = \lambda / N\mu\). Notice that for these two systems, the normalized load is the same, since for the \((T = 0)\) system it is just \(\rho = L\lambda / LN\mu = \lambda / N\mu\). In this figure, \(L = 1, 10\) and \(100\), where \(L = 1\) corresponds to the independent channels \((T = \infty)\) case. One can immediately see that the blocking probability decreases significantly when the channels are combined (i.e. \(T = 0\)) and is dependent on the number of channels combined. However, one should also note that for any given value of \(N\), there
Figure 2.4: Blocking Probability with and without Slow Tuning
exists a load region where the use of slow tuning is most useful, and outside of which the slow tuning effect is either insignificant or unimportant. Specifically, for very high loads the difference in blocking probability becomes very small and is thus insignificant, while at extremely light loads the blocking probability of the \((T = \infty)\) case is so low that decreasing it further is of no importance.

It thus appears that the load region of interest is:

\[
\{ \rho : B(N, \rho) = 10^{-3} \} < \frac{\lambda}{N\mu} < 1
\]  
(2.6)

The upper bound is a usual and reasonable assumption in optical fiber networks. It states that the load on the system is always less than capacity. The value of the lower bound is somewhat arbitrary, and simply states that this system works well with loads that result in \(B(N, \rho)\) being greater than \(10^{-3}\).

### 2.3 Discussion

From the above two extreme cases, it can then be reasonably conjectured that finite (non-zero) tuning time between channels improves performance over infinite tuning time (no cooperation between channels). This is in fact what we aim to prove in this thesis. However, it is important to note that with finite but non-zero \(T\), we obtain a system that is quite different from either of the above two methods, as we shall show next. Hence, one should not expect the performance of the new system to approach that of the two extremes as \(T \rightarrow \infty\) or zero.

When we allow \(T\) to be finite but nonzero, we are essentially allowing retrials. That is, a user that is blocked from a channel may retry in another channel. This necessitates building more sophistication into the system since receivers must know where to locate the corresponding transmitters as we have shown in Section 2.1.2.

The key assumption regarding this system is that the average time it takes to tune from one channel to another is much smaller than the arrival rate into each channel. That is,

\[
T_{ave} \ll 1/\lambda
\]  
(2.7)

where \(T_{ave}\) is defined to be the average tuning period between channels. Its value depends on the retrial scheme used, as we shall see in chapters 3 and 4.
CHAPTER 2. SYSTEM MODEL

Equation (2.7) must always hold, since if it were otherwise, it would be more sensible to retry in the same channel as a new customer with rate \( \lambda \). In that case, no thermal tuning would be used, and the \( L \) channels would be independent of each other.

Equations (2.6) and (2.7) are the key equations in this analysis. If we combine them we get

\[
T_{ave} \ll 1/N\mu < 1/\lambda
\]  

(2.8)

This puts a more severe restriction on the the thermal tuning period. It says that \( T_{ave} \) must be small enough so that no change in the system state (i.e. arrivals or departures) occurs during the tuning period. We assume however that call holding times will tend to be quite long (on the order of minutes or tens of minutes) as is typically the case in circuit switched networks. Hence \( T \) would not have to be extremely small, which technically is unrealistic. We will see in Chapter 3 that (2.8) puts a severe constraint on the maximum number of users when using the random retrial method.

2.4 Retrial Schemes

In this Section, we discuss the alternative retrial schemes that may be used in this system. We comment also on the solution technique used in the analysis of this system.

We consider two types of retrial schemes in this thesis: Nearest Neighbor and Random. In the former, retrials are limited to adjacent channels, whereas in the latter, blocked users from any channel randomly tune to any of the other \( L - 1 \) channels.

By definition, there is only one random retrial scheme\(^a\). On the other hand, several different nearest neighbor retrial schemes may be envisioned. For example, a blocked user might perform either a random walk or a one direction walk until reaching a nonblocked channel, or reaching either end of the bandwidth.

Any retrial schemes should be fair in the sense of not discriminating between different users on the basis of their home channel. As an example of an unfair retrial scheme, consider the nearest neighbor retrial scheme with random walk:

A blocked user chooses one of the adjacent channels with equal probability. Channels 1 and \( L \) are different from the other \( L - 2 \) channels: A blocked user from either of these

---

\(^a\)One may generalize random retrial to a retrial scheme with non equal channel transition probabilities, in which case the nearest neighbor retrial scheme can be include as a special case with nonzero probabilities for nearest neighbor transitions, and zero probabilities for all other channels.
two channels that decides to tune out of the \( W_T \) is lost. This system can be modeled as a random walk with \( L \) different starting points and with two absorbing barriers. Let \( R_i \) be defined as the expected number of retrials that a subscriber assigned to channel \( i \) is allowed before being lost (absorbed). It can be shown [Cox 65] that

\[
R_i = i(L + 1 - i)
\] (2.9)

In particular, \( R_1 = L \) and \( R_{L/2} = \frac{L}{2} \left( \frac{L}{2} + 1 \right) \) which clearly shows the unfairness of this system when \( L \) is large for the users assigned to "border" channels.

As a result of the above, we will restrict our attention to nearest neighbor retrial schemes with one way (i.e. deterministic or with memory) tuning. This is the subject of Chapter 4.

It should be clear that using retrial schemes causes some complex interaction between different traffic streams and the retrial users. An exact analysis of this interaction can be made by describing the system as a multidimensional Markov process. We opt however in this thesis to use approximate methods to describe this system.

We should justify here the reason for using these approximate methods rather than an exact method such as a Markovian description of the system: We are interested mainly in the cases where the number of channels \( L \) is very large, and where the number of users in each channel \( N \) may also be large. A Markovian description of this network would require a state variable of size \( 2L \), with an infinite state-space, since one would need to keep track of the number of customers in each channel (\( L \) variables each of which can vary from 0 to \( N \)) plus the number of users that are retrying in each channel (\( L \) variables each of which can vary from 0 to \( \infty \)). The Chapman-Kolmogorov equations can be easily (albeit tediously) set up, but solving them appears to be a very imposing problem.

Markovian analysis gives us much more information than we need. So we opt for an approximate, but insightful approach, which uses some common tools in traffic engineering. Furthermore, for the case of Poisson inputs with the random retrial scheme (Section 3.2), the results are asymptotically exact with increasing \( L \). In the case of the random retrial scheme with non-Poisson inputs (Section 3.3) a lower bound and an approximation are derived, whereas in the case of the nearest neighbor scheme with Poisson input (chapter 4), a lower bound is found and used to compute a lower bound on the subchannel access delay period.
Chapter 3

The Random Retrial Scheme

In this chapter we analyze in detail the random retrial scheme, described briefly before in Section 2.4. The following Section presents the particular features of this scheme, specifically average tuning time, the number of retrials allowed, and the constraints imposed by this scheme on both the tuning period $T$ and the channel bandwidth $W_C$. In Section 3.3 we analyze this scheme when the offered traffic is Poisson, and present asymptotically exact blocking probabilities with $L \to \infty$. In Section 3.4, a lower bound and an approximation for the blocking probability are presented in the case of non-Poisson traffic.

Throughout this chapter, traffic streams will be described in terms of the parameter pair $(\rho, z)$, where $\rho$ and $z$ denote the load and peakedness of the traffic stream respectively. The significance of this approximate description and these parameters is described in Section 3.2.

3.1 System Model

3.1.1 Average Tuning Time

In this model, a user that is blocked from using a channel, randomly (i.e. with equal probability) chooses one of the other $L-1$ channels to tune to for the following attempt. See Fig. 3.1. This results in tuning times between attempts that are dependent on the distance (frequency separation) between the present and future channel. For example, if the present channel is channel 1 (the leftmost channel), then the tuning time can range from $T/(L-1)$ (to channel 2) to $T$ (to channel $L$, the rightmost channel), where $T$ is defined as being the tuning time from one end of $W_T$ to the other. In general, we have
CHAPTER 3. THE RANDOM RETRIAL SCHEME

Figure 3.1: Random Retrial Method

\[ T_{ij} = \text{Tuning time from channel } i \text{ to channel } j \]
\[ = |i-j| \frac{T}{L-1} \quad 1 \leq i, j \leq L \quad i \neq j \]  

Straightforward calculation shows that

\[ T_{ave} = E(T_{ij}) \]
\[ = \frac{1}{L(L-1)} \sum_{i=1}^{L} \sum_{j=1}^{L} |i-j| \frac{T}{L-1} \]  
\[ = \frac{(L+1)T}{(L-1)^3} \]  
\[ \approx \frac{T}{3} \quad L \gg 1 \]  

and

\[ Var(T_{ij}) = \frac{T^2}{L(L-1)^3} \sum_{i=1}^{L} \sum_{j=1}^{L} (i-j)^2 - (E(T_{ij}))^2 \]
\[ = \frac{(L+1)(L-2)T^2}{18(L-1)^2} \]
CHAPTER 3. THE RANDOM RETRIAL SCHEME

\[ \Rightarrow \text{Var}(T_{ij}) \approx \frac{T^2}{18} \quad L \gg 1 \quad (3.9) \]

Thus for a system with a large number of channels, the average tuning period is independent of the number of channels, and grows linearly with \( T \). Similarly, the variance of the tuning period for \( L \) large is independent of \( L \), but has a quadratic dependence on \( T \).

3.1.2 Number of Retrials

The number of retrials allowed in this scheme is infinite. This leads to dependencies between exogenous traffic and the retrial traffic for each channel, but in the case of very large \( L \), these dependencies will tend to be negligible. The reasoning behind this statement is as follows: By allowing an infinite number of retrials, a fraction of blocked users may retry in their original (home) channel, thus creating a "feedback" effect. This effect is very weak, however, since a user will retry in the same channel only if blocked twice or more, and if he chooses his home channel in the second or further retrial (see fig. 3.2). The probability of this occurring in the second retrial is \( \frac{1}{L-1} P_{bi} P_{bj} \), where \( P_{bi}, P_{bj} \) are the blocking probabilities of the first trial channel \( i \), and the second trial channel \( j \) respectively. Thus as \( L \to \infty \), the probability of this event occurring \( \to 0 \). Notice that we need not assume that the channel blocking probabilities are equal for this argument to hold. By extension of this argument, the probability of the user coming back in further retrials for large \( L \) is even smaller.
CHAPTER 3. THE RANDOM RETRIAL SCHEME

From the above reasoning, and assuming that the number of available channels is sufficiently large, we make the approximation that the exogenous stream and the retrial stream of each channel are uncorrelated. Blocking probabilities of different channels on the other hand, are clearly correlated, since a high blocking probability in one channel results from, and produces high blocking probabilities in other channels.

3.1.3 Constraints Imposed by the Random Retrial Scheme

As shown in Section 2.3, (2.8) must hold in order for the random retrial method to be practical. Using (3.3) in (2.8) we get

\[
\frac{1}{3} T < \frac{(L+1)T}{3(L-1)} < 1/N\mu
\]

or,

\[
T < \frac{3}{N\mu} = \frac{3W_R}{\mu W_C}
\]

Equation (3.11) is a constraint on the total tuning period \(T\) that must be met if the random retrial scheme is to be useful. One can immediately conclude that the random retrial scheme is effective only in networks with relatively large call holding times, as was observed earlier in Chapter 2.

As a numerical example consider the following possible values: \(W_R = 100\) MHz, \(W_C = 10\) Ghz, \(1/\mu = 10\) min = 600 sec. Using these values in (3.11), we get that \(T < 18\) sec.

Note that the constraint in the above form assumes that \(W_C\) is fixed (such that (2.1) is satisfied), and \(T\) is to be determined. One may also use (3.10) to bound \(W_C\) in terms of \(W_R, T\) and \(\mu\):

\[
W_C < \frac{3W_R}{T\mu}
\]

In this case, (2.1) must be modified to

\[
W_R \leq W_C \leq \min\{W_{C_{\text{max}}}, \frac{3W_R}{T\mu}\}
\]

\(W_R\) is usually two or three orders of magnitude smaller than \(W_{C_{\text{max}}}\). For example, \(W_R\) is usually on the order of 100 MHz, whereas \(W_{C_{\text{max}}}\) is around 10 GHz. Hence in view of (3.11) and the fact that making \(T\) extremely small is not practical, \(\frac{3W_R}{T\mu}\) usually is the
CHAPTER 3. THE RANDOM RETRIAL SCHEME

upper bound of $W_C$. Again, for an example consider the following case: $T = 10$ sec, $W_R = 100$ MHz, $1/\mu = 5$ min $= 300$ sec, $W_{C_{\text{max}}} = 10$ GHz. With these numbers, (3.12) gives 9 GHz as the upperbound of $W_C$.

In the following two Sections we will assume that the above constraints on $W_C$ and $T$ hold, thus making random retrial useful in the sense of satisfying (2.8).

Before investigating the blocking probabilities of this retrial scheme, it is necessary to present the reader with an overview of the approximation method used to obtain these blocking probabilities.

3.2 Overflow Traffic and the Equivalent Random Method

In analyzing the performance of communication networks with blocking, one frequently needs to characterize various traffic flow processes. In general, these processes are usually not even renewal processes [DisK 85], and hence one frequently needs to resort to approximate methods due to the nontractability of an exact description. In this Section, we present an overview of some important concepts as well as the approximate methods that are used to characterize a special class of traffic flow streams, namely overflow streams $^1$.

3.2.1 Overflow Streams

In a loss system, users that arrive that find all servers busy are lost (or cleared). Thus loss systems are alternatively called “blocked calls cleared” systems, as opposed to queuing systems which are “blocked calls held” systems.

The output process of a loss system thus consists of two streams: a departure stream, and an overflow stream. The former consists of those users who were admitted to the system exiting the system after being served (no reneging is allowed). The latter consists of those customers who were blocked from entering the system.

When the input process is renewal, the overflow process is also a renewal process [Rio 62], but clearly not Poisson. The $N$-server interarrival distribution function $F_N(t)$ is given by the following recurrence equation [Rio 62]:

$$F_N(t) = \int_0^t [e^{-\mu t} + (1 - e^{-\mu t})F_N(t - r)]dF_{N-1}(r)$$

(3.14)

and

$^1$We use the terms process and traffic stream interchangeably in this thesis.
\[ F_0(t) = F(t) \]  \hspace{1cm} (3.15)

where \( F(t) \) is the interarrival distribution function, and \( 1/\mu \) is the mean service time.

Clearly, this description of the overflow process is not easily amenable to analysis, and an alternate description is needed.

One method commonly used to describe an overflow process is to present the overflow stream to an infinite server system with negative exponentially distributed service times (see Fig. 3.3). The probability distribution function of the number of busy servers then serves as an alternative description of the overflow stream \(^2\).

The probability generating function of the distribution \( P(j,k) \), where \( j \) denotes the number of busy servers in the \( N \)-server group and \( k \) denotes the number of busy servers in the infinite group, was derived by Kosten [Coo 81]. The derivation is straightforward, but quite long and tedious. Hence only the key steps will be shown here. For a complete derivation, the interested reader may refer to [Coo 81].

To begin, we first write down the Chapman-Kolmogorov equations for the partial system description \((j,k)\), where \( j \) represents the number of busy servers in the primary system \((0 \leq j \leq N)\), and \( k \) represents the number of busy servers in the infinite overflow system.

We have two cases:

1. \( j \leq N -1 \)

\[
(\rho + j + k) P(j,k) = \rho P(j-1,k) + (j+1) P(j+1,k) + (k+1) P(j,k+1) \]  \hspace{1cm} (3.16)

\[
P(-1,k) = 0 \quad ; \quad 0 \leq j \leq N - 1 \quad ; \quad 0 \leq k.
\]

2. \( j = N \)

\[
(\rho + N + k) P(N,k) = \rho P(N-1,k) + \rho P(N,k-1) + (k+1) P(N,k+1) \]  \hspace{1cm} (3.17)

with \( P(-1,k) = 0 \quad ; \quad 0 \leq k. \)

\(^2\)Another common method of describing an overflow process is by matching its first three moments with the first three moments of an interrupted Poisson process [See [Kuc 73], [Kaw 86], [McN 84]].
As usual, one also needs

\[ \sum_{j=0}^{N} \sum_{k=0}^{\infty} P(j, k) = 1 \]  

(3.18)

The solution of the system of equations (3.16)-(3.18) is found by using probability generating functions. The distribution is given by

\[ P(j, k) = (-1)^k \phi_0(N) \sum_{\nu=k}^{\infty} (-1)^\nu \binom{\nu}{k} \frac{\rho^\nu}{\nu!} \frac{\phi_\nu(j)}{\phi_{\nu+1}(N)\phi_\nu(N)} \]  

(3.19)

where \( \phi_\nu(j) \) is given by

\[ \phi_\nu(j) = \begin{cases} \frac{\rho^j}{j!} e^{-\rho} & \nu = 0 \\ e^{-\rho} \sum_{i=0}^{j} \binom{\nu+i-1}{i} \frac{\rho^j}{(j-i)!} & \nu > 0 \\ \end{cases} \]  

(3.20)

0 ≤ j ≤ N and 0 ≤ k.

To find the mean \( \alpha \) and variance \( \nu \) of the number of busy servers in the infinite server system, we use the following two formulas, which follow after several steps from the previous two equations

\[ \alpha = \rho \frac{\phi_0(N)}{\phi_1(N)} \]  

(3.21)

\[ z = \alpha - \alpha^2 + 2\alpha c_2 \phi_3(N) \]  

(3.22)

where \( c_i \) is defined by

\[ c_i = (-1)^i \frac{(\rho)^i}{i!} \frac{\phi_0(N)}{\phi_{i+1}(N)\phi_i(N)} \]  

(3.23)

Straightforward substitution yields

\[ \alpha = \rho B(N, \rho) \]  

(3.24)

\[ \nu = \alpha \left(1 - \alpha + \frac{\rho}{N + 1 + \alpha - \rho}\right) \]  

(3.25)

It turns out that describing an overflow process by just the mean and variance of the number of busy servers yields a surprisingly accurate description. This is surprising since
in general one cannot represent an arbitrary renewal process with another renewal process having the same first two moments. The reasons behind this behavior have been thoroughly investigated in [Hol 73].

It is common to use the term *peakedness* to denote how varied the process is with respect to its mean. That is, the peakedness, $z$, is defined as $v/\alpha$. The pair $(\alpha, z)$ thus describe the overflow process.

It can be shown that the peakedness of an overflow process is 1 if and only if the process is Poisson. See Appendix 3.A for a proof. In addition, Kuczura [Kuc2 73] showed that in a system with mixed renewal and Poisson inputs, the Poisson inputs observe a smaller blocking probability than the renewal inputs. Thus one may lower bound the blocking probability of the renewal users by the blocking probability of the Poisson users. These two facts will be used in the following Sections.

To summarize: an approximate, yet accurate, description of an overflow stream may be obtained by specifying the mean and variance of the number of busy servers in a fictitious infinite server system whose input is the overflow stream.

### 3.2.2 The Equivalent Random Method

The equivalent random method is a technique used for approximating the blocking probabilities for non-Poisson traffic streams. In particular, when the input stream is a superposition of overflow streams, the E.R.M can be used with good accuracy to find the blocking probability.

This method works as follows: A non-Poisson stream described by its mean $\rho$ and
CHAPTER 3. THE RANDOM RETRIAL SCHEME

Figure 3.4: Notation of the Equivalent Random Method

peakedness $z$ is given as an input to an $N$-server system. First, one finds an equivalent system with $S$-servers and a load $\gamma$, such that the resulting overflow process has the same mean and peakedness as the given non-Poisson stream (see fig. 3.4). The equations relating $\rho, z, S$, and $\gamma$ are as follows:

$$\gamma = 3z(z - 1) + z\rho$$  \hspace{1cm} (3.26)

$$S = \left(\frac{\rho + z}{\rho + z - 1}\right)\gamma - \rho - 1$$  \hspace{1cm} (3.27)

Note that the resulting value of $S$ from the previous equation is not necessarily an integer, and hence is usually rounded off. These equations give slight overestimates for $\gamma$ and $S$.

The blocking probability for the original $(\rho, z)$ stream is then easily seen to be

$$P_{b} = \frac{B(N + S, \gamma)}{B(S, \gamma)}$$  \hspace{1cm} (3.28)

3.2.3 Random Switch Effect

In this subsection, we show the following:

A traffic stream with mean $\alpha$ and peakedness $z$ is offered to a random switch with one input and $L - 1$ outputs. The random switch places each input onto one of the $L - 1$ output

\footnote{Note that we are concerned with the call blocking probability, or the probability that a call will be blocked, not the time blocking probability, which specifies the fraction of time the $S$ servers are all busy. They are different here since the input process is not Poisson.}
CHAPTER 3. THE RANDOM RETRIAL SCHEME

lines with equal probability \( q = \frac{1}{L-1} \). See Fig. 3.5. Then the mean \( \alpha' \) and peakedness \( z' \) of each of the resulting \( L-1 \) traffic streams is given by

\[
\alpha' = q\alpha
\]

\[
z' - 1 = q(z - 1)
\]

where \( \alpha \) and \( z \) are given by (3.21) and (3.22) respectively. It should be noted that these formulas were first given by Descloux in an unpublished paper [Nea 71]. We prove them here by a simple extension of the method used to prove (3.21) and (3.22).

The Chapman-Kolmogorov equations are almost identical to those given in (3.16) and (3.18):

(1) \( j \leq N - 1 \)

\[
(\rho + j + k)P(j, k) = \rho P(j - 1, k) + (j + 1)P(j + 1, k) + (k + 1)P(j, k + 1)
\]

\[
P(-1, k) = 0 \quad 0 \leq j \leq N - 1 \quad 0 \leq k.
\]

(2) \( j = N \) :

\[
(\rho + N + k)P(N, k) = \rho P(N - 1, k) + q\rho P(N, k - 1) + (k + 1)P(N, k + 1)
\]

\[
P(-1, k) = 0 \quad 0 \leq k.
\]

Note that the only effect of the random switch comes through \( q \) in the second term of the RHS of (3.32).

The solution of (3.31)-(3.32) is found by using probability generating functions in exactly the same manner as for (3.16) and (3.18). The distribution is given by the same formula as for the \( q = 1 \) case:

\[
P(j, k) = (-1)^k \phi_0(N) \sum_{\nu=k}^{\infty} (-1)^{\nu} \binom{\nu}{k} \frac{\rho^\nu}{\nu!} \frac{\phi_\nu(j)}{\phi_{\nu+1}(N)\phi_\nu(N)}
\]

(3.33)
where $\phi_\nu(j)$ is given by

$$
\phi_\nu(j) = \begin{cases} 
\frac{\rho^j}{j!} e^{-\rho} & \nu = 0 \\
\sum_{i=0}^{j} \frac{(\nu+i-1)!}{(j-i)!} \frac{\rho^j}{j!} & \nu > 0
\end{cases}
$$

\hspace{1cm} (3.34)

$0 \leq j \leq N$ and $0 \leq k$.

To find the means of the output processes $\alpha'$ and their variances $z'$, we use the following formulas which follow from (3.33)-(3.34)

$$
\alpha' = q \rho \frac{\phi_0(N)}{\phi_1(N)}
$$

\hspace{1cm} (3.35)

$$
z' = 1 - \alpha' + 2 c_2 \phi_3(N)
$$

\hspace{1cm} (3.36)

where $c_i$ is defined by

$$
c_i = (-1)^i (q \rho)^i \frac{\phi_0(N)}{i! \phi_{i+1}(N) \phi_i(N)}
$$

\hspace{1cm} (3.37)

Notice that (3.35)-(3.36) are just a generalized version of (3.21)-(3.22)($q = 1$). From (3.35) and (3.21), one immediately obtains (3.29). Similarly, one gets from (3.36) and (3.37)

$$
z' - 1 = -\alpha' + (-1)^2 \frac{2 (q \rho)^2}{2!} \frac{\phi_0(N)}{\phi_2(N)}
$$

\hspace{1cm} (3.38)

$$
= -\alpha' + \frac{q \rho}{N + 1 + \alpha - \rho}
$$

\hspace{1cm} (3.39)
Figure 3.6: Channel Input/Output Processes

\[
\begin{align*}
q(-\alpha + \frac{\rho}{N+1+\alpha - \rho}) &= q(\nu - 1) \\
= q(\nu - 1)
\end{align*}
\]

which is the desired result.

Descloix also derived a formula for the cross-covariance of the output streams [Nea 74], which we state here without proof:

\[c = q^2(\nu - \alpha)\]  

Thus for large \(L\), the output traffic streams from the random switch become almost uncorrelated.

### 3.3 Random Retrial with Poisson Inputs

From the outset one can use the symmetry of the system to observe only one channel. Figure 3.6 shows an abstraction of channel \(i\) \((1 \leq i \leq L)\) with all the input and output traffic streams. The figure also shows the \(N\) subchannels abstracted as servers. The input stream is the superposition of the exogenous traffic stream, which is independent of everything else, and the retrial traffic stream, which as shown in Fig. 3.6 is a superposition of all the
CHAPTER 3. THE RANDOM RETRIAL SCHEME

overflow streams directed from all the other channels towards channel \( i \). The output traffic stream is the superposition of the overflow traffic stream, and the departure traffic stream.

One can see from the above that the input traffic stream is, in general, not a renewal process (or GI process in queuing theory notation). This follows from the known fact (see [KaT 81] for example), that the superposition of two renewal processes is not a renewal process unless both are Poisson processes, in which case their superposition is a Poisson process as well. Thus strictly speaking, the input stream can only be described as a point (or G) process.

At the output, similar problems occur. The overflow process is not a renewal process unless the input process is a Poisson process, which is clearly not the case in general. And hence, retrial streams are not renewal processes either.

However, when the number of channels \( L \) is large, one can make the following observation: Since the retrial stream consists of a superposition of \( L \) streams each with a load inversely proportional to \( L \), one can use a well-known theorem, whose proof may be found in [KaT 81], that states that the superposition of infinitely many uniformly sparse renewal processes tends to a Poisson process.

We have seen in Section 3.2.3 that a random \( (L - 1) \) switch produces \( L - 1 \) output processes characterized by loads and peakednesses inversely proportional to \( L \), and with covariances inversely proportional to \( L^2 \) as well. However, the retrial stream at each channel input is the superposition of \( L - 1 \) streams from different channels, which we conjecture to have cross-covariances that are inversely proportional to a power higher than \( L^2 \). Thus for large \( L \), we can apply this theorem to these streams and deduce that the superposition of \( L \) of these streams is closely approximated by a Poisson process.

Using this approximation, we get the system consisting of \( L \) channels such as that shown in Fig. 3.7.

Since the input process in this case is just a superposition of two independent Poisson processes with loads \( \rho \) and \( \rho_r \) (and therefore a Poisson process itself with load \( \rho + \rho_r \)), the blocking probability can be found from the following equations:

\[ P_b = B(N, \rho_r + \rho) \]  \hspace{1cm} (3.43)

and

\[ \rho_r = (\rho + \rho_r)B(N, \rho + \rho_r) \]  \hspace{1cm} (3.44)
Figure 3.7: Channel Input/Output Processes for Large $L$ and Poisson Input

where $B(N, \rho)$ is the Erlang-B function defined in (2.4). For a comprehensive discussion of the properties of this function, the interested reader may refer to [Jag 74] and [Jag 84].

An explicit formula for $\rho_r$ in terms of $\rho$ and $N$ may be written using a continuous fraction representation:

$$\rho_r = \frac{-\rho B(N, \rho) (1 + \frac{\frac{B(N, \rho)(1 + \cdots)}{1-B(N, \rho)(1 + \cdots)}}{1-B(N, \rho)(1 + \cdots)})}{1-B(N, \rho)(1 + \cdots)}$$

(3.45)

Substituting (3.45) into (3.43), we get after some simplification,

$$P_b = B(N, \frac{\rho}{1-P_b})$$

(3.46)

whose solution can again be written in a continuous fraction format:

$$P_b = B(N, \frac{\rho}{1-B(N, \frac{\rho}{1-P_b})})$$

(3.47)

Figures 3.8, 3.9, and 3.10 show three plots of (3.47) for several values of $N$ in comparison with the blocking probability resulting without retrials ($B(N, \rho)$). One can easily see the rapid increase in blocking probability at high loads due to the large contribution of retrials. One should note, however, that due to the necessity of computing a finite number of terms in (3.47)$^4$, one obtains a lower bound on the exact value of $P_b$ which gets increasingly tight.

$^4$50 terms were used for the figures above
with the number of terms used.

At low loads, the blocking probability is almost identical to the case where the $L$ channels are independent, and no retrials are allowed. This is mainly due to the rareness of retrials at lower loads, as well as to the appearance of the retrial customers as a poisson process, and thus not increasing the peakedness of the input stream.

### 3.3.1 Subchannel Access Delay with Random Retrials / Poisson Inputs

The subchannel access delay $D$ is simply defined to be the time spent by the transmitter tuning between channels until a free subchannel is found. The mean subchannel access delay is then just (using (3.3))

$$
\bar{D} = \sum_{i=1}^{\infty} T_{\text{ave}} P_b^{i-1} (1 - P_b)
$$

(3.48)

$$
\approx \frac{T P_b}{3(1 - P_b)} \quad L \gg 1
$$

(3.49)

For convenience, we define the normalized subchannel access delay $D_N$ as
CHAPTER 3. THE RANDOM RETRIAL SCHEME

Figure 3.9: Blocking Probability for Random Retrial Scheme with Poisson Inputs II

Figure 3.10: Blocking Probability for Random Retrial Scheme with Poisson Inputs III
\\[ D_N \triangleq \frac{D}{T} \]\\

(3.50)

Thus, for random retrial with Poisson inputs,

\\[ \overline{D}_N = \frac{P_b}{3(1-P_b)} \]\\

(3.51)

where \( P_b \) is given by (3.47). This value should be compared to

\\[ \overline{D}_{N}^* = \frac{P_b}{T\lambda(1-P_b)} \]\\

(3.52)

where \( D^* \) is defined to be the subchannel access delay with no retrials allowed. Clearly, \( \overline{D}_{N}^* \) is greater than \( \overline{D}_N \) (from (2.8)).

Thus, for moderate loads we have reduced \( \overline{D} \), \textit{without affecting the blocking probability}. This is the significant advantage of random retrial.

We now take a look at the behavior of \( \overline{D} \) with respect to the channel bandwidth \( W_C \), or equivalently the number of subchannels in each channel.

Figures 3.11 and 3.12 show the mean subchannel access delay for several loads and for varying channel sizes. From these figures one may immediately conclude the following:

1. For all loads of interest, random retrial always performs better in terms of subchannel access delay with channels that are as large as possible.

2. The decrease in the subchannel access delay is substantial for lighter loads.

At first glance, the first conclusion may seem somewhat surprising since if we allow only one user/channel, we reduce the channel size to its minimum, and therefore reduce the tuning delay between channels as well. However, this is not the case, since \( T_{\text{ave}} \) in random retrial is independent of the number of channels \( (T/3) \). Thus, to reduce the access delay, \( P_b \) must be reduced by increasing \( W_C \).

### 3.4 Random Retrial with Non-Poisson Inputs

In this Section we investigate the performance of the random retrial scheme when the input traffic streams are not Poisson. Specifically, we assume that all input traffic streams have a peakedness \( z \) greater than one, thus signifying some "burstiness", as explained in Section 3.2.
CHAPTER 3. THE RANDOM RETRIAL SCHEME

Figure 3.11: Mean Subchannel Access Delay With Random Retrial / Poisson Input I

Figure 3.12: Mean Subchannel Access Delay With Random Retrial / Poisson Input II
Figure 3.13: Channel Input/Output Processes for Large $L$ and Non-Poisson Input

We desire to find the blocking probabilities suffered by both exogenous users and retrial users in this system. Akimaru et. al. have analyzed a closely related structure in [AkT 83]. However, they do not assume that overflows return to the same channels in the form of retrials, but instead are presented to new (overflow) channels.

Finding an exact formula for the blocking probabilities in this case is very difficult, due to the non-Poisson nature of the exogenous traffic stream. Strictly speaking, since the input traffic is non Poisson, the overflow process cannot be approximated using mean and peakedness values. However, we can always upper bound the resulting overflow stream by a “normal” overflow stream (one which results from a Poisson input) with a large enough peakedness. Then we use can use (3.30) to show that, for a large number of channels, the retrial stream can still be approximated by a Poisson process. In other words, regardless of the peakedness of the exogenous traffic, the retrial stream for a large number of channels can be approximated by a Poisson process.

The input traffic stream, then, is just a superposition of a Poisson process with load $\rho_r$, and a non Poisson process with load $\rho$ and peakedness $z$.

A simple lower bound to $P_b$ may be immediately found by solving for the case $z = 1$. Here again we use the fact that Poisson streams alway see blocking probabilities that are smaller than non Poisson (bursty) streams.

In this case, we simply return to the system of the previous Section (random retrial with
Poisson input). Hence the formula given in (3.47) can be considered as a lower bound for the true blocking probabilities of the exogenous and retrial streams.

For a more accurate approximation of the blocking probability of this system, we use the equivalent random method. First, we use the uncorrelatedness of the retrial stream and the exogenous stream to combine them into one non-Poisson input stream with parameters \((\rho', z')\), where \(\rho'\) and \(z'\) are given by:

\[
\rho' = \rho + \rho_r \tag{3.53}
\]

\[
z' = \frac{z \rho + \rho_r}{\rho'} \tag{3.54}
\]

We then use these values in the equivalent random method:

\[
\gamma = 3 z' (z' - 1) + z' \rho' \tag{3.55}
\]

\[
S = \frac{\rho' + z}{\rho' + z - 1} \gamma - \rho' - 1 \tag{3.56}
\]

We also have the additional restriction that results from specifying the overflow load \((\rho_r)\):

\[
\rho_r = \gamma B(N + S, \gamma) \tag{3.57}
\]

Using (3.28), the blocking probability is immediately seen to be
<table>
<thead>
<tr>
<th>$z = 1$</th>
<th>$z = 2$</th>
<th>$z = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>w/o retrials</strong></td>
<td><strong>w/ retrials</strong></td>
<td><strong>w/o retrials</strong></td>
</tr>
<tr>
<td>$\rho = 10$</td>
<td>1.5e-19</td>
<td>1.5e-19</td>
</tr>
<tr>
<td>$\rho = 20$</td>
<td>7.6e-09</td>
<td>7.6e-09</td>
</tr>
<tr>
<td>$\rho = 50$</td>
<td>0.10478</td>
<td>0.6743</td>
</tr>
</tbody>
</table>

Table 3.1: Blocking Probabilities for $N = 50$

\[
Pr(\text{blocking}) = \frac{B(N + S, \gamma)}{B(S, \gamma)} \quad (3.58)
\]

\[
= \frac{\rho_r}{\rho_r + \rho} \quad (3.59)
\]

Obtaining an explicit formula for the blocking probability from the set of equations (3.55)-(3.58) is quite difficult, since they are coupled. However, we give in Table 3.1 values for one specific case ($N = 50$) that was numerically computed from (3.55)-(3.58).

As expected, the blocking probabilities increase with increasing peakedness and with increasing load. In addition, the blocking probabilities increase significantly with retrials only at high loads. Note, however, that at high peakedness values, the blocking probability starts increasing at lower loads.

Thus, one may conclude that for input processes with moderate peakedness values a smaller delay may be achieved with no cost in blocking probability, as in the case of Poisson inputs. The behavior of the subchannel access delay for this system should parallel that of Poisson inputs, but with different absolute values (i.e., large channels are better than small ones here as well).
CHAPTER 3. THE RANDOM RETRIAL SCHEME

3.A Proof that $z = 1$ iff the Overflow Process is Poisson

($\Leftarrow$) If the overflow process is Poisson, we just have (refer to Fig. 3.3) an $M/M/\infty$ system. It is well known [Kle 75] that the distribution of the number of busy servers in an $M/M/\infty$ system is Poisson distributed (in steady state). Hence the mean and variance are equal, and $z = 1$.

($\Rightarrow$) It is shown in [Rio 62] that the binomial moments $B_i$ of the number of busy servers in a $GI/M/\infty$ system are given by the following (lengthy) set of equations:

$$B_i = \frac{\rho}{i} C_{i-1} \quad i > 0 \quad \rho = \lambda/\mu \quad (3.60)$$

$$C_k = \lambda_1 \lambda_2 \cdots \lambda_k \quad k > 0 \quad C_0 = 1 \quad (3.61)$$

$$\lambda_j = \frac{F(j\mu)}{1 - F(j\mu)} \quad j > 0 \quad (3.62)$$

where $F(s)$ is the Laplace Transform of the interarrival probability density function $f(t)$:

$$F(s) = \int_0^\infty f(t) e^{-st} dt \quad (3.63)$$

The mean of the number of busy servers is just given by $B_1 = \rho$. The variance is given by

$$2B_2 + B_1 - B_1^2 = \rho \frac{F(\mu)}{1 - F(\mu)} + \rho - \rho^2 \quad (3.64)$$

If we require mean = variance, then $\rho = \frac{F(\mu)}{1 - F(\mu)}$

or

$$F(\mu) = \int_0^\infty f(t) e^{-\mu t} dt = \frac{\lambda}{\lambda + \mu} \quad (3.65)$$

whose unique solution is

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

which shows that the input process interarrival time is negative exponentially distributed. Thus the input process is Poisson.
Chapter 4

The Nearest Neighbor Retrial Scheme

In this chapter we discuss the second type of retrial scheme, namely the nearest neighbor retrial. The structure of this Chapter closely follows that of the previous Chapter. However, for reasons of mathematical expediency, we limit our discussion to the case of Poisson input streams.

4.1 System Model

Nearest neighbor retrial schemes are much more attractive from the practical point of view than random retrial schemes, since tuning occurs in much shorter steps, thus presenting less of a burden on the lasers' tuning capabilities.

The term "nearest neighbor" only restricts the choice of retrial channels to the two neighboring channels of the current (blocked) channel. One must, in addition, specify whether the choice between either channel is made with equal probability, and whether the retrial scheme has memory or not. By memory we mean knowledge of the direction of previous tunings. The example of nearest neighbor retrial with random walk given in Chapter 2 is an example of memoryless tuning.  

Several different nearest neighbor schemes may be conceived, depending on the probability distribution and choice of memory / no memory. We therefore make a further restriction on the scheme, namely fairness.

We have already shown that the nearest neighbor retrial scheme with random walk was unfair to the subscribers assigned to remote channels (near either end of $W_T$) in the sense

\footnote{Note that, by definition, random retrial is memoryless as well.}
of allowing them a much smaller ($\approx L$) number of retrials than those near the center of $W_T$ ($\approx L^2/4$).

We use here another nearest neighbor retrial scheme based on one way tuning that is more fair. In this method a user that is blocked from his original channel tunes to either his left or right neighboring channel. If blocked in the neighboring channel, the user tunes in the same direction (i.e. away from his original channel). See for example Fig. 4.1. In this figure a channel 2 subscriber was blocked from that channel. He chose to tune to the right (to channel 3) with probability $p_r$. However, he was blocked from channel 3 as well. Hence he tunes to channel 4 with probability 1. Similarly for the channel $L$ subscriber. Thus, only the first retrial channel is unknown; all following retrial channels (if necessary) are known.

It should be evident that even this scheme is not exactly fair either, since it allows $L$ retrials for subscribers of end channels, whereas only $L/2$ retrials are allowed for subscribers of center channels. However, the ratio of the maximum and the minimum number of retrials is approaches 1/2 with increasing $L$, whereas in the random walk scheme it approached $\infty$.

The choice of left or right neighbor is not necessarily based on equal probability. Indeed, for a finite number of channels it should be clear that the direction choice should be dependent on the channel number. Specifically, the end channels have only one direction to choose from, and the other channels should choose either direction in accordance with the number of channels available in that direction: channel $i$ subscribers choose to tune to the
left with probability $\frac{i - 1}{L - 1}$ and to the right with probability $\frac{L - i}{L - 1}$. See Fig. 4.2. However, as the number of channels becomes very large, the effect of the end channels (absorption) becomes unimportant and equal probability directional tuning is equivalent to the unequal probability distribution given above. In this case, the number of allowed retrials becomes infinite (there is no “absorption” at the boundaries).

One major advantage of one-direction tuning is that it eliminates any first-order feedback in the form of users retrying in the same channel. Second order feedback still exists, however, and does not decrease with increasing $L$ as in the random retrial case, since users have only two channels to tune to in each step.

4.1.1 Average Tuning Time

The expected value and variance of tuning time are almost trivial here, since in this system $T_{ij}$ is defined only for $j = i \pm 1$, and is equal to

$$T_{ij} = \frac{T}{L - 1}$$

(4.1)

Obviously then, $T_{ave}$ is just $\frac{T}{L - 1}$ and the variance is just zero.
4.1.2 Number of Retries

As in random retrial, the number of allowed retrials is unconstrained. Here again as in random retrial, the infinite number of retrials assumption leads to complex interactions between traffic streams. In particular, due to the special structure of this system imposed by nearest neighbor tuning and "one way" retrials, the retrial stream of channel $i$ is the superposition of the following traffic streams:

(1) The fraction of channel $i - 1$ ($i + 1$) users whose original channel access trial was blocked, and who tuned to channel $i$ with probability $\frac{L-i+1}{L-1}$ ($\frac{i}{L-1}$).

(2) The fraction of channel $i - 2$ ($i + 2$) users whose original channel access trial was blocked, and who tuned to channel $i - 1$ ($i + 1$) with probability $\frac{L-i+2}{L-1}$ ($\frac{i+1}{L-1}$).

Their first retrial was also blocked, they tuned with probability 1 to channel $i$.

(3) And so on $\cdots$

See Fig. 4.3 for a graphical explanation of this superposition.

It should be clear that the superposition of these infinite number of traffic streams does not approach a Poisson process, since they do not satisfy the requirement of vanishingly small rates. Hence retrial streams in this system cannot be considered Poisson.
CHAPTER 4. THE NEAREST NEIGHBOR RETRIAL SCHEME

However, one may use here again the fact that Poisson streams see the smallest blocking probability to lower bound the blocking probability. This is our approach in the next Section.

Unfortunately, finding an upper bound on the blocking probability proves to be a much harder problem. The reason behind this is simply because one does not know the value of the maximum peakedness of all the streams composing a retrial stream, which could be used to upperbound the peakedness of the retrial stream, and hence the blocking probability.

4.1.3 Constraints Imposed by Nearest Neighbor Retrial

Applying (2.8) here just gives

\[
\frac{T}{L-1} < \frac{1}{N\mu} \tag{4.2}
\]

and hence for large \(L\),

\[
T < \frac{W_TW_R}{\mu W_C^2} \tag{4.3}
\]

This gives a necessary condition that must be satisfied in order for nearest neighbor retrial to be useful. Alternatively, one may use this to constrain \(W_C\) for a given \(T\) to get

\[
W_C < \sqrt{\frac{W_TW_R}{T\mu}} \tag{4.4}
\]

Hence in the case of nearest neighbor retrial, (2.1) must be modified to

\[
W_R \leq W_C \leq \min\{W_{C_{\text{max}}}, \sqrt{\frac{W_TW_R}{T\mu}}\} \tag{4.5}
\]

4.2 Nearest Neighbor Retrial with Poisson Inputs

As shown in Section 4.1.2, the retrial stream for each channel cannot be considered a Poisson process. However, we noted in that Section that making these retrial streams Poisson processes, provides a lower bound to the blocking probability.

Specifically, by observing Fig. 4.3, one may compute the load on channel \(i\) simply as follows:
\[ \rho_i = \rho + \sum_{k=1}^{i-1} \frac{L-k}{L-1} \rho P_b^{i-k} + \sum_{k=1}^{L-i} \frac{L-k}{L-1} \rho P_b^{L-i-k+1} \]  
\[ = \rho + 2 \sum_{k=1}^{i-1} \frac{L-k}{L-1} \rho P_b^{i-k} \]  
\[ \text{(4.6)} \]

As can be noted, the above expression assumes that all \( L \) channels are offered the same retrial load, since \( P_b \) is independent of \( i \). However, the same expression gives channel dependent loads (from the LHS). To overcome this inconsistency, we may either let \( L \to \infty \) thus making all channel loads identical and equal to

\[ \rho_{\text{input}} = \rho + 2\frac{\rho}{2} P_b + 2\frac{\rho}{2} P_b^2 + \cdots \]  
\[ = \frac{\rho}{1 - P_b} \]  
\[ \text{(4.8)} \]

\[ \text{(4.9)} \]

We may also average out \( \rho_i \) over the \( L \) channels thereby obtaining

\[ \bar{\rho} = \frac{1}{L} \sum_{i=1}^{L} \rho_i \]  
\[ \text{(4.10)} \]

\[ = \rho + 2 \sum_{i=1}^{L} \frac{1}{L} \sum_{k=1}^{i-1} \frac{L-k}{L-1} \rho P_b^{i-k} \]  
\[ = \rho + \frac{2\rho P_b X}{(L-1)(P_b-1)} - \frac{2\rho P_b^2 Y}{L(L-1)(P_b-1)^2} \]  
\[ \text{(4.11)} \]

\[ \text{(4.12)} \]

where \( X \) and \( Y \) are given by

\[ X = \left[ L - \frac{1 - P_b^{L+1}}{1 - P_b} \right] \]

\[ Y = \left[ \frac{1 - P_b^{L+1}}{1 - P_b} - \frac{L(L+1)}{2} - L + \frac{L(L+1)}{2P_b} \right] \]

Note that as \( L \to \infty \), (4.10) equals (4.8) as expected.

The blocking probability may then be computed as

\[ P_b = B(N, \bar{\rho}) = B\left(N, \frac{\rho}{1 - P_b}\right) \quad L \gg 1 \]  
\[ \text{(4.13)} \]

which is exactly the same formula obtained for random retrial with Poisson inputs. Note that in this case however, it is only a lower bound on the blocking probability.
4.3 Subchannel Access Delay with Nearest Neighbor Retrial

Using the lower bound on the blocking probability given in the previous Section, we may compute the expected value of the subchannel access delay $D$, and its variance. Using (3.48) here we get

$$
\overline{D} = \sum_{i=1}^{\infty} \frac{iT}{L-1} P_b^{i-1}(1 - P_b) \quad (4.14)
$$

$$
= \frac{TP_b}{L(1 - P_b)} \quad L \gg 1 \quad (4.15)
$$

Again, we define the normalized subchannel access delay $D_N$ as

$$
D_N \triangleq \frac{D}{T} \quad (4.16)
$$

Thus for the nearest neighbor retrial scheme with Poisson input,

$$
\overline{D}_N = \frac{P_b}{M (1 - P_b)} \quad (4.17)
$$

where $M = LN = W_T/W_R$ is a constant and $P_b$ is given by (3.47):

$$
P_b = B(N, \frac{\rho_s}{1 - B(N, \frac{\rho_s}{1 - ...}}) \quad (4.18)
$$

Similarly, the variance of the subchannel access delay $Var(D)$ can be computed to be

$$
Var(D) = \frac{T^2 P_b (1 + P_b)}{L^2 (1 - P_b)^3} - \left( \frac{TP_b}{L(1 - P_b)} \right)^2 \quad (4.19)
$$

and the normalized variance $Var_N(D)$ is just

$$
Var_N(D) = \frac{1}{T^2} Var(D) = \frac{P_b (1 + P_b)}{(M (1 - P_b))^2} \quad (4.20)
$$

Figures 4.4 through 4.7 show plots of $\overline{D}_N$ and $Var_N(D)$ for several loads and for various values of $N$ for two systems with different $W_{C_{max}}$.

From figures 4.4 and 4.6, we may conclude that making the channel bandwidth as large as possible such that it conforms with the upper limit in (4.5) is the best choice. It is clear, however, that in the case of the load exceeding capacity ($\rho > 1$), the optimum decision would
CHAPTER 4. THE NEAREST NEIGHBOR RETRIAL SCHEME

Figure 4.4: Normalized Expected Subchannel Delay - System I

Figure 4.5: Normalized Subchannel Delay Variance - System I
CHAPTER 4. THE NEAREST NEIGHBOR RETRIAL SCHEME

Figure 4.6: Normalized Expected Subchannel Delay - System II

Figure 4.7: Normalized Subchannel Delay Variance - System II
be to make the channels as small as possible (i.e. $W_R$). We have noted though in Chapter 2 that such loads are not likely in an optical fiber environment. But as noted before in Chapter 3, a case that would result in an "equivalent" load would be a sub-capacity load but with a high peakedness (burstiness).

One should also note that the variance of the subchannel access delay time decreases with increased channel bandwidth for even over-capacity loads. Hence, based on the variance, the choice would be to make channels with maximum bandwidth.

A third point to be noted is that all the above analysis was performed assuming the slow tuning time was independent of the channel bandwidth. In reality, however, this may not be the case, since slow tuning over $L$ small channels might take a longer period than over $L'$ large channels, where the $L$ channels and the $L'$ channels cover the same total bandwidth $W_T$. This is since slow tuning might have to make "stops" at each channel boundary, thus giving rise to an delay overhead which increases with the number of channels.

From all the above, we may conclude that channel bandwidth should always be as large as possible, except in the case where it is known that the user input streams are very bursty, in which case it would be advisable to choose the single user channel case.
Chapter 5

Conclusions

5.1 Summary

We have investigated in this thesis the performance improvement that can be achieved by using the combination of slow and fast tuning in single-mode fiber optic communication networks. The approach taken was a frequency division one with slow tuning acting as an inter-channel tuning technique. Thus by allowing slow tuning, all users had full accessibility to the total fiber bandwidth with delay, and zero-delay accessibility to only one channel bandwidth.

Slow tuning also made subchannel access retrials possible, since we specifically limited retrials to channels different from the channel in which the original retrial was made.

The finite duration of slow tuning resulted however in restrictions on the user arrival rate (or load) and the service (call holding) rate. It was found that retrial schemes are feasible and sensible only in networks with moderate load and large call holding times.

Two retrial schemes were investigated: Random Retrial and Nearest Neighbor Retrial. The motivation for analyzing the former was its inherent symmetry which simplified the analysis considerably. The latter was chosen for its attractiveness from the practical point of view, but as expected, was much harder to analyze.

It was found that for moderate loads, the subchannel access delay was significantly reduced by allowing random retrials, with no corresponding increase in blocking probability. Heuristically, this may be explained as follows: In any normal retrial system, the blocking probability increases significantly even at low loads due to the feedback effect caused by the retrials. However, this feedback effect is practically eliminated in this network, since we take

\[1\text{, i.e one in which the retrials try in the same channel}\]
advantage of the large bandwidth (and hence large number of channels) available. This is done by allowing users to retry only in other channels thus randomizing or "smoothing" the retrial traffic stream and making it appear as a Poisson stream as the number of channels grows. This is shown more rigorously in Chapter 3.

Furthermore, it was also shown in Chapter 3 that this method would also work for non-Poisson (i.e. \( z > 1 \)) exogenous streams. Blocking probabilities for this case were calculated using the equivalent random method. However, explicit formulas for the blocking probabilities could not be obtained, but numerical evaluation of one specific case was provided.

In the case of nearest neighbor retrial, the blocking probabilities could not be computed exactly, and only a lower bound on the blocking probability could be derived. Nonetheless some insight was achieved with this lower bound alone. In particular, it was shown that for nearest neighbor retrial, there exist cases where subchannel access delay is actually reduced by reducing the channel bandwidth to its smallest possible value (\( W_R \)).

5.2 Directions For Further Research

Two issues that were only lightly touched upon that need further study are:

1. **Protocols:** Clearly, one major issue we just glossed over is that of designing protocols that coordinate transmitters and receivers. Further investigation of this issue is necessary before designing a real system with slow and fast tuning. One related issue of concern is that of allowing a central controller in this system. For systems with small numbers of users, this might well be the most reasonable approach. However, as noted before, distributed control should be of interest in very large optical communication networks.

2. **Practical Costs:** A second topic not covered in this thesis is the actual cost of installing laser with slow and fast tuning capability. In other words, what is the cost (in terms of complexity and money) of using such laser? An indirect answer to this question may be given by providing a practical and economical laser design that incorporates both tuning methods.
Bibliography


[Hei 87] F. Heismann et al., "Electrooptically tunable narrow-linewidth InGaAsP-Ti:LiNbO extended cavity laser", Technical Digest, Optical Society of America OFC/IOOC '87, pp. 149.

BIBLIOGRAPHY


BIBLIOGRAPHY


