# CALL ADMISSION AND ROUTING IN TELECOMMUNICATION NETWORKS

By

Kit-Man CHAN

A Thesis

submitted in partial fulfillment of the requirements
for the Degree of Master of Philosophy
Division of Information Engineering
The Chinese University of Hong Hong
June 1994

# Acknowledgement

I would like to express my greatest gratitude and sincere thanks to my supervisor, Dr T.S. Yum, whose guidance has been invaluable to my learning process.

Ken Chan, Lawrence Yeung and Philip To have made the past two years very enjoyable for me. I wish all of them every success in their research works.

Kit-man Chan
May 1994

# Abstract

Call admission and routing are two of the most important network management functions in integrated service communication network. An integrated service communication network is a network that can support several classes of services, each with different bandwidth requirement, service characteristic and revenue earning rate. Such a network can be modeled as a multirate loss network. In this thesis, we first evaluate and compare five call admission policies under two different adaptive routing rules in a multirate loss network using the reduced load approximation method. The first routing rule is based on the traffic packing principle, while the second one is based on the least congestion principle. We find that routing based on the least congestion principle gives a smaller blocking probability and a lower revenue loss when compared to the one based on traffic packing. Call admission policies are designed to advoid the dominance of network capacity by a particular class of calls. Five policies are evaluated, namely, the Complete Sharing(CS), the Limited Occupancy(LO), the Guaranteed Bandwidth(GB), the External Blocking(EB) and the Direct-link Packing(DP) policies. Our study find that the LO, GB and EB policies can manipulate the relative blocking probabilities of different classes of calls provided that the bandwidth reservation parameters are suitably selected. However, all of them will increase the revenue loss of the network when compared to the Complete Sharing(CS) policy. The Direct-link Packing(DP) policy is found to give significant reduction in both the blocking probability and the revenue loss when compared to CS policy.

With the advent in switching technologies and the installation of Common Channel Signaling, it is now feasible to implement sophisticated adaptive routing scheme on networks which support different kind of services with heterogeneous bandwidth characteristics. Thus in the second part of this thesis, we analyze and compare two versions of the Least Congestion routing in fully connected multirate loss networks. The first routing rule is called the *Maximum mean Time to Blocking*(MTB) routing and is based on the *mean time to blocking* measure of a link. This measure captures the traffic rates, bandwidth characteristic and link capacity information and can reflect more accurately the congestion status of different paths. The second routing rule is called the $M^2$ routing. It is based on the link residual bandwidth measure. In our study, the MTB routing is found to outperform the $M^2$ routing for wide range of network load and different traffic composition. Aggregation of link status information can be used to significantly reduce signaling traffic. We show that with properly designed aggregation, both aggregated $M^2$ and MTB routings can have performance approach that of the respective non-aggregated schemes.

Fiber-optic transmission technology can offer very high bandwidth and very low bit error rate services. It is, however, very difficult to increase the electronic processing speed to that of the optical transmission. A natural solution is to develop all-optical networks. In the third part of this thesis, we extend our study of the Least Congestion routing rule to the Wavelength Division Multiplexed (WDM) lightwave networks. In such a network, each switching node may have a number of wavelength converters that can be used to resolve wavelength conflicts in multi-hop paths. We find that the use of wavelength converters can only provide a very small reduction of blocking probability.

# Contents

# Chapter 1

# Introduction

Traditional telephone networks which support only voice calls are inadequate for handling modern telecommunication needs. As demands for high bit rate transport services increase, it is no longer feasible to provide each services on a separate network with dedicated equipments. With the recent advent in fiber-optic technology, it is more economical to integrate all these services onto a common transmission facility.

The analysis of circuit-switched networks(e.g. telephone network) has been studied extensively in the past. But the emerging integrated networks present new problems in routing and call admission. The purpose of this thesis is to study the problems of call admission and routing in integrated service networks.

## 1.1 Overview of Integrated Service Digital Networks

With the advent of digital computer technology, it is evident to telephone companies that digital system operation is more advantageous than the traditional analog way of operation in terms of efficiency and cost-reduction. Digital systems promise efficient and cost-effective way of operation. Incoming voice signals at the local office are digitized using pulse-code modulation (PCM) and then time-division multiplexed (TDM) onto the outgoing link. At every intermediate switching office, only space division switching is needed to route the TDM signals. At the last office, demultiplexing and demodulation are done to recover the voice signals to the end users.

Transmission links and switching equipment have been replaced by their digital counterparts. The first digital transmission link, known as a T1 carrier, consisting of 24 voice channels of 64 kilobits per second (kbps) each, was installed in 1962 by AT&T. In 1976, the first time division digital switch, the AT&T's 4ESS, was installed. By 1980, 25% of the Bell System's switches employed digital technology and this figure has risen to about 80% by 1990 [3].

The digitalization of telecommunication network has made possible the transmission of data traffic such as interactive terminal-computer communication and information retrieval from business database. However, for applications such as interactive graphics and high-speed computer communications which require a much larger bandwidth, the current digital network is lacking the capabilities to support such applications. Consequently, dedicated links and terminals are required for every new service acquired by the user, and the cost involved is not economical. Furthermore, equipment vendors would develop their own standard if there is no standardization. Hence, new services will be difficult to become popular. To support new services economically and flexibly, the Integrated Services Digital Network (ISDN) is developed

To establish an ISDN which will satisfy all the communication needs, the major part is to set the necessary standards for the various aspects of ISDN. These aspects include network architecture, equipment, service, performance, and interfaces (user-network and network-network). ISDN will support digitized voice traffic via circuit switching which provides uninterrupted connection and data via either circuit switching or packet switching depending on the need. The bursty nature of data traffic favors the method of packet switching; however, some cases which require immediate allocation of dedicated bandwidth, will be better served by circuit switching.

Part of the established standards in ISDN limits the user-network access link to two standard packages. First, basic service, which is aimed at residential users consists of two B channels of 64 kbps each for carrying either PCM-encoded digitized voice via circuit switching or digital data via packet switching, and a D channel of 16 kbps for transport of control signal or lower-speed digital data packets. Second, primary service, which is aimed at business users comprises 23 B channels of 64 kbps each plus D channels of 64 kbps. It is commonly known as the T1 carrier advocated by United States, Canada and Japan. In European countries, however, a package consisting of 30 B channels of 64 kbps each plus a D channel of 64 kbps is used instead. The resulting overall bit rate is 2.048 Mbps.

The characteristic of different types of services varies in bandwidth requirement, duration of service time, and traffic intensity. The bandwidth requirement ranges from a few hundreds of bit per second, such as utility meter reading and security, to hundreds of mega bit per second, such as image applications. Service time varies from a fraction of a second for telemetry to a few hours for video teleconferencing. Furthermore, traffic rates can change from once every month for utility meter reading to once every second for critical monitoring.

Services demanding hundreds of mega bit per second transmission capacity are observed. To transmit full motion video signals through an ISDN primary-service access link, complex image-compression technique and complicated terminal equipment must be used, thus causing a degrade in picture quality. Moreover, broadband services requiring much larger bandwidth appear to have potential applications to both business and residential users. For instance, extended quality TV (EQTV) requires a transmission rate of 245 Mbps with straightforward coding, whereas high definition TV (HDTV) demands a bit rate of 1.2 gigabits per second (Gbps) or 300 Mbps, depending on the coding scheme adopted. As broadband services emerge, Broadband ISDN (BISDN), which consists of high bandwidth fiber-optic links and fast switching equipment, is developed. Two standard packages, 150 Mbps and 600 Mbps are proposed for the BISDN user-network access link [4].

There are two possible transport techniques for BISDN, namely the Synchronous Transfer Mode (STM) and Asynchronous Transfer Mode (ATM). STM is equivalent to multirate circuit switching, in which dedicated physical channels (or time slots) are assigned to a particular connection. ATM, on the other hand, is based on fast packet switching, packetizes user's information to be transferred into fixed size slots called cells, which are then transmitted to the destination according to the header information.

As the technologies of communication and computing progresses, it is found that large increase in the speed of electronic processing is becoming increasingly harder to achieve. With the advent in fiber optics, the low-loss wavelength window of a single-mode optical fiber provides a bandwidth of about 25 THz (200 nm). Optical transmission has a very low bit error rate and an excellent security which is unparalleled by the electronic counterpart. Provided a way to tap this huge amount of bandwidth, the cost of the transmission can be significantly reduced.

It was reported that at the end of 1990, over five million miles of fiber had been installed [1]. 50% of this installed base being "dark", meaning that no terminal equipment had been attached. While 2 million miles is in intercity trunking of the major carriers, 3 million miles is in the local exchange carriers and in the so-called "alternate-access" carriers. Another major provider is the cable television (CATV) industry. While the telephone companies have been investigating fiber to the curb, the CATV industry has decided similarly. Fiber is being run from the headend out to a "fiber node" on a utility pole, from which the distribution to home is by coaxial cable. On this increasingly pervasive base, it is possible to build a huge infrastructure of broadband transport network provided that the technology of all-optical networks becomes cost-effective, and that applications requiring only all optical solution are identified.

Recently, the AT&T, Digital Equipment Corporation (DEC), and Massachusetts Institute of Technology (MIT) [2] have formed a precompetitive consortium to address the challenges of utilizing the evolving terahertz bandwidth capacity of optical fiber technology to develop a national information infrastructure capable of providing flexible transport, common conventions and common servers.

The baseline architecture will potentially allow frequency division multiplexing to access the 25 THz of fiber bandwidth. Three basic services will be supported. First, point-to-point or point-to-multipoint high-speed circuit-switched multi-gigabits-per-second digital or analog sessions; second, time division multiplexed (TDM) circuit-switched sessions in the range of a few Mb/s to the full channel rate of multi-gigabits-per-second, and third, a service used for control, scheduling and network management. To achieve scalability in terms of dimensions of geographic span, number of users and data rates, the architecture employ hierarchical structure that includes local area networks (LAN's), metropolitan area networks (MAN's) and wide area networks (WAN's).

## 1.2 Multirate Loss Networks

Consider a network supporting various classes of traffic, where each class has a bandwidth requirement, a service rate and a revenue rate. We assume that connection requests from various classes arrive according to independent Poisson processes. If there is not enough bandwidth to satisfy a connection request, then the request is blocked and lost. Otherwise, the connection is accepted and hold an amount of bandwidth for its duration. The call duration is assumed to be exponential. An example of such network is shown in Figure 1.1 where there are two classes of call.

We refer to the model described above as a multirate loss network. It can be used to study the performance of broadband telecommunication networks that integrate disparate services, for example voice, data, video, etc, over the same transmission and switching facilities. Maximizing the revenue while satisfying the quality of services requirements of individual class of service has been a very important design objective of the routing and call admission policies for a multirate loss network.

Alternate routing of calls has long been regarded in the telephone industry as a means of increasing call throughput and robustness in the telephone network. Call throughput is increased by setting up calls on alternate paths when all circuits are occupied on the direct link. Robustness, measured in terms of the network's ability to respond to equipment failure and to unexpected surges of traffic, is made available by

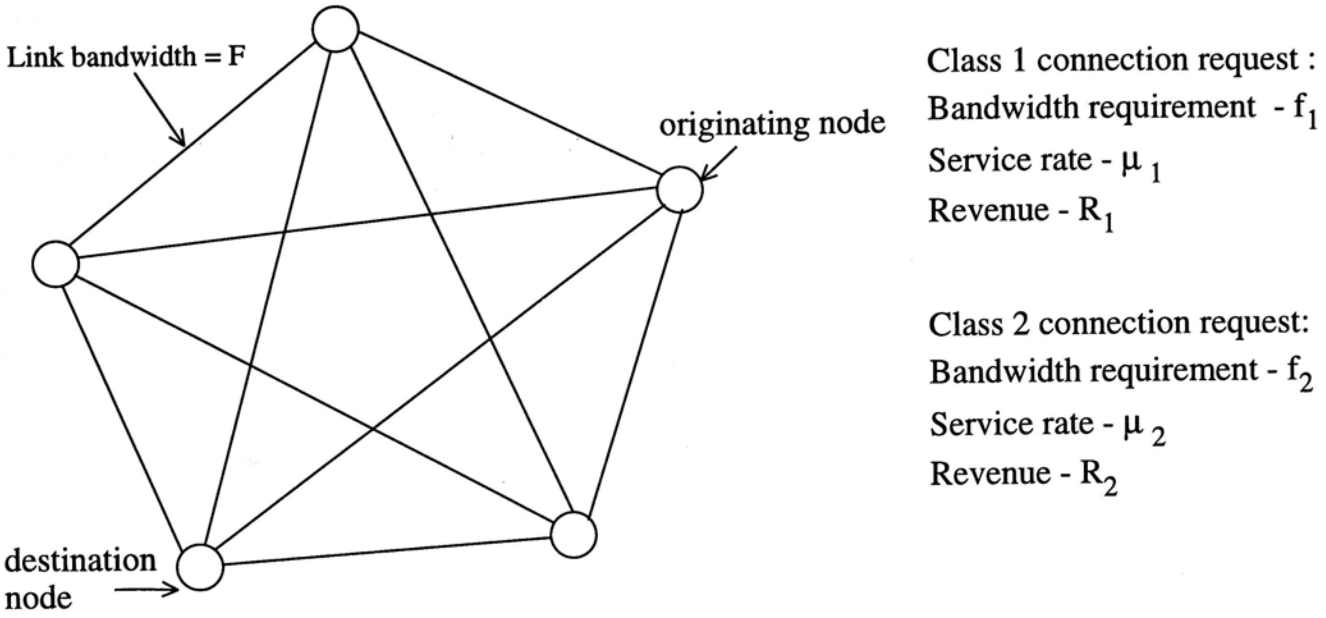


Figure 1.1: Example of a multirate loss network

transferring flows to alternate routes.

Since the early 1980s, the trend in the telephone industry is to implement dynamic routing in nonhierarchical networks because of its lower network management cost when compared to the hierarchical counterpart. Given the great gains in performance and reduction in cost offered by alternate routing in nonhierarchical telephone networks, it is natural to adopt adaptive routing in multirate loss network.

Routing and call admission in multirate traffic environment is a complex problem which should take into account two concerns. First, it may be desirable to pack narrowband calls within certain routes so that the remaining routes have enough capacity to support additional wideband calls. Second, it may be necessary to protect certain services from the domination of others.

Although the two transfer mode for broadband networks - Synchronous Transfer Mode (STM) and Asynchronous Transfer Mode (ATM) - are fundamentally different, the multirate loss network is a good model for the study of both cases. STM is equivalent to multirate circuit switching and is in fact the switching method in multirate loss networks. ATM uses virtual-circuit packet switching and blocks a connection if its admission would degrade the quality of service of the on-going connections. Here, a large number of traffic sources having a broad range of burstiness characteristics are multiplexed onto a transmission facilities. Recent research reported that an effective bandwidth[23, 24], which depends only on the source characteristic and the cell loss probability, can be assigned to a traffic source. This allows circuit-switched type call acceptance and routing scheme to be implemented in ATM network. Therefore, the multirate loss network can also be employed to model the performance of ATM networks at the connection level.

## 1.3 Previous Work

Call admission policies that maximizes the channel utilization of a single link network supporting multirate circuit-switched traffic was formulated by Gopal and Stern in [5] as a Semi-Markov Decision Process. However, the technique appeared infeasible for problems of realistic size due to the large dimensionality of the state space involved. An earlier work by Foschini, Gopinath, and Hayes [6] considered a similar problem of optimally allocating servers to two classes of user where each user requires only one server. The optimal policy that maximizes some general revenue function was proved to be a simple threshold policy, in which the maximum number of users of one of the classes is restricted while that of the other class is not.

Kraimeche and Schwartz compared two different strategies for managing the access of two types of traffic, a blockable wide-band(WB) type of traffic and a queueable narrow-band(NB) type of traffic, sharing a common transmission facilities[15]. The first strategy assign preemptive priority to the WB traffic over the NB traffic whereas the second strategy employs a wide-band to narrow-band bit rate compression mechanism. They derived exact models for both strategies.

Tsang and Ross located the optimal circuit access policies from among the class of coordinate convex policy, and proposed algorithm to determine the revenue for stochastic knapsack problem with two types of objects[16]. They also developed finite-stage dynamic programming algorithm for locating the optimal static control, for the general $K$ classes of objects, where for each class a portion of the knapsack is dedicated. In [17], they studied the optimal circuit access policies for Integrated Service Digital Networks employing fixed routing. A Markov Decision Process(MDP) approach was employed to obtain optimal access policies for three models: the flexible scheme access-port model, the contiguous scheme access-port model and the network-access model. Both linear programming and value-iteration MDP algorithms are coupled with a novel state descriptor to locate the optimal policy.

The solution for the blocking probability of a single trunk group, with single-rate calls having exponentially distributed interarrival and holding times, was developed by Erlang, who first published an exact expression in 1915, known as the Erlang-B formula. Erlang and later Jensen [7] generalized the result to handle networks with tree topologies; however, it is difficult to compute the normalization constant for problems of practical size.

For some specific network topologies, the product form solution has been employed to develop efficient combinatorial algorithms to calculate blocking probabilities. The first significant result for single-link networks supporting multirate traffic was independently reported by Kaufman [14] and Roberts [8] in 1981. A computationally efficient algorithm for determining the blocking probabilities in a single-link network topologies was proposed. Ross and Tsang developed algorithms to handle tree and hierarchical tree topologies with multirate connections [9]. However, it appears difficult, to develop efficient combinatorial algorithms for more general topologies and routing rule due to the complicated state space involved.

Since exact computations of blocking probabilities are exceedingly difficult for general network topologies and routing scheme, it is therefore natural to consider approximation schemes unless the particular network has some special structure that can be exploited to come up with efficient algorithms as in the single-link networks.

The reduced-load approximation for single-rate traffic was first proposed by Katz in 1967 [10]. It assumes that links are independent and that offered load to a link is reduced by blocking on the others. Whitt [11] suggested using successive approximations to solve the corresponding set of nonlinear equations iteratively and further proved convergence properties of the proposed method. Kelly [12] extended the reduced load approximation to include multirate traffic and proved the existence of a unique solution to the approximation. Furthermore, he showed that the approximation is asymptotically correct as both the offered loads and the capacities of the links go to infinity together.

In [29], Chung and Ross studied the approximate formulae for computing the loss probability of multirate loss networks under fixed routing. They studied two reduced load approximation called the Knapsack approximation and the Pascal approximation. Both were found to be more accurate than the generalization of Kelly's approximation to multirate traffic. They also studied the sensitivity of the average revenue to the changes of offered load and link capacity based on these approximations.

With the advent of stored program control and the installation of out-of-band signaling, it is now possible to implement a network that support different class of services with heterogeneous bandwidth requirement. The AT&T's Real Time Network Routing(RTNR) is an example of such adaptive routing scheme which implements new *class-of-service routing* capabilities for dynamic networks[13]. RTNR is a modification of the aggregated Least Busy Alternate Routing for circuit-switched networks studied by Mitra *et al.* in [32].

In [19] , random alternate routing in circuit-switched networks supporting two classes of services having the same bandwidth requirement was analyzed. Gupta and Ross have proposed a circuit-switched type routing algorithm for virtual-path-based ATM networks based on the fluid approximation of the buffer overflow probability [20].

Recently, adaptive call admission and routing schemes based on the Markov Decision Process(MDP) were proposed. The complexity of the algorithm, however, is unmanageable in multirate loss networks. In [21], Dziong and Mason reduced the complexity of the problem by decomposing the network reward process into a set of link reward processes. In [22], Hwang et al. employed the MDP approach and generalized the State Dependent Routing for multirate loss networks. They reduced the complexity of the problem by modeling each link as a one-dimensional birth/death process and derived a set of expressions to evaluate the state-dependent link shadow prices.

The Broadband Integrated Services Digital Networks currently under active development should meet the stringent requirements of integrated communication, such as bandwidth requirement, bit error rate and delay requirement. To satisfy the huge bandwidth requirement, optical fiber is considered as a promising transmission medium. Wavelength Division Multiplexing(WDM) [34, 35, 41, 43] is a promising technology to tap the huge bandwidth in optical fiber links. However, the existing switching, processing, and storage technologies lag behind the optical transmission capabilities. The switching nodes become bottelneck and limit the effective throughput of the network. This leads to the development of all-optical networks.

The design of new architectures for all-optical networks has received considerable interest in recent years. Bellcore's LAMBDANET [39], which is among the earliest, uses WDM with fixed-tuned optical transmitters. IBM Research has developed a 32-station circuit-switched WDM prototype network called Rainbow [40]. In both systems, a broadcast-and-select transmission is implemented using a single star-coupler. In [41, 42], systems based on fixed wavelength routing were proposed. An $NXN$ wavelength crossconnect is used to achieve full connectivity between the $N$ inputs and $N$ outputs.

Bala and Stern[36, 37] proposed the linear lightwave network which uses the linear divider and combiner(LDC) for link-selective routing. To simplify hardware requirement and reduce the number of optical switches, a set of wavelengths can be grouped into a waveband [38]. The problem of finding the best routing paths however is made more complicated.

In [43], the lightnet architecture based on the lightpath(an all-optical path) concept was proposed. It eliminates processing and buffering at intermediate nodes through the establishment of lightpath between node pairs. A lightpath establishment algorithm was also proposed in [43]. In [44], a switch architecture and a shortest path routing algorithm were proposed for a wavelength convertible optical network.

## 1.4 Organization

In Chapter two, we shall focus on the problem of packing and protection of calls in integrated-service networks. In particular, we shall evaluate and compare five call admission policies under two different adaptive routing rules in a multirate loss network. The first routing rule is based on the Traffic Packing principle, while the second one is based on the Least Congestion principle. The five call admission policies being studied are the Complete Sharing (CS), the Limited Occupancy (LO), the Guaranteed Bandwidth (GB), the External Blocking (EB) and the Direct-link Packing (DP) policies. The reduced load approximation is used to evaluate the blocking probability of different classes of calls.

With the installation of Common Channel Signaling network and the advent in electronic digital switching, it is now feasible to implement sophisticated adaptive routing schemes in integrated-service networks. Adaptive routing can increase the network throughput by routing calls to less congested paths. It can also be used to bypass transmission facility failures. In Chapter three, we shall analyze and compare two versions of Least Congestion routing. The first one is called the *Maximum mean Time to Blocking*(MTB) routing, which is based on the *mean time to blocking* measure of a link. This measure captures the traffic rates, bandwidth characteristic and link capacity information and reflects more accurately the congestion status of different paths. The second one is the $M^2$ routing and is based on the residual bandwidth measure. Aggregation of link status information can significantly reduce signaling traffic. We shall also study the aggregated $M^2$ and MTB routings. The use of complete sharing and limited occupancy policies under MTB routing are also discussed.

In Chapter four, we extend the scope of our study to the Least Congestion routing in WDM lightwave networks. Each switching node in the network may have a number of wavelength converters which can be used to resolve wavelength conflicts in multi-hop paths. We shall analyze the performance of the Least Congestion routing with and without the use of wavelength converters. We shall also study a modified version of the

LC routing. In this routing, priority is not given to shorter path during the alternate path selection process.

Finally, we conclude this study and suggest possible future research in Chapter 5.

## 1.5 Publications

Part of this study have been submitted for publication. In particular, the result of Chapter 3 has been published in IEEE GLOBECOM 1993 [26] and that of Chapter 4 has been accepted for publication in IEEE INFOCOM 1994 [27].

# Chapter 2

# Call Admission in Multirate Loss Networks

## 2.1 Introduction

Networks supporting different services with different traffic characteristics are called multirate loss networks. Maximizing the revenue while satisfying the quality of services requirements has been a very important design objective of the routing and call admission policies. In recent years, the design and analysis of routing rules and call admission policies have received considerable attention.

Previous works that study the congestion control policy on a single transmission facility includes [14, 16, 15]. In [15], two different traffic control policies are analyzed for a blockable wide-band(WB) and a queueable narrow-band(NB) type of traffic. In [16], two different call admission policies, namely, the complete sharing and the threshold-type policy are analyzed by formulating it as the stochastic knapsack problem. In [17], the optimal circuit access policy under fixed routing rule is determined by the Markov decision process approach. In [21], Dziong and Mazon proposed a sub-optimal call admission and routing strategy for multirate circuit switched networks based on maximization of total revenue from the network. In [19], Wang *et al.* have studied various trunk congestion control schemes based on restricted access and preemptive priority for single rate circuit-switched networks.

Both circuit-switching and virtual circuit packet switching can be modeled by multirate loss networks. The latter is the well-known ATM-based ISDN. In such a network, a large number of traffic sources having a broad range of burstiness characteristics are multiplexed. Recent research reported that an effective bandwidth[23, 24], which depends only on the source characteristic and the cell loss probability, can be assigned to a traffic source. This allows circuit-switched type call acceptance and routing scheme to be implemented in ATM network. Thus, our analysis can also be modified for ATM networks by adopting the effective bandwidth measures.

In general, there are two types of routing rules in multirate loss networks. The first one is based on the least congestion(LC) strategy while the second one is based on the traffic packing(TP) strategy. The LC strategy routes a call to the Least Congested Path to avoid blocking. The TP strategy packs a call to the most congested admissible path so as to reserve bandwidths on the less congested paths for calls with larger bandwidth requirements.

The purpose of call admission policy is to avoid the dominance of network capacity by a particular class of calls. Complete sharing(CS) policy imposes no restriction on the admission of calls aside from individual links capacity constraints. It serves as a benchmark where all other policies should measure against. Aside from the CS policy, we identify two other types of admission policies. The first type, called preemptive blocking policies, blocks certain classes of calls in order to leave capacities for other classes of call. The Limited Occupancy(LO), Guaranteed Bandwidth(GB) and the External Blocking(EB) are examples of preemptive blocking type of policies. The second type is called call-rerouting policies. An example is the Direct-link Packing(DP) policy. The DP policy tries to unblock the direct link for a particular class of calls by rerouting on-going calls on the direct link to other paths. We will give a detail description of this five policies in Section 2.3.

In this chapter, we attempt to give a comparative analysis of the CS, LO, GB, EB and DP policies in multirate loss networks under the LC and TP adaptive routing strategies. Section 2.2 and 2.3 present the two routing schemes and the five call admission policies respectively. Section 2.4 contains an analysis of the call admission policies studied under the LC or TP routing rules, and Section 2.5 presents the results of performance comparison. Finally, we conclude this chapter in Section 2.6.

## 2.2 Two Adaptive Routing Rules

In a typical multirate network, let there be $K$ classes of traffic where the class $k$ call has a mean holding time(or service time) $1/\mu_k$, bandwidth requirement $f_k$ and average revenue earning $R_k, k = 1, \ldots, K$.

Define the residual bandwidth $\gamma_k^{(l)}$ for the class $k$ traffic in link $l$ as the remaining free capacity on that link minus the capacity reserved for the other classes of calls. The residual bandwidth for the class $k$ traffic, $\beta_k(q)$, of path $q$ composed of link set $\mathcal{L}(q)$ is defined as the residual bandwidth of the link on path $q$ having the smallest residual bandwidth. In other words,

$$\beta_k(q) = \min_{l \in \mathcal{L}(q)} (\gamma_k^{(l)}) \tag{2.1}$$

To maintain network stability, certain amount of bandwidth, $r_k$, is reserved for direct class $k$ traffic. We call it bandwidth reservation, which is similar to the trunk reservation in circuit-switched networks. A link is **admissible** to class $k$ alternate path traffic if its residual bandwidth for that class of traffic is greater than $r_k + f_k$. A multi-hop path is admissible to alternate path traffic if all the constituting links are admissible.

When a connection request is made, the residual bandwidth of all the links constituting all the admissible paths are first collected. To choose a routing path, the set of admissible paths is divided into different subsets of different hop-count. $X_i$ denotes the set of admissible paths with $i$ hops. The connection request is established on a path selected from the first non-empty path set, starting from $X_1$[1].

The two routing rules studied here are the Least Congestion(LC) rule and the Traffic Packing(TP) rule. They are elaborated as follows.

1. The Least Congestion routing rule

   Under the Least Congestion principle, a call is first attempted on the direct link. If blocked it is overflowed to a path having the maximum amount of residual bandwidth and satisfying the bandwidth reservation requirement. When a path set $X_i$ is considered, the path having the maximum residual bandwidth $\beta$ is selected as the routing path. If there is a tie, the $\gamma$ values of the second most congested link of the paths in the tie set will be compared and the connection will be established on the path with the maximum such value. Further ties are broken in a similar manner.

   To illustrate the LC routing rule, suppose a particular node pair has four admissible alternate paths A, B, C, D and E with residual bandwidth {3,5}, {2,6}, {2,4}, {3,4} and {5,6,7} on its links respectively. The hop-count of the paths are given by the number of elements in the set. The residual bandwidth of these four paths are 3, 2, 2, 3 and 5 respectively. The routing rule will first consider paths A, B, C and D as they have smaller hop-counts. Among these three paths, path A and D have the largest residual bandwidth. Since there is a tie, the residual bandwidth of the second most congested links on paths A and D will be compared. The chosen route is therefore path A.

2. The Traffic Packing routing rule

   Using the Traffic Packing principle, alternate calls are routed onto paths having the least residual bandwidth and satisfying the bandwidth reservation requirement. If a call is blocked on the direct link, a path is selected from alternate path set starting from $X_2$. When a path set $X_i$ is considered, the path having

---

[1] In [27], it was shown that the use of shorter paths can reduce the call blocking probability.

the minimum residual bandwidth $\beta$ is selected. The tie breaking procedure is similar to that of the LC routing rule except that the selection criterion is based on the least residual bandwidth rather than the most residual bandwidth.

## 2.3 Call Admission Policies

Consider a particular link in a multirate loss network with $K$ classes of calls. Let $\mathbf{n} = \{n_1\ n_2\ \ldots\ n_K\}$ be the number of on-going class $k$ calls and $F$ be the bandwidth on that link. Also let $\lambda_i, i = 1, 2, \ldots, K$ be the arrival rate of class $i$ calls. On such a link with $K$ classes of services, five different call admission control policies can be readily identified:

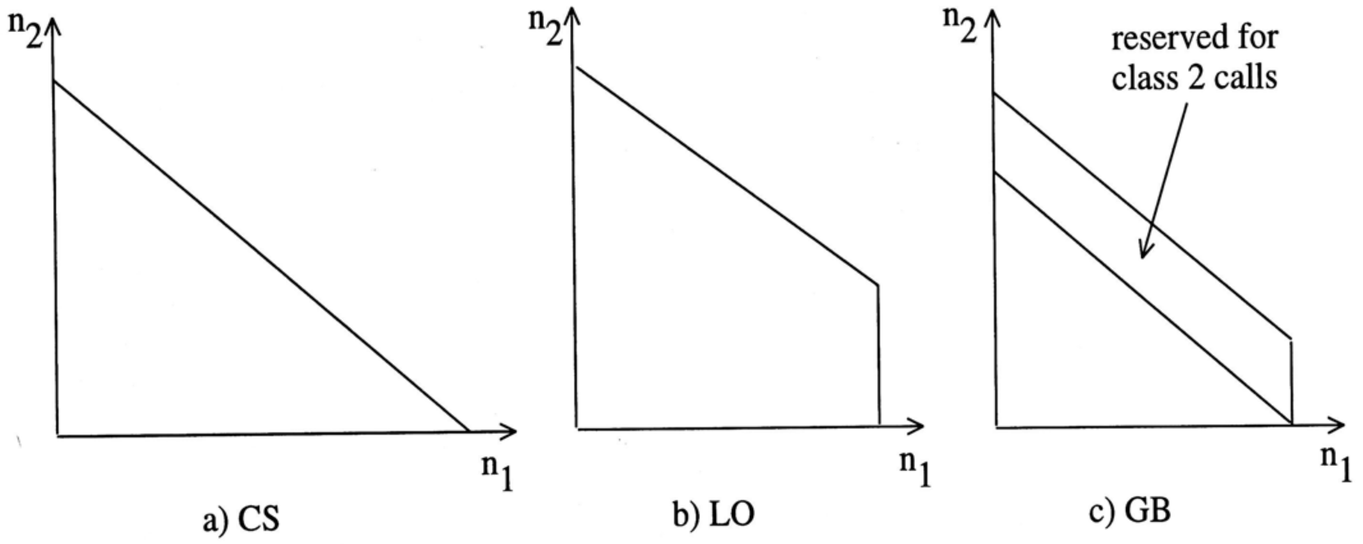


Figure 2.1: The CS, the LO and the GB policies

1. Complete Sharing(CS) policy

   If bandwidth is shared freely among different classes of calls, we have the so called Complete Sharing(CS) policy. To maintain stability of the network, a bandwidth of $r_k$ are reserved for direct class $k$ calls. For the CS policies, the state space $\Omega_{CS}$ is given by

$$\Omega_{CS} = \{\mathbf{n} : \mathbf{n} \cdot \mathbf{f} \leq F\}$$

At state $\mathbf{n}$, the residual bandwidth for class $k$ traffic is

$$\gamma_k(\mathbf{n}) = F - \mathbf{n} \cdot \mathbf{f} \tag{2.2}$$

The set of class $k$ direct and alternate call blocking states, $\Omega_{CS}^{(D)}(k)$ and $\Omega_{CS}^{(A)}(k)$ are given by,

$$\Omega_{CS}^{(D)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, \gamma_k(\mathbf{n}) < f_k\}$$

$$\Omega_{CS}^{(A)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, \gamma_k(\mathbf{n}) < f_k + r_k\}$$

Let $\mathbf{e}_k$ be a $K$-vector with a 1 at the $k$th position and 0 elsewhere. At state $\mathbf{n}$, the global balance equation relating the state probability, $P(\mathbf{n})$, and the state dependent traffic rates, $\Lambda_k(\mathbf{n})$ is given by

$$P(\mathbf{n}) \sum_k (\Lambda_k(\mathbf{n}) + n_k \mu) = \sum_k (n_k + 1)\mu_k P(\mathbf{n} + \mathbf{e}_k) + P(\mathbf{n} - \mathbf{e}_k)\Lambda_k(\mathbf{n} - \mathbf{e}_k) \tag{2.3}$$

with the understanding that $P(\mathbf{n}) = 0$ for $\mathbf{n} \notin \Omega$

The class $k$ direct link and alternate path blocking probability, $D_k$ and $A_k$ are given by

$$D_k = \sum_{\mathbf{n} \in \Omega_{CS}^{(D)}(k)} P(\mathbf{n}) \tag{2.4}$$

$$A_k = 1 - \left[ 1 - \sum_{\mathbf{n} \in \Omega_{CS}^{(A)}(k)} P(\mathbf{n}) \right]^2 \tag{2.5}$$

Using (2.4) and (2.5) and assuming that each node pair has $L$ possible alternate paths, the end-to-end call blocking probability $B_k$ is given in [30] as

$$B_k = D_k A_k^L \tag{2.6}$$

To analyze this policy, we can formulate the problem as a $k$-dimensional Markov Chain[26]. Figure 2.1(a) shows the state space of a link under the CS policy with $K$=2.

2. Limited Occupancy(LO) policy

   This is also known as the threshold type policy in [16] and the restricted access policy in [19]. In this policy, the class $k$ traffic is blocked when the number of on-going class $k$ calls reaches a threshold $h_k$. The state space of a link is given by

   $$\Omega_{LO} = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, n_k \leq h_k, h_k \geq 0\}$$

   At state $\mathbf{n}$, the residual bandwidth for class $k$ calls is given by

   $$\gamma_k(\mathbf{n}) = \min[F - \mathbf{n} \cdot \mathbf{f}, (h_k - n_k) f_k]$$

   The class $k$ direct and alternate call blocking states is given by

   $$\Omega_{LO}^{(D)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{LO}, \gamma_k(\mathbf{n}) < f_k\}$$

   $$\Omega_{LO}^{(A)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{LO}, \gamma_k(\mathbf{n}) < f_k + r_k\}$$

   The global balance equation at state $\mathbf{n} \in \Omega_{LO}$ is given by (2.3). The $D_k$ and $A_k$ are given by (2.4) and (2.5) with $\Omega_{CS}^{(D)}(k)$ and $\Omega_{CS}^{(A)}(k)$ replaced by $\Omega_{LO}^{(D)}(k)$ and $\Omega_{LO}^{(A)}(k)$. The call blocking probability is given by (2.6).

   Figure 2.1(b) shows the state space of a link under the LO policy. In this example, the maximum number of class 1 calls in a link is limited.

3. Guaranteed Bandwidth(GB) policy

   Under GB policy, an amount of bandwidth $c_k$ is reserved for class $k$ traffic while the rest are shared by traffic of all classes. Figure 2.1(c) shows the state space of a link under a kind of GB policy. In this example, certain bandwidth is reserved for class 2 calls only and none are reserved for class 1 calls.

   GB policy is usually employed when the blocking probability of a particular class of service does not meet the specified QOS requirement. To simplify the evaluation of this policy, we restrict our study to the case where only one class of calls has the guaranteed bandwidth. Without loss of generality, let that be the class $K$ traffic. Hence $c_k = 0 \; for \; k = 1, \ldots, K-1$.

   Excluding the bandwidth reserved for class $K$ calls, the maximum bandwidth available for the rest of class $k(1 \leq k < K)$ calls is $F - c_K$. Therefore the maximum allowable number of class $k(1 \leq k < K)$ on a link, defined as $h_k$ is, $h_k = (F - c_k)/f_k$ and for class $K$, $h_K = F/f_K$. With these, the state space of a link is given by

$$\Omega_{GB} = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, n_k \leq h_k, k = 1, \ldots, K\}$$

   The residual bandwidth for class $k$ traffic at state $\mathbf{n}$ is given by

$$\gamma_k(\mathbf{n}) = \begin{cases} F - \mathbf{n} \cdot \mathbf{f} - c_K & 1 \leq k < K \\ F - \mathbf{n} \cdot \mathbf{f} & k = K \end{cases}$$

   And the class $k$ direct and alternate path blocking states are given by

$$\Omega_{GB}^{(D)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{GB}, \gamma_k(\mathbf{n}) < f_k\}$$

$$\Omega_{GB}^{(A)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{GB}, \gamma_k(\mathbf{n}) < f_k + r_k\}$$

The global balance equation at state $\mathbf{n} \in \Omega_{GB}$ is given by (2.3). The $D_k$ and $A_k$ are given by (2.4) and (2.5) with $\Omega_{CS}^{(D)}(k)$ and $\Omega_{CS}^{(A)}(k)$ replaced by $\Omega_{GB}^{(D)}(k)$ and $\Omega_{GB}^{(A)}(k)$. The end-to-end call blocking probability is given by (2.6).

4. External Blocking(EB) policy

   In this policy, when a class $k$ call is offered to a link, it has a probability $\theta_k$ of being rejected. This concept is similar to the external blocking studied in [28] for overflow traffic control. When a call is rejected from the direct call, it can attempt alternate paths. If no alternate path is available, the call will be blocked and cleared. By manipulating the set of $\theta_k$'s, the relative blocking probabilities of different class of calls can also be manipulated. The state space, direct and alternate call blocking states of the EB policy are the same as that of the CS policy. The residual bandwidth of a link at state $\mathbf{n}$ is given by (2.2). The global balance equation at state $\mathbf{n}$ is given by (2.3).

   Under the EB policy, at the class $k$ blocking states, all class $k$ calls are blocked. But at other states, it also has a probability $\theta_k$ of being rejected. Therefore the direct and alternate path blocking probability is given by

$$D_k = \sum_{\mathbf{n} \in \Omega_{CS}^{(D)}(k)} P(\mathbf{n}) + \sum_{\mathbf{n} \in \Omega_{CS} \backslash \Omega_{CS}^{(D)}(k)} P(\mathbf{n})\theta_k \tag{2.7}$$

$$A_k = 1 - \left[ 1 - \sum_{\mathbf{n} \in \Omega_{CS}^{(A)}(k)} P(\mathbf{n}) - \sum_{\mathbf{n} \in \Omega_{CS} \backslash \Omega_{CS}^{(A)}(k)} P(\mathbf{n})\theta_k \right]^2 \tag{2.8}$$

   The end-to-end call blocking probability is given by (2.6).

   We will analyze the CS, LO, GB and the EB policies in Section 2.4.1.

5. Direct-link Packing(DP) policy

   In this policy, when a call with a large bandwidth requirement is blocked on the direct link, a number of on-going calls on the direct link are rerouted to alternate paths so as to empty up enough capacity in the direct link for that large bandwidth call. A Sophisticated Direct-link Packing(SDP) policy is described as follows.

When a class $j$ call is blocked on the direct link, the overflowed calls on the direct link is first rerouted to the alternate paths in order to unblock the link. If the direct link is still blocked, then direct calls are rerouted. In order to reduce the number of rerouting, the routing rule will first reroute the class of calls with the largest $f_k$ smaller than $f_j$. If more than one classes are found, then the one with the smallest *bandwidth* × *mean holding time* product are selected. If the direct link is still blocked, then the call will attempt alternate paths.

It may be a heavy burden for a switching node to reroute too many calls. To reduce the loading, a call will be rejected on the direct link if the number of rerouting needed exceeds certain number. The call will then attempt alternate paths.

The SDP policy is analytically unmanageable. We choose to analyze a simple but effective DP policy as follows. Without loss of generality, assume $f_1 \leq f_2, \ldots, \leq f_K$. If a class $k$ connection request is made and the direct link has a residual bandwidth $i < f_k$, a number of direct class 1 calls on direct link will be rerouted to the alternate paths based on either the LC or the TP routing rules. Since only direct calls is rerouted, signaling traffic can be significantly reduced.

There are many ways to select alternate paths to carry rerouted calls. To simplify the analysis, we assume that each alternate path accepts one rerouted call at a time. The number of direct class 1 calls need to be rerouted(or the minimum number of admissible alternate paths required) is $u_k(i) = \lceil (f_k - i)/f_1 \rceil$. A class $k(k > 1)$ call is blocked if i) a link has less than $f_k$ residual bandwidth and there is not enough admissible alternate paths or ii) there is not enough direct class 1 calls for rerouting. Let $\tau_k(i), 0 \leq i < f_k$ be the probability that *rerouting fails* given that the blocked direct link has residual bandwidth $i$ and it has enough direct class 1 calls for rerouting. $\tau_k(i)$ is given by

$$\tau_k(i) = \begin{cases} \sum_{i=0}^{u_k(i)-1}(1-A_1)^i(A_1)^{L-i} & u_k(i) \leq L \\ 1 & u_k(i) > L \end{cases}$$

For the DP policy, the state, $\mathbf{n}$ is modified to $\mathbf{n} = \{n_1\ n_2\ \ldots\ n_K\ n_{K+1}\}$, where $n_k, 2 \leq k \leq K$ is the number of on-going class $k$ calls, $n_1$ is the number of direct class 1 calls and $n_{K+1}$ is the number of on-going overflowed class 1 calls on a link. The state space, $\Omega_{DP}$ is given by

$$\Omega_{DP} = \{\mathbf{n} : \mathbf{n} \cdot \mathbf{f}' \leq F\}$$

where $\mathbf{f}' = \{f_1\ f_2\ \ldots\ f_K\ f_1\}$.

The residual bandwidth at state $\mathbf{n}$ for class $k$ traffic is

$$\gamma_k(\mathbf{n}) = F - \mathbf{n} \cdot \mathbf{f}' \qquad \mathbf{n} \in \Omega_{DP}^{(D)}$$

The class $k$ direct call blocking states is given by

$$\Omega_{DP}^{(D)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{DP}, \gamma_k(\mathbf{n}) < f_k\}$$

Since there is no alternate routing for class $k(k > 1)$ calls, the set of class $k$ alternate call blocking states are empty. The class 1 alternate calls blocking states is given by

$$\Omega_{DP}^{(A)}(1) = \{\mathbf{n} : \mathbf{n} \in \Omega_{DP}, \gamma_k(\mathbf{n}) < f_1 + r_1\}$$

The class 1 direct link, alternate path and call blocking probability is given by (2.4), (2.5) and (2.6) with $\Omega_{CS}^{(D)}$ and $\Omega_{CS}^{(A)}(1)$ replaced by $\Omega_{DP}^{(D)}(1)$ and $\Omega_{DP}^{(A)}(1)$.

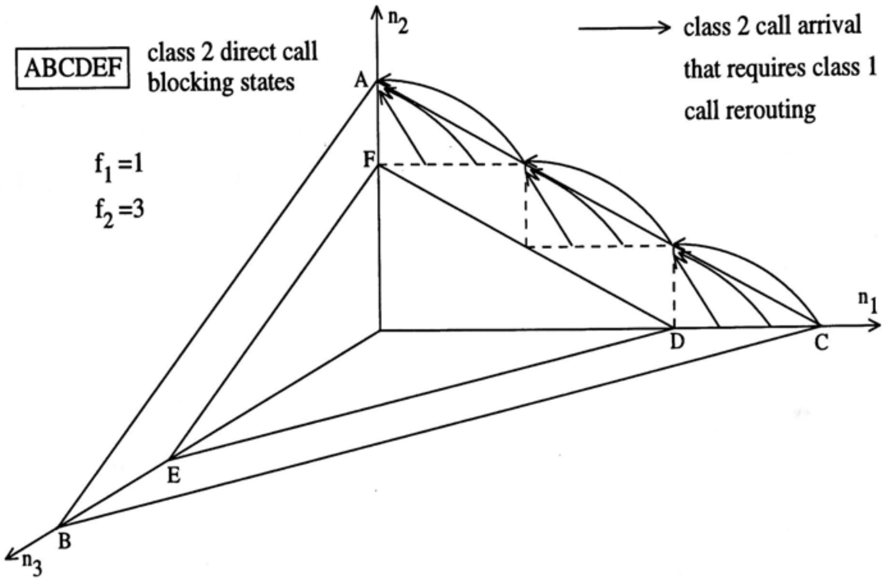


Figure 2.2: The state space of DP policy

Figure 2.2 shows the state space of a DP policy with two classes of calls. Direct class 1 calls are rerouted if the direct link does not have enough bandwidth for class 2 calls.

Let $J_k(\mathbf{n})$ be 1 if $\mathbf{n} \in \Omega_{DP}^{(D)}(k)$ and $n_1 \geq \lceil f_k - \gamma_1(\mathbf{n}) \rceil / f_1$, and 0 otherwise. The global balance equation is modified as follows,

$$
\begin{aligned}
& P(\mathbf{n}) \left\{ \sum_{k=1}^{K+1} [\Lambda_k(\mathbf{n}) + n_k \mu_k] + J_k(\mathbf{n}) \lambda_k [1 - \tau_k(\gamma_1(\mathbf{n}))] \right\} = \\
& \sum_{k=1}^{K+1} (n_k + 1) \mu_k P(\mathbf{n} + \mathbf{e}_k) + P(\mathbf{n} - \mathbf{e}_k) \Lambda_k(\mathbf{n} - \mathbf{e}_k) \\
& \sum_{k=2}^{K} \sum_{i=1}^{f_k} J_k(\mathbf{n} - \mathbf{e}_k + i\mathbf{e}_1) P(\mathbf{n} - \mathbf{e}_k + i\mathbf{e}_1) \cdot \lambda_k [1 - \tau_k(\gamma_1(\mathbf{n} - \mathbf{e}_k + i\mathbf{e}_1))] \quad (2.9)
\end{aligned}
$$

where $\mu_{K+1} = \mu_1$ for simplifying the presentation.

For class $k (k > 1)$ calls, the call blocking probability is given by,

$$
\begin{aligned}
B_k &= Prob\{Direct\ link\ blocked\ and\ Rerouting\ failed\} \\
&= \sum_{\substack{\mathbf{n} \in \Omega_{DP}^{(D)}(k) \\ n_1 < \lceil f_k - \gamma_1(\mathbf{n}) \rceil / f_1}} P(\mathbf{n}) + \sum_{\substack{\mathbf{n} \in \Omega_{DP}^{(D)}(k) \\ n_1 \geq \lceil f_k - \gamma_1(\mathbf{n}) \rceil / f_1}} P(\mathbf{n}) \tau_k(\gamma_1(\mathbf{n}))
\end{aligned}
$$

We will analyze the DP policy in Section 2.4.2.

## 2.4 Analysis of Call Admission Policies

In this Section, we analyze different call admission policies in a fully connected network with uniform traffic. The analysis can be extended to cover arbitrary network topology and asymmetric traffic rates. The issues of the generalization of state dependent routing(base on residual capacity) in asymmetric single rate networks are addressed in [29].

Assume that links are independent and overflowed traffics are independent Poisson processes. Our analysis is based on the fixed point iteration method and gives the numerical solution of $P(\mathbf{n})$ and $\Lambda_k(\mathbf{n})$.

Solving the equilibrium state probabilities, the class $k$ call blocking probabilities $B_k$ and the fractional revenue loss $\Psi$ can be computed. Assuming that each call generates an expected revenue of $R_k = f_k/\mu_k$, the fractional revenue loss is

$$\Psi = \sum_{k=1}^{K} B_k R_k \lambda_k / \sum_{k=1}^{K} R_k \lambda_k$$

### 2.4.1 The CS, LO, GB and the EB Policies

In an $N$ nodes fully connected network, each node pair has a direct link and $L = N-2$ alternate paths. We first derive the state dependent traffic rates under the LC routing principle.

Two alternate path ABC and ADC of a link AC is shown in Figure 2.3. The overflowed class $k$ calls of rate $\lambda_k[D_k + (1-D_k)\theta_k]$ will be routed to one of the paths according to the least congestion criterion. Suppose there are $\alpha$ such paths. The overflowed class $k$ calls of AC will be routed to one of these path at rate $\lambda_k[D_k + (1-D_k)\theta_k]/\alpha$. Let this path be a Least-congested Admissible Path(LAP).

Let $\omega_k(i)$ be the set of link states with residual bandwidth $i$ for class $k$ traffic and

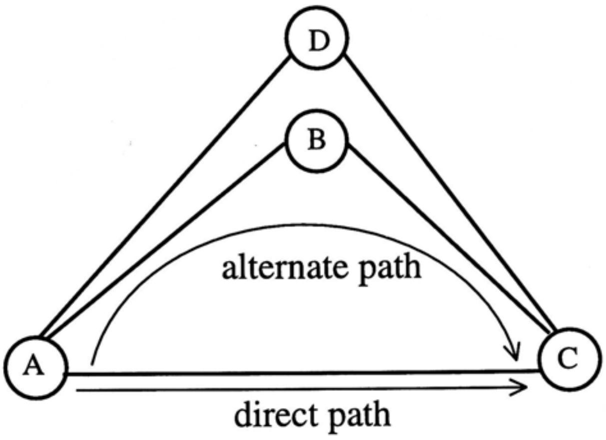


Figure 2.3: An alternate path of a node pair

$\pi_k(i)$ be the probability that *a link has an amount of residual bandwidth i for class k traffic and that the link is admissible to class k alternate traffic(or class k admissible)*

$$\pi_k(i) = (1-\theta_k) \sum_{\mathbf{n}\in\omega_k(i)\setminus\Omega^{(A)}(k)} P(\mathbf{n}) \tag{2.10}$$

Let $b_k(u,v)$ be the probability that *the two links of an alternate path have residual bandwidth u and v where $u \geq v$ for class k traffic* and that *the path is class k admissible.* As a path is class $k$ admissible if and only if the two links constituting it are both class $k$ admissible, by (2.10) we have

$$b_k(u,v) = \begin{cases} \pi_k(u)^2 & u = v \\ 2\pi_k(u)\pi_k(v) & u > v \end{cases} \tag{2.11}$$

Suppose the two links of path ADC have residual bandwidth $i$ and $j$ with $i \geq j$. Given that the two links of path ABC has residual bandwidth $x$ and $y(\leq x)$, define the two disjoint events $E_1, E_2$ as

$E_1$: $j < y$

$E_2$: $j = y$ *and* $i < x$

Let $V_k(x,y)$ be the probability that *path ABC is less congested than path ADC for class k traffic under the least congestion criterion.* This is just the probability of either

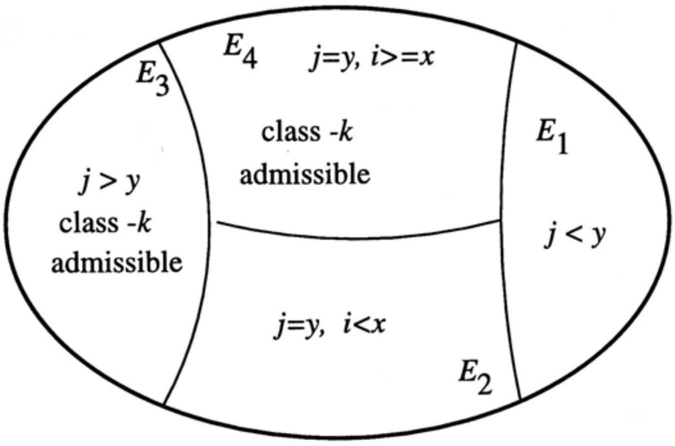


Figure 2.4: Different events of a two-hop path under LCP Routing

one of the two events $E_i$'s, $i = 1, 2$. Since these events are disjoint, from Figure 2.4, $V_k(x, y)$ is given by

$$\begin{aligned} V_k(x,y) &= Prob[E_1] + Prob[E_2] \\ &= 1 - Prob[E_3] - Prob[E_4] \\ &= 1 - \sum_{j=y+1}^{F} \sum_{i=j}^{F} b_k(i,j) - \sum_{i=x}^{F} b_k(i,y) \end{aligned} \tag{2.12}$$

Suppose AC is full, the two links of path ABC are class $k$ admissible and have residual bandwidth $x$ and $y$ for class $k$ traffic, the probability that *ABC is a LAP for class k traffic and there are* $\alpha - 1$ *other such paths*, $f_k(\alpha \mid x, y)$ is obtained by using (2.11) and (2.12)

$$f_k(\alpha \mid x,y) = \begin{pmatrix} L-1 \\ \alpha-1 \end{pmatrix} b_k(x,y)^{\alpha-1} V_k(x,y)^{L-\alpha}$$

Therefore, the overflowed class $k$ traffic rate that get routed from path AC to path ABC, denoted by $s_k(x, y)$, is given by

$$\begin{aligned} s_k(x,y) &= \sum_{\alpha=1}^{L} \frac{\lambda_k[D_k + (1-D_k)\theta_k]}{\alpha} f_k(\alpha \mid x,y) \\ &= \lambda_k[D_k + (1-D_k)\theta_k] \frac{[b_k(x,y) + V_k(x,y)]^L - V_k(x,y)^L}{L b_k(x,y)} \qquad (y > r_k) \end{aligned}$$

Given that link AB is in state $\mathbf{n} \in \Omega \setminus \Omega^{(A)}(k)$, the total overflowed class $k$ traffic, $A_k(\mathbf{n})$, obtained by removing all the conditioning is

$$A_k(\mathbf{n}) = 2L(1-\theta_k)^2 \sum_{i=r_k+f_k}^{F} s_k(\max[\gamma_k(\mathbf{n}), i], \min[\gamma_k(\mathbf{n}), i])\pi_k(i) \tag{2.13}$$

Next, we compute the state dependent overflow traffic rates under the TP routing principle. When a class $k$ call is blocked on AC, a path having the least residual bandwidth and is class $k$ admissible is chosen as the routing path. Let this path be the Most-congested Admissible Path(MAP).

Given that AC is blocked and ABC has residual bandwidth $i$ and $j(\leq i)$ on its two links for class $k$ traffic, the probability that *ABC is the MAP* is

$$X_k(i,j) = \sum_{i=0}^{L-1} \frac{1}{i+1} \binom{L-1}{i} [b_k(i,j)]^i [Z_k(i,j)]^{L-i-1}$$

where $Z_k(i,j)$ is the probability *a path is less congested than path ABC or is reserved for class-k calls.* Refer to Figure 2.5, $Z_k(i,j)$ is given by,

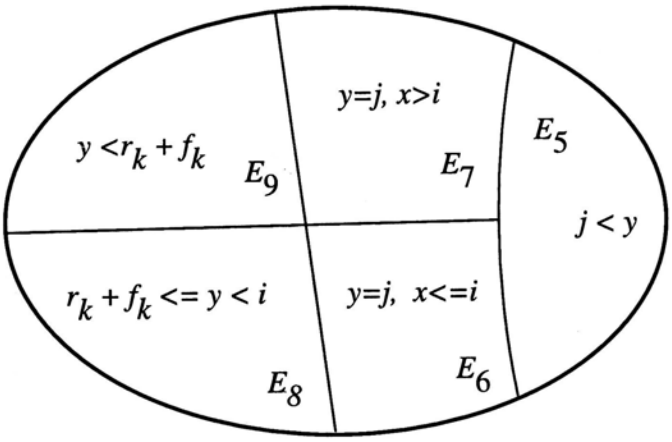


Figure 2.5: Different events of an alternate path under TP routing

$$\begin{aligned} Z_k(i,j) &= E_5 + E_7 + E_9 \\ &= 1 - E_6 - E_8 \\ &= 1 - \sum_{y=r_k+f_k}^{j-1} \sum_{x=y}^{F} b_k(x,y) - \sum_{x=j}^{i} b_k(x,j) \end{aligned}$$

At state $\mathbf{n} \in \Omega \setminus \Omega^{(A)}(k)$, the total class $k$ overflow rate is given by

$$A_k(\mathbf{n}) = 2L\lambda_k[D_k + (1 - D_k)\theta_k](1 - \theta_k)^2 \sum_{i=r_k+f_k}^{F} X_k(\max[\gamma_k(\mathbf{n}), i], \min[\gamma_k(\mathbf{n}), i])\pi_k(i)$$

At state $\mathbf{n}$, the class $k$ call arrival rate including direct and overflowed traffic is

$$\Lambda_k(\mathbf{n}) = \begin{cases} (1 - \theta_k)\lambda_k + A_k(\mathbf{n}) & \mathbf{n} \in \Omega \setminus \Omega^{(A)}(k) \\ (1 - \theta_k)\lambda_k & \mathbf{n} \in \Omega^{(A)}(k) \setminus \Omega^{(D)}(k) \\ 0 & \mathbf{n} \in \Omega^{(D)}(k) \end{cases} \tag{2.14}$$

Let $\Lambda$ denotes the set of $\Lambda_k(\mathbf{n})$ and $\mathcal{P}$ denotes the set of $P(\mathbf{n})$. Then (2.14), (2.3) can be expressed in the fixed point model form[26]: $\mathcal{P} = X(\Lambda)$ and $\Lambda = Y(\mathcal{P})$, where $X$ is a function defined by (2.3), and $Y$ is a function given by (2.14). The $P(\mathbf{n})$'s can be computed by the Successive Over-Relaxation (SOR) method with the set of alternate traffic rates obtained from (2.13) in each iteration. In the examples quoted in Section 2.5, all $P(\mathbf{n})$s are obtained with relative error less than $10^{-6}$.

### 2.4.2 The DP Policy

To simplify the presentation, we assume that $f_1 = 1$. Class 1 overflow traffic rate is composed of the overflow traffic rate due to the alternate routing and the DP policy. The overflow traffic rate due to the alternate routing policy is derived in Section 2.4.1. In this subsection, we will derive the overflow traffic rate due to the DP policy.

First, we will derive the overflow traffic rates under the LC routing rule. Suppose a call is blocked on the direct link, let $LAP(l)$ be the set of $l$ alternate paths selected to carry rerouted class 1 calls under the LC routing principle. Also let $\zeta_k(i)$ be the probability that *AC has residual bandwidth $i$ and has more than $f_k - i - 1$ on-going direct class 1 calls.*

$$\zeta_k(i) = \sum_{\substack{\mathbf{n} \in \omega_1(i) \ \Omega^{(A)}(k) \\ n_1 > f_k - i - 1}} P(\mathbf{n})$$

If AC has less than $f_k$ residual bandwidth, and that path ABC is class 1 admissible and has $\beta > r_1$ residual bandwidth, the overflow traffic rate $Y_k^{(1)}(\beta)$, from node pair AC to path ABC due to the DP policy is given by

$$\begin{aligned} Y_k^{(1)}(\beta) &= \lambda_k \sum_{l=1}^{\min(f_k,L)} Prob\{ABC \ \in \ LAP(l)\} \cdot \\ &\quad Prob\{AC \ has \ f_k - l \ residual \ bandwidth \ and \ has \ \geq \ l \ class \ 1 \ calls\} \end{aligned}$$

The probability that AC has $f_k - l$ residual bandwidth and has more than $l-1$ class 1 calls is given by $\zeta_k(f_k - l)$. Let $\mathcal{Q}_k(\beta, l)$ be the probability that $ABC \in LAP(l)$, i.e.

$$\mathcal{Q}_k(\beta, l) = Prob\{path \ ABC \ is \ among \ the \ l \ least \ cogested \ admissible \ paths\} \quad (2.15)$$

Let $E_5$ be the event that *among the* $L-1$ *alternate paths,* $i$ *of which are less congested than* $ABC$ *and* $j$ *of which are equally congested to* $ABC$ and $E_6$ be the event that *among the* $L-1$ *alternate paths,* $i$ *of which are less congested than* $ABC$*,* $j$ *of which are equally congested to* $ABC$ *and* $m$ *of which are more congested than* $ABC$ *but are class 1 admissible*, (2.15) can be written as

$$\begin{aligned} \mathcal{Q}_k(\beta, l) &= \sum_{i=0}^{l-1} \left\{ \sum_{j=0}^{l-i-2} \sum_{m=l-i-j-1}^{L-i-j-1} Prob\{ABC \ is \ chosen \ |E_6\} Prob\{E_6\} + \right. \\ &\qquad \left. \sum_{j=l-i-1}^{L-i-1} Prob\{ABC \ is \ chosen \ |E_5\} Prob\{E_5\} \right\} \\ &= \sum_{i=0}^{l-1} \left\{ \sum_{j=0}^{l-i-2} \sum_{m=l-i-j-1}^{L-i-j-1} Prob\{E_6\} + \sum_{j=l-i-1}^{L-i-1} Prob\{E_5\} \right\} \\ &= \sum_{i=0}^{l-1} \left\{ \sum_{j=0}^{l-i-2} \sum_{m=l-i-j-1}^{L-i-j-1} \begin{pmatrix} L-1 \\ i \ j \ m \end{pmatrix} [Q_1(\beta)]^i [U_1(\beta)]^j [W_1(\beta)]^m [W(\beta) - W_1(\beta)]^{L-i-j-} \right. \\ &\qquad \left. + \sum_{j=l-i-1}^{L-i-1} \frac{l-i}{j+1} \begin{pmatrix} L-1 \\ i \ j \end{pmatrix} [Q_1(\beta)]^i [U_1(\beta)]^j [W(\beta)]^{L-i-j-1} \right\} \qquad (2 \end{aligned}$$

where $Q_1(x)$, $U_1(x)$ and $W_1(x)$ are the probability that an alternate path has a residual bandwidth more than, equal to and less than $x$ and is class 1 admissible and $W(x)$ is the probability that an alternate path has a residual bandwidth less than $x$. They are defined by

$$U_1(x) = \sum_{i=x}^{F} b_1(i,x)$$

$$Q_1(x) = \begin{cases} \sum_{i=x+1}^{F} U_1(i) & x < F \\ 0 & x = F \end{cases}$$

$$W_1(x) = \begin{cases} \sum_{i=0}^{x-1} U_1(i) & x > 0 \\ 0 & x = 0 \end{cases}$$

$$W(x) = 1 - \sum_{i=x}^{F} \sigma(i)$$

At state $\mathbf{n} \in \Omega_{DP} \setminus \Omega_{DP}^{(D)}(1)$, the total class 1 overflow traffic rate due to DP is

$$A_1'(\mathbf{n}) = 2L \sum_{k=2}^{K} \sum_{i=r_1+f_1}^{F} Y_k^{(1)}(\min[\gamma_1(\mathbf{n}), i])\pi_k(i)$$

Next, we will compute the overflowed traffic rates due to the DP policy under the TP routing rule. Suppose a call is blocked on the direct link, let $MAP(l)$ be the set of $l$ alternate paths selected to carry the retouted class 1 calls under the TP routing principle.

Given that AC has less than $f_k$ residual bandwidth, and that path ABC is class 1 admissible and has $\beta > r_1$ residual bandwidth, the overflow traffic rate, $Y_k^{(2)}(\beta)$ from node pair AC to path ABC is given by

$$Y_k^{(2)}(\beta) \;=\; \lambda_k \sum_{l=1}^{f_k} Prob\{ABC \;\in\; MAP(l)\} \cdot \zeta_k(f_k - l)$$

Using a similar method in deriving (2.16), we can obtain $Prob\{ABC \in MAP(l)\}$ as

$$\sum_{i=0}^{l-1}\left\{\sum_{j=0}^{l-i-2}\sum_{k=l-i-j-1}^{L-i-j-1}\begin{pmatrix} L-1 \\ i\ j\ k \end{pmatrix}[W_1(\beta)]^i[U_1(\beta)]^j[Q_1(\beta)]^k[W(\beta)-W_1(\beta)]^{L-i-j-k-1}\right.$$
$$\left.+\sum_{j=l-i-1}^{L-i-1}\frac{l-i}{j+1}\begin{pmatrix} L-1 \\ i\ j \end{pmatrix}[W_1(\beta)]^i[U_1(\beta)]^j[Q_1(\beta)+W(\beta)-W_1(\beta)]^{L-i-j-1}\right\}$$

At state $\mathbf{n} \in \Omega_{DP} \setminus \Omega_{DP}^{(D)}(1)$, the total class 1 overflow traffic rate due to the DP policy is

$$A_1'(\mathbf{n}) = 2L\sum_{k=2}^{K}\sum_{i=r_1+f_1}^{F} Y_k^{(2)}(\min[\gamma_1(\mathbf{n}), i])\pi_k(i)$$

At state $\mathbf{n}$, the class 1 call arrival rate including direct and overflowed traffic is

$$\Lambda_1(\mathbf{n}) = \begin{cases} \lambda_1 + A_1(\mathbf{n}) + A_1'(\mathbf{n}) & \mathbf{n} \in \Omega_{DP} \setminus \Omega_{DP}^{(A)}(1) \\ \lambda_1 & \mathbf{n} \in \Omega_{DP}^{(A)}(1) \setminus \Omega_{DP}^{(D)}(1) \\ 0 & \mathbf{n} \in \Omega_{DP}^{(D)}(1) \end{cases} \qquad (2.17)$$

and that for class $k(>1)$ call is

$$\Lambda_k(\mathbf{n}) = \begin{cases} \lambda_k & \mathbf{n} \in \Omega_{DP} \setminus \Omega_{DP}^{(D)}(k) \\ \lambda_k[1-\tau_k(\gamma_1(\mathbf{n}))] & \mathbf{n} \in \Omega_{DP}^{(D)}(k), n_1 \geq f_k - \gamma_1(\mathbf{n}) \\ 0 & \mathbf{n} \in \Omega_{DP}^{(D)}(k), n_1 < f_k - \gamma_1(\mathbf{n}) \end{cases} \qquad (2.18)$$

## 2.5 Performance Comparisons

To evaluate the performance of various call admission policies and routing rules, we first consider a 12-node fully-connected network with all links having a bandwidth of twenty-four units. Let there be two classes of service. Class 1 requires one unit of bandwidth and has a service time exponentially distributed with mean 1. Class 2 requires six units of bandwidth and has a service time exponentially distributed with mean 10.

Figure 2.6(a) shows the performance of the CS policy with LC routing as a function of the bandwidth reservation parameters of class 1 and class 2 calls. The traffic rates of class 1 and class 2 calls are 12 and 0.2. We find that the revenue loss decreases with $r_1$ and $r_2$ and then increases. We also find that for minimum revenue loss, $r_1$ should be around 6 and $r_2$ should be at least 12. This shows that for optimal performance, the overflow traffic of each class should be regulated differently.

Figure 2.6(b) shows that $r_1$ and $r_2$ can affect the relative blocking performance of the two classes of traffic. But their optimal choice can only be made when the blocking are translated into revenue loss as shown in the last figure. As will be shown later, the relative blocking can be manipulated to a larger extend by the choice of call admission policies as compared to that by manipulating the $r_1$, $r_2$ parameters.

Define $\rho_k = \lambda_k f_k / \mu_k$ as the load of class-$k$ traffic and $\rho = \sum_k \rho_k$ as the total load to the network. Figure 2.7(a) shows the performance of TP and LC routing under CS policy for network load ranges from 11 to 15 in a network of size $N = 8$ and the use of optimal bandwidth reservation parameters (chosen to optimize the revenue loss). Consider two classes of traffic with $\rho_1 = \rho_2$. Curve a shows the fractional revenue loss when both classes use the LC routing and curve d shows that when both classes use TP routing. We find that the LC routing rule always performs better than the TP routing rule under a diverse range of traffic load.

Figure 2.7(b) shows the performance of LC and TP routings under different traffic loading ratios with network load kept at 14. We find again that LC routing gives lower revenue loss than TP routing under all loading ratios. Therefore, for simplicity, we shall restrict our study to that of the LC routing rule in the following performance comparisons.

Figure 2.8 shows the performance of the LO policy as a function of $H_1$, the maximum number of class 1 calls allowed, in a 12-node network. The base traffic rate is 7.5 for class 1 traffic and 0.125 for class 2 traffic. In (a), we find that class 2 blocking increases slightly when $H_1$ decreases from 24 to 14, which is counter intuitive. When $r_1$ is increased to 12 as in (b), such phenomenon is not observed. Hence, the LO policy cannot work properly if the bandwidth reservation parameters is not set appropriately.

Figure 2.9 shows the performance of the GB policy as a function of $c_2$, the bandwidth reserved for class 2 calls, under the same conditions in Figure 2.8. We observe a similar behavior to that of the LO policy, ie, the relative blocking of the two classes of traffic, $B_1$ and $B_2$, can be manipulated at will by varying $c_2$ provided that $(r_1, r_2)$ are appropriately chosen. Since both the LO and GB policies try to limit the number of class 1 calls, it is intuitive that they have similar performance.

Figure 2.10 shows the performance of the EB policy as a function of $\theta_1$, the class 1 external blocking parameter, under the same conditions in Figure 2.8. We find that when $r_1$ is set at 4, $B_1$, $B_2$ and $\Psi$ all increase with $\theta_1$ as shown in (a). This is due to the increase in class 1 overflow traffic. If $r_1$ is increased to 12, $B_2$ can again be traded with $B_1$ and $\Psi$ by varying $\theta_1$ as shown in (b). Collectively, Figure 2.8, 2.9 and 2.10 show that the use of LO, GB and EB policies can all reduce the blocking probability of a particular class of calls provided that the bandwidth reservation parameters are suitably selected. Even then, the revenue loss is increased in all cases.

Figure 2.11 shows the performance of the DP and CS policies as a function of network load in a 12-node network with $\rho_1$ equals to $\rho_2$. The bandwidth reservation parameters is chosen to minimize the revenue loss. We find that by rerouting class 1 calls to alternate paths, DP policy yields a smaller blocking as well as revenue loss than the CS policy. At $\Psi = 0.05$, the DP policy can increase the maximum load of the network from 15.3 to 19.3(26% increase) when compared to the CS policy. Figure 2.12 shows the increase of the expected number of rerouting as a result of load increase using the DP policy. This increase of rerouting represents the increase of processing load needed for the use of this policy and is therefore a tradeoff of the revenue and performance gain shown in the last figure.

Figure 2.13 shows the performance of the DP and CS policies for different traffic composition in the same network with $\rho_1 = \rho_2 = 7.5$. The DP policy is shown to perform uniformly better than the CS policy under a wide range of loading ratios.

To compare the SDP with the CS policy, and also to study a multirate network with more than two classes of services, we built a detailed simulation model. We simulate the SDP policy in a 12 node fully connected network with four classes of traffic having bandwidth requirement 1, 2, 4 and 8, and service rate 1, 0.5, 0.3, 0.2. Their load are equal. The network will be simulated for 10000 seconds and results are divided into ten batchs. The first batch is discarded.

Figure 2.14 shows the performance of CS and SDP policies with maximum number of rerouting equals 2 and 8. We find that the SDP(2) and SDP(8) policies can increase the network capacity by 15% and 50% at 2% fractional revenue loss. Comparing Figure 2.15, we find that the blocking probability of calls of large bandwidth consumption(such as class 4) is reduced significantly.

Figure 2.16 shows the average rerouting performed when a class 2, 3 and 4 calls arrive under the SDP(2) and SDP(8) respectively. It is found that by reducing the maximum number of rerouting from 8 to 2, the processing load incurred by the SDP policy (measured by the average number of rerouting) can be reduced significantly. But then the performance also degrades.

## 2.6 Concluding Remarks

In this chapter, we have analyzed five different call admission policies, namely the LO, GB, EB and DP policies and studied them the LC and TP routing rules. We find that the routing rule bases on traffic packing cannot reduce the blocking probability and the revenue loss of the network. We also show that the call admission policies based on the preemptive blocking, such as the LO, GB and EB policies, can protect a particular class of calls if the bandwidth reservation parameters are chosen appropriately. But at the same time, the revenue loss is also increased. The DP policy, on the other hand, gives a significantly smaller blocking probability and revenue loss when compared to the CS policy.

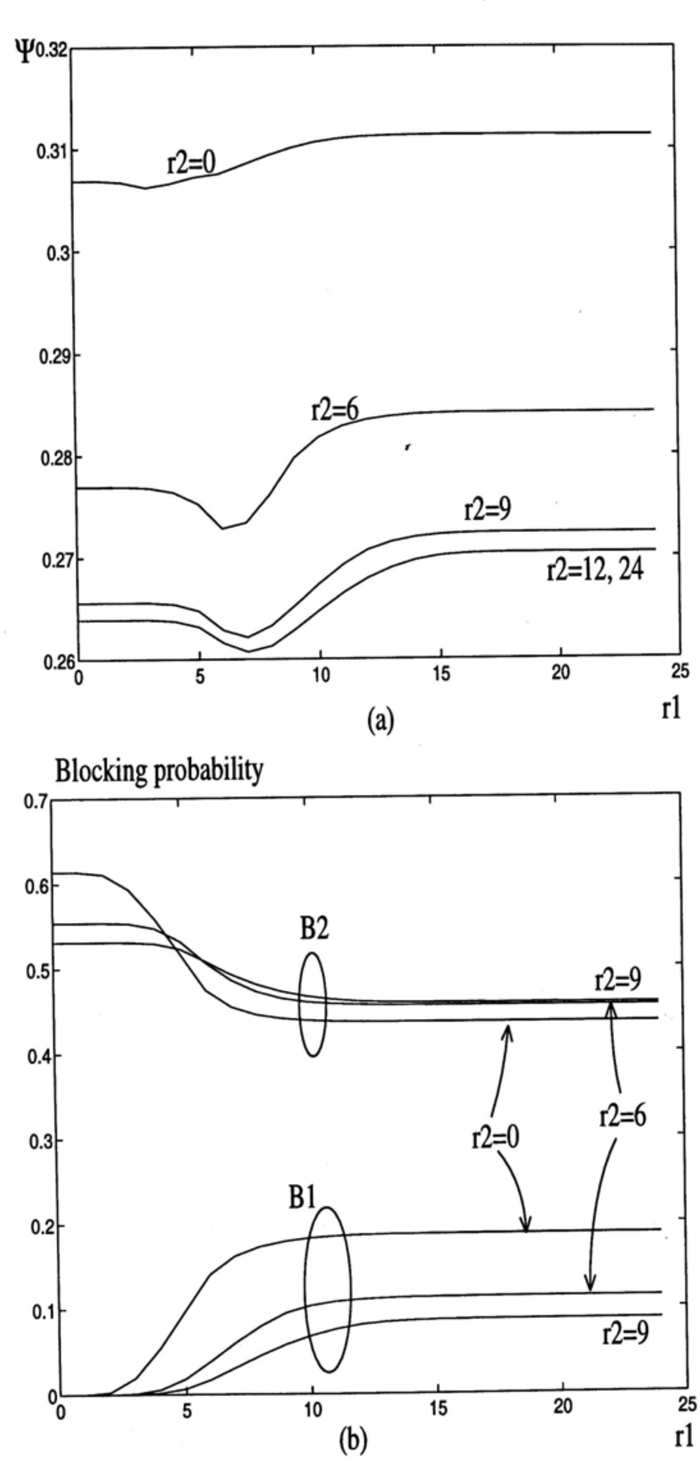
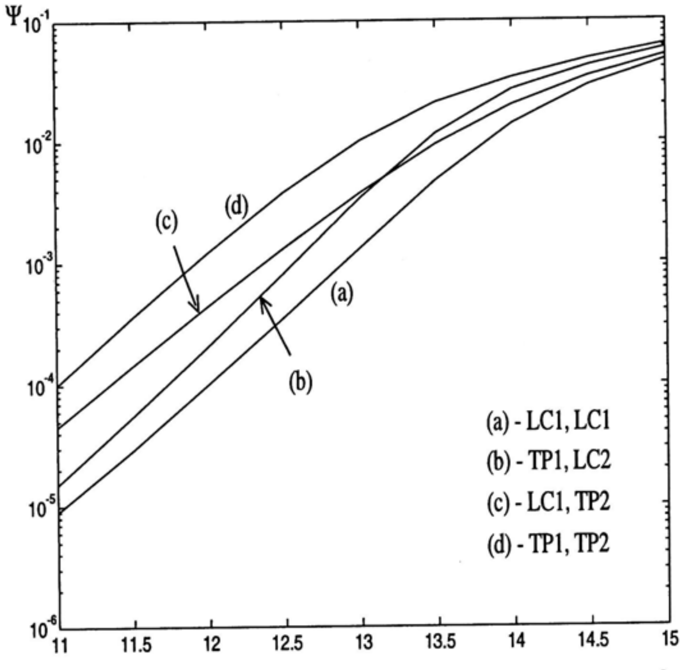
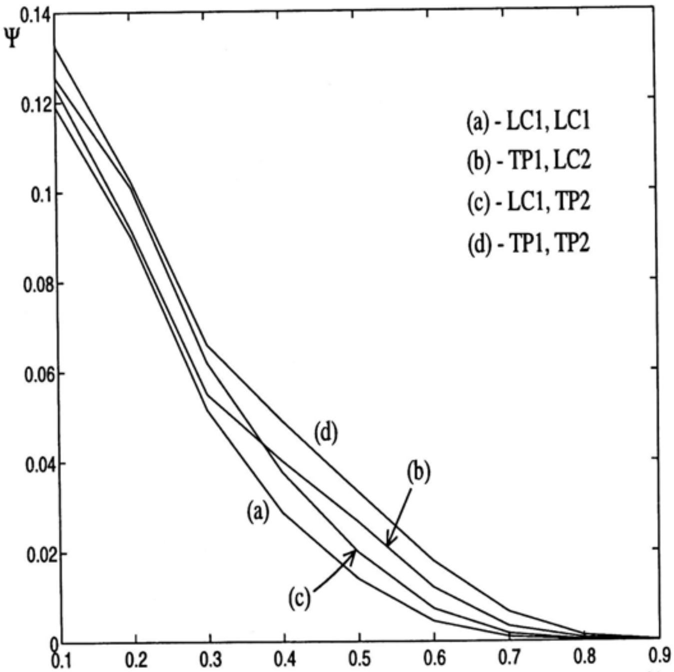


Figure 2.6: Interplay of TR parameters

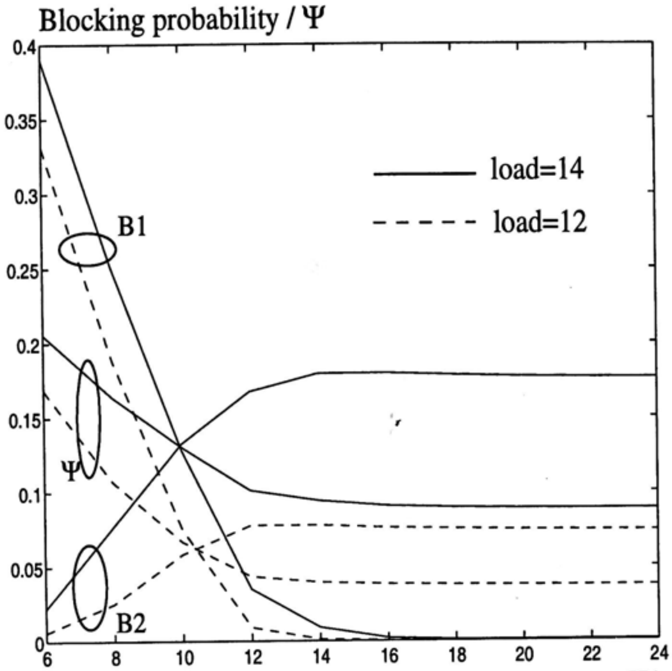


(a) different network loads

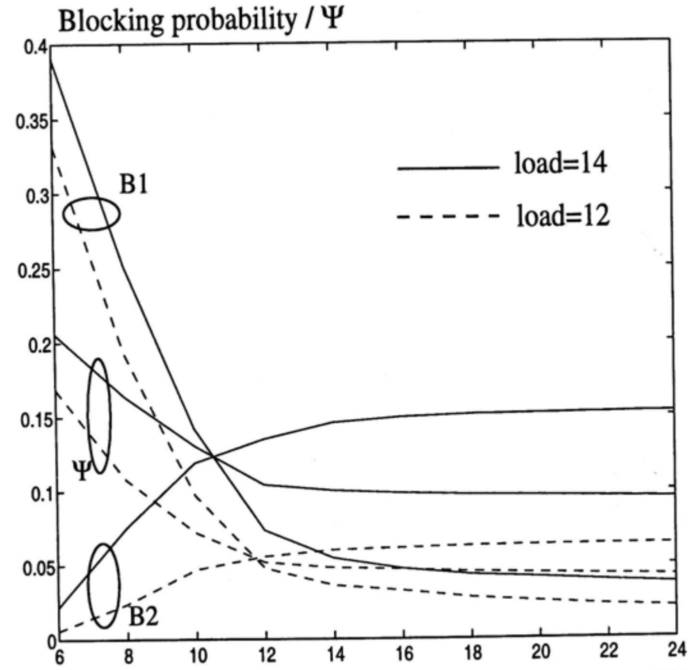


(b) different loading ratios

Figure 2.7: Comparison of LC and TP routing rules

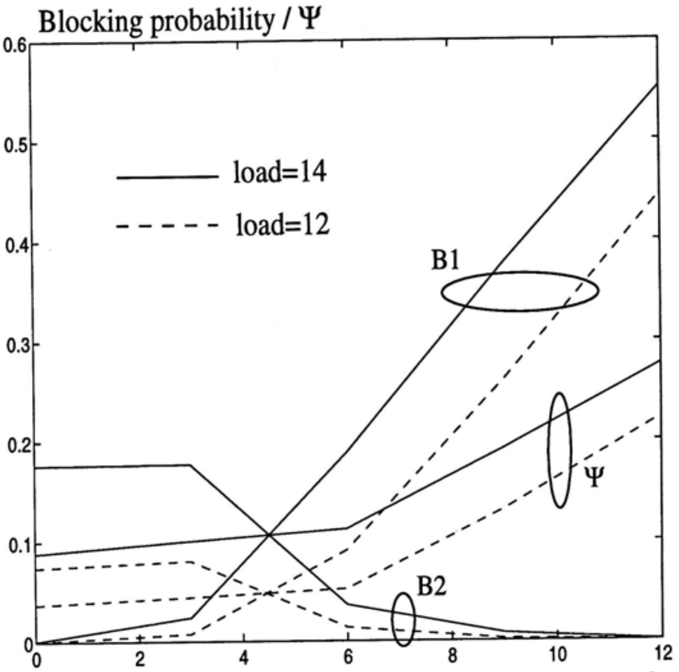


a) r1=4, r2=4

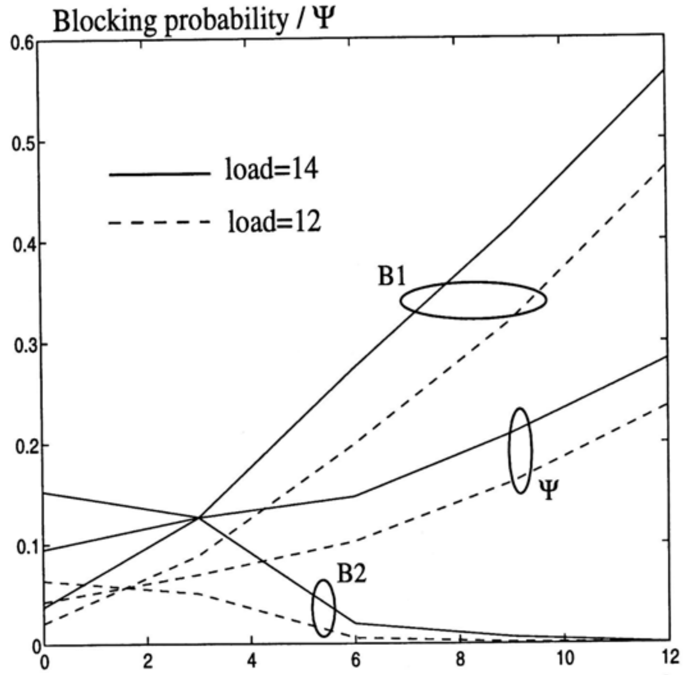


b) r1=12, r2=4

Figure 2.8: Performance of LO policy

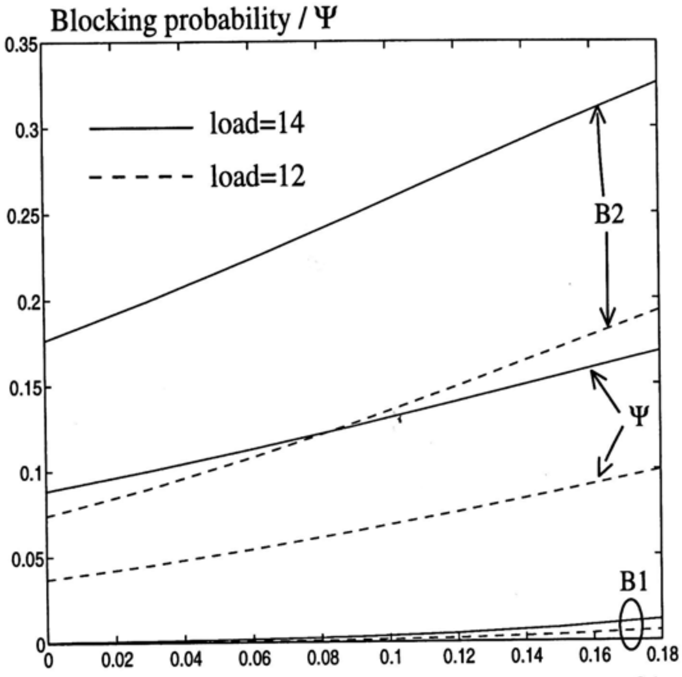


a) r1=4 r2=4

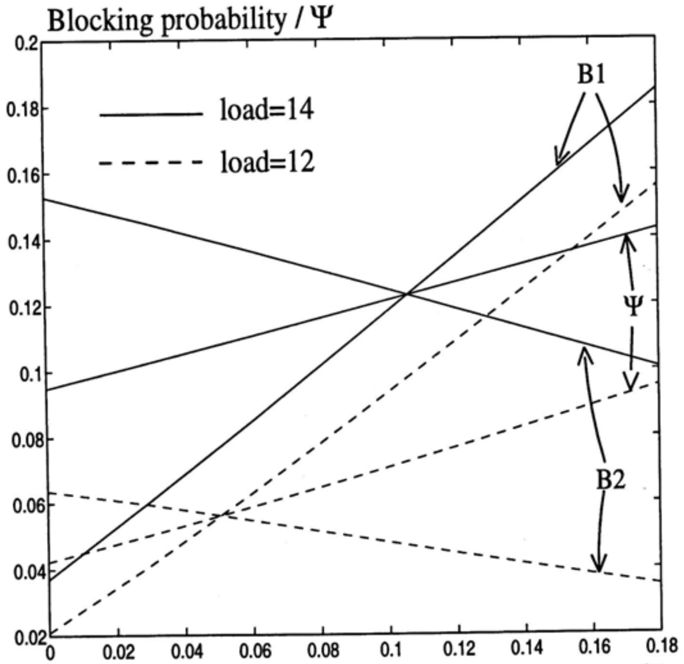


b) r1=12 r2=4

Figure 2.9: Performance of GB policy



a) r1=4 r2=4



b) r1=12 r2=4

Figure 2.10: Performance of EB policy

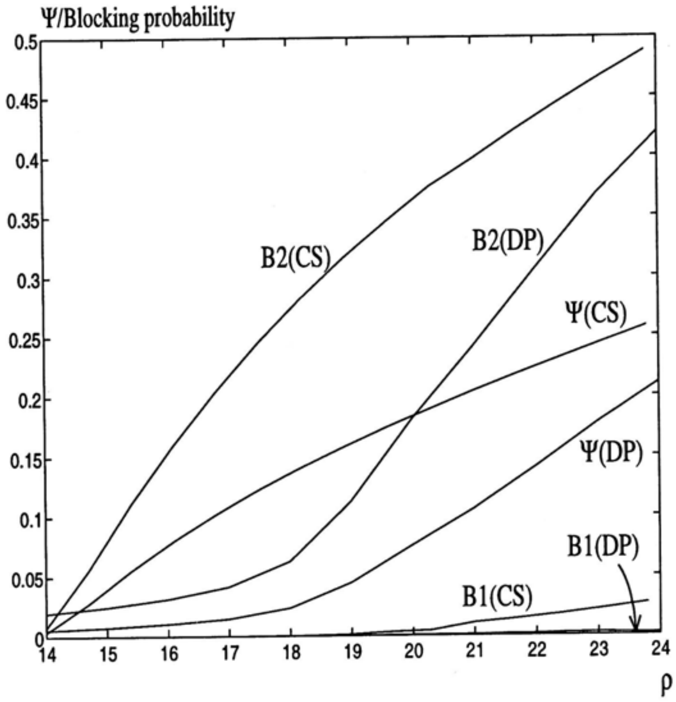


Figure 2.11: Comparison of DP and CS policies for different network loads

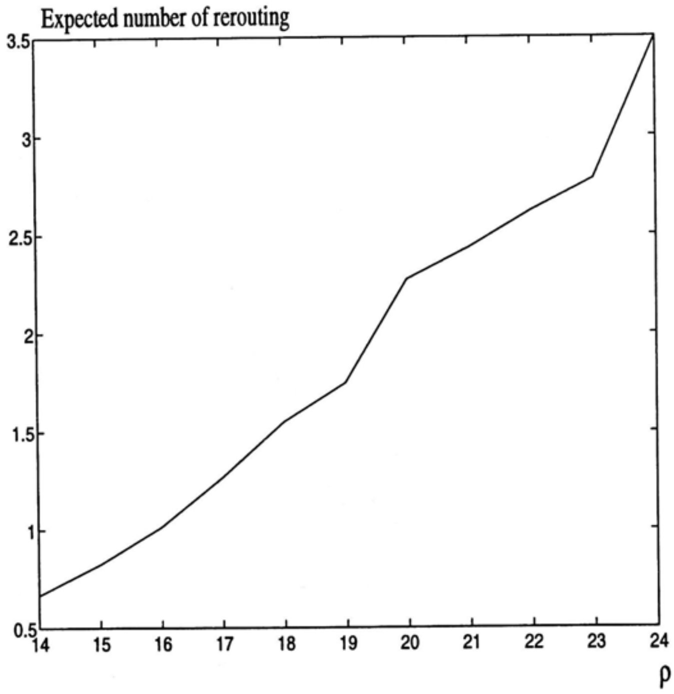


Figure 2.12: Average number of rerouting

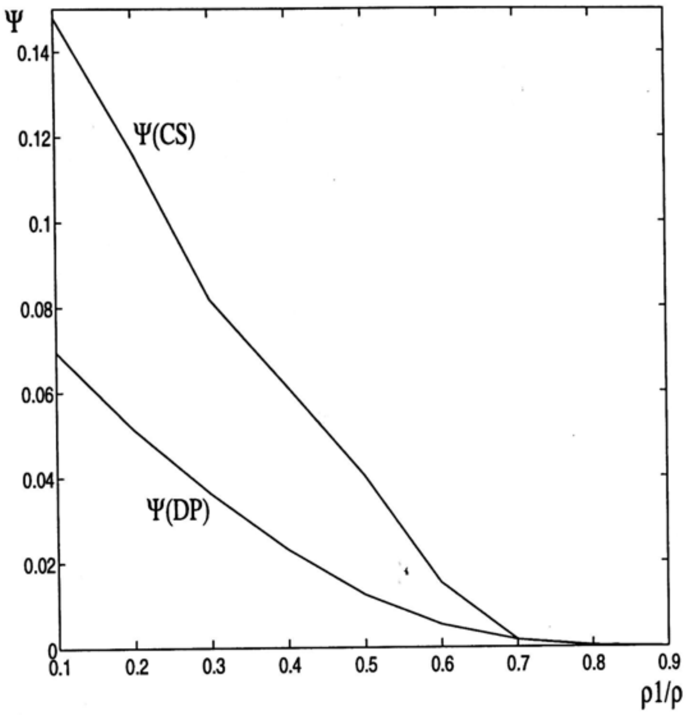
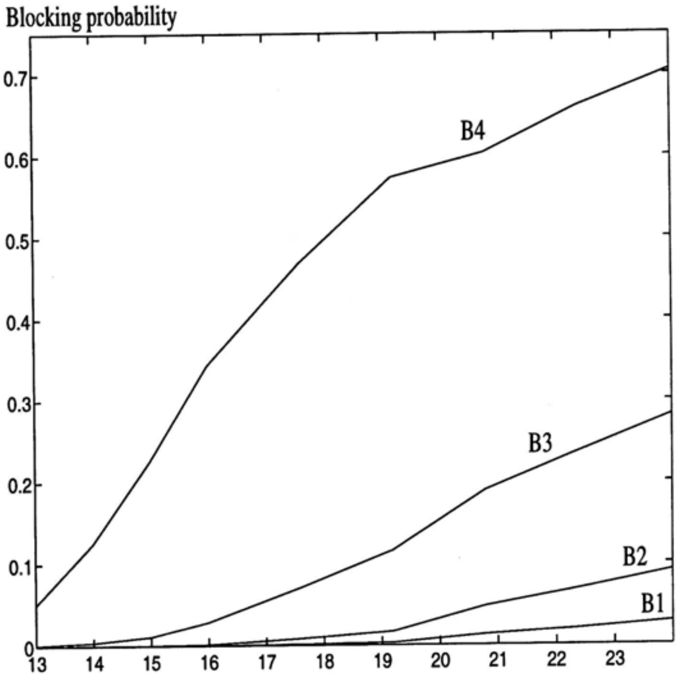
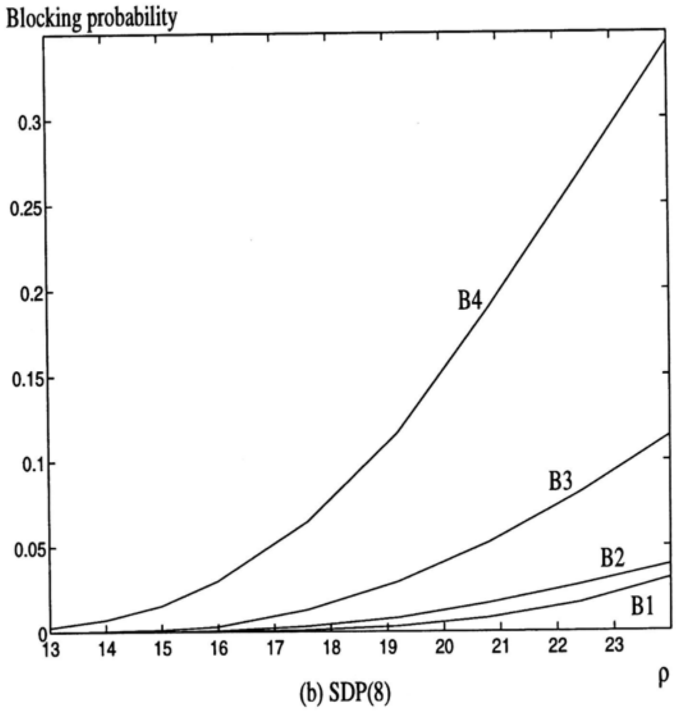


Figure 2.13: Comparison of DP and CS policies for different loading ratios



(a) CS



(b) SDP(8)

Figure 2.15: Blocking probability with CS and SDP(8) policies

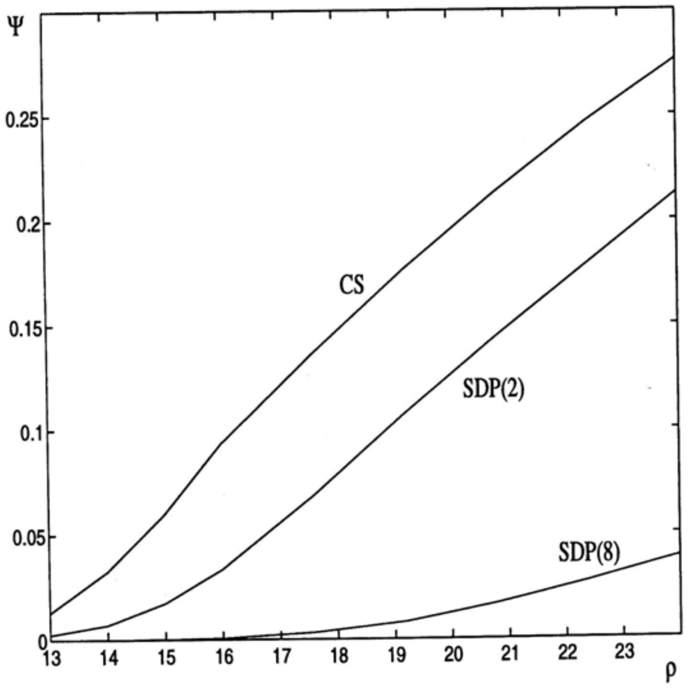
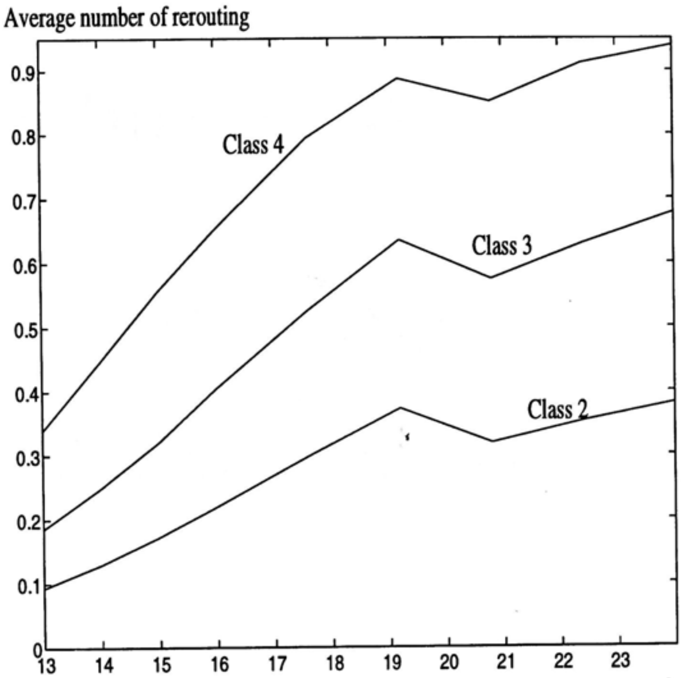
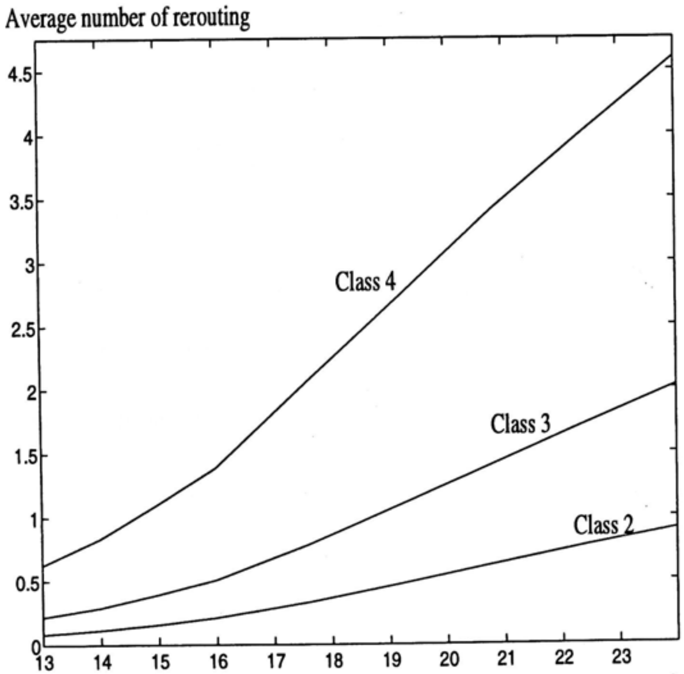


Figure 2.14: Revenue loss of SDP and CS policies



(a) SDP(2)



(b) SDP(8)

Figure 2.16: Average number of rerouting with SDP

# Chapter 3

# Least Congestion Routing in Multirate Loss Networks

## 3.1 Introduction

With the advent of stored program control and the installation of out-of-band signaling, it is now possible to implement a network that support different class of services with heterogeneous bandwidth requirement. With proper design, dynamic routing can reduce the blocking of calls and can adapt to facility failures and traffic pattern changes. The AT&T's Real Time Network Routing is an example of such adaptive routing scheme which implements new *class-of-service routing* capabilities for dynamic networks[13].

In [29], Chung and Ross studied various approximate formulae for computing the loss probability of multirate loss networks under fixed routing rule. They also studied the sensitivity of the average revenue to the changes of offered load and link capacity. In [19], random alternate routing in circuit-switched networks supporting two classes of services having the same bandwidth requirement are analyzed. In [20], Gupta et al. proposed a routing algorithm for virtual-path-based ATM networks based on the fluid approximation of the buffer overflow probability.

Recently, adaptive call admission and routing schemes based on the Markov Decision Process(MDP) were proposed. The complexity of the algorithm, however, is unmanageable in multirate loss networks. In [21], Dziong and Mason reduced the complexity of the problem by decomposing the network reward process into a set of link reward processes. In [22], Hwang et al. employed the MDP approach and generalized the State Dependent Routing for multirate loss networks. They reduced the complexity of the problem by modeling each link as a one-dimensional birth/death process and derived a set of expressions to evaluate the state-dependent link shadow prices.

In this chapter, we will compare two versions of Least Congestion routing. The first one is called the $M^2$ routing and is based on the residual bandwidth measure. The second one is called the *Mean Time to Blocking*(MTB) routing and is based on the *Mean Time to Blocking* measure. The *mean time to blocking* measure incorporates the link capacity and traffic rates information. The rationale for using such a measure is that in multirate traffic environment, the link occupancy is actually the sum of the occupancies of a number of traffic types, each having a different *mean time to blocking.* Moreover, in asymmetric traffic environments, the amount of residual bandwidth in a path does not reflect very accurately the congestion level since paths have different loadings. The *mean time to blocking* measure, however, does not have that problem.

In Section 3.2, we present the two version of the LC routing. Then in Section 3.3, we discuss bandwidth sharing policies and aggregation of link status information. In Sections 3.4 and 3.5, the aggregated $M^2$ routing and the aggregated MTB routing are analyzed. Numerical results and discussions are presented in Section 3.6. Finally, conclusions are drawn in Section 3.7.

## 3.2 The $M^2$ and MTB Routings

Consider a network supporting $K$ classes of services where each class is characterized by its arrival rate $\lambda_i$, mean holding time(or service time) $1/\mu_i$, revenue $R_i$ and bandwidth requirement $f_i$, $i = 1, 2, ..., K$. Let $F^{(i)}$ be the amount of bandwidth on link $i$ and let the state of a link be represented by vector $\mathbf{n} = (n_1, n_2, \ldots, n_K)$, where $n_k$ is the number of class $k$ call on the link. A path $q$ in the network is specified by a link set $\mathcal{L}_q$. Each node pair has a direct path and we consider only two-hop alternate paths, ie, $|\mathcal{L}_q|=2$.

### 3.2.1 $M^2$ Routing

In $M^2$ routing, an overflowed call is routed to the path with the maximum amount of residual bandwidth and satisfying the bandwidth reservation requirement. The residual bandwidth $\beta(q)$ of path $q$ is defined by

$$\beta(q) = \min_{i \in \mathcal{L}_q}(F^{(i)} - \sum_{k=1}^{K} n_k^{(i)} f_k^{(i)})$$

where the superscript $(i)$ is the link index. In case two or more paths have the same $\beta$ value, the call is routed to path $q$ with maximum residual bandwidth on the less busy link. If there is a tie, the overflowed call will be routed to one of the candidate paths randomly.

As the maximum residual bandwidth criterion is used twice in selecting the best alternate path, we call this routing scheme the $M^2$ routing. This is a generalization of the $M^2$ routing in [25] for multirate loss networks. The corresponding simpler rule, $M$ routing, in multirate loss networks would route an overflowed call to one of the paths having the same maximum residual bandwidth in a random manner. In the aggregated $M^2$ routing, several link occupancies are lumped into an aggregate-state and the routing rule is similar.

### 3.2.2 MTB Routing

The *mean time to blocking* measure on a path incorporates the link loadings, the traffic rates and link bandwidth information and can give a better measure of the degree of congestion of a path. Before we specify the routing rule, we define the *mean time to blocking* measure as follows

**Definition 1** *The mean time to blocking, $T_k(\mathbf{n})$ for class $k$ traffic on a link at state $\mathbf{n}$ is the mean first passage time from state $\mathbf{n}$ to the set of blocking states of class $k$ traffic.*

Consider a Markov chain with state space $\Omega$. Let $S(\mathbf{n})$ be the mean sojour time in state $\mathbf{n}$, $p_{\mathbf{nj}}$ be the transition probability from state $\mathbf{n}$ to state $\mathbf{j}$ and

$$\Omega^{(D)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega, \mathbf{n} + \mathbf{e}_k \notin \Omega\}$$

be the set of blocking states for class $k$ traffic where $\mathbf{e}_k$ is a $K$-vector with a '1' at the $k$th position and zero elsewhere. The *mean time to blocking* at state $\mathbf{n}$ for class $k$ traffic is given in [33] as

$$T_k(\mathbf{n}) = \begin{cases} S(\mathbf{n}) + \sum_{\mathbf{j}\in\Omega} p_{\mathbf{nj}} T_k(\mathbf{j}) & \mathbf{n} \in \Omega \setminus \Omega^{(D)}(k) \\ 0 & \mathbf{n} \in \Omega^{(D)}(k) \end{cases} \tag{3.1}$$

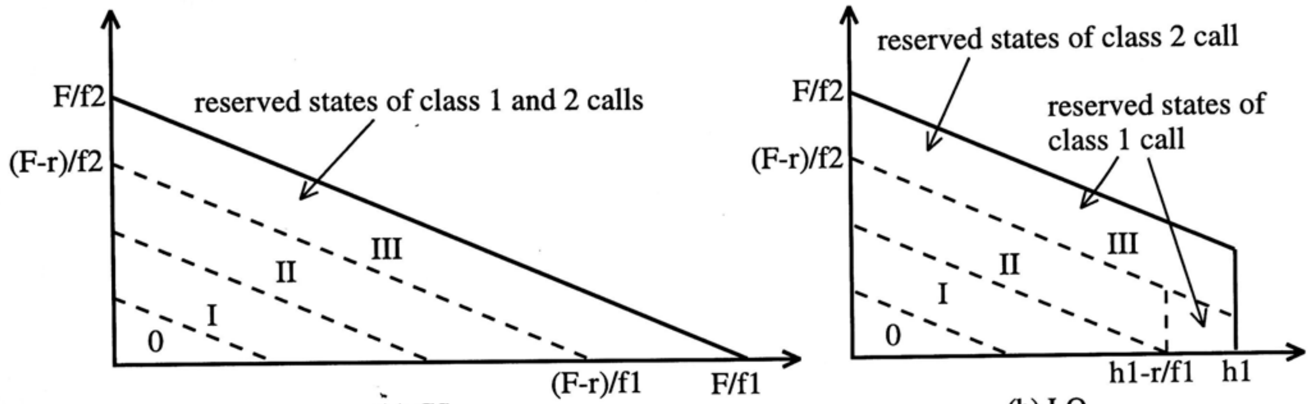
**Definition 2** *An upper bound $\Gamma_k(q)$ of the mean time to blocking on path $q$ for class $k$ traffic is the smallest of such link measures among all links in $\mathcal{L}_q$*

$$\Gamma_k(q) = \min_{i\in\mathcal{L}_q}\{T_k^{(i)}(\mathbf{n}^{(i)})\}$$

*where the superscript $(i)$ is the link index.*

In MTB routing, the direct path is tried first. If the direct path is full, the call will be directed to the path having the largest $\Gamma_k(q)$ value and satisfying the bandwidth reservation requirement. In the aggregated MTB routing, the time axis is divided into several regions and each region is called an aggregate-state. When the direct path is congested, the routing scheme will compare the *mean time to blocking* of the busiest links of those alternate paths satisfying the bandwidth reservation requirement and select one having the largest such value. If there is more than one such path, the overflowed call will be routed to the path having the largest such measure on the second busiest links. If there is a tie, the overflowed call will be routed to one of the candidate paths randomly.

## 3.3 Bandwidth Sharing Policies and State Aggregation



(a) CS

(b) LO

Figure 3.1: Two different bandwidth sharing policies

Link bandwidth is shared in two ways. First, it is shared between different classes of calls. If bandwidth is shared freely among different classes of calls, we have the so called Complete Sharing(CS) policy. On the other hand, if the number of calls in a link is limited, we have the Limited Occupancy(LO) policy(also known as the threshold-type policy in [17]). To maintain the stability of the network, an amount of bandwidth $r_k$ are reserved for direct link traffic only. Figure 3.1 shows the state space of a link having a bandwidth of $F$ under the CS and the LO policies. Two classes of calls and four-level state aggregations are considered. For the CS and the LO policies, the state space $\Omega_{CS}$ and the set of class $k$ alternate call blocking states $\Omega_k^{(A)}$ are given by,

$$\Omega_{CS} = \{\mathbf{n} : \mathbf{n} \cdot \mathbf{f} \leq F\}$$

$$\Omega_{CS}^{(A)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, (\mathbf{n} + \mathbf{e}_k) \cdot \mathbf{f} + r_k > F\}$$

$$\Omega_{LO} = \{\mathbf{n} : \mathbf{n} \in \Omega_{CS}, n_k \leq h_k, h_k \geq 0\}$$

$$\Omega_{LO}^{(A)}(k) = \{\mathbf{n} : \mathbf{n} \in \Omega_{LO}, (\mathbf{n} + \mathbf{e}_k) \cdot \mathbf{f} + r_k > F \textit{ or } n_k \geq h_k - \lceil r_k / f_k \rceil, h_k \geq 0\}$$

Second, bandwidth can be shared between the direct and overflowed calls. In order to prevent the unstable behavior at heavy loading conditions, a certain amount of bandwidth $r_k$ can be reserved for direct calls only. This kind of bandwidth reservation is similar to the trunk reservation in the conventional telephone network.

During a call set-up, the originating switch requests status information of all via links through the Common Channel Signaling (CCS) network. These information can often be compressed: The practice of lumping several link occupancies into an aggregate-state is called state aggregation [32]. The rationale behind state aggregation is that instead of soliciting state information on a call by call basis, individual links can simply broadcast the change of their aggregate-state information in case of $M^2$ routing (or the *mean time to blocking* in case of MTB routing). As the change of aggregate-state is much less often than the change of state. This could significantly reduce the signaling traffic. Moreover, the reduction of the number of states could drastically reduce the route computation time in case routes are computed on-line or drastically reduce the routing table size in case route are computed off-line.

In the following sections, we analyze the aggregated $M^2$ and the MTB routings in an $N$-node fully connected symmetric multirate loss networks with the assumptions that links are independent and overflowed traffics are independent Poisson processes. Each node pair has $L = N - 2$ alternate paths. Our analysis gives the numerical solution of the equilibrium state probabilities and the alternate traffic rates. With these, the call blocking probabilities of individual classes can be computed.

Let $P(\mathbf{n})$ be the equilibrium state probability. Then the class $k$ direct call blocking probability $D_k$ and the alternate call blocking probability $A_k$ are

$$D_k = \sum_{\mathbf{n} \in \Omega^{(D)}(k)} P(\mathbf{n})$$

$$A_k = \sum_{\mathbf{n} \in \Omega^{(A)}(k)} P(\mathbf{n})$$

The class $k$ call blocking probability $B_k$ is given in [30] as

$$B_k = D_k \left[1 - (1 - A_k)^2\right]^L$$

Assuming that each call brings an average revenue of $R_k = f_k/\mu_k$, the fractional revenue loss $\Psi$ is the weighted average lost of revenue[21]

$$\Psi = \frac{\sum_{k=1}^{K} B_k R_k \lambda_k}{\sum_{k=1}^{K} R_k \lambda_k}$$

## 3.4 Analysis of $M^2$ Routing

Consider a general state space $\Omega$ of a link, let the link occupancy information be aggregated into $\Upsilon$ levels. Let $g(\mathbf{n})$ be the aggregate-state in which $\mathbf{n}$ falls into and $\Omega(i), i = 0, 1, \ldots, \Upsilon - 1$ be the set of link states falling into aggregate-state $i$.

In Figure 3.2, alternate paths of a link AC are shown. Let $u$ be the maximum and $v$ be the minimum of the aggregate-states of the two links AB and BC. When AC is full, the overflowed class $k$ calls of rate $\lambda_k D_k$ will be routed to one of the paths having the maximum residual bandwidth. If there is a tie, the overflowed call is routed to an alternate path having the largest link residual bandwidth on the less busy link. Let this path be a $M^2$ path. Suppose there are $\alpha$ such paths. The overflowed class $k$ calls of AC will be routed to one of these path, say ABC, at rate $\lambda_k D_k / \alpha$.

Let $\pi_k(i)$ be the probability that *a link is in aggregate-state $i$ and that it is admissible to class $k$ alternate calls(class-$k$ admissible)*, or

$$\pi_k(i) = \sum_{\mathbf{n} \in \Omega(i) \setminus \Omega^{(A)}(k)} P(\mathbf{n})$$

Let $b_k(u, v)$ be the probability that *the two links of an alternate path is in aggregate-states $u$ and $v$ where $u \geq v$ and that the path is class-$k$ admissible.* As a path is class-$k$

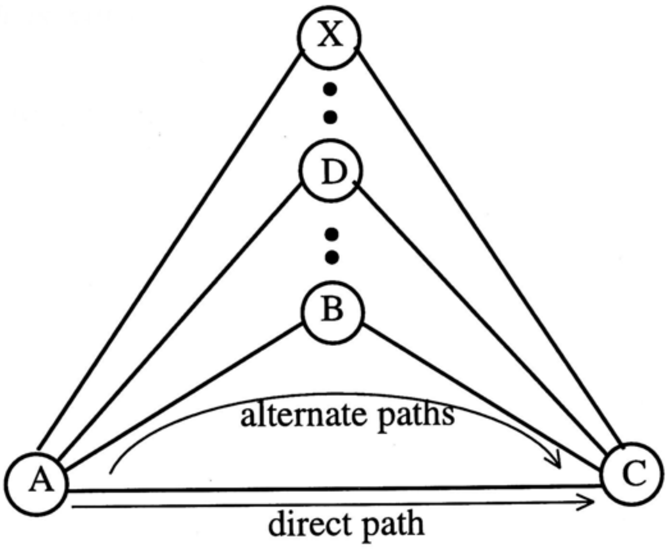


Figure 3.2: An alternate path of a node pair

admissible if and only if the two links constituting it are both class-$k$ admissible, we have

$$b_k(u,v) = \begin{cases} \pi_k(u)^2 & u = v \\ 2\pi_k(u)\pi_k(v) & u > v \end{cases} \tag{3.2}$$

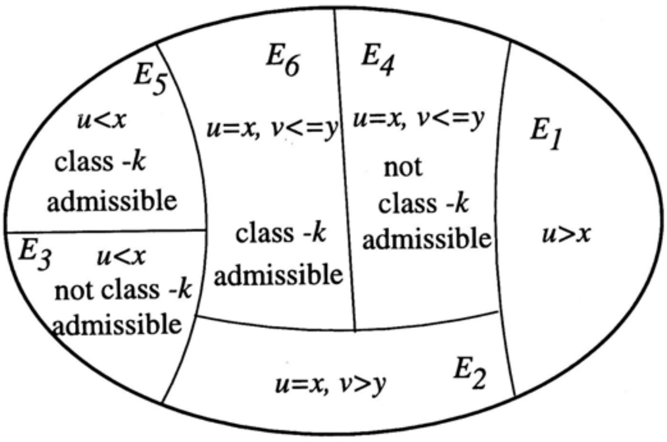


Figure 3.3: Different events of an alternate path under $M^2$ routing

Consider an alternate path with parameters $u$ and $v$ and define the four disjoint events $E_1, E_2, E_3$ and $E_4$ for a given pair of threshold $x$ and $y (y \leq x)$ as

$E_1$: $u > x$

$E_2$: $u = x$ *and* $v > y$

$E_3$: *$u < x$ and the path is not class-$k$ admissible*

$E_4$: *$u = x$, $v \leq y$ and the path is not class-$k$ admissible.*

Let $V_k(x, y)$ be the probability that path ADC is busier than path ABC given that AB and AC are in aggregate state x and y under the $M^2$ criterion. Let $E_0 = E_1 \vee E_2 \vee E_3 \vee E_4$. $V_k(x, y)$ is the probability of $E_0$. Since the $E_i$'s, $i = 1, ..., 4$ are disjoint, from Figure 3.3, we have

$$\begin{aligned} V_k(x,y) &= Prob[E_1] + Prob[E_2] + Prob[E_3] + Prob[E_4] \\ &= 1 - Prob[E_5] - Prob[E_6] \\ &= 1 - \sum_{u=0}^{x-1}\left[\sum_{v=0}^{u-1} 2\pi_k(u)\pi_k(v)\right] + \pi_k(u)^2 - \left[\sum_{v=0}^{y} 2\pi_k(x)\pi_k(v) - \delta(x-y)\pi_k(x)^2\right] \end{aligned}$$

where $\delta(i)$ is one for $i = 0$ and is zero otherwise.

Next, let $E_7$ be defined as

$E_7$: *$\alpha - 1$ alternate paths have the same aggregate-states $x$ and $y$ in their two links and the two links are both class-$k$ admissible*

Then from (3.2), $Prob[E_5] = [b_k(x, y)]^{\alpha-1}$. Now, suppose AC is full and the two links of path ABC are in aggregate-states $x$ and $y$, the probability that *ABC is a $M^2$ path and there are $\alpha - 1$ other such paths*, $f_k(\alpha \mid x, y)$ is given by

$$\begin{aligned} f_k(\alpha \mid x,y) &= \begin{pmatrix} L-1 \\ \alpha-1 \end{pmatrix} Prob[E_7 \wedge E_0] \\ &= \begin{pmatrix} L-1 \\ \alpha-1 \end{pmatrix} b_k(x,y)^{\alpha-1} V_k(x,y)^{L-\alpha} \end{aligned}$$

If the two links of a class-$k$ admissible alternate path ABC is in aggregate-state $x$ and $y$, the overflowed class $k$ traffic rate that get routed from path AC, denoted by $s_k(x, y)$, is given by

$$\begin{aligned} s_k(x,y) &= \sum_{\alpha=1}^{L} \frac{\lambda_k D_k}{\alpha} f_k(\alpha \mid x,y) \\ &= \lambda_k D_k \frac{[b_k(x,y)+V_k(x,y)]^L - V_k(x,y)^L}{L b_k(x,y)} \end{aligned}$$

Given that link AB is in state $\mathbf{n} \in \Omega \setminus \Omega^{(A)}(k)$, the total overflowed class $k$ traffic, $A_k(\mathbf{n})$, obtained by removing the conditioning on the second link is

$$A_k(\mathbf{n}) = 2L \sum_{i=0}^{\Upsilon-2} s_k(\max[g(\mathbf{n}),i], \min[g(\mathbf{n}),i])\pi_k(i) \tag{3.3}$$

At state $\mathbf{n}$, the class $k$ call arrival rate including direct and overflowed traffic is

$$\Lambda_k(\mathbf{n}) = \begin{cases} \lambda_k + A_k(\mathbf{n}) & \mathbf{n} \in \Omega \setminus \Omega^{(A)}(k) \\ \lambda_k & \mathbf{n} \in \Omega^{(A)}(k) \setminus \Omega^{(D)}(k) \\ 0 & \mathbf{n} \in \Omega^{(D)}(k) \end{cases} \tag{3.4}$$

Therefore for state $\mathbf{n} \in \Omega$, the global balance equation is given by

$$\sum_k [\Lambda_k(\mathbf{n}) + n_k\mu_k] P(\mathbf{n}) = \sum_k (n_k+1)\mu_k P(\mathbf{n}+\mathbf{e}_k) + \Lambda_k(\mathbf{n}-\mathbf{e}_k)P(\mathbf{n}-\mathbf{e}_k) \tag{3.5}$$

with the understanding that $P(\mathbf{n}) = 0$ for $\mathbf{n} \notin \Omega$.

Let $\Lambda$ denotes the set of $\Lambda_k(\mathbf{n})$ and $\mathcal{P}$ denotes the set of $P(\mathbf{n})$. Then (3.4) and (3.5) can be expressed in the fixed point model form[32]: $\mathcal{P} = f_1(\Lambda)$ and $\Lambda = f_2(\mathcal{P})$. The $P(\mathbf{n})$'s can be computed by the Successive Over-Relaxation(SOR) method with the set of alternate traffic rates obtained from (3.3) in each iteration. In the examples quoted in Section 3.6, all $P(\mathbf{n})$s are obtained with relative error less than $10^{-8}$.

## 3.5 Analysis of MTB Routing

In MTB routing, aggregation is done on the *mean time to blocking.* This means that the time axis is partitioned into $\Upsilon$ regions numbered 0 to $\Upsilon - 1$ with the $i$th region covers the interval $[l_i, l_{i+1})$ where $l_\Upsilon = \infty$.

Consider path AC. When it is full, the MTB routing scheme will route overflowed calls to the alternate path having the longest *mean time to blocking.* In the aggregated MTB routing, it may happen that two paths are in the same aggregate-state. In this case, the routing scheme will compare the *mean time to blocking* value of the less busy links of these paths and route the call to the path having the largest such value. We call this path the MTB path. Suppose there are $\alpha$ such paths, the overflowed class $k$ call of AC is randomly routed to one of these paths, say ABC, at rate $\lambda_k D_k/\alpha$.

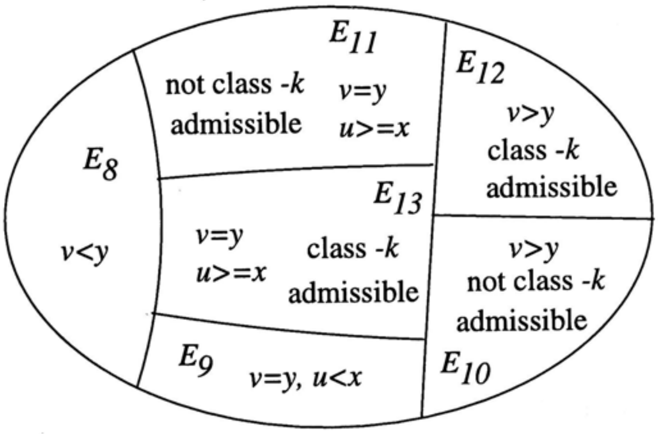


Figure 3.4: Different events of an alternate path under MTB routing

Consider an alternate path with parameters $u$ and $v (v \leq u)$ and define the following events for given thresholds $x$ and $y (y \leq x)$ as follows

$E_8$: $v < y$

$E_9$: $v = y$ and $u < x$

$E_{10}$: $v > y$ and *the path is not class-k admissible*

$E_{11}$: $v = y$, $u \geq x$ and *the path is not class-k admissible*

Let $Y_k(x, y)$ be the probability that path ADC is busier than path ABC given that path ABC is in aggregate state $x$ and $y$ under the MTB criterion. $Y_k(x, y)$ is given by the probability of $E_8 \vee E_9 \vee E_{10} \vee E_{11}$. From Figure 3.4, it can be expressed as

$$\begin{aligned} Y_k(x,y) &= 1 - Prob[E_{12}] - Prob[E_{13}] \\ &= 1 - \sum_{v=y+1}^{\Upsilon-1} \left[ \sum_{u=v+1}^{\Upsilon-1} 2\pi_k(u)\pi_k(v) \right] + \pi_k(v)^2 - \\ & \left[ \sum_{u=x}^{\Upsilon-1} 2\pi_k(y)\pi_k(u) - \delta(x-y)\pi_k(x)^2 \right] \end{aligned}$$

Now, suppose AC is full and the two links of a class-$k$ admissible alternate path ABC is in aggregate-state $x$ and $y (y \le x)$, the probability that *it is a MTB path and there are* $\alpha - 1$ *other such paths*, $f_k(\alpha \mid x, y)$ is

$$f_k(\alpha \mid x, y) = \begin{pmatrix} L-1 \\ \alpha - 1 \end{pmatrix} b_k(x,y)^{\alpha-1} Y_k(x,y)^{L-\alpha}$$

Given that the two links of a class-$k$ admissible path ABC is in aggregate-state $x$ and $y$, the amount of class $k$ traffic, $s_k(x,y)$, that gets routed from AC is

$$\begin{aligned} s_k(x,y) &= \sum_{\alpha=1}^{L} \frac{\lambda_k D_k}{\alpha} f(\alpha \mid i) \\ &= \lambda_k D_k \frac{[b_k(x,y) + Y_k(x,y)]^L - Y_k(x,y)^L}{L b_k(x,y)} \end{aligned}$$

Therefore, given that link AB is in state $\mathbf{n} \in \Omega \setminus \Omega^{(A)}(k)$, the total overflowed class $k$ traffic obtained by removing the conditioning on the second link is

$$A_k(\mathbf{n}) = 2L \sum_{i=0}^{\Upsilon-1} s_k(\max[g(\mathbf{n}), i], \min[g(\mathbf{n}), i]) \pi_k(i) \tag{3.6}$$

Denote $\mathcal{T}$ as the set of *mean time to blocking*, then (3.5), (3.1) and (3.6) can also be expressed in the fixed point model form as: $\mathcal{P} = f_1(\Lambda)$, $\mathcal{T} = f_3(\Lambda)$, and $\Lambda = f_4(\mathcal{T}, \mathcal{P})$. The state probabilities can be solved by the same SOR method with $\mathcal{T}$ and $\Lambda$ given by the solution of (3.1) and (3.6) in each iteration.

## 3.6 Numerical Results and Discussions

Consider a twelve nodes fully connected network supporting two classes of calls. Let their bandwidth requirements be $(f_1, f_2) = (1, 2)$ and their mean service rates be $(\mu_1, \mu_2) = (1, 0.5)$. Limited Occupancy policy is employed on class 1 call with threshold $h_1 = 8$. The amount of bandwidth of each link is twelve and the bandwidth reservation parameter is three. The base traffic rates is $(\lambda_1, \lambda_2) = (4, 0.6)$. Figure 3.5 shows the class 1 alternate traffic rate for the CS policy under MTB routing at 60% overload. We find that the overflowed traffic rates decreases with increasing $n_1$ and $n_2$. This behavior of decreasing overflowed traffic rates with increasing occupancies is similar to that observed in the state dependent routing of single rate networks.

Figure 3.6 shows the analytic results of the MTB routing as a function of the percentage overload from the base load. Using the CS policy, we find that class 2 call suffers a very large blocking probability when compared to that of class 1 call. This unfair condition can be improved by the LO policy as shown. On the other hand, the LO policy also increases the fractional revenue loss.

Next, we study a seven-node network with a bandwidth of 24 in each link. There are two classes of calls with $(f_1, f_2)$=(1,3), $(\lambda_1, \lambda_2)$ =(12,1.6) and $(\mu_1, \mu_2)$=(1,0.5). Figure 3.7(a) shows the blocking probabilities and Figure 3.7(b) shows the fraction revenue loss as a function of $h_1$ for $r_k = 0$ and $r_k = 4$. It is found that the LO policy is in fact trading the decrease of the blocking probability with the increase of the average revenue loss.

Figure 3.7(a) also shows that when $r = 0$, the blocking of class 2 calls increases when $h_1$ decreases from 24 to 14 and then decreases abruptly when $h_1$ decreases further to 10. The reason for this phenomenon is that by restricting the number of class 1 calls, the free bandwidth for class two calls are shared by both direct and alternate traffic. Since the alternate traffic uses twice as much resource as the direct traffic, the blocking probability of class 2 call would increase if bandwidth reservation is not imposed. This phenomenon vanishes when $h_1$ is further decreased to 10 at the expense of higher blocking of the class 1 calls.

Figure 3.8 shows the blocking probabilities as a function of the percentage overload for the $M^2$ and the MTB routing. The network studied has twelve nodes and a bandwidth of fifteen in each link. The bandwidth requirement of the two classes are (1,3). The bandwidth reservation parameter is three and $h_1$ is eleven. The base traffic rates are (5,0.4) and mean service rates are (1,0.4). It is observed that the use of MTB routing can reduce the blocking probability and the average revenue loss of both classes of traffic as compared to $M^2$ routing. For instance, at the same 2% revenue loss level, MTB routing can tolerate 3.5% more overload than $M^2$ routing.

This result can be appreciated intuitively. As external loading increases, the increased blocking of the direct calls causes an increase of overflowed rate. How to choose alternate paths becomes crucial as it determines how the remaining network resources are allocated. In $M^2$ routing, a particular link occupancy can represent many different traffic composition. MTB routing, on the other hand, uses the *mean time to blocking* as a measure of the congestion status of a link which takes into consideration different traffic composition. MTB routing therefore can identify more accurately the best alternate path.

Next, we change to a five node network with a bandwidth of 30 on each link. There are two classes of calls with bandwidth requirement of $(f_1, f_2) = (1, 10)$, mean service rate of $(\mu_1, \mu_2) = (1, 0.5)$ and the base traffic rates of $(\lambda_1, \lambda_2) = (6, 0.3)$. The bandwidth reservation parameters are chosen to minimize the revenue loss. Figure 3.9(a) shows the comparison of fractional revenue loss of the MTB and $M^2$ routing as a function of overload. Again we find that the MTB routing outperforms the $M^2$ routing. Figure 3.9(b) shows the relative performance of the MTB routing when compared to the $M^2$ routing. We find that at light loading, MTB routing yields a larger relative loss reduction than in the heavy load region. This is because at loading increases, less alternate routing are performed due to the bandwidth reservation.

Figure 3.10 shows the performance of MTB and $M^2$ routing for a wide range of traffic composition. The network parameters are the same as that used in Figure 3.9.

Network load are 12 and bandwidth reservation parameters are 3. Again, we find that MTB routing outperforms $M^2$ routing in a wide range of different traffic compositions.

Next, we study the effect of state aggregation. Table 3.1 shows the performance of the aggregated $M^2$ routing with the same control and network parameters as the LO policy in Figure 3.6. For the 2-level aggregation, the state partitioning is $[0, 8]$ and [9,12], while that of the 3-level aggregation is $[0, 4]$, $[5, 8]$, $[9, 12]$ and the 4-level aggregation is $[0, 2]$, $[3, 5]$, $[6, 8]$, $[9, 12]$. The last aggregate-state is treated as the reserved state. We observe that the performance of the 3-level aggregated $M^2$ routing is already close to that of the non-aggregated scheme. We also found that at light loading, more levels is needed to approach the non-aggregated scheme because the alternate traffic rate decreases more slowly when compared to that in heavy loading conditions.

Table 3.1: BLOCKING PROBABILITY OF THE AGGREGATED $M^2$ ROUTING

| overload | non-aggregated $M^2$ | | $\Upsilon = 2$ | |
|---|---|---|---|---|
| (%) | $B_1$ | $B_2$ | $B_1$ | $B_2$ |
| 0 | 2.603e-3 | 5.121e-4 | 3.025e-3 | 7.589e-4 |
| 10 | 1.553e-2 | 6.440e-3 | 1.652e-2 | 7.598e-3 |
| 20 | 4.190e-2 | 2.780e-2 | 4.275e-2 | 2.927e-2 |
| 30 | 7.367e-2 | 6.170e-2 | 7.422e-2 | 6.281e-2 |
| 40 | 1.061e-1 | 1.001e-1 | 1.064e-1 | 1.008e-1 |
| 50 | 1.375e-1 | 1.387e-1 | 1.377e-1 | 1.392e-1 |
| 60 | 1.675e-1 | 1.758e-1 | 1.677e-1 | 1.762e-1 |
| | $\Upsilon = 3$ | | $\Upsilon = 4$ | |
| 0 | 2.729e-3 | 5.479e-4 | 2.606e-3 | 5.257e-4 |
| 10 | 1.592e-2 | 6.608e-3 | 1.560e-3 | 6.464e-3 |
| 20 | 4.235e-2 | 2.803e-2 | 4.205e-2 | 2.787e-2 |
| 30 | 7.403e-2 | 6.189e-2 | 7.382e-2 | 6.175e-2 |
| 40 | 1.063e-1 | 1.003e-1 | 1.062e-1 | 1.001e-1 |
| 50 | 1.377e-1 | 1.389e-1 | 1.376e-1 | 1.387e-1 |
| 60 | 1.676e-1 | 1.760e-1 | 1.676e-1 | 1.759e-1 |

Table 3.2 shows the performance of the aggregated MTB routing with the same control and network parameters as the LO policy in Figure 3.6. Only uniform aggregation, ie, the time-axis is divided into intervals of equal length, are considered. With

the time axis divided into 0.5 second's intervals, the performance of the aggregated MTB routing is similar to that of the non-aggregated scheme. It is also found that at heavy loading conditions, the *mean time to blocking* becomes smaller. Therefore, small intervals are needed for better resolution.

Table 3.2: BLOCKING OF THE AGGREGATED MTB ROUTING

| overload (%) | non-aggregated MTB $B_1$ | non-aggregated MTB $B_2$ | 0.5s-interval $B_1$ | 0.5s-interval $B_2$ |
|---|---|---|---|---|
| 0 | 1.901e-03 | 3.376e-04 | 1.912e-03 | 3.388e-04 |
| 10 | 9.017e-03 | 3.018e-03 | 9.223e-03 | 3.127e-03 |
| 20 | 2.686e-02 | 1.458e-02 | 2.724e-02 | 1.462e-02 |
| 30 | 5.431e-02 | 3.893e-02 | 5.527e-02 | 3.932e-02 |
| 40 | 8.325e-02 | 7.142e-02 | 8.369e-02 | 7.264e-02 |
| 50 | 1.180e-01 | 1.073e-01 | 1.201e-01 | 1.101e-01 |
| 60 | 1.507e-01 | 1.442e-01 | 1.534e-01 | 1.453e-01 |
| | 1s-interval | | 2s-interval | |
| 0 | 2.115e-03 | 3.404e-04 | 2.507e-03 | 3.623e-04 |
| 10 | 9.319e-03 | 3.273e-03 | 9.462e-03 | 3.384e-03 |
| 20 | 2.912e-02 | 1.481e-02 | 3.064e-02 | 1.532e-02 |
| 30 | 5.778e-02 | 4.164e-02 | 5.985e-02 | 4.356e-02 |
| 40 | 8.464e-02 | 7.382e-02 | 8.764e-02 | 7.623e-02 |
| 50 | 1.264e-01 | 1.211e-01 | 1.363e-01 | 1.315e-01 |
| 60 | 1.587e-01 | 1.498e-01 | 1.667e-01 | 1.639e-01 |

## 3.7 Concluding Remarks

The MTB routing and the $M^2$ routing are analyzed under symmetric traffic conditions in this chapter. Numerical results show that the MTB routing gives a better performance than the $M^2$ routing. We have also studied the aggregated version of these two routing schemes and have found that with properly designed aggregation, they can perform closely to the non-aggregated schemes. We have also studied the Limited Occupation(LO) policy and showed that under the MTB routing, the LO policy can manipulate the relative blocking probabilities of different classes but it also incurs a larger fractional revenue loss.

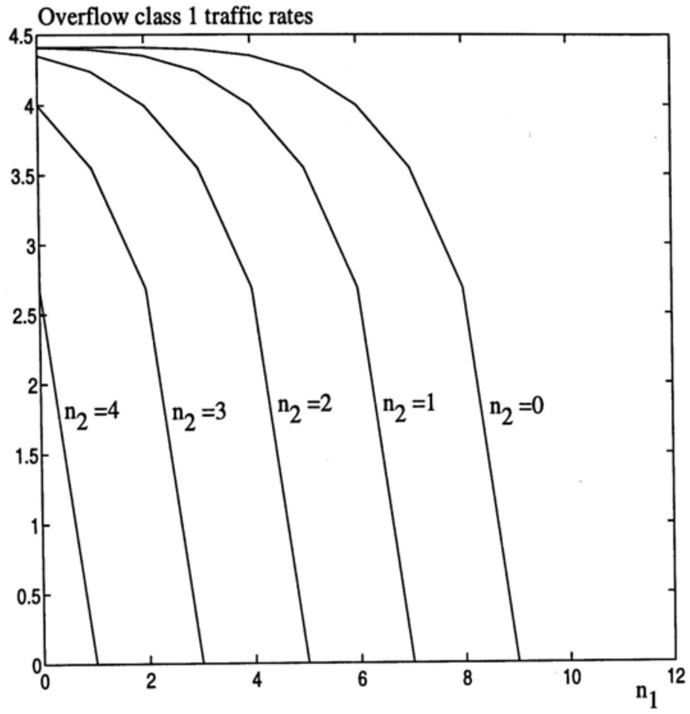
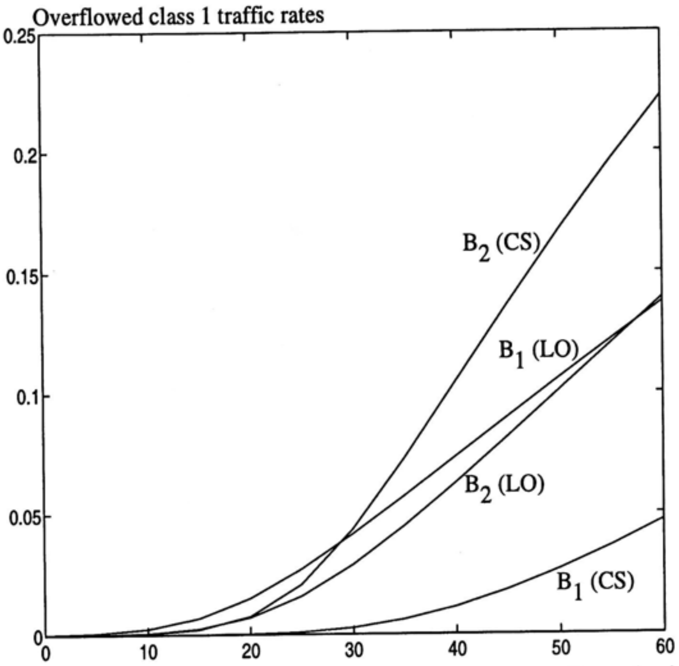
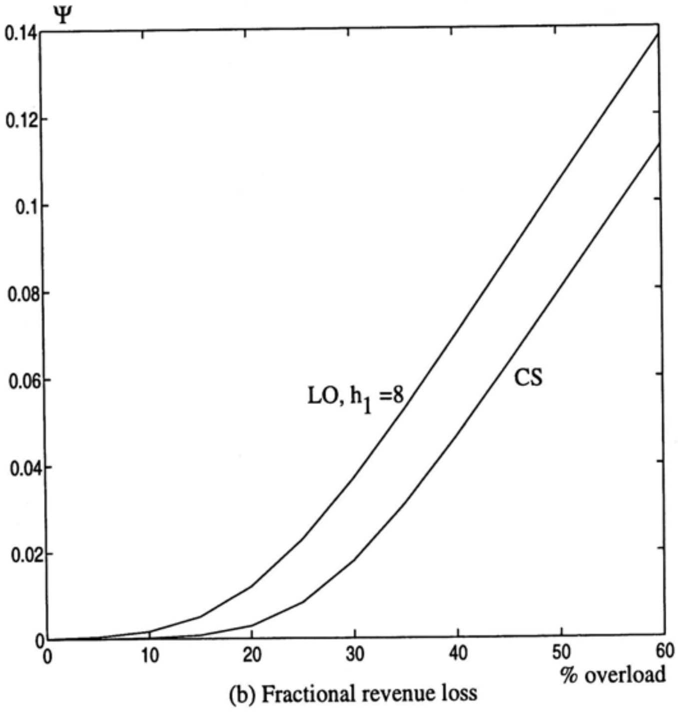


Figure 3.5: Overflowed class 1 traffic rate against $n_1$

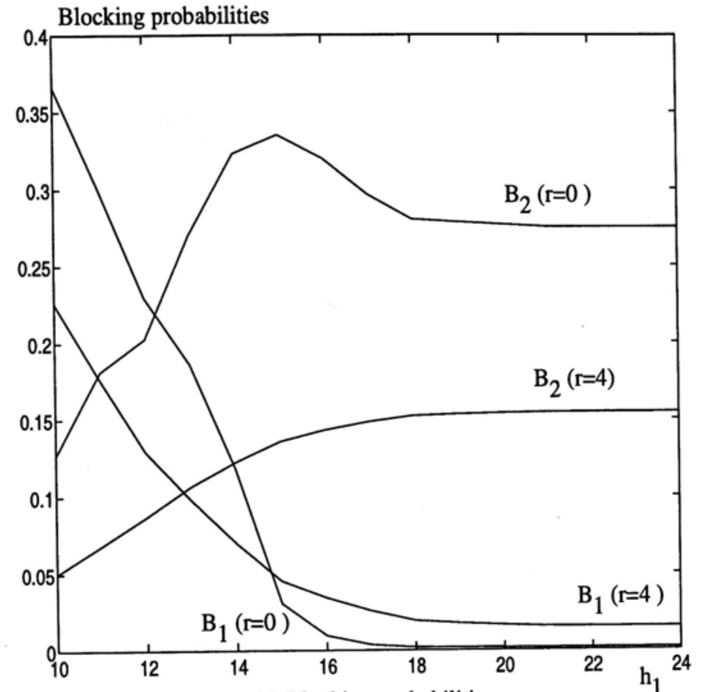


(a) Blocking probabilities

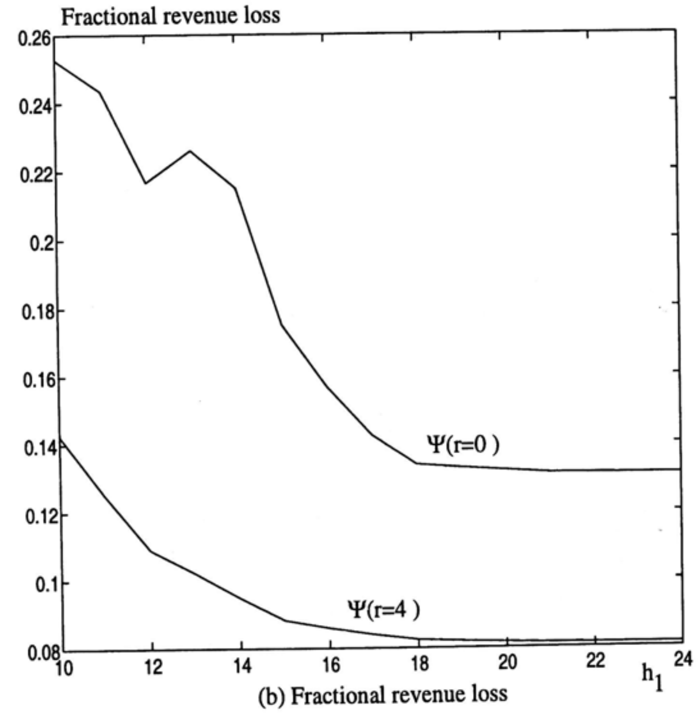


(b) Fractional revenue loss

Figure 3.6: Performance of CS and LO policies under MTB routing

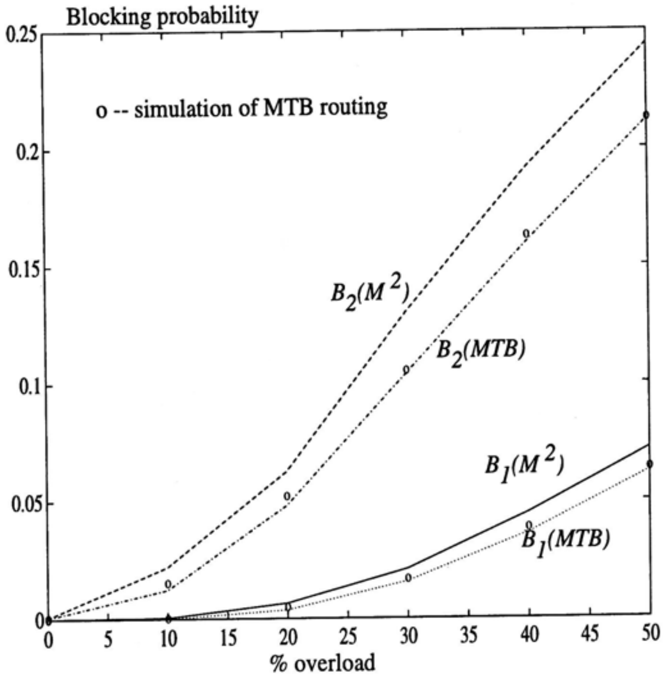


(a) Blocking probabilities

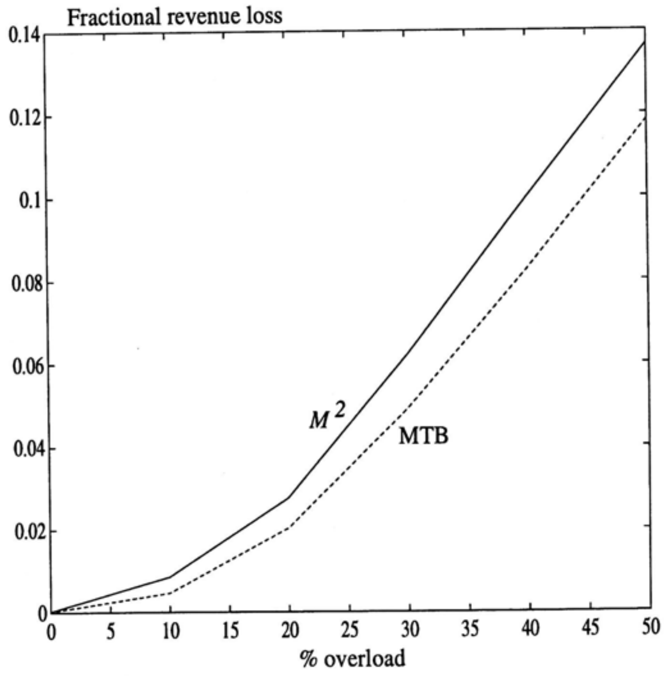


(b) Fractional revenue loss

Figure 3.7: The trading of blocking and revenue loss for LO policy under MTB routing

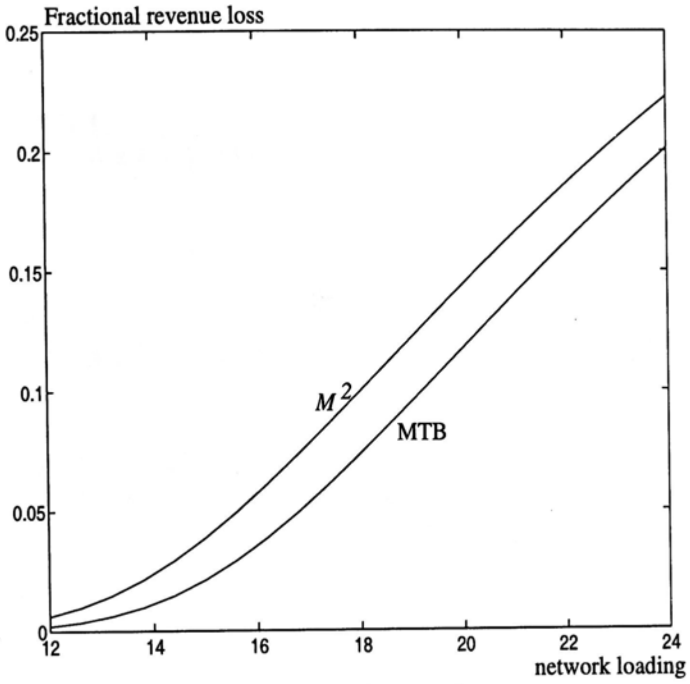


(a) Blocking probability

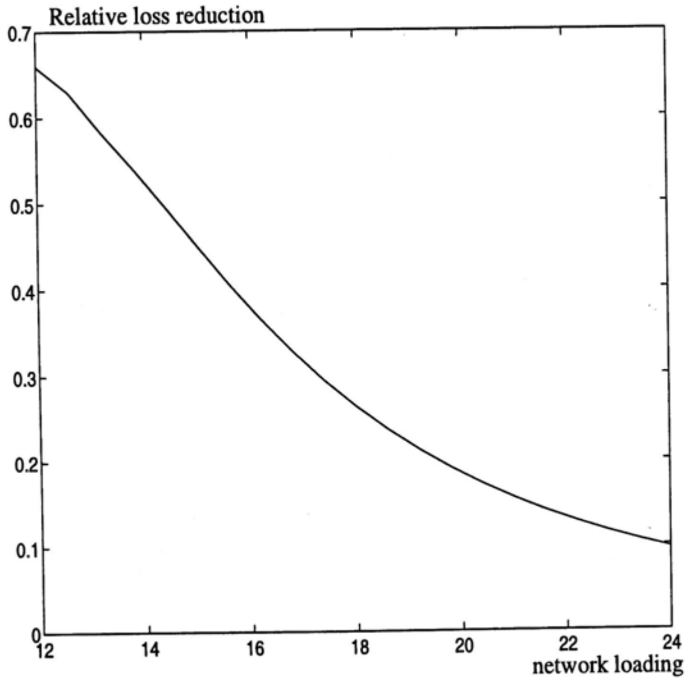


(b) Fractional revenue loss

Figure 3.8: Comparison of $M^2$ and MTB routing with $(f_1, f_2) = (1, 3)$ in a 12 node network



a) Fractional revenue loss



b) Relative loss reduction

Figure 3.9: Comparison of $M^2$ and MTB routing with $(f_1, f_2) = (1, 10)$ in a 5 node network

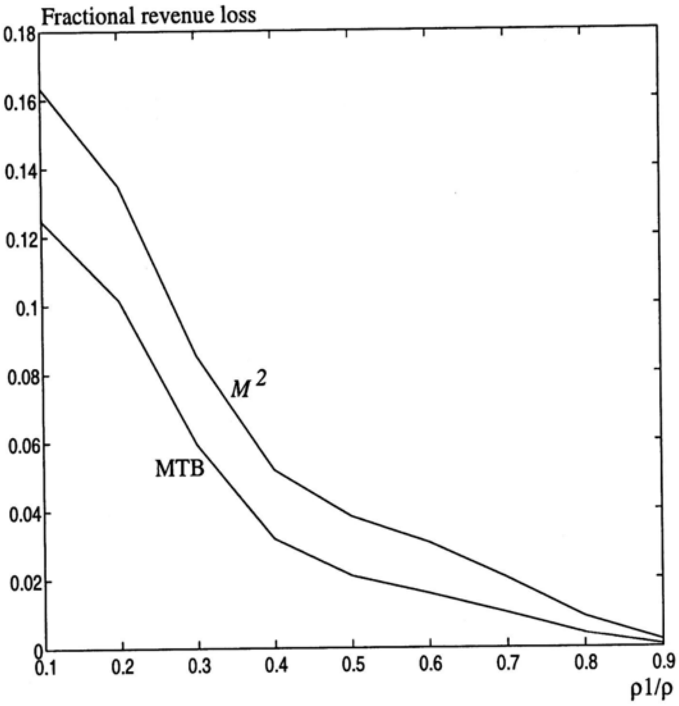


Figure 3.10: Comparison of $M^2$ and MTB routing with different traffic composition

# Chapter 4

# The Least Congestion Routing in WDM Lightwave Networks

## 4.1 Introduction

The high speed networks currently under active development should meet several stringent requirements of integrated communication applications. First, sufficient bandwidth must be provided for switched services such as multi-media conferencing, HDTV and super-computer communications. Second, low latency and low error rate are extremely important to real time interactive services. Third, reliability and availability are critical issues for some of these services such as on-line medical applications.

Wavelength Division Multiplexing(WDM) [34, 35, 41, 43] is a promising technology to tap the huge bandwidth in optical fiber links. The existing switching, processing, and storage technologies lag behind the optical transmission capabilities. This limits the effective throughput of the network and makes the switching nodes a bottleneck. This leads to the development of all-optical networks.

The design of new architectures for all-optical networks has received considerable interest in recent years. Bellcore's LAMBDANET [39], which is among the earliest, uses WDM with fixed-tuned optical transmitters. IBM Research has developed a 32-station circuit-switched WDM prototype network called Rainbow [40]. In both systems, all signals from the transmitters are combined and broadcast to all receivers using a single star-coupler.

The broadcast-and-select approach offers no wavelength reuse capability and has excessive splitting loss. This limits the geographical coverage of the network and make this architecture difficult to use in WANs. In [41, 42], systems based on fixed wavelength routing were proposed. An $NXN$ wavelength crossconnect is used to achieve full connectivity between the $N$ inputs and $N$ outputs.

Bala and Stern[36, 37] proposed the linear lightwave network which uses the linear divider and combiner(LDC) for link-selective routing. The signal of an outgoing link is a linear combination of the signals from the incoming links. This network supports point-to-point as well as multicast transmission. To simplify hardware requirement, a set of wavelengths can be grouped into a waveband [38] with each optical switch within the routing node operates on a different waveband, the number of optical switches can be reduced. The problem of finding the best routing paths however is made more complicated.

In [43], the lightnet architecture based on the lightpath(an all-optical path) concept was proposed. It eliminates processing and buffering at intermediate nodes through the establishment of lightpath between node pairs. Connections are setup on lightpaths. Such connection can support stream oriented real time traffic and bursty non-real-time traffic. Real time traffic is sensitive to delay and is best delivered by dedicated circuits. Non-real-time traffic, on the other hand, can share a channel in a packet interleaved way since it can tolerate delay. A lightpath establishment algorithm was also proposed in [43]. In [44], a switch architecture and a shortest path routing algorithm are proposed for a wavelength convertible optical network.

In this chapter, we study the routing problem in all-optical networks equipped with wavelength switches. A wavelength switch can switch a wavelength from one of the input port to an arbitrary output port if there is no contention on that output port. If a wavelength converter is used in a wavelength switch, a wavelength can be converted to another wavelength to resolve a contention. We will describe this architecture in detail in Section 4.2.

The problem of routing in this kind of all-optical WDM lightwave networks involves the setting up of a lightpath from the origin to the destination at a particular wavelength. To reduce blocking probability, the least congested path, or the path with the maximum number of remaining idle wavelength channels, would appear to be a good choice. We will present this adaptive routing rule in Section 4.3. In Section 4.4, we analyze the performance of the routing rule and a variation of it in a fully connected WDM lightwave network. We find that under low blocking (about 1%) conditions, wavelength conflicts do not significantly affect the blocking probability whereas at heavy traffic conditions, wavelength conflicts exhibits an inherent congestion control even without channel reservation. When suitable channel reservation is used, the blocking probability of a network without using wavelength converters is very close to the one with abundant wavelength converters and optimal channel reservation.

## 4.2 Architecture and Some Design Issues

A typical WDM lightwave network is shown in Figure 4.1. Connection requests are set up by establishing circuits from the origin to the destination. A circuit occupies a wavelength channel when its signal passes through a link. At each node, optical switching functions are performed. These include connecting a circuit from an incoming link to an outgoing link and converting the wavelength of a circuit if needed. Although each wavelength channel can accommodate several circuits by means of subcarrier multiplexing, we assume for simplicity that each wavelength channel accommodates only one circuit.

The wavelength assigned to a circuit can be fixed or can be changed along the path. The latter requires switching nodes to possess wavelength conversion capability. If the wavelength used is fixed throughout the path, the switches complexity and the implementation cost can be reduced. However, connection requests are blocked if there

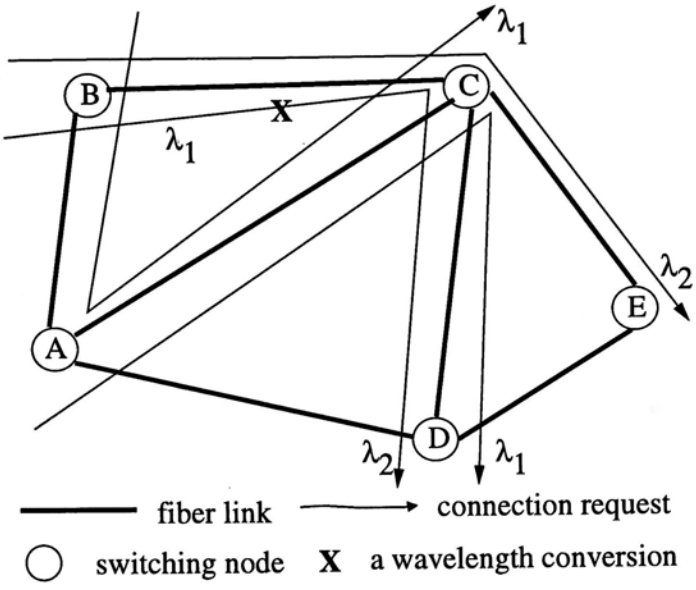


Figure 4.1: The architecture of a WDM lightwave network

is no common free wavelength channel on the links constituting the path.

As shown in Figure 4.1, three connections are setup in the network on node pairs BC, AD and BE. The paths being used are indicated by the arrows. Two wavelength channels are supported on each fiber. Suppose a connection is to be setup on node pair BD and only a two-hop path BCD is available for call setup on this node pair, then if switching nodes have no wavelength conversion capability, the connection request will be blocked. This blocking can be avoided if a wavelength converter is installed at node C to convert the $\lambda_1$ to $\lambda_2$.

Several hardware-related issues must be examined in this network. These include the amplifiers required to compensate for the path, the switching and the multiplexing losses, the multiplexing device technology and the wavelength switches. To compensate for losses, the Erbium-doped fiber amplifiers operated at 1550 nanometer window are considered as a promising technology[45]. The high gain($> 40$ dB), high power($> 100$ mW) and the near-ideal noise performance is unparalleled by any competing amplifier technology. The only restriction on the number of wavelength channels that can be supported is the limited spectral width. Moreover, the non-uniform gain profile that spans roughly from 1525 to 1565 nm must also be compensated.

The wavelength division multiplexing technique allows thousand of wavelength channels each operated at a moderate speed, say 1Gb/s to be multiplexed on the low loss window of a single mode fiber link and transmitted for a long distance [46]. To maximize the number of channels, WDM requires narrow spectral width lasers, typically the distributed feedback (DFB) semiconductor lasers or distributed Bragg reflector (DBR) semiconduction lasers [47], optical filter and optical multiplexer/demultiplexer to distinguish between wavelength channels. Further multiplexing within each wavelength channel, such as subcarrier multiplexing is possible. A detail discussion of the WDM technology can be found in [41].

The wavelength switch is a critical element in this architecture. A possible implementation of a 2X2 wavelength switch is shown in Figure 4.2(a). It has two incoming and two outgoing fibers. Each fiber supports two wavelength channels. The switch can be divided into three stages: wavelength demultiplexers(ODEMUX), space division optical switches(OSS) and wavelength multiplexers(OMUX). Wavelength demultiplexers and multiplexers can be implemented by means of a diffraction grating. The middle space switch can be implemented by using optical crossconnects or acoustic tunable optical filters[48].

To illustrate the function of this switch, suppose a circuit carried by wavelength channel $1(\lambda_1)$ in incoming fiber 1 is to be switched to outgoing fiber 2. $\lambda_1$ is first separated from the other wavelength by ODEMUX, then it is switched spatially to output port 2 of an OSS in the middle stage. An OMUX in the final stage then multiplexes $\lambda_1$ and other wavelength channel into the outgoing link 2. If $\lambda_1$ in outgoing link 2 has already carried a circuit, then this connection cannot be established.

A second implementation of a 2X2 wavelength switch with one wavelength converter is shown in Figure 4.2(b). The added wavelength converter can resolve one wavelength conflict. An optical wavelength converter can be implemented by absorption saturation of semiconduction laser diode [49], four-wave mixing [50] or gain saturation of optical amplifier [51]. Signal which needs to be converted to another wavelength will be switched to the wavelength converter by an OSS immediately after the ODEMUX. Having been converted to another wavelength, it will be switched to the desired OSS

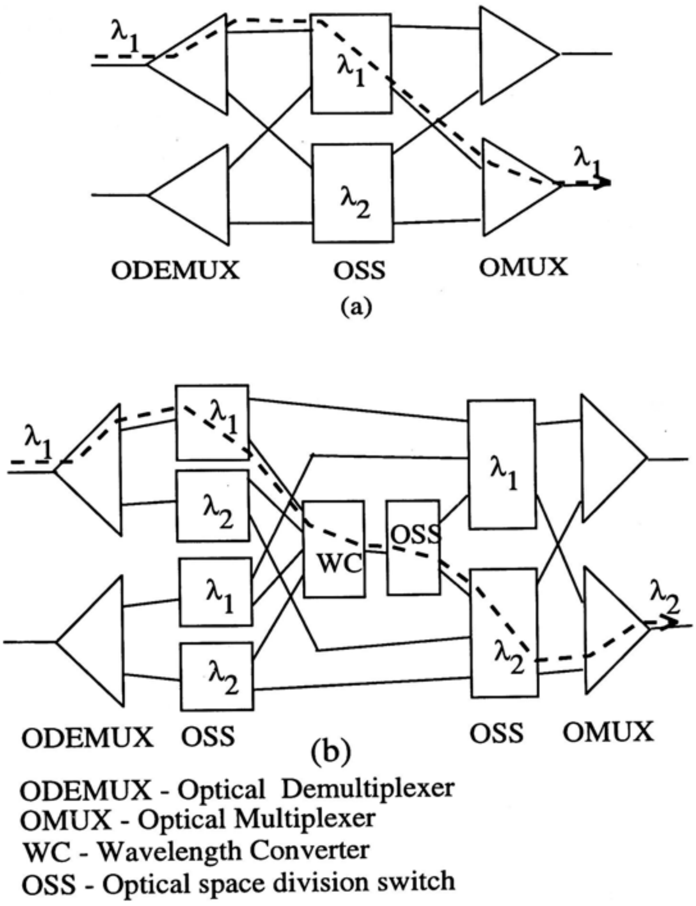


Figure 4.2: Two implementations of a wavelength routing switch

module and then multiplexed into the required outgoing link.

We now illustrate how this switch can overcome the wavelength conflict. Suppose wavelength channel $\lambda_2$ in outgoing link 2 is free, then $\lambda_1$ in incoming link 1 can be switched to the wavelength converter by the space switch and get converted to $\lambda_2$. The optical signal will be transmitted to OSS and then switched to an OMUX in the final stage and then multiplexed into the outgoing link 2.

Based on this switch architecture, a connection request can be established on an all-optical path. To minimize the end-to-end blocking probability, alternate path routing is used on the network. Since alternate routing introduces instability [30], a few wavelength channels are usually reserved for direct traffic use only. The number of channels being reserved, $r$ is referred to as the channel reservation parameter.

A link is admissible to alternate connection requests if the number of free wavelength channels is larger than the channel reservation parameter, $r$. An alternate path is admissible if i) *all links constituting the alternate path are admissible* and ii) *there are at least one common free wavelength in all the links or there are at least one free wavelength converter.* Therefore, a connection request on an alternate path can be blocked by *the channel reservation*, or *the wavelength conflict and the depletion of wavelength converters.*

## 4.3 The Routing Rule

Consider a node pair $p$ in a WDM lightwave network. Let $\mathcal{S}_p$ be the set of paths connecting $p$. The degree of congestion of a link is determined by the number of free channels. The larger the number of free channels, the less congested the link is. When a connection request is made, the number of free channels of all links constituting all the admissible paths are collected first. The number of free channels of a $k$-link path in $\mathcal{S}_p$ with free channels $x_1, x_2, \ldots x_k$ is defined as $\min_k(x_k)$. In words, it is the minimum of the number of free channels of all links constituting the path. The connection request will be established on the path having the largest number of free channels so defined. If there is a tie, the path having the largest number of free channels on the second most congested links is chosen from the tie set. If two paths with different hop-count have the same occupancies on their links, the smaller hop-count one is chosen. Further tie is broken arbitrarily. A free wavelength channel is then chosen randomly from the pool of wavelength channels on the chosen path to carry the call. Since this routing rule seeks the least congested path to route a connection request, we call it the Least Congestion (LC) routing. In Section 4.4.2, we will analyze this routing rule.

To illustrate the routing rule, suppose a particular node pair has four admissible paths A,B,C and D with number of free channels $\{2\}, \{3,4\}, \{2,4\}, \{3,4,5\}$ respectively. The length of the path is given by the number of elements in the set. The number of free channels of these four paths are 2, 3, 2 and 3 respectively. Path B and D have the largest number of free channels. This tie is broken by considering the number of free channels on the second most congested link of them, which are 4 for both of them. Since path B is a two-hop path which is shorter than path D, so it is chosen as the routing path.

Alternatively, we can always give priority to shorter paths. In this way, we first divide the set of admissible paths into different subsets according to their number of hop-count. Let $\{X_i\}$ denotes the set of admissible paths with $i$ hops. A connection request is established on a path selected from the first non-empty path set, starting from $\{X_1\}$. When the $X_i$ path set is considered, the path having the maximum number of free channels on the most congested link is selected as the routing path. If there is a tie, the same criterion is applied on the second most congested link and so on until a routing path is identified. Consider the previous example, the connection request will be established on path A, which is the direct link. If path A has no free channel, then the set of two-hops paths B and C will be considered. Since path B has 3 free channels while path C has only 2, so the call is established on path B. To complete the call setup, a free wavelength channel on the path is assigned randomly to carry the call. This is a version of the LC routing with priority given to shorter paths. We will analyze it in Section 4.4.3.

## 4.4 Analysis of the LC Routing Rule

### 4.4.1 Fixed Point Model

We first describe the analytic model being used. Consider an asymmetric network with $J$ links, assume links are independent and overflow traffic are Poisson(these two assumptions are widely used in the literatures). Each link can be modeled by a M/M/N/N queue. The state is the number of on-going connections. Let there be $M_j$ wavelength channels in link $j$, and let $r_j$ wavelength channels be reserved for link $j$ traffic only. The state dependent arrival rates on link $j$ is given by,

$$\Lambda_j(x) = \begin{cases} 0 & x = M_j \\ A_j(x) & 0 \le x < M_j \end{cases} \tag{4.1}$$

where $A_j(x)$ is the amount of traffic given that link $j$ is at state $x$, which depends on the routing rule being used. Deriving $A_j(x)$ is the aim of Section 4.4.2 and 4.4.3.

Let $\mu^{-1}$ be the mean call holding time. Using the set of state-dependent arrival rates, the limiting state probability of link $j$ at state $x$, $\pi_j(x)$ can be computed for each links by the global balance equation,

$$\begin{aligned}\left[\Lambda_j(x) + x\mu\right]\pi_j(x) = \\ (x+1)\mu\pi_j(x+1) + \Lambda_j(x-1)\pi_j(x-1)\end{aligned} \tag{4.2}$$

with the understanding that $\pi_j(x) = 0$ for $x < 0$ and $x > M_j$.

Let $\mathcal{A}$ denotes the set of $\Lambda_j(x)$ and $\mathcal{P}$ denotes the set of $\pi_j(x)$. Assuming an initial set of $\mathcal{P}$, we can use (4.1) to compute a set of $\mathcal{A}$. Using this new set of $\mathcal{A}$, we can compute a new set of $\mathcal{P}$ from (4.2). This iterative procedure continues until a certain convergence requirement is achieved for $\mathcal{P}$. In the examples quoted in Section 4.5, the iteration stops when all $\pi_j(x)$s achieve a relative error of less than $10^{-8}$.

### 4.4.2 Without Direct-link Priority

In this subsection, we analyze the LC routing with no direct-link priority in an $N$-node fully connected symmetric WDM network with $M$ wavelength channels on each link and $K$ wavelength converters in each node. Each node pair has a direct link and $L = N - 2$ two-hops alternate paths. This analysis can be generalized to apply on asymmetric network with arbitrary topology as well as on multirate loss network[27].

Before we proceed to derive the state dependent arrival rate, we need to compute the probability that *all wavelength converters are busy.* This probability is needed for the computation of the probability that *an alternate path is blocked.* The exact derivation is very complicated. In the following, we introduce an approximation method.

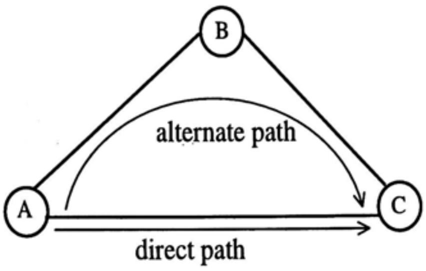


Figure 4.3: An alternate of a direct link in the network

Consider a particular alternate path ABC of a node pair AC as shown in Figure 4.3. Suppose that the state of its links are AB and BC are $i$ and $j(\leq i)$ and denote the rate of overflowing connection request to path ABC as $s(i,j)$. The derivation of it will be given shortly. If path ABC is blocked due to wavelength conflict, one of the wavelength converters in node B will be used. When the connection is released, the wavelength converter is freed. Therefore we can model this situation by a M/M/K/K queue (Figure 4.4). The state is the number of busy wavelength converters. Since there are $\binom{N-1}{2}$ node pairs that can overflow traffic through node B, the arrival rate of the queuing model, $\kappa$, is given by

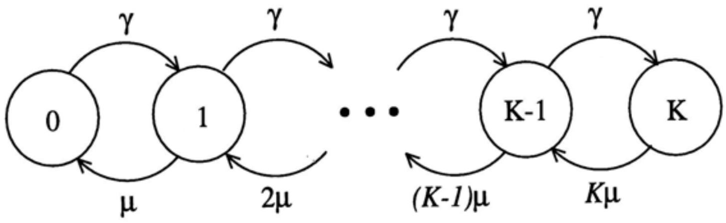


Figure 4.4: A M/M/K/K queuing model for the wavelength converter

$$\kappa = \binom{N-1}{2} \sum_{i=0}^{M-r-1} \sum_{j=0}^{i} s(i,j) H(i,j)$$

where $H(i,j)$ is the probability that *the two links of a path is in state $i$ and $j$ and that the path is blocked due to wavelength conflict* and is given by

$$H(i,j)=\begin{cases} 0 & i+j<M \\ \frac{2\pi_i\pi_j i!j!}{(i+j-M)!M!} & i+j\geq M, i\neq j \\ \frac{(\pi_i i!)^2}{(2i-M)!M!} & i+j\geq M, i=j \end{cases}$$

Solving this model, the probability that *all wavelength converters are busy*, $\phi_K$ is given by,

$$\phi_K=\left\{\sum_{i=0}^{K}\frac{K!}{i!}\left(\frac{\mu}{\kappa}\right)^{K-i}\right\}^{-1}$$

Let the probability that *path ABC is blocked* be $C(i,j)$ and let $E_1$ be the event $\{M-i\leq r \text{ or } M-j\leq r\}$. $C(i,j)$ is given by

$$C(i,j)=\begin{cases} 1 & E_1 \\ 0 & i+j<M \text{ and } \overline{E_1} \\ \frac{i!j!}{(i+j-M)!M!}\phi_K & i+j\geq M \text{ and } \overline{E_1} \end{cases}$$

Let $D(x,y)$ be the probability that *the two links of an alternate path are in state $x$ and $y$ and that the path is admissible.* Then, with random assignment of wavelength channels, $D(x,y)$ is given by,

$$D(x,y)=\begin{cases} 2\pi_x\pi_y[1-C(x,y)] & x\neq y \\ \pi_x^2[1-C(x,y)] & x=y \end{cases}$$

Consider an alternate path with link occupancies $i$ and $j$(assume $i\geq j$). Define the two independent events $E_2$ and $E_3$ for a given pair of thresholds $x$ and $y(y\leq x)$ as follows

$E_2$: $i>x$

$E_3$: $i=x$ and $j>y$

Let $F(x,y)$ be the probability of the union of all these events. Since all $E_2$ and $E_3$ are independent,

$$\begin{aligned} F(x,y) &= 1 - Prob\{i < x\} - Prob\{i = x, j \le y\} \\ &= 1 - \sum_{i=0}^{x-1} \sum_{j=0}^{i} D(i,j) - \sum_{j=0}^{y} D(x,j) \end{aligned}$$

Given that AC is full and the two links of path ABC are in states $x$ and $y (x \ge y)$ and path ABC is admissible, the probability $G(\alpha, x, y)$ that *there are* $\alpha - 1$ *other such alternate paths and the remaining* $(L - \alpha)$ *alternate paths are neither admissible nor better than* $ABC$ is given by

$$G(\alpha, x, y) = \begin{pmatrix} L-1 \\ \alpha-1 \end{pmatrix} D(x,y)^{\alpha-1} F(x,y)^{L-\alpha}$$

Since no priority is given to the direct link, the direct traffic is not always established on the direct link even if it has idle capacity. If a direct link is at state $i$, or has $M - i$ free wavelength channels, then a connection request will be established on it if *all the alternate paths have less than or equal to* $M - i$ *free channels* or *have more than* $M - i$ *free channels and the path is not admissible.*

Let $Q(x)$ be the probability that i) *the number of free channels of a two-hop path is less than or equal to M-x* or ii) *the number of free channels of a two-hop path is greater than M-x and the path is not admissible.* Then,

$$\begin{aligned} Q(x) &= 1 - Prob\{number\, of\, free\, channels > M - x \\ &\quad\; and\ the\ path\ is\ admissible\} \\ &= \begin{cases} 1 - \sum_{i=0}^{x-1} \sum_{j=0}^{i} D(i,j) & x > 0 \\ 1 & x = 0 \end{cases} \end{aligned}$$

Therefore the probability that *a connection is established on the direct link given that its occupancy is* $x$ is $[Q(x)]^L$.

If the two links of an admissible alternate path ABC are in state $x$ and $y (\le x)$, the amount of overflowed traffic that get routed from path AC, denoted by $s(x, y)$, is given by

$$s(x,y) = \sum_{\alpha=1}^{L} \left(\frac{\lambda}{\alpha}\right) G(\alpha,x,y) \left[\sum_{i=x+1}^{M} \pi_i\right]$$

The term in $[\cdot]$ is the probability that *the number of free channels of the direct link is smaller than the number of free channels of the alternate path.*

Given that link AB is in state $x (0 \leq x \leq M-r-1)$, the total overflowed traffic $B(x)$, obtained by removing all the conditioning is

$$B(x) = 2L \sum_{i=0}^{M-r-1} [1 - C(x,i)] s(\max[x,i], \min[x,i]) \pi_i \tag{4.3}$$

At state $x$, the call arrival rate including the direct and overflowed traffic is therefore

$$\Lambda(x) = \begin{cases} \lambda \cdot [Q(x)]^L + B(x) & 0 \leq x < M-r \\ \lambda \cdot [Q(x)]^L & M-r \leq x < M \end{cases} \tag{4.4}$$

Knowing the state probabilities, the end-to-end blocking probability is given by

$$\pi_M \left[1 - \sum_{i=0}^{M} \sum_{j=0}^{i} D(i,j)\right]^L \tag{4.5}$$

where $[\cdot]$ in (4.5) is the probability that an alternate path is blocked.

### 4.4.3 With Direct-link Priority

With direct-link priority, a connection is always established on the direct link if it is not full. Therefore, if the two links of an admissible alternate path ABC are in states $x$ and $y$, the amount of overflowed traffic that get routed from path AC, denoted by $s(x,y)$, is given by

$$\begin{aligned} s(x,y) &= \sum_{\alpha=1}^{L} \frac{\lambda \pi_M}{\alpha} G(\alpha,x,y) \\ &= \frac{\lambda \pi_M \{[D(x,y) + F(x,y)]^L - F(x,y)^L\}}{L D(x,y)} \end{aligned} \tag{4.6}$$

Given that link AB is in state $x$, the total overflowed traffic, $A(x)$, is given by (4.3) with $s(x, y)$ replaced by (4.6).

At state $x$, the call arrival rate including direct and overflowed traffic is therefore

$$\Lambda(x) = \begin{cases} \lambda + A(x) & 0 \leq x < M - r \\ \lambda & M - r \leq x < M \end{cases} \tag{4.7}$$

Therefore the fixed point iteration method described in Section 4.4.1 can be used to obtain all $\pi_i$ and $\Lambda(x)$. The end-to-end call blocking probability is given by (4.5).

## 4.5 Performance Comparisons

We study the performance of the routing rule on a seven-nodes fully connected optical network with thirty wavelength channels on each link. We first focus on the LC routing with priority given to the direct link. Figure 4.5 shows the end-to-end blocking probability as a function of direct traffic rates for the two cases with and without the use of wavelength converters. For the former case, we assume that the number of wavelength converters is sufficiently large so that connection requests will not be blocked due to wavelength conflict. The channel reservation parameters used are chosen to minimize the end-to-end blocking probability. If channel reservation is not used(data not shown), the network exhibits unstable behavior with much larger blocking probability. For the latter case, ie, with no wavelength converters, channel reservation is not used. We observe that at heavy loading, the blocking probability is only a bit larger than the counterpart. This shows that wavelength conflict has a build-in flow control mechanism at heavy loading. At small blocking(less than 2%), we find that the wavelength conflict causes only insignificant increase in the blocking probability.

Figure 4.6 shows the state dependent traffic rates. We observe that without wavelength converters, the overflow rates at low occupancy is higher than the case with wavelength converters. The overflow rates decrease at higher occupancy, which shows that less overflow connections are established because of the inherent alternate path blocking due to wavelength conflict.

Figure 4.7 shows a decomposition of the overall end-to-end blocking probability into the direct and alternate blocking probability. It is found that without wavelength converters, there is a large blocking of the alternate calls. As a result, the direct call is protected in heavy loading even without channel reservation.

Figure 4.8 shows the blocking for different number of wavelength converters without channel reservation. In the two figures shown, we observe that at light loading(low traffic rate), the use of wavelength converters can reduce blocking probability. However, at heavy loading, we observe that the blocking probability increases with the number of wavelength converter. This can be appreciated intuitively. At light loading, the use of alternate routing paths can reduce the blocking probability [30]. Without wavelength converters, some alternate path traffic is blocked by wavelength conflict. The wavelength converters can resolve wavelength conflict and consequently reduce blocking probability. As network loading increases, the blocking of direct traffic increases and the alternate traffic rate increases accordingly. Without the wavelength converters, an inherent alternate path blocking mechanism mentioned previously blocks most of the alternate path traffic. The use of the wavelength converters, on the other hands, allows more alternate path traffic and causes network instability.

Next we study the blocking performance if the direct link have no priority. Figure 4.9 shows the blocking probability of a twelve nodes network with twenty wavelength channels in each link. Here channel reservation is not used. We find that the routing rule that gives priority to direct links always outperforms its counterpart. This is because the routing rule that gives no priority to direct links can establish connections on alternate paths even when the direct link has idle capacity. As load increases, this increased amount of alternate traffic increases the direct blocking probability, which in turns induces more alternate traffic and higher overall blocking.

Table 4.1 shows the end-to-end blocking probability of both cases with the use of optimal channel reservation parameter. The network parameters of Figure 4.5 are begin used. The optimal channel reservation parameter is found by minimizing the end-to-end blocking probability. We see that with a suitable channel reservation parameter, the increase in blocking probability(as shown in Figure 4.5) is negligible at heavy loading for the case without wavelength converters.

## 4.6 Concluding Remarks

In this chapter, we have analyzed the effect of the wavelength conflict in fully connected WDM lightwave networks employing wavelength routing switches. We find that the use of wavelength converters causes insignificant reduction in blocking probability at light loading whereas at heavy loading, the wavelength conflict causes an inherent alternate path blocking function. The use of wavelength converters to resolve this conflict will only increase blocking probability if channel reservation parameter is not set appropriately.

We also find that without any wavelength converter but with the use of optimal channel reservation parameters, the end-to-end blocking probability will be very close to the case where abundant wavelength converters and optimal channel reservation parameters are used.

We analyze two versions of Least Congestion routing and find that the one giving direct-link priority outperforms that with no direct-link priority. Therefore, during alternate path selection process, admissible paths with less hop-count should be considered first.

Table 4.1: BLOCKING PROBABILITY UNDER OPTIMAL CHANNEL RESERVATION

| $\lambda$ | With wavelength converters | Without wavelength converters |
|---|---|---|
| 24.0 | 0.0041 | 0.0064 |
| 24.5 | 0.0140 | 0.0160 |
| 25.0 | 0.0254 | 0.0265 |
| 25.5 | 0.0363 | 0.0373 |
| 26.0 | 0.0476 | 0.0480 |
| 26.5 | 0.0584 | 0.0589 |
| 27.0 | 0.0689 | 0.0693 |
| 27.5 | 0.0795 | 0.0797 |
| 28.0 | 0.0900 | 0.0900 |
| 28.5 | 0.1001 | 0.1002 |
| 29.0 | 0.1100 | 0.1101 |
| 29.5 | 0.1198 | 0.1199 |
| 30.0 | 0.1296 | 0.1296 |

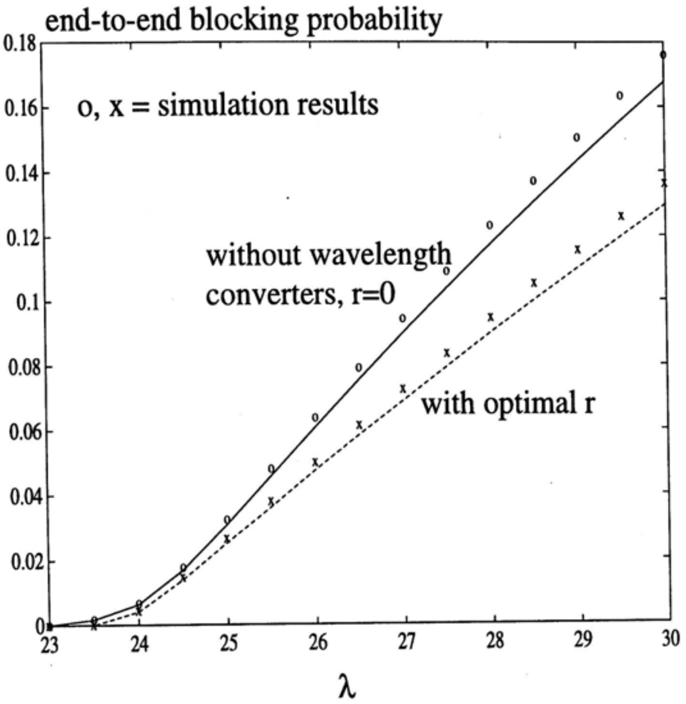


Figure 4.5: Blocking as a function of occupancy

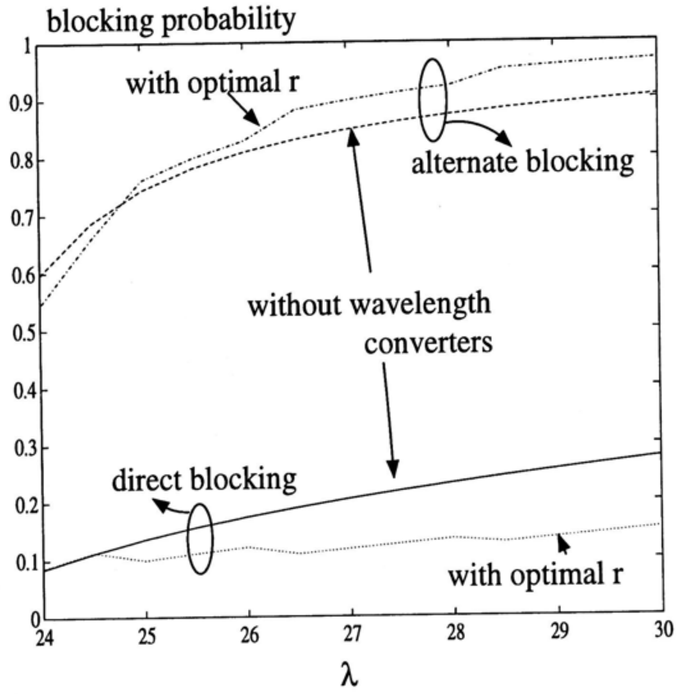


Figure 4.7: Decomposition of the overall blocking

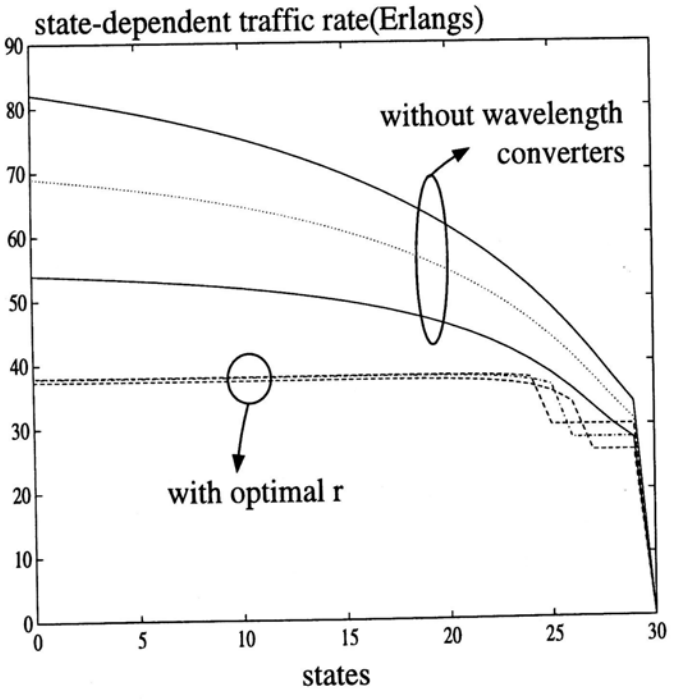
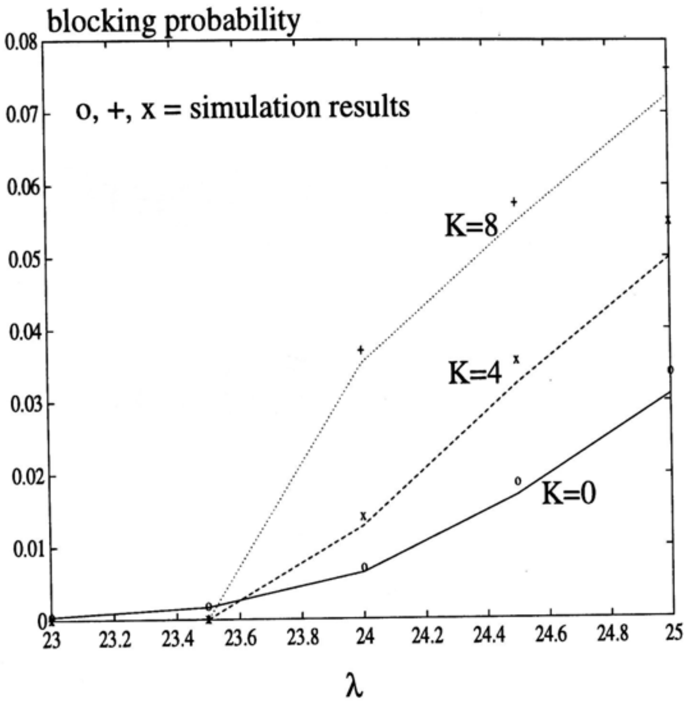
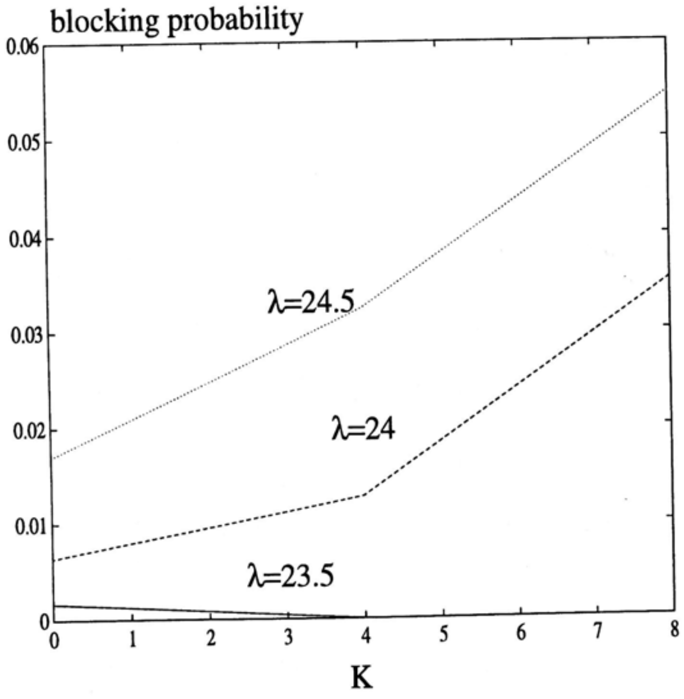


Figure 4.6: State-dependent traffic rates



(a) Blocking versus $\lambda$ for different K



(b) Blocking versus K with L=5

Figure 4.8: Blocking with the use of wavelength converters

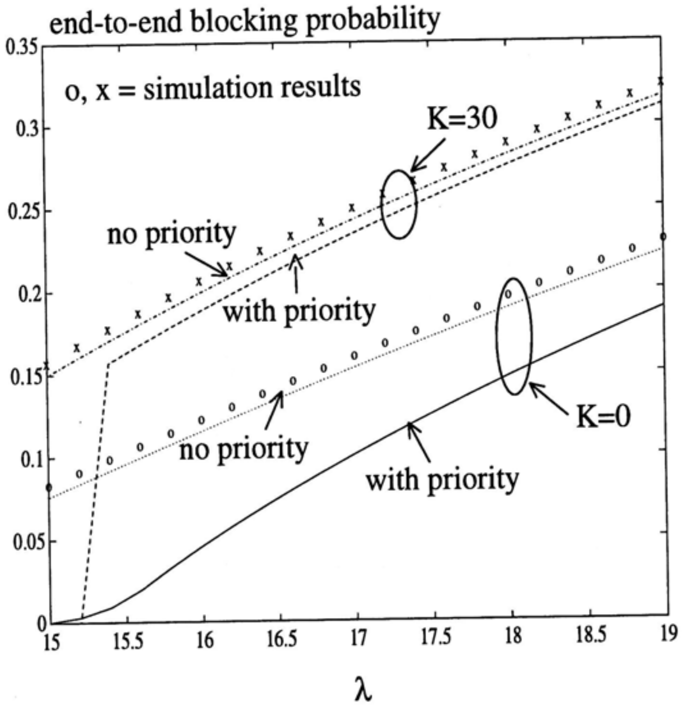


Figure 4.9: Blocking for the two versions of LC routing

# Chapter 5

# Conclusions and Future Work

The goal of this study is to study some of the problems in routing and call admission in integrated-services networks. We have employed the multirate loss network to model an integrated-services network and have used the reduced load approximation to solve the limiting state probabilities for different routing schemes and call admission policies.

We first focus on the problem of call packing in integrated networks. Routing rules based on the traffic packing and least congestion principle are analyzed. We find that the routing rule based on traffic packing cannot reduce the blocking probability and the revenue loss of the network when compared to that using the least congestion principle. We then concentrate on the problem of protecting network resources from being dominated by a particular class of calls. Five different call admission policies, namely the CS, LO, GB, EB and DP policies are analyzed under the LC and TP routing rules. We show that the LO, GB and EB policies, which is based on preemptive blocking principle, can protect a particular class of calls if the bandwidth reservation parameters are chosen appropriately; however, they tend to increase the revenue loss. The DP policy is found to give a significantly smaller blocking probability and revenue loss when compared to the CS policy.

Next, we study two versions of Least Congestion routing. We have analyzed the aggregated MTB and the $M^2$ routing under symmetric traffic conditions in fully connected networks. MTB measure incorporates the traffic characteristics of different classes of services, link bandwidth and occupancy information. It reflects more accurately the link busy status. Numerical results show that the MTB routing gives a better performance than the $M^2$ routing. To reduce signaling traffic, the link congestion measure are lumped into aggregate states. We have analyzed the aggregated version of these two routing schemes and have found that with properly designed aggregation, they can perform closely to the non-aggregated schemes.

Finally, we have analyzed the Least Congestion routing in all-optical WDM lightwave networks. We find that the wavelength conflict inherent in the wavelength switch exhibits blocking to alternate traffic. The use of wavelength converters can resolve wavelength conflict. On the other hand, it also increases the alternate traffic rates which is not desirable in high loading. We find that under the uniform traffic condition and with the use of optimal channel reservation parameter, the blocking probability of a fully connected network without wavelength converters is very close to the one having abundant wavelength converters. It is also found that the Least Congestion routing with direct-link priority outperforms the one without direct-link priority. This means that in the alternate path selection process, the routing rule should consider shorter paths first.

## 5.1 Future Work

1. We find that the use of MTB routing in symmetric fully connected networks with uniform traffic provides a small gain in capacity. In real time operation, MTB values are computed from the direct and alternate traffic rates measured. Error in measurement or estimation of traffic rates may affect the performance of MTB routing. Therefore it is worth to study the routing rule in asymmetric network with non-uniform traffic condition.

2. Aggregation of link congestion measure can reduce signaling traffic. However, inappropriate aggregation may deteriorate the performance of LC routing. It is worth to study an optimal or suboptimal aggregation for MTB and $M^2$ routing.

3. The use of wavelength converters causes insignificant reduction in blocking probability at light loading whereas at heavy loading, the wavelength conflict possesses an inherent alternate path blocking function. The use of wavelength converters to resolve this conflict will only increase blocking probability. However, it is interesting to study whether this phenomenon occurs in other scenarios. For example, the use of wavelength converters in asymmetric network having traffic hot-spots is a typical case in real network. Another example is the use of fixed routing in asymmetric networks. Since node pairs may not have a direct link, the use of wavelength converters can reduce the blocking of the traffic in these node pairs.

4. It would also be interesting to consider a version of Least Congestion routing without direct-link priority in asymmetric networks having a traffic hot-spot. Releasing the priority on direct links allows a connections to be established in the under-loaded region of the network. This act may improve the overall blocking probability.

# Bibliography

[1] J.M. Kraushaar, "Fiber deployment update-End of year, 1990," *Federal Commun. Commission, Common Carrier Bureau,* Washington, DC 20554, March 1991.

[2] Stephen B. Alexander *et al.*,, "A Precompetitive Consortium on Wide-Band All-Optical Networks," *IEEE JLT*, vol 11, no. 5/6, May/June 1993.

[3] R. Cooke, "Intercity limits: looking ahead to all-digital networks and no bottlenecks," *Data Communications*, March 1984.

[4] R. Handel, "Evolution of ISDN towards broadband ISDN," *IEEE Network*, January 1989.

[5] I.S. Gopal and T.E. Stern, "Optimal Call Packing Policies in an Integrated Services Environment," *Conference on Information Sciences Systems*, Johns Hopkins University, 1983.

[6] G.J. Foschini, B. Gopinath, and J.F. Hayes, "Optimum Allocation of Servers to Two Types of Competing Customers," *IEEE Trans. on Commun.,* COM-29, 1051-1055, 1981.

[7] E. Brockmeyer, H.L. Halstrom, and A. Jensen, *The Life and Works of A.K. Erlang,* The Copenhagen Telephone Company., Copenhagen, 1984.

[8] J.W. Roberts, "A Service System with Heterogeneous User Requirements," *Performance of Data Communications Systems and their Applications,* 423-431, 1981.

[9] D. Tsang and K.W.Ross, "Algorithm for determining exact blocking probabilities in tree networks," *IEEE Trans. on Commun.,*

[10] S.S. Katz, "Statistical Performance Analysis of a Switched Communications Network," *Proc. 5th ITC*, 566-575, New York, Rockefeller University, 1967.

[11] W. Whitt, "Blocking When Service is Required from Several Facilities Simulataneously," *AT&T Technical Journal,* 1807-1856, 1985.

[12] F.P. Kelly, "Blocking Probabilities in Large Circuit-switched Networks," *Adv. Apply Prob.,* 473-505 1986.

[13] G.R.Ash, et al. "Real-Time Network Routing in a Dynamic Class-of-Service Network," *Proc. IEEE Globecom*, 1992.

[14] J.S.Kaufman, "Blocking in a Shared Resource Environment," *IEEE Trans. on Commun.,* vol. com-29, no. 10 Oct. 1981.

[15] B. Kraimeche, M.Schwartz "Analysis of Traffic Access Control Strategies in Integrated Service Networks," *IEEE Trans. on Commun.*, vol. com-33, Oct. 1985.

[16] D.Tsang, K.W.Ross "The Stochastic Knapsack problem," *IEEE Trans on Commun,* vol. 37, no. 7, July 1989.

[17] K.W.Ross, D.Tsang "Optimal Circuit Access Policies in an ISDN Environment: A Markov Decision Approach," *IEEE Trans. on Commun.,* vol. 37, Sept. 1989.

[18] Shun-Ping Chung, Keith W. Ross, "Reduced Load Approximations for Multirate Loss Networks," to appear in *IEEE Trans. on Commun.*

[19] Weilin Wang, Tarek N. Saadawi, "Trunk Congestion Control in Heterogeneous Circuit-Switched Networks," *IEEE Trans. on Commun.,* vol. 40, no. 7, Jul. 1992.

[20] Sanjay Gupta, Keith W. Ross, Magda El Zarki, "Routing in Virtual Path Based ATM networks," *Proc. IEEE Globecom,* 1992.

[21] Zbigniew Dziong, Lorne Mason, "An Analysis of Near Optimal Call Admission and Routing Model for Multi-Service Loss Networks," *Proc. IEEE Infocom,* 1992.

[22] Ren-Hung Hwang, James F. Kurose, Don Towsley, "State Dependent Routing for Multirate Loss Networks," *Proc. IEEE Globecom,* 1992.

[23] A.I. Elwalid, D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High-Speed Networks," *IEEE/ACM Transactions on Networking*, June 1993.

[24] G. Kesidis, J. Walrand and C. S. Chang, "Effective Bandwidth for Multiclass Markov Fluids and Other ATM sources," *IEEE/ACM Transactions on Networking*, August 1993.

[25] Eric Wing-ming Wong, Peter Tak-shing Yum and Kit-man Chan, "Analysis of the $M$ and $M^2$ Routings in Circuit-Switched Networks," *Proc. IEEE Globecom,* 1992.

[26] Kit-man Chan, Tak-shing Yum, "Analysis of Adaptive Routing Schemes in Multirate Loss Networks," *Proc. IEEE Globecom,* 1993.

[27] Kit-man Chan, Tak-shing Yum, "Analysis of Least Congested Path Routing in WDM Lightwave Networks," *Proc. IEEE Infocom,* 1994.

[28] Tak-kin G. Yum, Mischa Schwartz, "Comparison of Routing Procedures for Circuit-Switched Traffic in Nonhierarchical Networks," *IEEE Trans. on Commun.*, vol. com-35, no. 5, May 1987.

[29] S.P.Chung, A.Kashper, and K.W.Ross, "Computing Approximate Blocking Probabilities for Large Loss Networks with State-Dependent Routing," *IEEE/ACM Transactions on Networking*, Feb. 1993.

[30] M. Schwartz, *Telecommunication Network: Protocols, Modeling and Analysis*, Addison Wesley, 1988.

[31] Roch Guérin, Hamid Ahmadi, Mahmoud Naghshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks," *IEEE J. Select. Areas Commun.,* vol. 9, no. 7, Sep. 1991.

[32] Debasis Mitra, Richard J. Gibbens, B.D. Huang, "Analysis and Optimal Design of Aggregated-Least-Busy-Alternate Routing on Symmetric Loss Network with Trunk Reservations," *proc. ITC-13,* 1991.

[33] Ronald A. Howard, *Dynamic Probabilistic System, Vol II: Semi-Markov and Decision Processes,* John Wiley & Sons, Inc., 1971.

[34] I.P.Kaminow, "Photonic Multiple-Access Network: Routing and multiplexing," *AT&T Technical Journal*, March 1989.

[35] I.P.Kaminow, "Photonic Multiple -Access Network : Topology ," *AT&T Technical Journal*, March 1989.

[36] K. Bala *et al.*, "Algorithm for Routing in a Linear Lightwave Network," *Proc. INFOCOM* 1991.

[37] K. Bala, K. Petropoulos and T. E. Stern, "Multicasting in LLNs," *Proc. INFOCOM* 1993.

[38] T. E. Stern, "Linear Lightwave Networks: How Far Can They Go?" *Proc. GLOBECOM* 1990.

[39] M. S. Goodman *et al.*, "The LAMBDANET multiwavelength network: Architecture, applications and demonstrations," *IEEE JSAC,* vol. 8, no. 6, 1990.

[40] N. Dono *et al.*, " A Wavelength Division Multiple Access Network for Computer Communication," *IEEE JSAC*, vol. 8, Aug. 1990.

[41] C. A. Brackett, "Dense Wavelength Division Multiplexing Networks: Principles and Applications," *IEEE JSAC*, vol. 8, no. 6, Aug. 1990.

[42] G. R. Hill, "A Wavelength Routing Approach to Optical Communications Networks," *Proc. INFOCOM* 1988.

[43] Imrich Chlamtac, Aura Ganz, Gadi Karmi "Lightpath communications: An approach to High Bandwidth Optical WAN's," *IEEE Trans. on Commun.* vol. 40, Jul. 1992.

[44] Kuo-chun Lee, Victor O.K.Li, "Routing and Switching in a Wavelength Convertible Optical Network," *Proc. INFOCOM*, 1993.

[45] John L. Zyskind *et al.*, "Erbium-Doped Fiber Amplifiers and the Next Generation of Lightwave Systems," *AT & T Technical Journal*, Jan/Feb, 1992.

[46] R. Ramaswami "Multiwavelength Lightwave Networks for Computer Communication," *IEEE Commun. Mag.*, Feb., 1993.

[47] T.P.Lee "Recent Advances in Long-wavelength Semiconductor Lasers for Optical Fiber Communication," *Proc. IEEE*, vol. 79, no. 3, March 1991.

[48] K.W.Cheung, "Switch For Selectively Switching Optical Wavelengths," U.S. Patent No. 4,906,064, March 6, 1990.

[49] P. Barnsley and P. Chidegy, "All-optical wavelength switching from $1.3\mu m$ to $1.55\mu m$ WDM wavelength routed network: System results," *IEEE Photon. Technol. Letter,* 4(1):91-94, Jan. 1992.

[50] K. Inoue and T. Toba, "Wavelength conversion experiment using fiber four-wave mixing," *IEEE Photon. Technol. Letter,* 4(1): 69-72, Jan. 1992.

[51] B. Glance *et al.* "Broadband optical wavelength shifter," *CLEO '92* pp. PD-10, 1992.