Analysis of Rerouting in Circuit-Switched Networks

Eric W. M. Wong, *Member, IEEE*, Andy K. M. Chan, and Tak-Shing Peter Yum, *Senior Member, IEEE*

***Abstract*—Dynamic routing has been adopted in circuit-switched networks in many parts of the world. Most of the routing algorithms used are least loaded routing (LLR) based for its simplicity and efficiency. Rerouting is the practice of routing calls on alternate paths back to direct paths or to other less congested alternate paths. It allows the continuous redistribution of network loads for the relief of the congestion on direct paths. In this paper, we present an original analysis of an LLR-based rerouting scheme. Through numerical examples and confirmation by computer simulation, the throughput gain of rerouting is established.**

***Index Terms*—Circuit-switched networks, dynamic routing, least loaded routing, rerouting.**

I. INTRODUCTION

THE BASIC idea of dynamic routing in a circuit-switched network is to increase throughput by routing calls to alternate paths when the direct path is blocked. However, the use of alternate paths usually consumes more network resources as the path length, in hops, is usually longer. Therefore, indiscriminate use of alternate paths could lead to the decrease of the network throughput and even network instability, as many of the previous research studies had shown [22], [2], [28], [27].

In recent years, a variety of approaches to alternate routing networks have been developed. AT&T has used a decentralized nonhierarchical routing strategy, called dynamic nonhierarchical routing (DNHR) [3] for a number of years. DNHR is a time-dependent routing scheme that increases network efficiency by taking advantage of the noncoincidence of busy hours in a large-toll network. Dynamically controlled routing (DCR), a routing scheme developed by Bell Northern Research, uses a central processor to track the busy–idle status of network trunks and determine the best alternate route choices based on status data collected every ten seconds [9], [8]. Dynamic alternate routing (DAR) was developed by British Telecom and uses a simple decentralized learning approach to adaptive routing [15], [29]. When the direct trunk group is busy, the two-link alternate path used last time is chosen for the overflow call. If the alternate path is busy, the overflow call is blocked and a new alternate path is selected at random for the next overflow call. Taking advantage of the fact that it is feasible to monitor channel occupancies and make routing decisions on a call-by-call basis, AT&T recently adopted a new routing scheme called real-time network routing (RTNR) [4]. In RTNR, if a direct path is blocked, the call will be routed to the least loaded two-link alternate path. It was shown that this practice can improve the network connection availability while simultaneously reducing the network costs.

One major focus of dynamic routing research is to determine the "right" choice of alternate paths for overflow calls [27]. The idea is to balance the traffic load among the alternate paths as much as possible while not loading any of them into congestion. Over the years, researchers have learned that the simple least loaded routing (LLR) with trunk reservation can provide significant throughput gain over fixed routing while other more elaborate approaches (e.g., Markov decision process) can only provide marginal throughput gain [17], [18]. This is probably the reason LLR was chosen as the basis of RTNR in the AT&T Network.

In dynamic routing, a routing decision must be made at call-arrival time based on the network information available at that time. However, a decision once made is final. One method to increase the throughput of the traditional dynamic routing is to redistribute network load to eliminate traffic hot spots or bottlenecks. Rerouting (also called call repacking) is the practice whereby calls on alternate paths can be rerouted back to direct paths or to other less congested alternate paths as situation warrants. Previous studies on rerouting for network of different kinds can be found in [1], [16], [18], [31], [14], [39], [23], [7], [37], [20], [32]. In [18] and [23], rerouting is studied in virtual-circuit packet-switched networks and in circuit-switched wavelength-division-multiplexed all-optical networks, respectively. In [14], an "intentional" rerouting, which is not "forcibly" triggered by a failure, was considered in asynchronous transfer mode (ATM) networks. The paper showed that rerouting can be implemented *smoothly* in ATM networks. This means that rerouting can ensure not only first-in-first-out (FIFO) but also integrity of cells. Thus, it can be completely transparent to the upper layers of the communicating processes. In [39], a dynamic virtual-path rearrangement scheme was proposed and a strategy based on the scheme was presented for ATM network provisioning. The robustness of this scheme was demonstrated by simulations not only against the forecast errors, but also against changes in transport-network and virtual-path-network structures. Note that as far as fixed route and dedicated bandwidth to users are concerned, a *virtual path* in ATM networks is very similar to a *circuit* in circuit-switched networks.

In advanced cellular networks, rerouting is realized in the handoff process [21]. In other words, during a handoff, a call is rerouted from one base-station/mobile-switching-center to another base-station/mobile-switching-center. In an overlaying cellular network [7], [20] with a mix of micro- and macro-cells, a call in a macro-cell which was previously overflowed from a

Manuscript received June 8, 1998; revised June 22, 1999; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor U. Shankar. This work was supported by a Competitive Earmarked Research Grant under Project Number 9040427.

E. W. M. Wong and A. K. M. Chan are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: ewong@ee.cityu.edu.hk; kmchan@ee.cityu.edu.hk).

T.-S. P. Yum is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: yum@ie.cuhk.edu.hk).

Publisher Item Identifier S 1063-6692(00)04996-7.

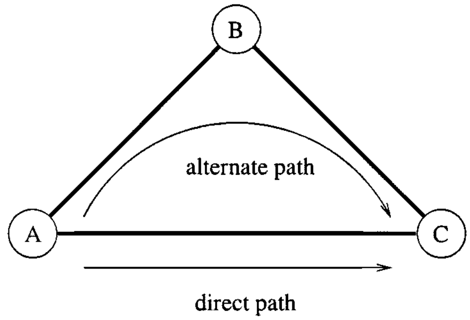


Fig. 1. Alternate path of a node pair.

micro-cell is rerouted back to the micro-cell whenever possible. In addition, the technique of rerouting was also used on the Internet [32]. The University of Southern California took a test of rerouting many Internet address queries to a computer in the university, instead of letting them go to a US government-sponsored central facility in Herndon, Virginia, that usually handles the traffic. Research for rerouting in circuit-switched networks can be found in [16], [37], [38]. A simple rerouting scheme was studied in [16] through simulation while in [37] another simple rerouting scheme was analyzed. In [38], a taxonomy of rerouting was studied.

One purpose of rerouting is to redistribute network load from time to time so as to free up more capacity for direct path calls. For example, if a channel has just been freed up in link AC due to a call departure (Fig. 1), then a call on an alternate path (path ABC) can be rerouted back to link AC. Doing so would free up one channel each on links AB and AC and thus throughput can be increased. The focus of this paper is on an original analysis of the LLR-based rerouting rule in a symmetrical fully connected network. This analysis is based on the Erlang fixed-point model first applied to state-dependent dynamic routing by Wong and Yum [33], in which LLR was analyzed. While significantly reducing the computational effort, the model was shown to be quite accurate. Subsequent applications of this approach appeared in [24]–[26], [5], [10], [11], [34], [35], [30], [36], [12], [13]. This approach has also been applied for analyzing the AT&T network [6], and the simulation studies of RTNR [6] revealed that RTNR single-service voice network can be designed adequately using the approach for preplanned dynamic routing. Two assumptions were used in these models and remain used in this paper: link independence and Poisson overflow traffic, that is, the direct-path traffic overflowing to alternate paths. In addition, we assume the rerouted traffic, that is, the alternate path traffic rerouted back to direct paths, is also Poisson. As before, verification by computer simulation is done here as well.

Many versions of rerouting are possible [38]. Our focus is on a very simple one [38] as follows. A new arriving call will be routed to its (one-link) direct path X first if there is a free channel on that path. If path X is blocked, one alternate call on path X, if any, is picked randomly and rerouted back to its direct path (if possible) to make room for the new call. If the rerouting of this alternate call is unsuccessful, another alternate call on path X is picked at random and so on. If none of the alternate calls on path X can be rerouted, the overflow call is routed to a (two-link) alternate path with the maximum number of free circuits, i.e., the least loaded alternate path, as in LLR routing. If all alternate paths are full, the call is blocked.

The implementation of rerouting in a network requires the making of routing decisions on a call-by-call basis. This was shown to be feasible by the AT&T's RTNR [4] using stored-program control switches interconnected through a common channel signaling (CCS) network. Aside from a slightly more complicated software control and more signaling traffic, no hardware modification to the switching system is envisaged for implementing rerouting.

II. Analysis of Rerouting

Consider an M-node fully connected and uniformly loaded network where all links consist of N channels. Let $\mathbf{n} = (d, a)$ be the occupancy state of a link where d and a are the numbers of direct and alternate calls on a link, respectively. Let $P_{\mathbf{n}} = P_{d,a}$ be the state probability. Let λ_D be the rate of direct-path offered traffic to a link and λ_O be the rate of overflow traffic from a link to other links. Let both these traffic streams be Poisson processes and let the call service time be exponentially distributed with mean equal to one time unit. We restrict our choice of alternate paths to those of two links only and let m be the number of two-link alternate paths. Note that m is a system selection parameter and the maximum value of m is $M-2$ for the fully connected network. For $m < M-2$, intermediate nodes of overflowed paths are selected at random. The resulting network is, therefore, still symmetrical. It is well known that, if no suitable control is taken, the number of alternate path calls may dominate, resulting in significant reduction of network capacity [28]. It is also well known that trunk reservation, i.e., reserving the last r unoccupied channels in a link for direct-route calls only, is an effective means of maintaining the network in the high capacity mode. Let Ω, Ω_D, and Ω_A be the state space, the set of direct call blocking states, and the set of alternate call blocking states, respectively:

$$\Omega = \{\mathbf{n}\colon d, a \geq 0, a \leq N - r, d + a \leq N\} \quad (1)$$
$$\Omega_D = \{\mathbf{n}\colon \mathbf{n} \in \Omega, d + a = N\} \quad (2)$$
$$\Omega_A = \{\mathbf{n}\colon \mathbf{n} \in \Omega, d + a \geq N - r\}. \quad (3)$$

With the symmetric and uniform traffic assumption, our analysis can be simplified to a single-link Poisson process with N servers. Fig. 2(a) shows the system state transition diagram and Fig. 2(b) shows all possible transitions and their rates for an arbitrary state (d, a). Consider link AB at state $\mathbf{n}$. Let $D_{\mathbf{n}}$ and $A_{\mathbf{n}}$ represent the direct and alternate call arrival rates on a link and let $\mu_{\mathbf{n}}^{(D)}$ and $\mu_{\mathbf{n}}^{(A)}$ represent the direct and alternate call departure rates. When a new call arrives and finds link AB full, alternate calls (e.g., call ABC) on that link will be randomly picked up and examined to see if rerouting can be performed so as to accommodate the new call. Let $\psi_{\mathbf{n}}^{(R)}$ represents the rerouting traffic rates from a link. On the other hand, a rerouting introduces an alternate call departure on the other leg of the alternate path, i.e., link BC, as well as a direct call arriving on link AC. Let $\psi_{\mathbf{n}}^{(A)}$ represent the alternate call departure rate due to rerouting and let $\psi_{\mathbf{n}}^{(D)}$ represent direct call arrival rate due to

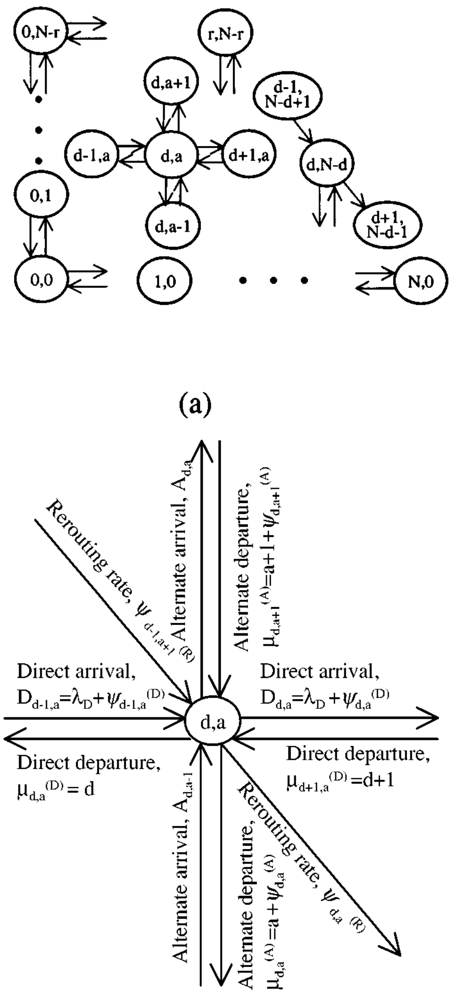


Fig. 2. State transition diagram for rerouting.

rerouting. In the following sections, we will drive the state transition rates exiting from state $\mathbf{n}$, i.e., $D_{\mathbf{n}}$, $A_{\mathbf{n}}$, $\mu_{\mathbf{n}}^{(D)}$, $\mu_{\mathbf{n}}^{(A)}$, $\psi_{\mathbf{n}}^{(D)}$, $\psi_{\mathbf{n}}^{(A)}$ and $\psi_{\mathbf{n}}^{(R)}$, respectively.

A. Alternate Traffic Rate A_n

To derive $A_{\mathbf{n}}$ in terms of $P_{\mathbf{n}}$'s, we need to obtain the the intermediate system parameters λ_O, H_a, and $\beta_{a,k}$, as follows.

1) Deriving λ_O: When a new call arrives and finds the direct path blocked, the call will overflow from that link. The overflow traffic from a link with rate λ_O consists of two components. The first component represents new calls finding all channels on the direct link being occupied by direct path calls. The second component comes from new calls finding $a\ (>0)$ alternate path calls and $N-a$ direct path calls but the direct paths of these a calls are all full. Under this situation, these a alternate path calls cannot be rerouted to their direct paths to make room for the new calls and these new calls have to overflow onto their alternate paths. Let H_a be the probability that all the direct paths corresponding to the a alternate path calls are full. Then the overflow rate λ_O can be expressed as

$$\begin{aligned}\lambda_O &= \lambda_D \ \mathrm{Pr}\ [\text{direct call is blocked on the direct link}] \\ &= \lambda_D P_{N,0} + \lambda_D \sum_{a=1}^{N-r} P_{N-a,a} H_a. \qquad (4)\end{aligned}$$

2) Deriving H_a: Consider a tagged link carrying a alternate calls. Let $\xi_{a,k}$ be the event that the a alternate calls come from k direct paths. At first sight, this is equivalent to the classical combinational problem of finding the probability that k boxes are occupied when a balls are thrown into $2M-4$ boxes at random. But a closer observation reveals that the balls are less likely to fall into empty boxes. Let us see why. We can distinguish two types of direct paths contributing traffic to the alternate paths. The first type, denoted as *ordinary* direct paths, are those which do not currently have alternate path calls on the tagged link. These correspond to the empty boxes. The second type, denoted as the *overflow* direct paths, are those currently having alternate path calls carried on the tagged link. These correspond to the occupied boxes. It was found that these a alternate calls have some kind of *self-aggregate* property in the sense that the overflow calls are more likely to come from those overflow direct paths. Therefore, some modification to the urn problem is needed to capture this self-aggregate effect. We start from the following recursive argument by relating the case of "a alternate calls" to that of the "$a-1$ alternate calls" as follows:

$$\begin{aligned}\mathrm{Pr}\ [\xi_{a,k}] =& \ \mathrm{Pr}\ [\text{the additional alt. call comes from an ordinary direct path } |\xi_{a-1,k-1}]\ \mathrm{Pr}\ [\xi_{a-1,k-1}] \\ &+ \mathrm{Pr}\ [\text{the additional alt. call comes from an overflow direct path } |\xi_{a-1,k}]\ \mathrm{Pr}\ [\xi_{a-1,k}] \\ =& \ \frac{(2M-k-3)\lambda_O/\lambda_D}{k-1+(2M-k-3)\lambda_O/\lambda_D}\ \mathrm{Pr}\ [\xi_{a-1,k-1}] \\ &+ \frac{k}{k+(2M-k-4)\lambda_O/\lambda_D}\ \mathrm{Pr}\ [\xi_{a-1,k}]. \qquad (5)\end{aligned}$$

where we assume that the overflow direct paths remain full for the additional call, and λ_O/λ_D is the probability that an ordinary direct path is chosen to join the existing set of the overflow direct paths for the additional alternate call. This equation can be solved using the initial condition $\mathrm{Pr}[\xi_{1,1}]=1$ and $\mathrm{Pr}[\xi_{1,k}]=0$ for $k\neq 1$.

Returning to the derivation of H_a, we have

$$\begin{aligned}H_a =& \ \mathrm{Pr}\ [\text{all direct paths corresponding to the } a \text{ alt. path calls on the tagged link are full}] \\ =& \sum_{k=1}^{\min(a,2M-4)} \mathrm{Pr}\ [\text{all } k \text{ direct paths are full } |\xi_{a,k}] \\ & \mathrm{Pr}\ [\xi_{a,k}]. \qquad (6)\end{aligned}$$

Let $\beta_{a,k}$ denote the probability that a particular direct path L corresponding to one (or more) of the a alternate paths calls on the tagged link are full given that there are k such direct paths. Recall that links are assumed to be statistically independent. Therefore, H_a can be further expressed as

$$H_a = \sum_{k=1}^{\min(a,2M-4)} (\beta_{a,k})^k\ \mathrm{Pr}\ [\xi_{a,k}]. \qquad (7)$$

3) Deriving $\beta_{a,k}$: To derive $\beta_{a,k}$, let us consider the rerouting of alternate calls back to the direct path L. Note that when a rerouting attempt occurs, the direct (path) link is more likely to be in a high occupancy state, as otherwise alternate routing would not take place in the first place. Therefore, the direct link occupancy as seen by a rerouting attempt is no longer $P_{\mathbf{n}}$. Here, for simplicity, we choose to approximate it to be

$$P^*_{d,a} = \begin{cases} \dfrac{P_{d,a}}{\sum\limits_{i+j=N-1}^{N} P_{i,j}} & d+a \in \{N-1, N\} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

by assuming the occupancy is always at either $N-1$ or N. Therefore, we have $\beta_{a,k} = \Sigma_{i+j=N} P^*_{i,j}$. Note that for all cases we tried, $\Sigma_{i+j=N} P^*_{i,j}$ matches the simulation results well and also much better than the results obtained from the steady state distribution, i.e., $\Sigma_{i+j=N} P_{i,j}$ (the probability of the direct link being fully occupied). In other words, $\beta_{a,k}$ can be better estimated through $P^*_{d,a}$'s. We have obtained so far $\beta_{a,k}$ and hence H_a and λ_O in terms of $P_{\mathbf{n}}$'s. We are now ready to derive $A_{\mathbf{n}}$.

4) Deriving $A_{\mathbf{n}}$: Consider a particular alternate path. Let the number of occupied channels on the first link be i and that on the second link be j. Then, the number of available circuits k on that path is $k = N - \max(i,j)$. When the direct path is full, the LLR routing will direct the call to the alternate path with the maximum number of free circuits. When there is more than one such path, choose one at random.

Consider a particular link AC as shown in Fig. 1. If this link is full, the overflow calls of rate λ_O will be routed randomly to one of the LLR's. Let there be a total of α such paths. Then the alternate path load of AC that falls on a particular LLR, say path ABC, is λ_O/α.

Let Z_i be the probability that a two-link alternate path has i or more occupied circuits. Then, with the assumption that links are independent, we have

$$Z_i = 1 - \left\{ \sum_{\forall \mathbf{n}: n<i} P_{\mathbf{n}} \right\}^2 \tag{9}$$

where $n = d + a$.

Given that path ABC has i occupied circuits, the probability $f(\alpha|i)$ that $\alpha - 1$ other alternate paths also have i occupied circuits each **and** each of the remaining $m - \alpha$ alternate paths has more than i occupied circuits is given by

$$f(\alpha|i) = \binom{m-1}{\alpha-1} (Z_i - Z_{i+1})^{\alpha-1} Z_{i+1}^{m-\alpha} \tag{10}$$

where $Z_i - Z_{i+1}$ is the probability that an alternate path has exactly i occupied circuits. Therefore, given that path ABC has i occupied circuits, the amount of traffic y_i that gets routed from AC to alternate path ABC is

$$\begin{aligned} y_i &= \sum_{\alpha=1}^{m} \frac{\lambda_O}{\alpha} f(\alpha|i) \\ &= \frac{\lambda_O}{m} \frac{Z_i^m - Z_{i+1}^m}{Z_i - Z_{i+1}}. \end{aligned} \tag{11}$$

Given that link AB is in state $\mathbf{n} \in \Omega \backslash \Omega_A$, where $A \backslash B$ denotes set A minus set B, the alternate traffic rate at state $\mathbf{n}$, denoted as $A_{\mathbf{n}}$, can be obtained by removing the condition on the second link to be

$$A_{\mathbf{n}} = 2m \sum_{\mathbf{i} \in \Omega \backslash \Omega_A} y_{\max(n^{(AB)}, i^{(BC)})} P_{\mathbf{i}} \tag{12}$$

where $x^{(y)}$ represents the number of calls in link y at state $\mathbf{x}$.

B. Rerouting Traffic Rate $\psi^{(R)}_{\mathbf{n}}$

If a new call between nodes A and B finds link AB full but has in it one or more alternate calls, i.e., the link is in state $(N - a, a > 0)$, then it will cause an alternate call to be rerouted to its direct path if there is a vacant channel in that direct path. In this way, the new direct call can be accommodated on link AB. This occurs with probability $1 - H_a$. Therefore, the corresponding transition rate $\psi^{(R)}_{\mathbf{n}}$ from $(N-a, a)$ to $(N-a+1, a-1)$ is

$$\psi^{(R)}_{\mathbf{n}} = \begin{cases} \lambda_D(1 - H_a) & \mathbf{n} \in \Omega_R \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where $\Omega_R \equiv \{\mathbf{n}\colon \mathbf{n} \in \Omega, a > 0, d + a = N\}$ is the set of rerouting states. As far as link AB is concerned, one direct call comes in and one alternate call departs, leaving the total number of calls unchanged.

C. Direct-Path Traffic Rate $D_{\mathbf{n}}$

The rerouted traffic from alternate paths transforms into "extra" direct-path traffic in direct links. Assuming the alternate calls for rerouting are randomly chosen, the rerouted traffic rate ψ in a link is simply $\psi = \Sigma_{\mathbf{k} \in \Omega_R} \lambda_D (1 - H_a) P_{\mathbf{k}}$. There are two issues to note. First, when a rerouting occurs, the distribution of trunk occupancy on the direct link seen by a rerouted call can be obtained from (8) and is given by

$$P^{**}_{d,a} = \begin{cases} \dfrac{P_{d,a}}{\sum\limits_{i+j=N-1} P_{i,j}} & d+a = N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

This means that the states for which rerouting can take place are $\{(d,a) | d + a = N-1\}$. Therefore, the effective rerouted traffic rate ψ' seen by these states is

$$\psi' = \frac{\psi}{\sum\limits_{i+j=N-1} P_{i,j}} \tag{15}$$

Second, the extra direct traffic in a link comes from $2m$ possible links and each link contributes part of the direct traffic which is equivalent to the rerouted traffic rate of that link, i.e., ψ. But each of these $2m$ links also has $2m$ possible links for rerouted traffic to go into, or, in other words, has a probability $1/2m$ to choose the tagged link. Therefore, the extra direct traffic rate due to rerouting is

$$\begin{aligned} \psi^{(D)}_{\mathbf{n}} &= 2m \left(\frac{1}{2m} \right) \psi' \\ &= \psi'. \end{aligned} \tag{16}$$

The direct-path traffic rate at state $\mathbf{n}$ is the sum of direct-path offered rate and extra direct traffic rate due to rerouting and is given by

$$D_{\mathbf{n}} = \begin{cases} \lambda_D + \psi_{\mathbf{n}}^{(D)} & d+a = N-1 \\ \lambda_D & 0 \le d+a < N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

D. Service Rates $\mu_{\mathbf{n}}^{(A)}$ and $\mu_{\mathbf{n}}^{(D)}$

The rerouting of alternate calls in a link, say AB, (due to new call arrivals on that link) causes "extra" departures in another link, say BC, since an alternate call lies on a two-link alternate path. However, since link BC has $2m$ possible alternate path traffic on it and its companion link such as AB also has $2m$ possible links on which its alternate paths can possibly lies, the alternate call departure rate due to rerouting, $\psi_{\mathbf{n}}^{(A)}$, is

$$\psi_{\mathbf{n}}^{(A)} = 2m\left(\frac{1}{2m}\right)\frac{\psi}{\sum_{j\neq 0} P_{i,j}} = \frac{\psi}{\sum_{j\neq 0} P_{i,j}} \tag{18}$$

where $\Sigma\, P_{i,j}$ is the normalized rate of "extra" departures since rerouting only happens under the condition $a \neq 0$. The service rate $\mu_{\mathbf{n}}^{(A)}$ for alternate calls at state $\mathbf{n}$ is increased to

$$\mu_{\mathbf{n}}^{(A)} = a + \psi_{\mathbf{n}}^{(A)}. \tag{19}$$

As a check, $\Sigma_{a\neq 0}\, (\mu_{\mathbf{n}}^{(A)} - a)P_{d,a} = \psi$ as it should be. Moreover, the service rate of direct calls at state $\mathbf{n}$ is simply $\mu_{\mathbf{n}}^{(D)} = d$.

E. System Equations

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

$$\begin{aligned}(\mu_{\mathbf{n}}^{(D)} + \mu_{\mathbf{n}}^{(A)} + D_{\mathbf{n}} + A_{\mathbf{n}} + \psi_{\mathbf{n}}^{(R)})P_{\mathbf{n}} &= \mu_{d+1,a}^{(D)} P_{d+1,a} + \mu_{d,a+1}^{(A)} P_{d,a+1} + D_{d-1,a}P_{d-1,a} \\ &\quad + A_{d,a-1}P_{d,a-1} + \psi_{d-1,a+1}^{(R)} P_{d-1,a+1}\end{aligned} \tag{20}$$

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

Let $\mathcal{P}$ denote the set of $P_{\mathbf{n}}$ and Λ denote the set of traffic rates $D_{\mathbf{n}}, A_{\mathbf{n}}, \mu_{\mathbf{n}}^{(D)}, \mu_{\mathbf{n}}^{(A)}, \psi_{\mathbf{n}}^{(D)}, \psi_{\mathbf{n}}^{(R)}$ and $\psi_{\mathbf{n}}^{(A)}$. Equation (20) can be expressed as $\mathcal{P} = f_1(\Lambda)$ and Λ can be expressed as $f_2(\mathcal{P})$ using the equations derived in this section. Thus, we have formulated our analysis as a fixed point model [27] which can be solved by the successive over-relaxation (SOR) method with the set of traffic rates and probabilities obtained by f_1 and f_2 in each iteration.

Finally, with the assumption of link independence, the end-to-end call blocking probability of LLR-based rerouting is given by

$$\begin{aligned}P_B &= \Pr\,[\text{direct call is blocked on the direct link}] \\ &\qquad \Pr\,[\text{all } m \text{ alternate paths are blocked}] \\ &= \lambda_O/\lambda_D \left[1 - \left(1 - \sum_{\mathbf{n}\in\Omega_A} P_{\mathbf{n}}\right)^2\right]^m.\end{aligned} \tag{21}$$

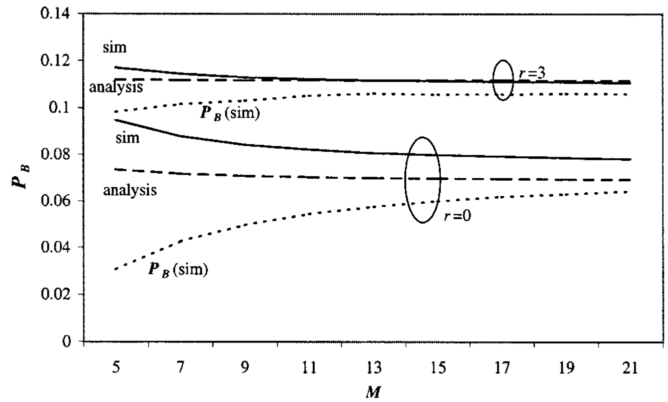


Fig. 3. Call blocking probability P_B of rerouting against M for $m = 3, \lambda_D = 19$ and $N = 20$.

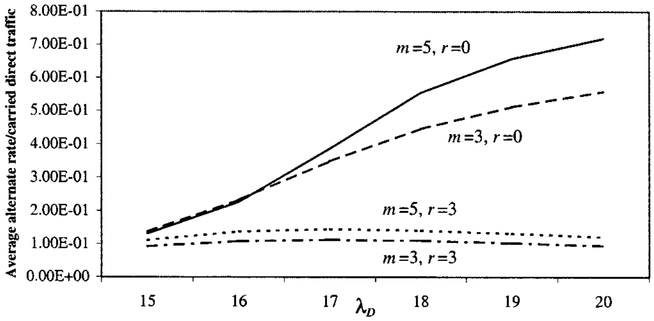
III. Discussion

We study the performance of rerouting under fully connected symmetric networks by numerical examples. Computer simulation results are also obtained and compared. The 95% confidence intervals have all been set to a size smaller than the markers on the figures (by varying the duration of simulation) and are therefore not shown. The results of LLR without rerouting based on the model in [33] are presented for comparison.

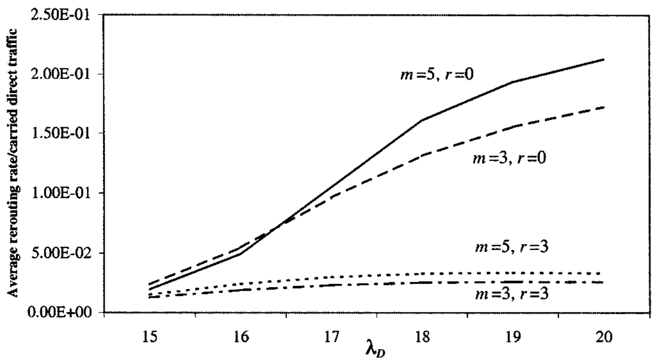
We will first demonstrate how link independency and Poisson traffic assumptions affect the accuracy of our analysis and see how the alternate and the rerouting traffic affect the validity of these assumptions. Then we will compare the performance of rerouting to that of LLR. Moreover, we will validate other assumptions in the paper by showing the network internal parameters from simulation. Finally, we will compare the computational effort of simulation and analysis.

A. Checking Link Independency and Poisson Traffic Assumptions

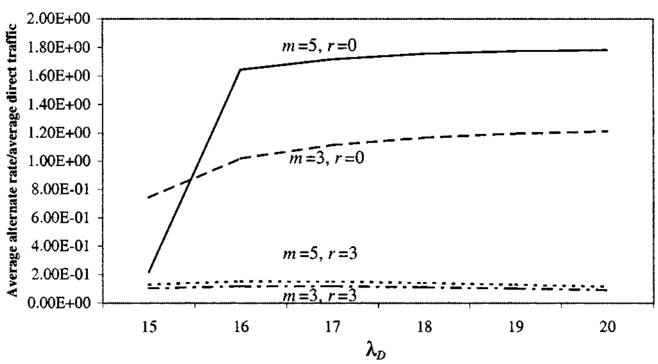
Fig. 3 shows the end-to-end call blocking probability P_B of LLR-based rerouting as a function of the number of network nodes M with a fixed number of alternate paths $m = 3$, $\lambda_D = 19$ and $N = 20$. Besides simulation and analytical results, we also obtain the end-to-end call blocking probability using (21), denoted as P_B(sim), but with the network variables (λ_O and $P_{\mathbf{n}}$'s) obtained by the simulation. This is for checking the impact of the link independence assumption to the accuracy of the analysis. We regard P_B(sim) as an appropriate measure for this test because 1) it contains multiplication terms of all alternate path blocking probabilities and will magnify the link dependence effect (if any), *and* 2) it avoids the effect of Poisson traffic assumption, which is the other main approximation in the paper. In other words, the accuracy of P_B(sim) is *only* affected by the link independence assumption. It is seen that for both $r = 0$ and $r = 3$ cases, the curve of simulation and the curve of P_B(sim) converge with the increase of M. This means that as M increases, the link independence assumption (or approximation) becomes *increasingly valid*. Interestingly, P_B exhibits the same trend (i.e., the analytical model is more accurate for large M) but is more accurate than P_B(sim). This indicates that the



(a)



(b)



(c)

Fig. 4. Average alternate traffic and average rerouting traffic rates against λ_D. (a) Average alternate rate/carried direct traffic against λ_D. (b) Average rerouting rate/carried direct traffic against λ_D. (c) Average alternate rate/carried direct traffic against λ_D for LLR.

two types of main approximations appear to compensate each other. In addition, the analysis is seen to be more accurate for $r = 3$. This will be explained in the next figure. In the following, we set $M = 20$ and $N = 20$ unless otherwise specified.

Fig. 4(a) and (b) shows the average alternate traffic rate (i.e., $\Sigma_{\text{all } \mathbf{n}} \psi_{\mathbf{n}}^{(A)} P_{\mathbf{n}}$) and the average rerouting traffic rate (i.e., $\Sigma_{\text{all } \mathbf{n}} \psi_{\mathbf{n}}^{(R)} P_{\mathbf{n}}$) as a fraction of carried direct traffic rate on a link, i.e., λ_D (1- Pr [direct call is blocked on the direct link]) or, from (4), $\lambda_D(1 - P_{N,0} - \Sigma_{a=1}^{N-r} P_{N-a,a} H_a)$, against λ_D for different combinations of $m = 3$, $m = 5$, $r = 0$ and $r = 3$. First, it is seen that for $r = 3$, the amount of alternate traffic is only 10% of the carried direct traffic while the amount of rerouting traffic is less than 3%. This explains why in Fig. 3, the analytical results always match well with simulation for $r = 3$ since, intuitively, the less alternate/rerouting traffic, the less the dependencies among the links. Consequently, both types of main approximations become more valid and hence the analytical model also become more accurate. On the other hand, it is found that for $r = 0$ the amount of alternate traffic increases from 10% to 55% for $m = 3$ and to 70% for $m = 5$, while the amount of rerouting traffic increases from 3% to 16% for $m = 3$ and 21% for $m = 5$. This increase of alternate and rerouting traffic induces higher link dependency and consequently higher discrepancy between simulation and analytical results. Moreover, the alternate and rerouting traffic types are known to be more bursty than Poisson traffic. Therefore, as these two traffic increase, the Poisson assumption causes more analytical error.

Fig. 4(c) shows the average alternate traffic rate as a fraction of carried direct traffic rate for LLR, i.e., without rerouting. Comparing Fig. 4(a)–(c), it is found that rerouting can reduce the amount of alternate traffic for all cases with $r = 0$. For $r = 3$, rerouting, however, does not help much in reducing alternate traffic.

B. Rerouting Performance

Fig. 5 shows the end-to-end call blocking probability P_B of LLR-based rerouting as a function of direct-path offered traffic rate λ_D for $m = 1$, $m = 3$, and $m = 5$, respectively, with and without trunk reservation. Note that lines represent analytical results while marks represent simulation results. It is seen that the analytical results match well with the simulation results for rerouting. It is also seen that the performance of LLR can always be improved by adding rerouting. Comparing the analytical and simulation results of rerouting in Fig. 5(a)–(c), it is found that as m increases, the discrepancy between simulation and analysis keeps very small for $r = 3$ but increases for $r = 0$. This is attributed to the increase in link dependency as m increases for $r = 0$ and will be further explained in later figures. Comparing the analytical results of LLR and rerouting at $r = 0$, it is seen that rerouting can maintain network stability without trunk reservation. In fact, minimum blocking probability occurs at $r = 0$ in the case of rerouting.

Fig. 6 shows the call blocking probability of rerouting as a function of m with $\lambda_D = 19$, and with $r = 3$ for LLR and $r = 0$ for rerouting, respectively. It is found that as m increases from 3 to 10, call blocking using rerouting is reduced at fixed loading while for LLR routing the blocking probability is nearly constant. This shows that rerouting has the added advantage of sizing efficiency in addition to trunking efficiency.

C. Checking the Accuracy of Internal Parameters

1) Checking the "self-aggregate" assumption: We now examine the "self-aggregate" assumption in deriving $\Pr[\xi_{a,k}]$'s. For simplicity, we look at the average value (i.e., $\Sigma_{\text{all } k}\ k \Pr[\xi_{a,k}]$), denoted as AV[$\xi$], for a given value of a. Fig. 7 compares the simulation and analytical results of AV[ξ] for $m = 5$ and $\lambda_D = 19$. It is seen that our modified urn model matches well with the simulation results. We also plotted the unbiased/normal urn model, which can be obtained by setting $\lambda_O/\lambda_D = 1$ in (5), and it is seen to be a less accurate description of $\xi_{a,k}$.

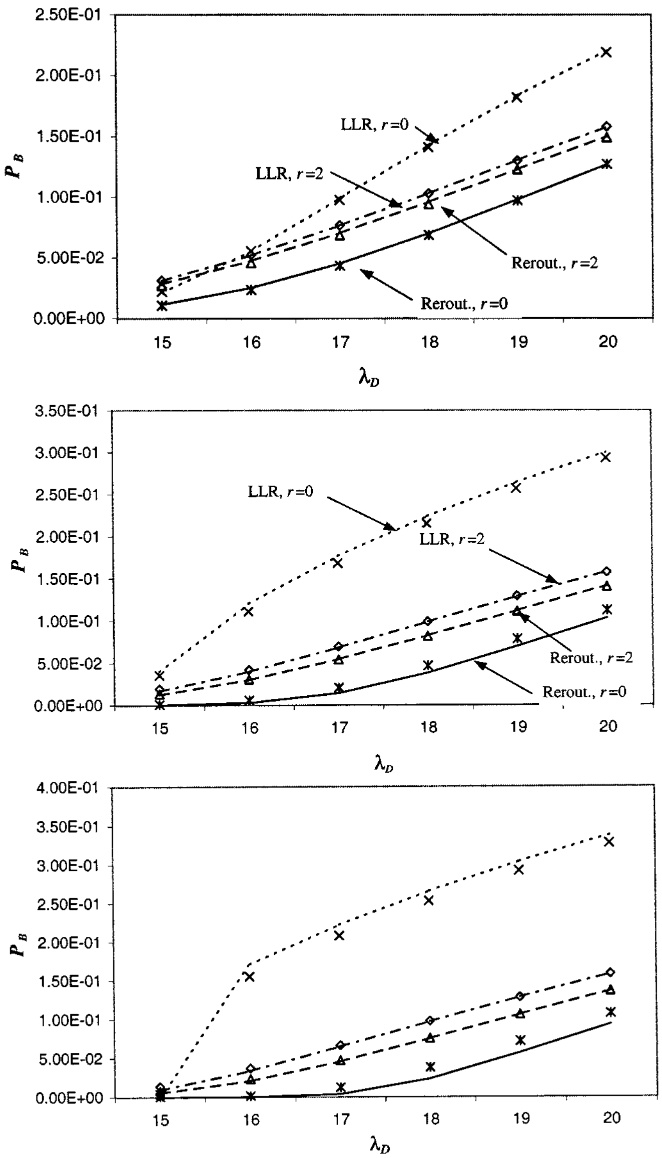


Fig. 5. Call blocking probability P_B of rerouting against offered traffic rate λ_D. (a) $m = 1$. (b) $m = 3$. (c) $m = 5$.

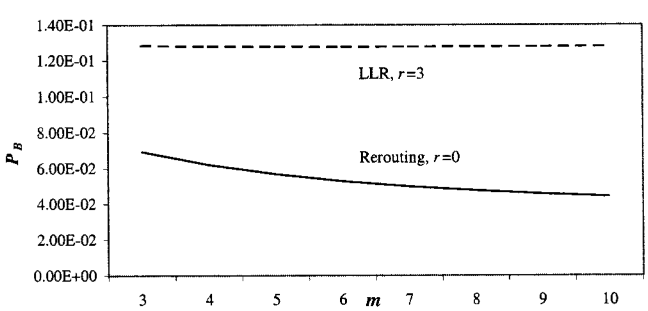


Fig. 6. Call blocking probability P_B against m for $\lambda_D = 19$.

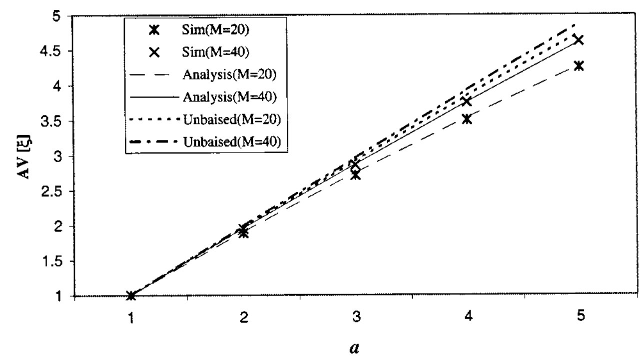


Fig. 7. AV[ξ] against a for $m = 5$ and $\lambda_D = 19$.

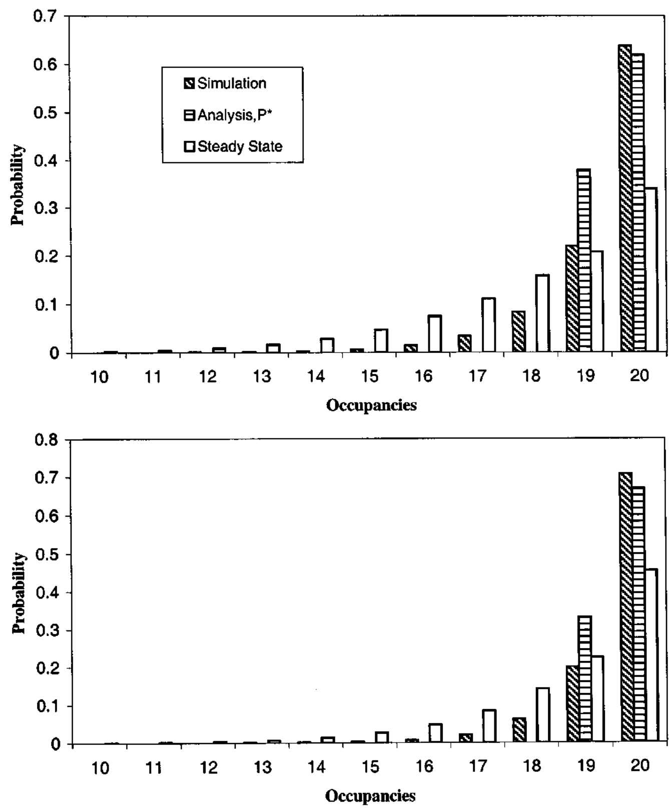


Fig. 8. Distribution of link occupancies for rerouting for $m = 3$ and $r = 0$. (a) $\lambda_D = 18$. (b) $\lambda_D = 20$.

2) Checking the accuracy of $\beta_{a,k}$: Fig. 8 shows the conditional distribution of link occupancy of a direct link when a rerouting is attempted to it. For $m = 3$ and $r = 0$, we have plotted the conditional distributions from simulation and from analysis with $\lambda_D = 18$ and $\lambda_D = 20$, respectively. Also shown is the distribution calculated from the steady state probabilities. It is found that under both loading conditions our analysis gives a much better approximation on the probability of the blocking state (i.e., occupancy $=$ 20) on a link during rerouting than the steady state distribution. This means that a better estimation of $\beta_{a,k}$ can be obtained by our analysis.

D. Comparison of Computational Efforts

In a fully connected symmetric network with M nodes and N trunks in a direct path, there are $M(M-1)/2$ direct paths. The simulation time is determined by the number of events generated during the simulation run which is proportional to the number of direct paths in the network. As a result, the simulation time grows at O(M^2) aside from other factors.

TABLE I
COMPUTATION TIME OF SIMULATION AND ANALYSIS UNDER DIFFERENT M AND r FOR $m = 3$, $N = 20$, AND $\lambda_D = 19$ (TIME IN hh:mm:ss)

	r=3		r=0	
M	Simulation	Analysis	Simulation	Analysis
7	0:20:25	0:00:53	0:20:24	0:01:43
9	0:42:44	0:00:53	0:45:20	0:01:44
11	1:20:54	0:00:54	1:26:21	0:01:46
15	6:14:54	0:01:49	7:11:01	0:03:33
19	15:49:11	0:01:52	19:11:57	0:03:38

TABLE II
COMPUTATION TIME OF SIMULATION AND ANALYSIS UNDER DIFFERENT m FOR $M = 13$, $N = 20$, $\lambda_D = 19$, AND $r = 3$ (TIME IN hh:mm:ss)

m	Simulation	Analysis
3	2:15:33	0:00:55
5	2:18:54	0:02:10
7	2:24:26	0:04:41
9	2:24:48	0:08:22
11	2:26:49	0:13:26

On the other hand, the computation of our analysis is based on the fixed-point iteration method. An initial set of link state probabilities, denoted as $P_{\mathbf{n}}(0)$, is set arbitrarily. Alternate and rerouting traffic rates are computed based on the derived formulae. With these traffic rates, a new set of link state probabilities $P_{\mathbf{n}}(1)$ can be computed and so on. This iteration repeats until the difference of two consecutive sets of $P_{\mathbf{n}}(i)$ is smaller than a predefined error threshold. Therefore, the computation time depends very much on the state space of $P_{\mathbf{n}}$ which is a 2-D array with size of N^2. As a result, the computation time of analysis grows at O(N^2) aside from other factors.

We program the analysis in C language and simulation in SIMSCRIPT on a SUN Sparc 20 workstation with two 75 MHz processors and 128M memory. Table I compares the computation time of simulation and analysis under different M and r for $m = 3$, $N = 20$ and $\lambda_D = 19$. It is shown that the run time for the analytical approach is much faster than that for the simulation for all M and r under consideration.

Table II shows the computation time of simulation and analysis under different m for $M = 13$, $N = 20$, $\lambda_D = 19$, and $r = 3$. It can be seen that the run time for the analytical approach increases rapidly with m while the simulation time is relatively insensitive with the increase of m. However, it is shown that the run time for the analytical approach is much faster than that for the simulation even when m is set to maximum (11 in this case).

Table III shows the computation time of simulation and analysis under different N for $M = 13$, $m = 3$, $\lambda_D = 0.95N$ and $r = 3$. It is again found that the run time for the analytical approach is much faster than that for the simulation. For networks with trunk size N large, the computation time of analysis (of complexity O(N^2)) can be reduced by grouping the occupancy states [26].

TABLE III
COMPUTATION TIME OF SIMULATION AND ANALYSIS UNDER DIFFERENT N FOR $M = 13$, $m = 3$, $\lambda_D = 0.95N$, AND $r = 3$ (TIME IN hh:mm:ss)

N	Simulation	Analysis
10	0:42:17	0:00:02
20	2:17:38	0:00:55
30	5:15:24	0:08:02
40	8:45:10	0:36:26
50	14:07:57	1:56:15

IV. CONCLUSIONS

A simple rerouting scheme based on LLR was studied in this paper. Numerical results showed that rerouting can provide a significant throughput increase over LLR under all conditions. More interestingly, rerouting is shown to be an effective means for maintaining the stability of the network under dynamic routing. Therefore, using rerouting, trunk reservation is no longer needed and for which additional throughput gain can be obtained. As using fixed-point-approximation technique, the proposed model is computationally efficient. Also, the approximate analysis is shown to be quite accurate, especially when the trunk reservation is used or the number of alternate paths is small compared to the number of network nodes.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their great efforts and valuable comments on improving the quality of the paper.

REFERENCES

[1] M. H. Ackroyd, “Call repacking in connecting networks,” *IEEE Trans. Commun.*, vol. COM-27, pp. 589–591, Mar. 1979.
[2] J. M. Akinpelu, “The overload performance of engineered networks with nonhierarchical and hierarchical routing,” *AT&T Bell Labs Tech. J.*, vol. 63, pp. 1261–1281, 1984.
[3] G. R. Ash and E. Oberer, “Dynamic routing in the AT&T network—Improved service quality at lower cost,” in *Proc. IEEE Global Telecommun. Conf.*, Dallas, TX, Nov. 1989.
[4] G. R. Ash, J.-S. Chen, A. E. Frey, and B. D. Huang, “Real-time network routing in a dynamic class-of-service network,” in *Proc. 13th Int. Teletraffic Congress*, Copenhagen, Denmark, June 1991.
[5] G. R. Ash and B.-S. D. Huang, “An analytical model for adaptive routing networks,” *IEEE Trans. Commun.*, vol. 41, pp. 1748–1759, Nov. 1993.
[6] G. R. Ash, *Dynamic Routing in Telecommunications Networks*. New York, NY: McGraw-Hill, 1998.
[7] R. Beraldi, S. Marano, and C. Mastroianni, “A reversible hierarchical scheme for microcellular systems with overlaying macrocells,” in *Proc. IEEE Infocom ’96*, 1996, pp. 51–58.
[8] Bell Northern Research, Special Issue, “Dynamic network controller family,” in *Telesis Mag.*, 1986.
[9] W. H. Cameron, J. Regnier, P. Galloy, and A. M. Savoie, “Dynamic routing for intercity telephone networks,” in *ITC-10*, Montreal, PQ, Canada, 1983.
[10] K. M. Chan and T.-S. P. Yum, “Analysis of adaptive routing schemes in multirate loss networks,” *Telecommun. Syst.*, vol. 5, pp. 341–359, 1996.
[11] ——, “Analysis of least congested path routing in WDM lightwave networks,” in *Proc. IEEE Infocom*, 1994, pp. 962–969.
[12] ——, “The maximum mean time to blocking routing in circuit switched networks,” *IEEE J. Select. Areas Commun.*, vol. 12, pp. 313–321, Feb. 1994.
[13] ——, “Analysis of adaptive routing in multirate loss networks,” *Telecommun. Syst.*, vol. 5, pp. 341–359, 1996.
[14] R. Cohen, “Smooth intentional rerouting and its applications in ATM networks,” in *Proc. IEEE Infocom*, 1994, pp. 1490–1497.

[15] R. J. Gibbens, “Some aspects of dynamic routing in circuit-switched telecommunications networks,” Cambridge Univ., Statist. Lab., Cambridge, U.K., Jan. 1986.
[16] A. Girard and S. Hurtubise, “Dynamic routing and call repacking in circuit-switched networks,” *IEEE Trans. Commun.*, vol. COM-41, pp. 1290–1294, Dec. 1983.
[17] R. H. Hwang, J. F. Kurose, and D. Towsley, “MDP routing in ATM networks using the virtual path concept,” in *Proc. IEEE Infocom*, 1994, pp. 1509–1517.
[18] R. H. Hwang, “LLR routing in homogeneous VP-based ATM networks,” in *Proc. IEEE Infocom*, Apr. 1995, pp. 587–593.
[19] R. H. Hwang and J. F. Kurose, “On virtual circuit routing and re-routing in packet-switched networks,” in *Proc. IEEE ICC’91*, 1991, pp. 1318–1323.
[20] V. B. Iversen, “Traffic engineering of cellular wireless systems,” *Network Inform. Processing Syst.*, Oct. 1997.
[21] S. Keshav, *An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network*. Reading, MA: Addison Wesley, 1997, pp. 345–350.
[22] R. S. Krupp, “Stabilization of alternate routing networks,” in *Proc. Int. Communications Conf.*, Philadelphia, PA, June 1982, pp. 31.2.1–31.2.5.
[23] K. C. Lee and V. O. K. Li, “A circuit rerouting algorithm for all-optical wide-area networks,” in *Proc. IEEE Infocom*, 1994, pp. 954–961.
[24] D. Mitra, R. J. Gibbens, and B. D. Huang, “Analysis and optimal design of aggregated-least-busy-alternate routing on symmetric loss network with trunk reservations,” presented at the 13th Int. Teletraffic Congress, Copenhagen, Denmark, June 1991.
[25] D. Mitra and R. J. Gibbens, “State-dependent routing on symmetric loss networks with trunk reservations, II: Asymptotic, Optimal Design,” *Ann. Oper. Res.*, vol. 35, pp. 3–30, 1992.
[26] D. Mitra, R. J. Gibbens, and B. D. Huang, “State-dependent routing on symmetric loss networks with trunk reservations, I,” *IEEE Trans. Commun.*, vol. 41, Feb. 1993.
[27] K. W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*. New York, NY: Springer-Verlag, 1995, ch. 7.
[28] M. Schwartz, *Telecommunication Network: Protocols, Modeling and Analysis*. Reading, MA: Addison Wesley, 1988.
[29] R. R. Stacey and D. J. Songhurst, “Dynamic alternative routing in the British telecom trunk network,” in *Proc. Int. Switching Symp.*, Phoenix, AZ, Mar. 1987.
[30] C. Vargas, M. V. Hegade, M. Naraghi-Pour, and P. S. Min, “Shadow prices for LLR and ALBA,” *IEEE/ACM Trans. Networking*, vol. 4, pp. 796–807, 1996.
[31] B. R. Venkatraman and R. E. Newman-Wolfe, “Performance analysis of a method for high level prevention of traffic analysis using measurements from a campus network,” in *Proc. 10th Annu. Computer Security Application Conf.*, 1994, pp. 288–297.
[32] I. Way, “Rerouting of internet traffic sparks concern,” *PC Market*, Feb. 5, 1998.
[33] E. W. M. Wong and T.-S. P. Yum, “Maximum free circuit routing in circuit-switched networks,” in *Proc. INFOCOM*, 1990, pp. 934–937.
[34] E. W. M. Wong, T.-S. P. Yum, and K. M. Chan, “Analysis of the M and M^2 routings in circuit-switched networks,” *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 613–619, Sept.–Oct. 1995.
[35] E. W. M. Wong, K. M. Chan, S. Chan, and K. T. Ko, “Bandwidth allocation and routing in virtual path based ATM networks,” in *Proc. IEEE ICC’96*, June 1996, pp. 647–652.
[36] ——, “VP formulation and dynamic routing in ATM networks,” in *Proc. IEEE GLOBECOM’96*, Nov. 1996, pp. 1715–1720.
[37] E. W. M. Wong, A. K. M. Chan, and T. S. P. Yum, “Re-routing in circuit switched networks,” in *Proc. IEEE Infocom 97*, pp. 1375–1381.
[38] E. W. M. Wong, T.-S. P. Yum, and A. K. M. Chan, “A taxonomy of re-routing in circuit-switched networks,” *IEEE Commun. Mag.*, vol. 37, pp. 116–122, Nov. 1999.
[39] A. Yamashita, R. Kawamura, and H. Hadama, “Dynamic VP rearrangement in an ATM network,” in *Proc. IEEE Globecom’95*, 1995, pp. 1379–1383.

Eric W. M. Wong (S’87–M’90) received the B.Sc. and M.Phil. degrees in electronic engineering from The Chinese University of Hong Kong, Shatin, in 1988 and 1990, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Massachusetts at Amherst in 1994.

He joined the City University of Hong Kong, Kowloon, as an Assistant Professor in the Department of Electronic Engineering in 1994. His research interests are in high-speed networks, video-on-demand, satellite communications and dynamic routing. The most notable of these involved the analytical modeling of the least loaded routing scheme in circuit-switched networks. The model drastically reduces the computational complexity of designing and dimensioning telephone systems. The model has also served as a core for the analysis of many other advanced routing schemes, such as the real-time network routing recently implemented in the AT&T telephone network.

Andy K. M. Chan was born in Hong Kong in 1970. He received the B.Eng. and M.Phil. degrees in information engineering from The Chinese University of Hong Kong, Shatin, in 1992 and 1994, respectively.

He is now working as a part-time Research Assistant at the City University of Hong Kong, Kowloon, and his research interest is in network performance analysis.

T.-S. Peter Yum (S’76–A’78–SM’86) worked in Bell Telephone Laboratories, Holmdel, NJ, for two and a half years and taught in National Chiao Tung University, Taiwan, for two years before joining The Chinese University of Hong Kong, Shatin, in 1982. He has published original research on packet-switched networks with contributions in routing algorithms, buffer management, deadlock-detection algorithms, message resequencing analysis and multiaccess protocols. In recent years, he branched out to work on the design and analysis of cellular network, lightwave networks and video distribution networks.