Analysis of adaptive routing schemes in multirate loss networks

Kit-Man Chan and Tak-Shing Peter Yum[a]

Department of Information Engineering, Chinese University of Hong Kong, Hong Kong

[a] E-mail: yum@ie.cuhk.hk

Received February 1994; in final form March 1995

As technology evolves, it is now feasible to implement sophisticated adaptive routing schemes on networks which support different kinds of services with heterogeneous bandwidth characteristics. 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 this paper, we analyze and compare two adaptive routing schemes. The first 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 is the M^2 routing, which is a modification of the least loaded routing (LLR). Aggregation of link status information can significantly reduce signalling traffic. We show in this paper that with properly designed aggregation, the aggregated M^2 and MTB routings can have performance that approach that of the non-aggregated schemes. The use of complete sharing and restricted access policies together with trunk reservation control in multirate loss networks are also studied.

1. Introduction

With the advent of the stored program control switching network and the installation of out-of-band signalling, it is possible to implement more sophisticated dynamic routing schemes through the exchange of link status information during call set-up. 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 an adaptive routing scheme which implements new class-of-service routing capabilities for dynamic networks [1].

Networks supporting different services with different traffic characteristics are called multirate loss networks. The design and analysis of routing rules and flow control methods in this kind of networks have received considerable attention in recent years. In [2], Chung and Ross studied various approximate formulae for computing the loss probability of multirate loss networks under fixed routing. They also studied the sensitivity of the average revenue to the changes of offered load and link capacity. In [3], random alternate routing in circuit-switched networks supporting two classes of services having the same bandwidth requirement are analyzed. In [4], 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 [5], Dziong and Mason reduced the complexity of the problem by decomposing the network reward process into a set of link reward processes. In [6], 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 modelling each link as a one-dimensional birth/death process and derived a set of expressions to evaluate the state-dependent link shadow prices.

In [7], a new decentralized state- and time-dependent adaptive routing scheme called the maximum mean time to blocking (MTB) Routing was proposed in telephone networks. This scheme makes use of the *mean time to blocking* as a measure of links congestion. The measure captures the link loading and capacity information and was shown to be more accurate in reflecting the busy status of a path. Extensive simulation showed that the *mean time to blocking* is a better measure of trunk group congestion than the number of free channels used in the maximum free circuit routing (M routing). In fact, MTB routing was shown to give better blocking performance than both SDR and the M routing.

In this paper, we present an analytical model of the M^2 routing and the *maximum mean time to blocking* (MTB) routing in multirate loss networks based on the fixed point iteration method. The M^2 routing is a modification of the M routing (or LLR) whereas the MTB routing is based on the *mean time to blocking* measure on links. 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 accurately reflect the congestion level since paths can have different loadings. The *mean time to blocking* measure, however, does not have that problem.

A multirate loss network can use either circuit switching or virtual-circuit packet switching. In circuit-switching networks, the bandwidth requirement of a call is defined as the peak bit rate (PBR) of that service and call admission is done based on the bandwidth sharing policy. On the other hand, in packet switching networks (ATM networks), the bandwidth requirement of a call is usually assigned to some smaller value than the PBR and whether a call is accepted or not depends on whether such admission would degrade the quality of service (QOS) of the connections in progress. The bandwidth requirement of such a call can be defined as the equivalent bandwidth under a given QOS requirement [9]. The routing rules studied in this paper are equally applicable to the call-level routing in the virtual-path-based ATM network.

This paper is organized as follows. In section 2, we present the two routing rules. Then in section 3, we discuss bandwidth sharing policies and aggregation of link status information. In sections 4 and 5, the M^2 routing and the MTB routing are analyzed. Numerical results and discussions are presented in section 6. Finally, conclusions are drawn in section 7.

2. Two dynamic routing rules

Consider a network supporting K classes of services where each class is characterized by a Poisson arrival process with arrival rate λ_i, holding time (or service time) exponentially distributed with mean $1/\mu_i$, revenue R_i and bandwidth requirement f_i, $i = 1, 2, \ldots, K$. Let $F^{(j)}$ be the number of channels on link j 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 $\mathscr{L}_q$. Each node pair has a direct path and we consider only two-hop alternate paths, ie, $|\mathscr{L}_q| = 2$.

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 trunk reservation requirement. The residual bandwidth $\beta(q)$ of path q is defined by

$$\beta(q) = \min_{j \in \mathscr{L}_q} \left(F^{(j)} - \sum_{k=1}^{K} n_k^{(j)} f_k^{(j)} \right),$$

where the superscript (j) is the link index.

In case two or more paths have the same β value, the call is routed to path q with the 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 [8] 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.

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. We specify the routing rule and 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 Ω. Let $S(\mathbf{n})$ be the mean sojour time in state $\mathbf{n}$, $p_{\mathbf{ni}}$ be the transition probability from state $\mathbf{n}$ to state $\mathbf{i}$ and

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

be the set of blocking states for class k traffic, where $\mathbf{e}_k$ is a K-vector with a "1" at the kth position and zero elsewhere. The *mean time to blocking* at state $\mathbf{n}$ for class k traffic is given in [13] as

$$T_k(\mathbf{n}) = \begin{cases} S(\mathbf{n}) + \sum_{\mathbf{i} \in \Omega} p_{\mathbf{ni}} T_k(\mathbf{i}), & \mathbf{n} \in \Omega \setminus \Omega_k^{(D)}, \\ 0, & \mathbf{n} \in \Omega_k^{(D)}. \end{cases} \tag{1}$$

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 $\mathcal{L}_q$

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

where the superscript (j) 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 trunk 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 trunk reservation requirement and select one having the largest such value. If there are 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. Bandwidth sharing policies and state aggregation

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 certain bandwidth is reserved for a particular class of calls, we have the restricted access (RA) policy (also known as the threshold-type policy in [10]). Fig. 1 shows the state space of a trunk group of F channels under the CS and the RA policies. Two classes of calls and four-level state aggregations are considered. For the CS policies, the state space Ω_{CS} and the set of class k alternate call blocking states $\Omega_k^{(A)}$ are given by

$$\Omega_{\mathrm{CS}} = \{\mathbf{n} : \mathbf{n} \cdot \mathbf{f} \leq F\},$$
$$\Omega_k^{(A)} = \{\mathbf{n} : \mathbf{n} \in \Omega_{\mathrm{CS}}, (\mathbf{n} + \mathbf{e}_k) \cdot \mathbf{f} + r > F\}.$$

For the RA policy, the class k traffic is blocked when the number of on-going class k calls reaches a threshold h_k. We have similarly

$$\Omega_{RA} = \{\mathbf{n} : \mathbf{n} \in \Omega_{\mathrm{CS}}, n_k \leq h_k, h_k \geq 0\},$$
$$\Omega_k^{(A)} = \{\mathbf{n} : \mathbf{n} \in \Omega_{\mathrm{RA}}, (\mathbf{n} + \mathbf{e}_k) \cdot \mathbf{f} + r > F \quad \text{or} \quad n_k \geq h_k - \lceil r/f_k \rceil, h_k \geq 0\}.$$

Note that $h_k - \lceil r/f_k \rceil$ is the maximum number of alternate class-k calls allowed in a link (after substracting the bandwidth reserved for class-k direct calls). Therefore $n_k \geq h_k - \lceil r/f_k \rceil$ is the condition on n_k for rejecting alternate class-k calls.

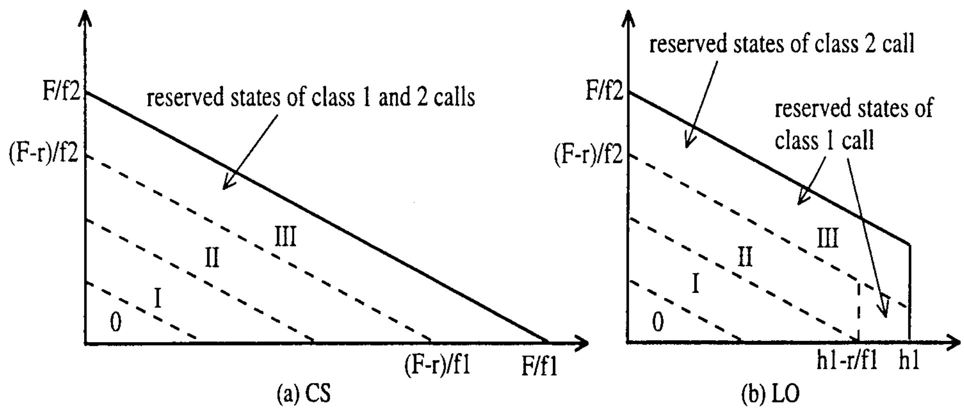


Fig. 1. The (a) CS and (b) RA bandwidth sharing policies.

Second, bandwidth can be shared between the direct and overflowed calls. In order to prevent the unstable behavior at heavy loading conditions, a certain number of channels r can be reserved for direct calls only. This is known as trunk reservation in the conventional telephone network [12].

During a call set-up, the originating switch requests status information of all via links through the common channel signalling (CCS) network. In some existing traffic management systems, this information can often be compressed by lumping several link occupancies into an aggregate-state [11]. 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 signalling 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 routes are computed off-line.

In the following sections, we analyze the 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 $m = N - 2$ alternate paths. Our analysis gives the numerical solution of the equilibrium state probabilities and the alternate traffic rates. With these, the call blocking probabilities and the average revenue loss 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

$$D_k = \sum_{\mathbf{n}\in\Omega_k^{(D)}} P(\mathbf{n}),$$

$$A_k = \sum_{\mathbf{n}\in\Omega_k^{(A)}} P(\mathbf{n}).$$

The class k call blocking probability B_k is given in [12] as

$$B_k = D_k[1 - (1 - A_k)^2]^m.$$

Assuming that each call brings a revenue of $R_k = f_k/\mu_k$, the average revenue loss Υ is the weighted average lost of revenue [5]:

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

4. Analysis of M^2 routing

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

In fig. 2, an alternate path ABC of a link AC is 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 α such paths. The overflowed class k calls of AC will be routed to one of these path, say ABC, at rate $\lambda_k D_k/\alpha$.

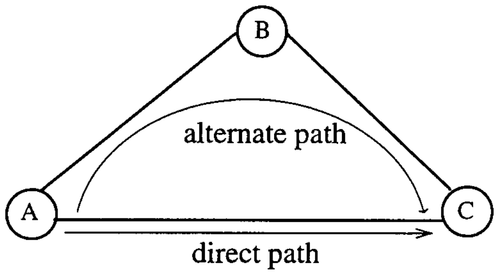


Fig. 2. An alternate path of a node pair.

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

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

Let $b_k(u, v)$ be the probability that *the two links of an alternate path are in aggregate-states u and v, respectively, where $u \geq v$ 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, we have

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

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 $E_0 = E_1 \vee E_2 \vee E_3 \vee E_4$ and $V_k(x,y)$ be the probability of E_0. Since the E_i's, $i = 1, \ldots, 4$ are disjoint, from fig. 3, we have

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

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 (2), $Prob[E_7] = [b_k(x,y)]^{\alpha - 1}$. Now, suppose AC is full and the two links

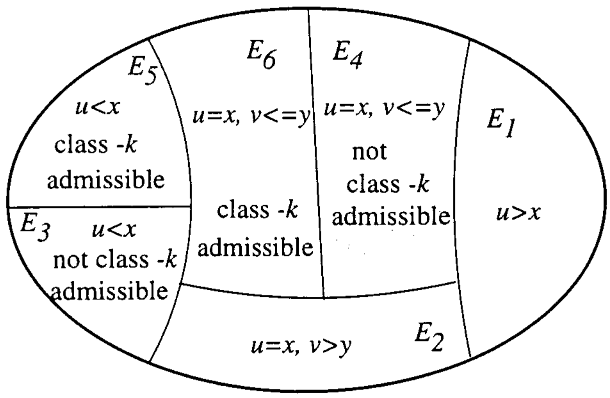


Fig. 3. Different events of a path under M^2 routing.

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,* $f_k(\alpha | x, y)$ is given by

$$f_k(\alpha | x,y) = \binom{m-1}{\alpha-1} Prob[E_7 \wedge E_0]$$
$$= \binom{m-1}{\alpha-1} b_k(x,y)^{\alpha-1} V_k(x,y)^{m-\alpha}.$$

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 over path ABC from node pair AC, denoted by $s_k(x,y)$, is given by

$$s_k(x,y) = \sum_{\alpha=1}^{m} \frac{\lambda_k D_k}{\alpha} f_k(\alpha | x,y)$$
$$= \lambda_k D_k \frac{[b_k(x,y) + V_k(x,y)]^m - V_k(x,y)^m}{m b_k(x,y)}.$$

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

$$A_k(\mathbf{n}) = 2m \sum_{i=0}^{L-2} s_k(\max[g(\mathbf{n}), i], \min[g(\mathbf{n}), i]) \pi_k(i)\mathbf{n}. \tag{3}$$

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

$$\Lambda_k(\mathbf{n}) = \begin{cases} \lambda_k + A_k(\mathbf{n}), & \mathbf{n} \in \Omega \setminus \Omega_k^{(A)}, \\ \lambda_k, & \mathbf{n} \in \Omega_k^{(A)} \setminus \Omega_k^{(D)}, \\ 0, & \mathbf{n} \in \Omega_k^{(D)}. \end{cases} \tag{4}$$

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

$$\sum_k (n_k + 1)\mu_k P(\mathbf{n} + \mathbf{e}_k) + \Lambda_k(\mathbf{n} - \mathbf{e}_k) P(\mathbf{n} - \mathbf{e}_k)$$
$$= \sum_k [\Lambda_k(\mathbf{n}) + n_k \mu_k] P(\mathbf{n}). \tag{5}$$

Let Λ denotes the set of $\Lambda_k(\mathbf{n})$ and $\mathscr{P}$ denotes the set of $P(\mathbf{n})$. Then (4) and (5) can be expressed in the fixed point model form [11]: $\mathscr{P} = f_1(\Lambda)$ and $\Lambda = f_2(\mathscr{P})$. The $P(\mathbf{n})$'s can be computed by the successive over-relaxation (SOR) method [14] with the set of alternate traffic rates obtained from (3) in each iteration. In the examples quoted in section 4, all $P(\mathbf{n})$'s are obtained with relative error less than 10^{-8}.

5. Analysis of MTB routing

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

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 α such paths, the overflowed class k call of AC is randomly routed to one of these paths, say ABC, at rate $\lambda_k D_k/\alpha$.

Consider an alternate path with parameters u and $v(v \le u)$ and define the following events for given thresholds x and $y(y \le x)$ 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}: $v = y, u \ge x$ *and the path is not class-k admissible.*

Let $Y_k(x, y)$ be the probability of $E_8 \vee E_9 \vee E_{10} \vee E_{11}$. From fig. 4, it can be

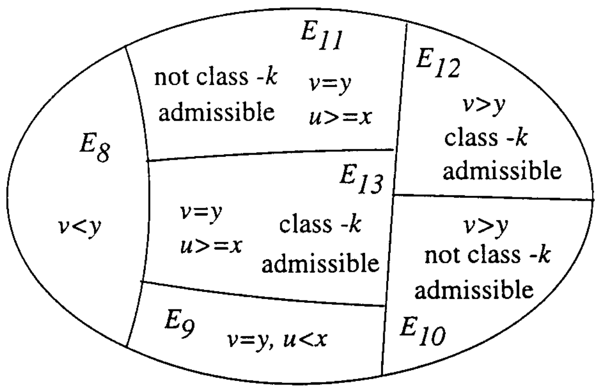


Fig. 4. Different events of a path under MTB routing.

expressed as

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

Now, suppose AC is full and the two links of a class-k admissible alternate path ABC are 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*, $f_k(\alpha | x, y)$ is

$$f_k(\alpha | x,y) = \binom{m-1}{\alpha-1} b_k(x,y)^{\alpha-1} Y_k(x,y)^{m-\alpha}.$$

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

$$s_k(x,y) = \sum_{\alpha=1}^{m} \frac{\lambda_k D_k}{\alpha} f(\alpha | i)$$
$$= \lambda_k D_k \frac{[b_k(x,y) + Y_k(x,y)]^m - Y_k(x,y)^m}{m b_k(x,y)}.$$

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

$$A_k(\mathbf{n}) = 2m \sum_{i=0}^{L-1} s_k(\max[g(\mathbf{n}), i], \min[g(\mathbf{n}), i]) \pi_k(i). \tag{6}$$

Denote $\mathcal{T}$ as the set of *mean time to blocking*, then (5), (1) and (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 Λ given by the solution of (1) and (6) in each iteration.

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 $(\mu_1, \mu_2) = (1, 0.5)$. Restricted access policy is employed on class 1 call with threshold $h_1 = 8$. The number of channels in each link is twelve and the trunk reservation parameter is three. The base traffic rates is $(\lambda_1, \lambda_2) = (4, 0.6)$. Fig. 5 shows the class 1 alternate traffic rate for the CS policy at 60% overload. As expected, the alternate traffic decreases with increasing n_1 and n_2.

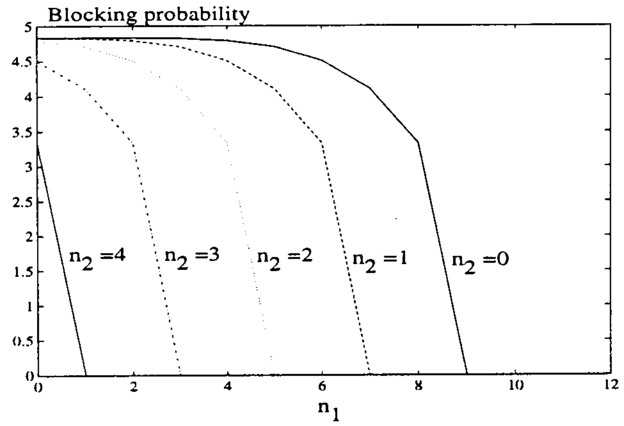


Fig. 5. Alternate traffic rate of class 1 traffic against n_1.

Fig. 6 shows the analytic and simulation results of the M^2 routing as a function of the percentage overload from the base load. We find that the analytic result matches well with the simulation result except at blocking level 10^{-4} or lower (which are not shown in the diagram) where the analytic result underestimates the blocking probability. The reason is that in this region, the overflowed traffic becomes very bursty and the assumption of Poisson overflowed traffic causes an underestimation of the blocking probability. 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 RA policy as shown.

Fig. 7 shows the blocking probabilities of a network with twelve nodes and ten channels per link. We let two channels be reserved for direct calls and consider two classes of calls with the same bandwidth requirement, i.e. $(f_1, f_2) = (1, 1)$. Let $(\lambda_1, \lambda_2) = (4, 1.2)$ and $(\mu_1, \mu_2) = (1, 0.5)$. We find that by reserving channels for a certain class of calls, the relative blocking probabilities of different classes can be manipulated at will. Note that the CS policy is the same as the RA policy with $h_1 = 10$.

Figs. 8(a) and 8(b) show the blocking probability and average revenue loss as a function of the number of channels reserved for class 2 calls. The network studied has twelve nodes and each link has fifteen channels. The bandwidth requirement of the two classes are $(1, 3)$ and the mean service rates are $(1, 0.4)$. The network is

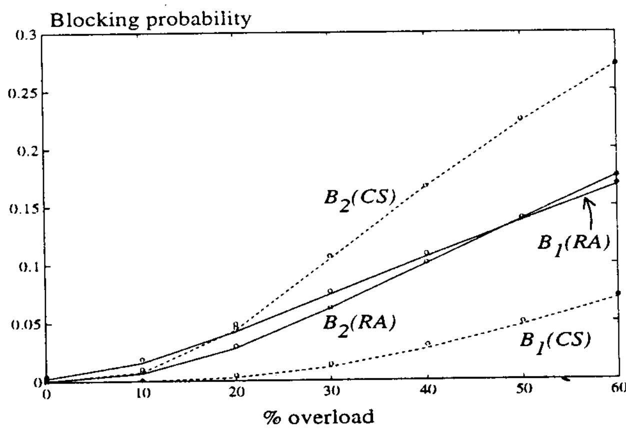


Fig. 6. Simulation and analytic results of M^2 routing.

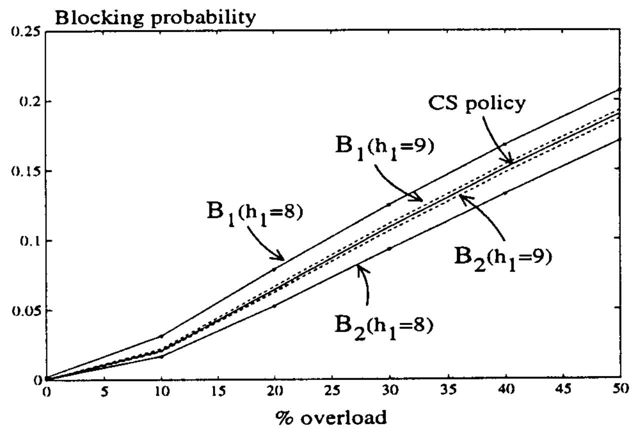
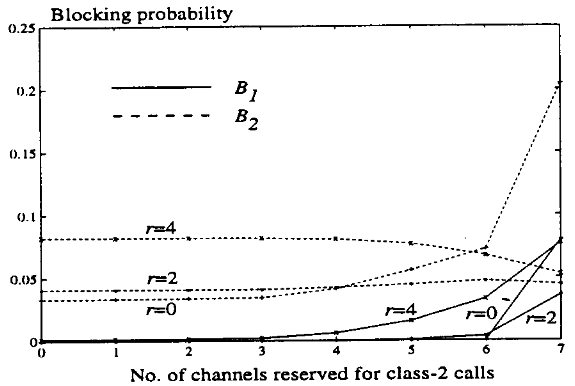
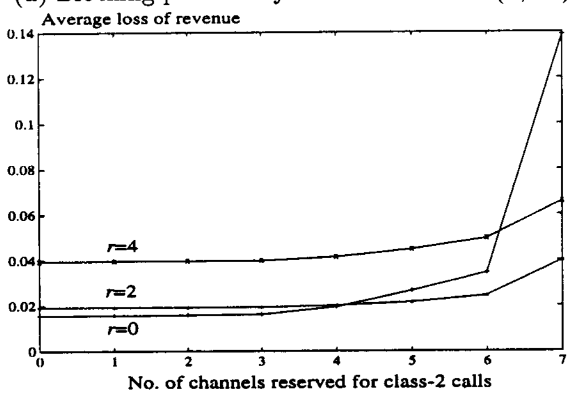


Fig. 7. Blocking of M^2 routing with $(f_1, f_2) = (1, 1)$.

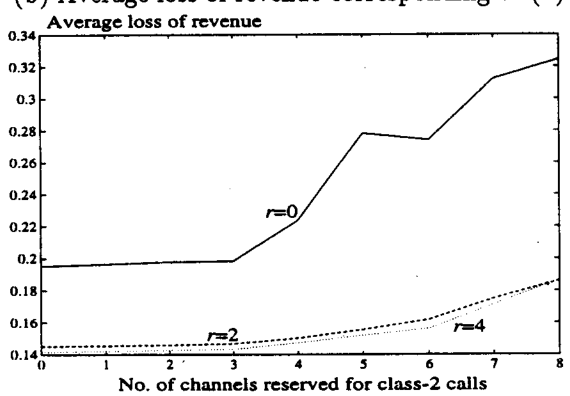
loaded with calls arrival rates $(5, 0.6)$. Note that when $r = 0$, the blocking probability of both classes of calls increase with the number of channels reserved for class 2 calls. This is obviously due to the presence of alternate traffic. We observe that with alternate traffic, the RA policy cannot reduce the blocking probabilities of the class 2 call without trunk reservation and the optimal trunk reservation parameter depends on the number of reserved channels used in the RA policy. For instance, if the number of channels reserved for class 2 call changes from 0 to 6, then the optimal trunk reservation parameter changes from 0 to 2 (fig. 8(b)). In fig. 8(c), the same network is now heavily loaded with calls arrival rates $(6, 0.8)$. We observe that the use of optimal trunk reservation ($r = 4$) can significantly reduce the average revenue loss.



(a) Blocking probability with traffic rates (5,0.6)



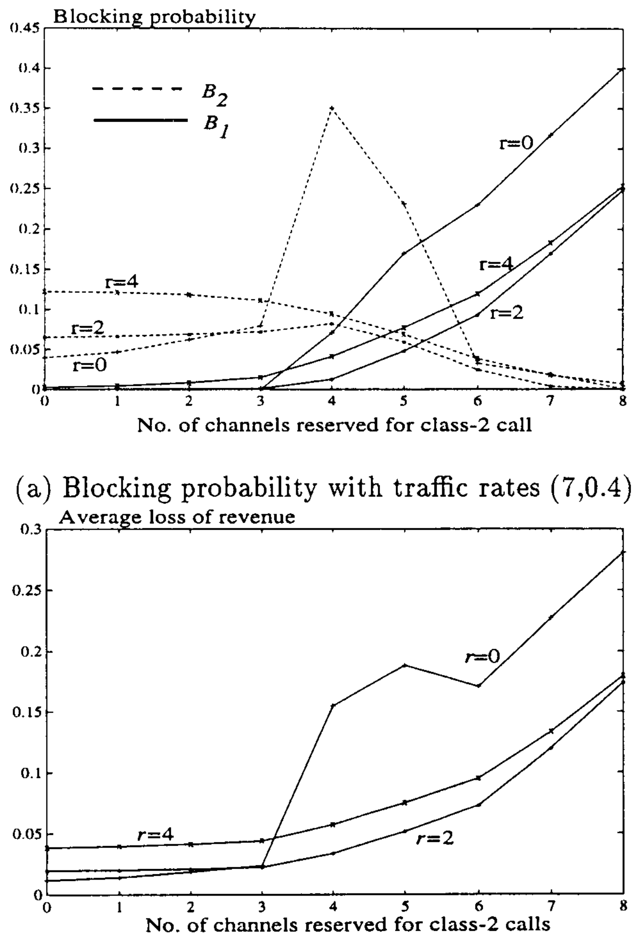
(b) Average loss of revenue corresponding to (a)



(c) Average loss of revenue with traffic rates (6,0.8)

Fig. 8. The interplay of RA policy and trunk reservation.

To show the trade-off of blocking probability and average revenue loss, we change the call arrival rate to (7, 0.4) in fig. 9. It is shown that the RA policy is in fact trading the decrease of the blocking probability with the increase of the average revenue loss. For instance, with $r = 4$ and $h_1 = 10$ (number of reserved channels = 5), the blocking of class 2 calls is significantly reduced and the average loss of revenue is only slightly increased when compared to that with $r = 4$ and $h_1 = 15$ (number of reserved channels = 0). In general, in a multirate network, the optimal control parameters (i.e., r and h_i) should reflect not only the maximum revenue gain (or minimum revenue loss), but should also take into consideration the blocking levels of different services requests.



(a) Blocking probability with traffic rates (7,0.4)

(b) Average revenue loss

Fig. 9. The trading of blocking and loss of revenue.

Fig. 9(a) also shows that when $r = 0$, the blocking of class 2 calls increases tremendously when the number of reserved channels for class 2 traffic increases from 3 to 4 and then decreases abruptly when it changes from 5 to 6. The reason for this phenomenon is that the channels reserved 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 trunk reservation is not imposed. This phenomenon vanishes when the number of reserved channels is further increased to 6 at the expense of higher blocking of the class 1 calls.

Fig. 10 shows the blocking probabilities as a function of the percentage

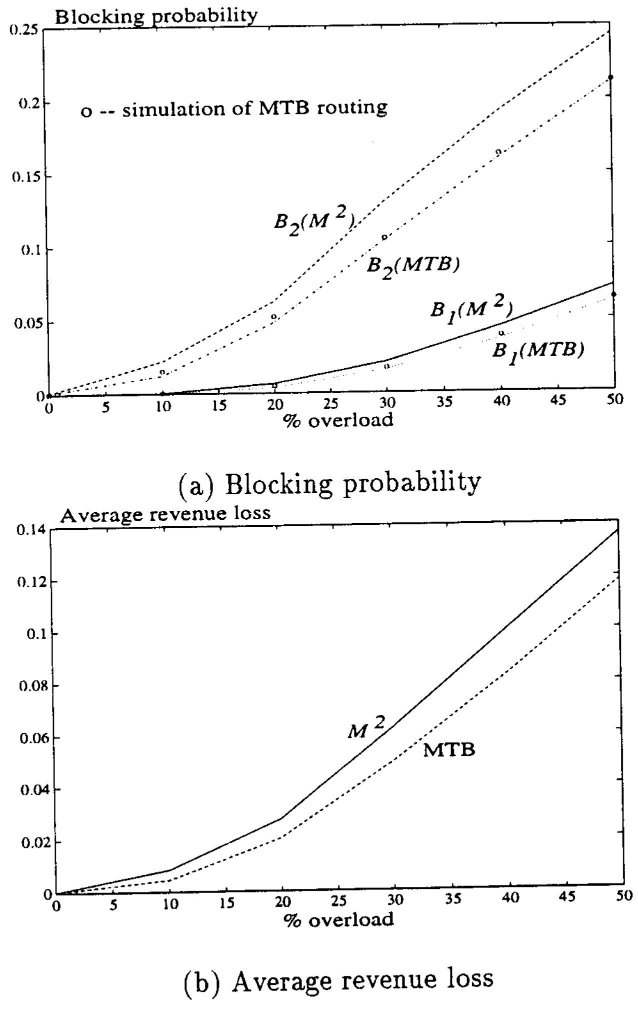


Fig. 10. Comparison of M^2 and MTB routings.

overload for the M^2 and the MTB routing. The network studied has twelve nodes and each link has fifteen channels. Three channels are reserved for direct traffic and four channels are reserved for class 2 calls. 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 overflow 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 compositions. 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 compositions. MTB routing therefore can identify more accurately the best alternate path.

Next, we consider the effect of state aggregation [11]. Table 1 shows the performance of the aggregated M^2 routing with the same control and network parameters as the RA policy in fig. 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

Table 1
Blocking of the aggregated M^2 routing.

Overload (%)	Non-aggregated M^2		$L = 2$	
	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
	$L = 3$		$L = 4$	
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

routing is already close to that of the non-aggregated scheme. We also find that at light loading, more levels are needed to approach the non-aggregated scheme because the alternate traffic rate decreases more slowly when compared to that in heavy loading conditions.

Table 2 shows the performance of the aggregated MTB routing with the same control and network parameters as the RA policy in fig. 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 2
Blocking of the aggregated MTB routing.

Overload (%)	Non-aggregated MTB		0.5 s-interval	
	B_1	B_2	B_1	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
	1 s-interval		2 s-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

7. Conclusions

The MTB routing and the M^2 routing are analyzed under symmetric traffic conditions in this paper. Numerical results show that the MTB routing gives a better performance than the M^2 routing.

The aggregated version of these two routing schemes have been analyzed. It is found that with properly designed aggregation, both schemes can perform closely to the non-aggregated schemes. It is interesting to study the performance of non-uniform aggregation in MTB routing. We have also studied the congestion control methods and showed that the control parameters should reflect not only the minimum average revenue loss but also the blocking probability of different service requests.

Accurate traffic rate estimation in real time is very difficult. How robust MTB routing is requires further study.

References

[1] G.R. Ash et al., Real-time network routing in a dynamic class-of-service network, *Proc. Globecom* (1992).

[2] S.-P. Chung and K.W. Ross, Reduced load approximations for multirate loss networks, to appear in IEEE Trans. Commun.

[3] W. Wang and N. Saadawi, Trunk congestion control in heterogeneous circuit-switched networks, IEEE Trans. Commun. 40(7)(1992).

[4] S. Gupta, K.W. Ross and M. El Zarki, Routing in virtual path based ATM networks, *Proc. Globecom* (1992).

[5] Z. Dziong and L. Mason, An analysis of near optimal call admission and routing model for multi-service loss networks, *Proc Infocom* (1992).

[6] R.-H. Hwang, J.F. Kurose and D. Towsley, State dependent routing for multirate loss networks, *Proc. Globecom* (1992).

[7] K.-M. Chan, T.-S.P. Yum, The maximum mean time to blocking routing in circuit-switched networks, IEEE J. Select. Areas Commun. 12 (Feb. 1994).

[8] E.W.-M. Wong, T.-S.P. Yum and K.-M. Chan, Analysis of the M and M^2 routings in circuit-switched networks, *Proc. Globecom* (1992).

[9] R. Guérin, H. Ahmadi and M. Naghshineh, Equivalent capacity and its application to bandwidth allocation in high-speed networks, IEEE J. Select. Areas Commun. 9(7)(1991).

[10] K.W. Ross and D.H.K. Tsang, The stochastic knapsack problem, IEEE Trans. Commun. 37(7)(1989).

[11] D. Mitra, R.J. Gibbens and B.D. Huang, Analysis optimal design of aggregated-least-busy-alternate routing on symmetric loss network with trunk reservations, *Proc. ITC-13* (1991).

[12] M. Schwartz, *Telecommunication Network: Protocols, Modeling and Analysis* (Addison-Wesley, 1988).

[13] R.A. Howard, *Dynamic Probabilistic System, Vol II: Semi-Markov and Decision Processes* (Wiley, 1971).

[14] K.E. Atkinson, *An Introduction to Numerical Analysis* (Wiley, New York, 1978) pp. 450–457.