An Analytical Model on the Blocking Probability of a Fault-Tolerant Network

Mathew P. Haynos and Yuanyuan Yang, *Senior Member, IEEE*

Abstract—The well-known Clos network has been extensively used for telephone switching, multiprocessor interconnection and data communications. Much work has been done to develop analytical models for understanding the routing blocking probability of the Clos network. However, none of the analytical models for estimating the blocking probability of this type of network have taken into account the very real possibility of the interstage links in the network failing. In this paper, we consider the routing between arbitrary network inputs and outputs in the Clos network in the presence of interstage link faults. In particular, we present an analytical model for the routing blocking probability of the Clos network which incorporates the probability of interstage link failure to allow for a more realistic and useful determination of the approximation of blocking probability. We also conduct extensive simulations to validate the model. Our analytical and simulation results demonstrate that for a relatively small interstage link failure probability, the blocking behavior of the Clos network is similar to that of a fault-free network, and indicate that the Clos network has a good fault-tolerant capability. The new integrated analytical model can guide network designers in the determination of the effects of network failure on the overall connecting capability of the network and allows for the examination of the relationship between network utilization and network failure.

Index Terms—Multistage interconnection networks, performance analysis, analytical model, fault tolerance, blocking probability, Clos network, random routing.

1 Introduction

Ongoing microprocessor developments have recently sparked interest in large-scale multiprocessors composed of hundreds or thousands of processors and in data communication networks allowing for delivery of advanced digital services. These developments have resulted in an increased focus on the connecting capabilities and the reliability of interconnection networks responsible for connecting the processing nodes in the network. To eliminate the need to support a direct connection from a given source node to each destination node, many interconnection networks are comprised of intermediate stages of switches used to route connection requests through. This type of interconnection network is generally referred to as *multistage interconnection networks* (MINs).

While there have been numerous designs proposed for multistage interconnection networks, each having its own merits, a network design proposed by Clos [1] originally for telephone networks continues to find applications today in multiprocessor interconnection networks and data communication networks. For example, the NEC ATOM switch for Broadband Integrated Services Digital Network (BISDN) is based on the three-stage Clos architecture [2] and, more recently, it was shown that the network in the IBM SP2 parallel computer is functionally equivalent to the Clos network [3].

It is not surprising that, as the number of components increases in a computing system (which an interconnection network is a significant part of), issues of fault tolerance and reliability as they relate to the interconnection network become ever more of a concern. Much of the study of fault-tolerant interconnection networks has been focused on network architectures which support some form of hardware redundancy, thereby allowing for limited types of network component failure. Examples include the Multipath Omega network [4], Enhanced IADM network [5], d-dilated square banyan network [6], and Multibutterflies network [7]. Key distinguishing characteristics of fault-tolerant interconnection network architectures are the scope of component failure they allow for, with only a few encompassing a complete set, and in the number of faults tolerated [8].

Additionally, the reliability analysis of interconnection networks has been concentrated on computing some measure of overall network reliability and there have been many techniques proposed [9]. For example, [10], [11], [12] presented algorithms on the determination of fault-free path availability in several multipath MINs. However, most models have not incorporated the various states an interstage link may be in. Of particular difficulty in the reliability analysis of interconnection networks is that many problems are computationally intractable.

Important work has also been done in establishing analytical models for understanding the connecting capability of an interconnection network; see, for example, [13], [14], [15], [16], [17], [18]. These models approximate the probability that an arbitrary connection request between a network input and a network output cannot be successfully routed through the network, i.e., the blocking probability, given various network component probabilities such as the probability that an input/output link is busy and the probability that an interstage link is busy. Few models,

- *M.P. Haynos is with IBM Corporation, 527 Encinitas Blvd. Encinitas, CA 92024. E-mail: haynosm@us.ibm.com.*
- *Y. Yang is with the Department of Electrical and Computer Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794. E-mail: yang@ece.sunsb.edu.*

For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number 110031.

1045-9219/99/$10.00 © 1999 IEEE

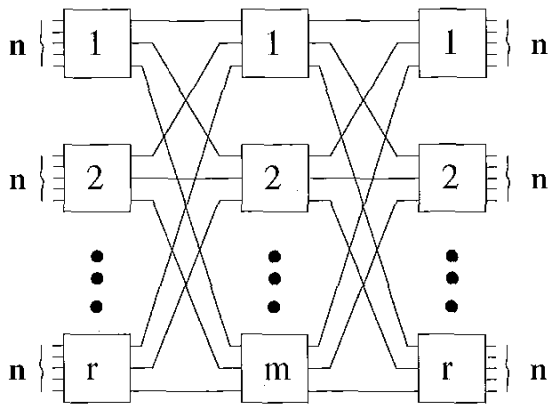

Fig. 1. General schematic of a three-stage Clos network.

however, have incorporated the probability of the various network components failing in their determination of blocking probability.

With the gain in importance of fault-tolerant and reliable interconnection networks then, analytical models which distinguish between both network failure and network utilization can provide a more realistic and useful measure of blocking probability. In this paper, we consider the routing between arbitrary network inputs and outputs in the Clos network in the presence of interstage link faults. In particular, we present an analytical model for the approximation of the routing blocking probability of the Clos network which incorporates interstage link failure probability. This new type of integrated analytical model can guide network designers in the determination of the effects of network failure on the overall connecting capability of the network and allows for the examination of the relationship between network utilization and network failure. We also conduct extensive simulations to validate the model. As can be seen later, our analytical and simulation results indicate that, for a small interstage link failure probability, the blocking behavior of the Clos network is similar to that of a fault-free network.

The rest of this paper is organized as follows. Section 2 provides some definitions used in this paper. Section 3 briefly describes previous related work. Section 4 derives the new analytical model for the fault-tolerant Clos network. Section 5 gives some further discussions on the new model. Section 6 describes the experimental simulations and compares the simulation results with the analytical ones. Section 7 concludes the paper.

2 Preliminaries

The general Clos network is comprised of an input stage, an output stage, and an odd number of middle stages. Each stage consists of multiple switch modules. We will concentrate on the basic three-stage Clos network in this paper since any odd number stage networks with various switch sizes can be built in a recursive fashion from the three-stage networks. The schematic of a three-stage Clos network is depicted in Fig. 1. A switch module with n input ports and m output ports is referred to as an $n \times m$ *switch*. The first stage in the three-stage network is called the *input stage* and consists of r *input stage switches* of size $n \times m$. The second stage in the network is referred to as the *middle stage* and consists of m *middle stage switches* of size $r \times r$. The third stage in the network is called the *output stage* and consists of r *output stage switches* of size $m \times n$. Each input stage switch has exactly one connection to each of the m middle stage switches and this connection is referred to as an *input-middle interstage link*. Additionally, each middle stage switch has exactly one connection to each of the r output stage switches and this connection is referred to as a *middle-output interstage link*. An interstage link is said to be *functional* if it is capable of transmitting data and *faulty* otherwise. Furthermore, an interstage link is *available* if it is functional and not busy. The *fault model* we will use assumes that the interstage links in the network may fail and that these failures are *permanent*.

We consider the network capable of *one-to-one* or *unicast communication*. A *connection request* is a request to transmit data from an input port on an input stage switch to an output port on an output stage switch. A *legal* connection request is a connection request for which both the input port and the output port are not busy. Given a legal connection request, a *routing algorithm* attempts to route the connection in the network. A *path* is uniquely defined by an input port, an input stage switch, a middle stage switch, an output stage switch, and an output port.

A legal connection request is said to be *satisfiable* if a path can be found for which both the input-middle interstage link and the middle-output interstage link are available. Otherwise, the connection request is *blocked*. The *blocking probability* is the probability that a legal connection request is blocked. Finally, the *network utilization* is the percentage of the network input ports servicing active connections at any time.

In attempting to satisfy a legal connection request, a routing algorithm must be used to find a path through the network. The routing algorithm we will consider in this paper is a *random-routing* algorithm. A random-routing algorithm attempts to satisfy a legal connection request by randomly selecting an input-middle interstage link from the set of all available input-middle interstage links emanating from the input switch of the connection request. Next, the middle-output interstage link from the chosen middle stage switch to the output stage switch of the connection request is checked to see if it is available. If it is, then a path is established in the network using the input-middle interstage link and the middle-output interstage link and the connection request is satisfied. If it is not, then another available input-middle interstage link emanating from the input stage switch of the legal connection request is chosen and the process is repeated. If a path cannot be established in the network, then the connection request is blocked.

3 Previous Related Work

The Clos network has been extensively studied in the literature; see, for example, [20], [21], [22], [23], [24], [25] for deterministic results, which focus on determining network structural parameters for a certain type of connecting capability, and [13], [14], [15], [16], [17], [18], [19] for

probabilistic results, which focus on analyzing the blocking probability of the network. Because the Clos network can potentially support many possible routings from a source node to a destination node, it is inherently more fault-tolerant of interstage link failure than many interconnection network designs. One may expect that the network has a good fault-tolerant capability in the presence of link faults. However, to our knowledge, no existing work has considered the fault-tolerant capability of the Clos network. In this paper, we address this issue in the context of probabilistic analysis. We will examine how interstage link failures affect the blocking probability of the network.

Much work has been done in the analysis of blocking probability for multistage interconnection networks. The work can be generally classified into two categories of interest. The first is from the viewpoint of reliability analysis and is referred to as the *terminal reliability* problem. Terminal reliability is defined as the probability that there is at least one operative path between a given pair of network input port and output port. For example, [10], [11], [12] presented algorithms for computing terminal reliability in several multipath MINs. The second category of work has been focused on establishing analytical models based on stochastic network parameters (i.e., network utilization) for the blocking probability of the network. Among the analytical models proposed for the Clos network in the literature that estimate blocking probability, two well-known and widely used models for random routing were proposed by Lee [14] and Jacobaeus [13], respectively. Both models assume that the incoming traffic is uniformly distributed over the m interstage links of the Clos network and the events that individual links are busy are independent. Lee gave the simplest method for analyzing the blocking probability of the Clos network. Let n be the number of ports on a particular input or output switch, m $(\geq n)$ be the number of middle stage switches, a be the network utilization, p be the probability that an interstage link is busy and be defined as $p = \frac{an}{m}$, q be the probability that an interstage link is idle and be defined as $q = 1 - p$. Then in Lee's model, the blocking probability, P_B, is given by

$$P_B = (1 - q^2)^m. \tag{1}$$

A more accurate model was provided by Jacobaeus [13], and the blocking probability of the three-stage Clos network is calculated by the following formula:

$$P_B = \frac{(n!)^2(2-a)^{2n-m}a^m}{m!(2n-m)!}. \tag{2}$$

Both models, however, do not meet the deterministic nonblocking condition set forth by Clos [1] that, when $m \geq 2n - 1$, the Clos network is nonblocking for arbitrary one-to-one communication. Recently, Yang [18] has presented a more accurate analytical model which still follows the same assumptions as in these two models but is proven to agree with the deterministic nonblocking condition as well. In this model, the probability that a connection request is not blocked for a three-stage Clos network is given by

$$\Pr\{\text{connection not blocked}\} = \frac{\sum_{n_1=0}^{n-1}\sum_{n_2=0}^{n-1}\sum_{k=\max\{0,n_1+n_2-m-1\}}^{\min\{n_1,n_2\}}\binom{m}{n_1}\binom{n_1}{k}\binom{m-n_1}{n_2-k}p^{n_1+n_2}q^{2m-n_1-n_2}}{\left[\sum_{j=0}^{n-1}\binom{m}{j}p^j q^{m-j}\right]^2} \tag{3}$$

and the blocking probability is given by

$$P_B = 1 - \Pr\{\text{connection not blocked}\}. \tag{4}$$

It is interesting to note that the categories seem somewhat discontinuous. Many models presented in the context of reliability analysis do not provide for a precise measure of blocking probability because of the view of network components as being either available or unavailable. The focus of these models has not been on incorporating network utilization parameters. Furthermore, analytical models presented for the blocking probability of the network incorporate network utilization, but do not support the notion of network component failure. We can expect then that analytical models which integrate the fact that network links can be either busy or faulty will provide a more realistic measure of network blocking probability.

4 A New Probabilistic Model for Incorporating Link Failure

In this section, we present an analytical model for the blocking probability of a three-stage Clos network which incorporates interstage link failure and allows for a more realistic measure of blocking probability. In general, determination of blocking probability in a multistage network is inherently complex and difficult. This is due to the fact that there are many possible paths to consider in a typical large network and the dependencies among links in the network lead to combinatorial explosion problems. Therefore, some approximations and assumptions are necessary for the calculations.

Let us first consider the network state depicted in Fig. 2, in which n_1 input-middle interstage links from input switch i are busy, f_1 input-middle interstage links from input switch i are faulty, n_2 middle-output interstage links to output switch j are busy, and f_2 middle-output interstage links to output switch j are faulty, where $0 \leq n_1, n_2 \leq n-1$, $0 \leq f_1 \leq m - n_1$, $f_2 \leq m - n_2$, and k pairs of these interstage links are overlapped. A pair of interstage links is said to be *overlapped* if both links are either busy or faulty. Note that by definition n_1 and f_1 (and, similarly, n_2 and f_2) are mutually exclusive, since an interstage link cannot be both busy and faulty. Further, it is important to note that the number of busy input-middle interstage links and the number of faulty input-middle interstage links are constrained by m, that is, $0 \leq n_1 + f_1 \leq m$. Similarly, for n_2 and f_2, we have $0 \leq n_2 + f_2 \leq m$.

Having established the relationship between busy and faulty interstage links, it is evident that a new analytical model which incorporates interstage link failures can give a more realistic and useful measure of blocking probability of the Clos Network. In accounting for interstage link failures,

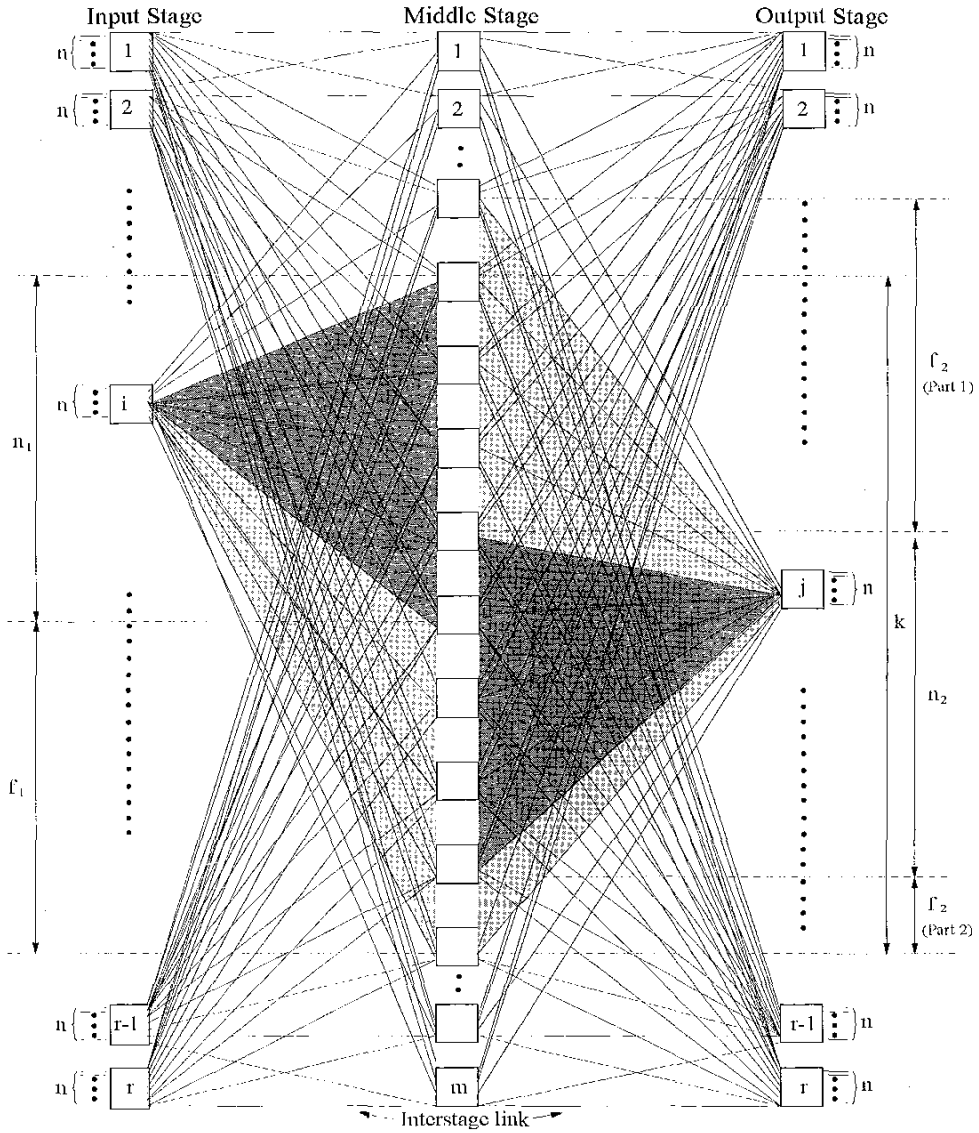

Fig. 2. A three-stage Clos network with n_1 busy input-middle interstage links from input stage switch i, n_2 busy middle-output interstage links to output stage switch j, f_1 faulty input-middle interstage links from input stage switch i, f_2 faulty middle-output interstage links to output stage switch j, and k pairs of links overlapped.

we need to define four events: f_1, the event that f_1 input-middle interstage links are faulty, f_2, the event that f_2 middle-output interstage links are faulty, n_1, the event that n_1 input-middle interstage links are busy and n_2, the event that n_2 middle-output interstage links are busy. Given these four events we can establish the probability that k pairs of interstage links are overlapped. Looking at Fig. 2 shows that there are four states that a pair of input-middle interstage link and middle-output interstage link can be in: busy and busy, faulty and faulty, busy and faulty, and faulty and busy. Having defined overlapped links, we now determine the probability that some k pairs of interstage links are overlapped, given events n_1, f_1, n_2, and f_2. We assume that the locations of faulty interstage links are independent and the faults are uniformly distributed over all interstage links. Under these assumptions and by taking advantage of the fact that an interstage link cannot both be busy and faulty, we have the following lemma concerning the overlapped interstage links.

Lemma 1. *Given events n_1, n_2, f_1, and f_2, the probability that k pairs of interstage links are overlapped is given by*

$$\Pr\{k \text{ pairs of links overlapped} \mid n_1, n_2, f_1, f_2\} = \frac{\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_2+f_2}} = \frac{\binom{n_2+f_2}{k}\binom{m-n_2-f_2}{n_1+f_1-k}}{\binom{m}{n_1+f_1}} \tag{5}$$

Proof. As shown in Fig. 2, there are a total of

$$\binom{m}{n_1+f_1}\binom{m}{n_2+f_2}$$

ways to choose n_1+f_1 busy or faulty input-middle interstage links and n_2+f_2 busy or faulty middle-output interstage links. We can construct k pairs of overlapped interstage links in the following way: Initially, we select the n_1+f_1 input-middle interstage links from a total of m input-middle interstage links and there are

$$\binom{m}{n_1+f_1}$$

ways to do this; then, k busy or faulty input-middle interstage links which are overlapped with k busy or faulty middle-output interstage links can be chosen from the n_1+f_1 input-middle interstage links and there are

$$\binom{n_1+f_1}{k}$$

ways to do this; finally, we must select the rest of the n_2+f_2-k busy or faulty middle-output interstage links from the remaining $m-n_1-f_1$ busy or faulty input-middle interstage links and there are

$$\binom{m-n_1-f_1}{n_2+f_2-k}$$

ways to do this. Therefore, the probability that k pairs of links are overlapped is

$$\frac{\binom{m}{n_1+f_1}\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_1+f_1}\binom{m}{n_2+f_2}}=\frac{\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_2+f_2}}.$$

The probability can also be attained symmetrically by constructing the k busy or faulty input-middle interstage links which are overlapped with k busy or faulty middle-output interstage links by initially selecting n_2+f_2 middle-output interstage links from a total of m middle-output interstage links, then the k overlapped links from n_2+f_2, and, finally, the remaining n_1+f_1-k from $m-n_2-f_2$. □

We have determined the probability that k pairs of links are overlapped. We now establish the relationship between this fact and the probability that a connection request is not blocked. A connection request from input switch i to output switch j is not blocked if there exists at least one path from input switch i to output switch j in which both the input-middle and middle-output interstage links are not busy and are functional. This condition can be represented by

$$n_1+n_2+f_1+f_2-k<m,$$

which implies

$$k\geq n_1+n_2+f_1+f_2-m+1$$

and

$$k\geq \max\{0,n_1+n_2+f_1+f_2-m+1\}.$$

Having established a lower bound for the k overlapped links, it is obvious from Fig. 2 that

$$k\leq \min\{n_1+f_1,n_2+f_2\}.$$

Therefore, the probability that a connection is not blocked, given events n_1, f_1, n_2, and f_2 is given by

$$\Pr\{\text{connection not blocked} \mid n_1,n_2,f_1,f_2\}=\sum_{k=\max\{0,n_1+n_2+f_1+f_2-m+1\}}^{\min\{n_1+f_1,n_2+f_2\}}\frac{\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_2+f_2}} \quad (6)$$

Given (6), we now need to determine the probability of the simultaneous occurrence of the four events n_1, f_1, n_2 and f_2. Since there is dependency between events n_1 and f_1, and between events n_2 and f_2, we cannot simply assume that the four events are independent. However, we can group events n_1 and f_1 together, and events n_2 and f_2 together. Similar to previous work [13], [14], in order to make the calculation possible it is reasonable to assume that the occurrence of events n_1 and f_1 is independent of the occurrence of events n_2 and f_2 for sufficiently large n and r. Under this assumption we have

$$\Pr\{n_1,n_2,f_1,f_2\}=\Pr\{n_1,f_1\}\cdot\Pr\{n_2,f_2\}. \quad (7)$$

Let us calculate $\Pr\{n_1,f_1\}$ and $\Pr\{n_2,f_2\}$. We know that we can have at most m input-middle (or middle-output) interstage links which may be busy or faulty. Furthermore, it is evident that an interstage link cannot be both busy and faulty, which implies that a busy link is functional. Therefore, an interstage link can have three possible states:

1. faulty
2. functional and busy
3. functional and idle.

Let p_f denote the probability that an interstage link is faulty, p_b denote the probability that an interstage link is functional and busy, and q denote the probability that an interstage link is functional and idle. p_f may be specified by the network designer and can be determined by examining historical data on interstage link failures. However, as we shall see, a constraint exists which must be satisfied. Also, it is expected that p_f will be rather small, as larger values indicate an increasingly unreliable network. Furthermore, we need to ensure that $p_b+p_f\leq 1$ because an interstage link cannot be both busy and faulty. As in [18] and [14], let a be the probability that a typical input or output port is busy. We assume that the faults are uniformly distributed over m interstage links and the incoming traffic is uniformly distributed over all functional interstage links. Thus, for a given p_f, the average number of functional interstage links is $(1-p_f)m$ and the probability that an interstage link is busy can be calculated by

$$p_b=\min\left\{\frac{an}{(1-p_f)m},1\right\}. \quad (8)$$

Clearly, p_b is a function of p_f. We are interested in the small values of p_f and consider the case $\frac{an}{(1-p_f)m}\leq 1$. Furthermore, it is important that values for n, m, a, and p_f conform to the constraint $p_b+p_f\leq 1$, which implies

$$\frac{an}{(1-p_f)m} + p_f \leq 1.$$

Therefore, p_f must be chosen such that

$$p_f \leq 1 - \sqrt{\frac{an}{m}}. \tag{9}$$

In addition, we represent the probability that an interstage link is idle by

$$q = 1 - p_b - p_f. \tag{10}$$

To start with, it is reasonable to assume that the number of functional and faulty input-middle interstage links (or middle-output interstage links) follows a binomial distribution and the number of busy and idle input-middle interstage links (or middle-output interstage links) also follows a binomial distribution. However, given the fact that, among the m input-middle (or middle-output) interstage links, at most $n-1$ of them can be busy, we can obtain a more accurate probability distribution. Therefore, we represent the probability of the joint occurrence of events n_1 and f_1 as

$$\Pr\{n_1, f_1\} = \frac{\binom{m}{f_1}\binom{m-f_1}{n_1} p_f^{f_1} p_b^{n_1} q^{m-n_1-f_1}}{\sum_{i=0}^{n-1}\sum_{j=0}^{m-i}\binom{m}{j}\binom{m-j}{i} p_f^{j} p_b^{i} q^{m-i-j}} \tag{11}$$

and the probability of the joint occurrence of events n_2 and f_2 as

$$\Pr\{n_2, f_2\} = \frac{\binom{m}{f_2}\binom{m-f_2}{n_2} p_f^{f_2} p_b^{n_2} q^{m-n_2-f_2}}{\sum_{i=0}^{n-1}\sum_{j=0}^{m-i}\binom{m}{j}\binom{m-j}{i} p_f^{j} p_b^{i} q^{m-i-j}}. \tag{12}$$

Given (6)-(12), we can now obtain the probability that a connection request is not blocked, accounting for interstage link failures. It is given by

See Fig. 3. (13)

and the blocking probability is

$$P_B = 1 - \{\Pr \text{ connection not blocked}\}. \tag{14}$$

Note that in (13) the summation indices for f_1 and f_2 are from 0 to $m - n_1$ and from 0 to $m - n_2$, respectively. This is to account for the constraints $0 \leq n_1 + f_1, n_2 + f_2 \leq m$ and $0 \leq n_1, n_2 \leq n-1$.

Of particular interest in the above model is the special case $p_f = 0$, where the probability that an interstage link is faulty is equal to 0, that is, a fault-free network. Let's define $0^0 = 1$. Then, when $p_f = 0$, (13) becomes

See Fig. 4. (15)

Thus, if $p_f = 0$, (13) reverts to the model for the non-fault-tolerant Clos network in (3).

5 Discussions on the Analytical Model

In this section, we take a closer look at the new analytical model under several network configurations. We are primarily interested in the effects of both an increasing failure rate and an increasing number of middle stage switches, given a constant failure rate, on the blocking probability given by (13).

Fig. 5 plots the blocking probability for three network configurations: $n = r = 16$ with $16 \leq m \leq 28$, $n = r = 32$ with $32 \leq m \leq 48$, and $n = r = 64$ with $64 \leq m \leq 84$ at a network utilization of 80 percent for four different interstage link failure rates. From Fig. 5, we observe that, for all network configurations, given a constant number of middle stage switches and a constant network utilization, as the probability of interstage link failure increases, the blocking probability increases. For smaller interstage link failure rates, say, $p_f \leq 0.03$, the blocking probability increases only slightly compared with that of $p_f = 0$. This indicates that the network has a good fault-tolerant capability in this case. However, the increase in blocking probability is more dramatical for larger values of p_f (> 0.03).

Furthermore, in Fig. 5, we show the effect of an increasing number of middle stage switches (m) on the blocking probability given by (13). For the four curves ($p_f = 0.0$, $p_f = 0.01$, $p_f = 0.03$, and $p_f = 0.07$) shown for each plot of Fig. 5, we see that as the number of middle stage switches increases, the blocking probability decreases sharply. Each plot also indicates that for two probabilities of interstage link failure, p_{f_1} and p_{f_2}, if $p_{f_1} > p_{f_2}$ then equivalent blocking probability occurs when the number of middle stage switches for the curve for p_{f_1} is greater than the number of middle stage switches for the curve for p_{f_2}. For all network configurations, given a constant interstage link failure probability and a constant network utilization, increasing the number of middle stage switches results in decreasing blocking probability.

Finally, in Table 1, we give more data for different network configurations and different values of p_f, which also shows the same trends as discussed for Fig. 5.

6 Experimental Simulations

An experimental study was performed to verify the approximations for the blocking probabilities given by the new analytical model (13). We developed a network simulator, which employed a random routing algorithm and allowed for interstage link failure, to determine the blocking probability of a Clos network given the same set of network conditions used to derive the analytical model.

6.1 Assumptions

The network for which simulation data was generated is a three-stage Clos network. Comparisons between analytical data and simulated data were based on similar network configurations. A *network configuration* is defined to be a unique set of five variables: r, the number of input stage and output stage switches in the network, n, the number of input/output ports on each input/output switch, m, the number of middle stage switches in the network, a, the network utilization, and p_f, the probability that an interstage link in the network is faulty.

6.2 The Network Simulator

The network simulator consists of two main parts: a request generator and a request processor. The request generator

$$\Pr\{\text{connection not blocked}\}$$

$$= \frac{\displaystyle\sum_{n_1=0}^{n-1}\sum_{f_1=0}^{m-n_1}\sum_{n_2=0}^{n-1}\sum_{f_2=0}^{m-n_2}\sum_{k=\max\{0,n_1+f_1+n_2+f_2-m+1\}}^{\min\{n_1+f_1,n_2+f_2\}} \frac{\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_2+f_2}}\binom{m}{f_1}\binom{m-f_1}{n_1}\binom{m}{f_2}\binom{m-f_2}{n_2}}{\left[\displaystyle\sum_{i=0}^{n-1}\sum_{j=0}^{m-i}\binom{m}{j}\binom{m-j}{i}p_f^j p_b^i q^{m-i-j}\right]^2} \cdot p_f^{f_1+f_2} p_b^{n_1+n_2} q^{2m-n_1-f_1-n_2-f_2} \tag{13}$$

Fig. 3. Equation (13).

$$\Pr\{\text{connection not blocked}\}$$

$$= \frac{\displaystyle\sum_{n_1=0}^{n-1}\sum_{n_2=0}^{n-1}\sum_{k=\max\{0,n_1+n_2-m+1\}}^{\min\{n_1,n_2\}} \frac{\binom{n_1}{k}\binom{m-n_1}{n_2-k}}{\binom{m}{n_2}}\binom{m}{0}\binom{m}{n_1}\binom{m}{0}\binom{m}{n_2}0^0 p_b^{n_1+n_2} q^{2m-n_1-n_2}}{\left[\displaystyle\sum_{i=0}^{n-1}\binom{m}{0}\binom{m}{i}0^0 p_b^i q^{m-i} + \sum_{i=0}^{n-1}\sum_{j=1}^{m-i}\binom{m}{j}\binom{m-j}{i}0^j p_b^i q^{m-i-j}\right]^2} +$$

$$\frac{\displaystyle\sum_{n_1=0}^{n-1}\sum_{f_1=1}^{m-n_1}\sum_{n_2=0}^{n-1}\sum_{f_2=1}^{m-n_2}\sum_{k=\max\{0,n_1+f_1+n_2+f_2-m+1\}}^{\min\{n_1+f_1,n_2+f_2\}} \frac{\binom{n_1+f_1}{k}\binom{m-n_1-f_1}{n_2+f_2-k}}{\binom{m}{n_2+f_2}}\binom{m}{f_1}\binom{m-f_1}{n_1}\binom{m}{f_2}\binom{m-f_2}{n_2}}{\left[\displaystyle\sum_{i=0}^{n-1}\binom{m}{0}\binom{m}{i}0^0 p_b^i q^{m-i} + \sum_{i=0}^{n-1}\sum_{j=1}^{m-i}\binom{m}{j}\binom{m-j}{i}0^j p_b^i q^{m-i-j}\right]^2} \cdot 0^{f_1+f_2} p_b^{n_1+n_2} q^{2m-n_1-f_1-n_2-f_2}$$

$$= \frac{\displaystyle\sum_{n_1=0}^{n-1}\sum_{n_2=0}^{n-1}\sum_{k=\max\{0,n_1+n_2-m+1\}}^{\min\{n_1,n_2\}}\binom{n_1}{k}\binom{m-n_1}{n_2-k}\binom{m}{n_1}p_b^{n_1+n_2}q^{2m-n_1-n_2}}{\left[\displaystyle\sum_{i=0}^{n-1}\binom{m}{i}p_b^i q^{m-i}\right]^2} \tag{15}$$

Fig. 4. Equation (15).

randomly generates a list of connection requests based on r, the number of input and output stage switches, and n, the number of input and output ports. A *connection request* is a four-tuple specifying the input port, the input-stage switch, the output-stage switch, and the output port. For routing purposes, we need only to be concerned with the input and output stage switches of the connection request. However, because we need to determine the legality of a connection request, the request generator must additionally generate the input and output ports of the connection request. The list of connection requests is generated before the actual simulation commences, as it is used as input to the request processor, which processes the list of connection requests. To strengthen comparisons among the simulation results, for a given n (and, likewise, r), the request processor utilized the same list of connection requests generated by the request generator.

Upon simulation startup, to account for interstage link failures, a certain portion of the interstage links are marked as faulty. This is accomplished for each interstage link by randomly generating a real number x such that $0 \leq x \leq 1$. If $x \leq p_f$, the interstage link is marked as faulty and cannot be used to route connection requests. These failures are considered *permanent*. To maintain a constant network utilization, as specified by a, the network simulator must release active connections from the network. This is accomplished by the request processor when processing a legal and satisfiable connection request. When the network utilization is at the prescribed utilization, after establishing a connection request in the network, the request processor randomly chooses an active connection for termination. By doing so, the network utilization is held constant throughout the rest of the simulation. Finally, the blocking probability for a network simulation is defined to be the number of blocked connection requests divided by the total number of legal connection requests generated.

6.3 Methodology and Network Configurations

We were primarily interested in analyzing the approximations for the blocking probability of the analytical model to

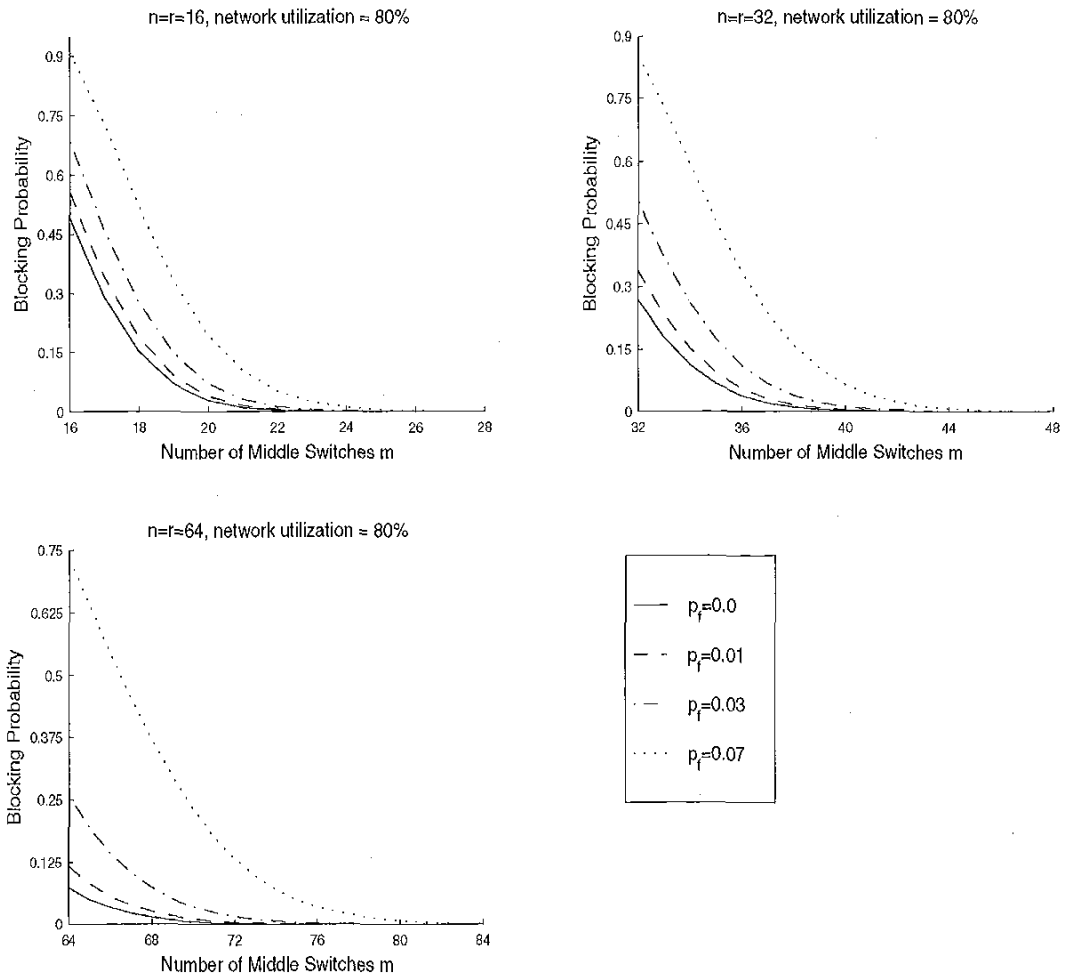

Fig. 5. Blocking probability of the Clos network for the analytical model for $n = r = 16$ with $16 \leq m \leq 28$, $n = r = 32$ with $32 \leq m \leq 48$ and $n = r = 64$ with $64 \leq m \leq 84$ at 80 percent network utilization for four different interstage failure rates.

see the effect of interstage link failure, using zero interstage link failures ($p_f = 0$) as a basis. We used data generated by simulation to verify that the analytical model is consistent with repeated trials.

We examined three network configurations: $n = r = 16$ with $16 \leq m \leq 45$, $n = r = 32$ with $32 \leq m \leq 80$, and $n = r = 64$ with $64 \leq m \leq 140$. For each network configuration, we let $50\% \leq a \leq 80\%$ in increments of 10 percent and we let $p_f = \{0, 0.001, 0.005, 0.01, 0.03, 0.07\}$. For the simulations, a represents the network utilization and p_f represents the probability that an interstage link is faulty. Therefore, blocking probabilities were obtained for the same percentages. Furthermore, the request processor processed $20,000$ legal connection requests from a list of $2,000,000$ connection requests generated by the request generator.

We chose values for m which would allow comparison to the Clos deterministic nonblocking condition, $m \geq 2n - 1$. Values for p_f were chosen on the basis of preliminary experimentation which suggested larger values of p_f yielded blocking probabilities indicative of a network with a minimal connecting capability. Of particular interest are the approximations for blocking probability given $0 \leq p_f \leq 0.01$, as these values represented failure rates which would seem to be the most practical.

6.4 Simulation Results

In this section, we present the data generated for the Clos network simulations utilizing a random routing algorithm. We were primarily interested in verifying that the two main hypotheses discussed in Section 5 were consistent with repeated trials.

Table 2 shows the blocking probabilities resulting from the experimental simulations for similar network configurations presented for the analytical model in Table 2. Also, the blocking probability curves in Fig. 6 show the relationship between the analytical model data and the data resulting from the network simulations. First, for all values of n and r, we see that the curves for the analytical model show the same trend as the curves for the network simulations, indicating that the analytical model is a good approximation of the actual blocking probability obtained by repeated trials. More specifically, those simulation results confirm our two main hypotheses discussed in Section 5 for the analytical model.

TABLE 1
Blocking Probabilities from Model Data for $a = 0.8$ and for $n = r = 16$, $n = r = 32$, and $n = r = 64$

$n = r = 16$					
p_f	$m = 16$	$m = 20$	$m = 24$	$m = 28$	$m = 31$
0.0	4.922×10^{-1}	2.818×10^{-2}	1.896×10^{-4}	6.055×10^{-8}	0.0
0.0001	4.928×10^{-1}	2.828×10^{-2}	1.911×10^{-4}	6.218×10^{-8}	2.939×10^{-13}
0.005	5.232×10^{-1}	3.339×10^{-2}	2.810×10^{-4}	1.884×10^{-7}	7.862×10^{-11}
0.01	5.547×10^{-1}	3.935×10^{-2}	4.051×10^{-4}	4.605×10^{-7}	5.237×10^{-10}
0.03	6.825×10^{-1}	7.199×10^{-2}	1.451×10^{-3}	6.194×10^{-6}	3.986×10^{-8}
0.07	9.093×10^{-1}	1.956×10^{-1}	1.050×10^{-2}	1.914×10^{-4}	5.298×10^{-6}
$n = r = 32$					
p_f	$m = 32$	$m = 40$	$m = 48$	$m = 56$	$m = 63$
0.0	2.697×10^{-1}	2.286×10^{-3}	4.479×10^{-7}	3.750×10^{-13}	0.0
0.0001	2.703×10^{-1}	2.300×10^{-3}	4.540×10^{-7}	3.946×10^{-13}	6.440×10^{-15}
0.005	3.032×10^{-1}	3.057×10^{-3}	8.454×10^{-7}	2.184×10^{-12}	3.830×10^{-14}
0.01	3.392×10^{-1}	4.051×10^{-3}	1.540×10^{-6}	9.049×10^{-12}	6.051×10^{-14}
0.03	5.033×10^{-1}	1.150×10^{-2}	1.288×10^{-5}	6.883×10^{-10}	1.227×10^{-13}
0.07	8.530×10^{-1}	6.500×10^{-2}	3.804×10^{-4}	2.600×10^{-7}	9.786×10^{-11}
$n = r = 64$					
p_f	$m = 64$	$m = 81$	$m = 98$	$m = 115$	$m = 127$
0.0	7.334×10^{-2}	5.570×10^{-6}	2.230×10^{-13}	1.110×10^{-16}	0.0
0.0001	7.369×10^{-2}	5.635×10^{-6}	2.588×10^{-13}	3.020×10^{-14}	3.086×10^{-14}
0.005	9.274×10^{-2}	9.868×10^{-6}	9.045×10^{-13}	1.585×10^{-13}	1.617×10^{-13}
0.01	1.160×10^{-1}	1.720×10^{-5}	2.694×10^{-12}	2.152×10^{-13}	2.424×10^{-13}
0.03	2.554×10^{-1}	1.360×10^{-4}	1.637×10^{-10}	4.369×10^{-13}	5.362×10^{-13}
0.07	7.316×10^{-1}	4.271×10^{-3}	1.298×10^{-7}	8.196×10^{-13}	8.562×10^{-13}

Also, Fig. 6 confirms that, for a constant network utilization and a constant number of middle stage switches, as the probability of link failure increases, the probability that a connection request cannot be satisfied increases. For small values of p_f, the increase in blocking probability is only within a narrow range. However, for larger values of p_f, the effect of interstage link failure probability is felt more dramatically. Furthermore, Fig. 6 demonstrates that as r (and likewise n) increases, the effect of interstage link failure is increasingly evident for the network simulations. For $n = r = 16$ and $n = r = 32$ the curves for the network simulations are almost identical for $p_f = 0.005$ and $p_f = 0.01$. However, for $n = r = 64$ (and somewhat for $n = r = 32$) the blocking probabilities obtained for the smaller of the two link failure probabilities ($p_f = 0.005$) are lower than the blocking probabilities shown for $p_f = 0.01$ and we begin to see the two simulation curves look comparatively close to the two curves shown for the analytical model. This further demonstrates the consistency of the analytical model with repeated trials.

From Fig. 6, we also observe that there is some gap between the analytical results and the simulation results. This is mainly because that in the analytical model, every input stage switch is assumed to have exactly $(1 - p_f)m$ functional interstage links; but in the simulation, it is impossible to enforce this assumption. In fact, the number of functional interstage links for each input stage switch may vary depending on where the faults are located. Thus, the simulation results which were obtained by averaging the blocking probabilities of all requests from different input stage switches are not exactly the same as those obtained by the analytical model. Since the interstage link busy probability is inversely proportional to the number of functional interstage links as shown in (8), the gap is more noticeable in the case of a larger p_f.

7 Conclusions

In this paper, we have presented a new analytical model for the routing blocking probability of the three-stage Clos network in the presence of interstage link faults. Because the Clos network has a powerful connecting capability and is inherently more fault-tolerant of interstage link failure than many proposed network designs, it continues to be an

TABLE 2
Blocking Probabilities from Network Simulations for $n = r = 16$, $n = r = 32$, and $n = r = 64$ with $a = 0.8$ and $p_f = \{0, 0.001, 0.005, 0.01, 0.03, 0.07\}$

$n = r = 16$				
p_f	$m = 16$	$m = 18$	$m = 20$	$m = 22$
0.0	3.266×10^{-1}	7.820×10^{-2}	9.900×10^{-3}	5.000×10^{-4}
0.0001	3.269×10^{-1}	8.200×10^{-2}	1.230×10^{-2}	8.000×10^{-4}
0.005	3.680×10^{-1}	9.610×10^{-2}	1.330×10^{-2}	7.000×10^{-4}
0.01	3.790×10^{-1}	8.440×10^{-2}	1.740×10^{-2}	1.500×10^{-3}
0.03	4.444×10^{-1}	1.408×10^{-1}	2.840×10^{-2}	4.400×10^{-3}
0.07	7.635×10^{-1}	2.295×10^{-1}	8.580×10^{-2}	2.480×10^{-2}
$n = r = 32$				
p_f	$m = 32$	$m = 36$	$m = 38$	$m = 40$
0.0	1.678×10^{-1}	1.470×10^{-2}	2.200×10^{-3}	6.000×10^{-4}
0.0001	1.687×10^{-1}	1.560×10^{-2}	2.600×10^{-3}	3.500×10^{-4}
0.005	1.810×10^{-1}	2.120×10^{-2}	4.100×10^{-3}	7.000×10^{-4}
0.01	2.182×10^{-1}	2.340×10^{-2}	5.000×10^{-3}	6.500×10^{-4}
0.03	2.334×10^{-1}	4.520×10^{-2}	7.600×10^{-3}	1.600×10^{-3}
0.07	4.807×10^{-1}	1.350×10^{-1}	5.110×10^{-2}	1.790×10^{-3}
$n = r = 64$				
p_f	$m = 64$	$m = 66$	$m = 68$	$m = 72$
0.0	4.050×10^{-2}	1.480×10^{-2}	5.550×10^{-3}	2.500×10^{-4}
0.0001	3.860×10^{-2}	1.650×10^{-2}	5.400×10^{-3}	6.000×10^{-4}
0.005	4.870×10^{-2}	1.910×10^{-2}	5.850×10^{-3}	7.000×10^{-4}
0.01	5.630×10^{-2}	2.200×10^{-2}	7.550×10^{-3}	1.000×10^{-3}
0.03	9.130×10^{-2}	5.350×10^{-2}	2.300×10^{-2}	3.350×10^{-3}
0.07	2.418×10^{-1}	1.420×10^{-1}	9.230×10^{-2}	2.120×10^{-2}

important part of the interconnection network landscape. By incorporating interstage link failure, we have seen that the proposed model allows for a more accurate and realistic measure of blocking probability. We have also simulated a Clos network where interstage links can fail randomly upon simulation startup and are permanent in nature. Results obtained from the simulations have confirmed that the analytical model is consistent with repeated trials, indicating that it is a reliable predictor of blocking probability for the network it is intended to model. Our analytical and simulation results demonstrate that for a smaller interstage link failure probability, say, $p_f \leq 0.03$, the blocking behavior of the Clos network is similar to that of a fault-free network and indicate that the Clos network has a good fault-tolerant capability. The new model presents a unified view of reliability analysis and traditional methods for the estimation of blocking probability. By doing so, network designers can measure the effect of interstage link failure on the overall connecting capability of the network. Future work may integrate additional network component failure probabilities to allow for a determination of the components which make the most sense to provide redundancy for, and consider other routing algorithms besides random routing. Another interesting issue is to generalize the model to collective communication.

Acknowledgments

Research was supported by the U.S. Army Research Office under Grant No. DAAH04-96-1-0234 and the U.S. National Science Foundation under Grant No. OSR-9350540.

References

[1] C. Clos, "A Study of Nonblocking Switching Networks," *The Bell System Technical J.*, vol. 32, pp. 406-424, 1953..

[2] A. Itoh et al., "Practical Implementation and Packaging Technologies for a Large-Scale ATM Switching System," *IEEE J. Selected Areas in Comm.*, vol. 9, no. 8, pp. 1,280-1,288, Oct. 1991.

[3] M.T. Bruggencate and S. Chalasani, "Equivalence between SP2 High-Performance Switches and Three-Stage Clos Networks," *Proc. 25th Int'l Conf. Parallel Processing*, pp. I-1–I-8, Bloomingdale, Ill., 1996.

[4] K. Padmanabhan and D.H. Lawrie,, "A Class of Redundant Path Multistage Interconnection Networks," *IEEE Trans. Computers*, vol. 32, no. 12, pp. 1,099-1,108, Dec. 1983.

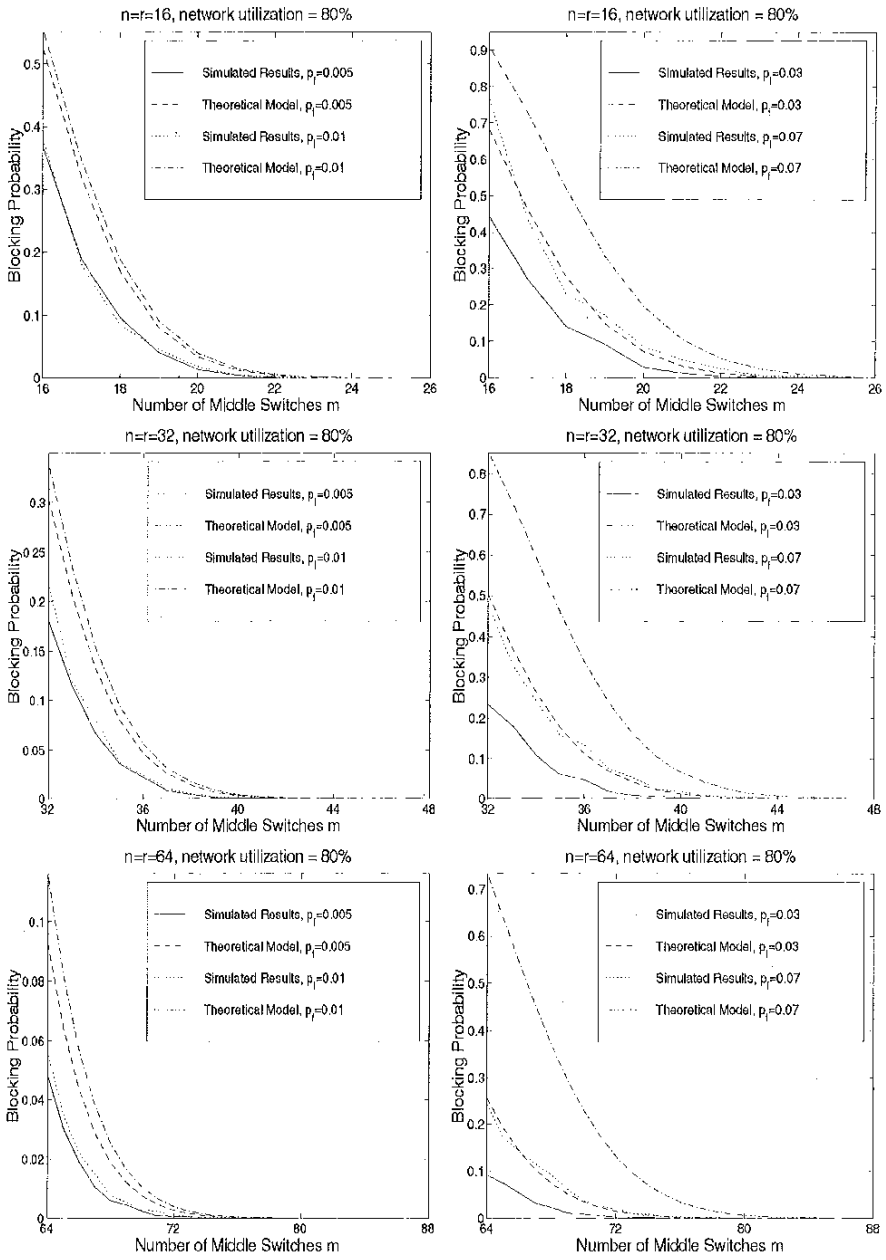

Fig. 6. Blocking probability of the Clos network comparing the analytical model results with the simulated results for $n = r = 16$ with $16 \leq m \leq 26$, $n = r = 32$ with $32 \leq m \leq 48$, and $n = r = 64$ with $64 \leq m \leq 88$ at 80 percent network utilization.

[5] R.J. McMillen and H.J. Siegel, "Performance and Fault Tolerance Improvements in the Inverse Augmented Data Manipulator Network," *Proc, Ninth Symp. Computer Architecture*, pp. 63-72, Apr. 1982.

[6] C.P. Kruskal and M. Snir, "The Performance of Multistage Interconnection Networks for Multiprocessors," *IEEE Trans. Computers*, vol. 32, no. 12, pp. 1,091-1,098, Dec. 1983.

[7] F.T. Leighton and B.M. Maggs, "Fast Algorithms for Routing around Faults in Multibutterflies and Randomly-Wired Splitter Networks," *IEEE Trans. Computers*, vol. 41, no. 5, pp. 578-587, May 1992.

[8] G.B. Adams, D.P. Agrawal, and H.J. Siegel, "A Survey and Comparison of Fault-Tolerant Multistage Interconnection Networks," *Computer*, vol. 20, no. 6, pp. 14-27, June 1987.

[9] S. Rai and D.P. Agrawal, *Distributed Computing Network Reliability*. Los Alamitos, Calif.: IEEE CS Press, 1990.

[10] A. Varma and C.S. Raghavendra, "Reliability Analysis of Multistage Interconnection Networks," *IEEE Trans. Reliability*, pp. 130-137, Apr. 1989.

[11] J.T. Blake and K.S. Trivedi, "Multistage Interconnection Network Reliability," *IEEE Trans. Computers*, vol. 38, no. 11, pp. 1,600-1,604, Nov. 1989.

[12] X. Cheng and O.C. Ibe, "Reliability of a Class of Multistage Interconnection Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 3, no. 2, pp. 241-246, Mar. 1992.

[13] C. Jacobaeus, "A Study on Congestion in Link Systems," *Ericsson Technics*, vol. 51, no. 3, 1950.

[14] C.Y. Lee, "Analysis of Switching Networks," *The Bell System Technical J.*, vol. 34, no. 6, pp. 1,287-1,315, Nov. 1955.

[15] M. Karnaugh, "Loss of Point-to-Point Traffic in Three-Stage Circuit Switches," *IBM J. Research and Development*, vol. 18, pp. 204-216, 1974.
[16] P.M. Lin, B.J. Leon, and C.R. Stewart, "Analysis of Circuit-Switched Networks Employing Originating Office Control with Spill Forward," *IEEE Trans. Computers*, vol. 26, no. 6, pp. 754-765, June 1978.
[17] Y. Mun, H.Y. Youn, "Performance Modeling and Evaluation of Circuit Switching Using Clos Networks," *IEEE Trans. Computers*, vol. 43, pp. 854-861, 1994.
[18] Y. Yang, "An Analytical Model on Network Blocking Probability," *IEEE Comm. Letters*, vol. 1, no. 5, pp. 143-145, Sept. 1997.
[19] Y. Yang and J. Wang, "On Blocking Probability of Multicast Networks," *IEEE Trans. Comm.*, vol. 46, no. 7, pp. 957-968, July 1998.
[20] V.E. Benes, *Math. Theory of Connecting Networks and Telephone Traffic*. New York: Academic Press, 1965.
[21] Y. Yang and J. Wang, "Wide-Sense Nonblocking Clos Networks under Packing Strategy," *IEEE Trans. Computers*, vol. 48, no. 3, pp. 265-284, Mar. 1999.
[22] F.K. Hwang and A. Jajszczyk, "On Nonblocking Multiconnection Networks," *IEEE Trans. Comm.*, vol. 34, pp. 1,038-1,041, 1986.
[23] Y. Yang and G.M. Masson, "Nonblocking Broadcast Switching Networks," *IEEE Trans. Computers*, vol. 40, no. 9, pp. 1,005-1,015, 1991.
[24] A. Varma and S. Chalasani, "Asymmetrical Multiconnection Three-Stage Clos Networks," *Networks*, vol. 23, pp. 427-439, John Wiley & Sons, 1993.
[25] Y. Yang, "A Class of Interconnection Networks for Multicasting," *IEEE Trans. Computers*, vol. 47, no. 8, pp. 899-906, Aug. 1998.

Matthew P. Haynos received the BA degree in computer science/applied mathematics and cognitive science (honors) from the University of Rochester in 1990, and the MS degree in computer science from the University of Vermont in 1998. He has been employed by the IBM Corporation since June 1990. His research interests include interconnection networks, fault-tolerant computing, algorithm design, and distributed computing.

Yuanyuan Yang received the BEng and MS degrees in computer engineering from Tsinghua University, Beijing, China, in 1982 and 1984, respectively, and the MSE and PhD degrees in computer science from Johns Hopkins University, Baltimore, Maryland, in 1989 and 1992, respectively. She is currently an associate professor of computer engineering at the State University of New York at Stony Brook. Before joining SUNY Stony Brook, Dr. Yang was a faculty member in the Department of Computer Science, University of Vermont, in Burlington, from 1992-1999 (as an associate professor from 1998-1999). Dr. Yang's research interests include parallel and distributed computing and systems, high speed networks, optical networks, high performance computer architecture, and fault-tolerant computing. She has published extensively in major journals and refereed conference proceedings related to these research areas. Dr. Yang holds two U.S. patents in the area of multicast communication networks, with four more patents pending. Her research has been supported by the U.S. Army Research Office and the U.S. National Science Foundation. She has served on the program/organizing committees of a number of international conferences. Dr. Yang is a senior member of the IEEE and a member of the ACM, IEEE Computer Society, and IEEE Communication Society.