12-1996

Performance Analysis of Preemption Algorithms in an IDNX Circuit Switch Communications Network

Eric C. Gumbs

Recommended Citation

Gumbs, Eric C., "Performance Analysis of Preemption Algorithms in an IDNX Circuit Switch Communications Network" (1996). *Theses and Dissertations*. 5921.
https://scholar.afit.edu/etd/5921

PERFORMANCE ANALYSIS OF PREEMPTION ALGORITHMS IN AN IDNX CIRCUIT SWITCH COMMUNICATIONS NETWORK

THESIS

Eric C. Gumbs, Captain, USAF

AFIT/GE/ENG/96D-05

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GE/ENG/96D-05

PERFORMANCE ANALYSIS OF PREEMPTION ALGORITHMS IN AN IDNX CIRCUIT SWITCH COMMUNICATIONS NETWORK

THESIS

Eric C. Gumbs, Captain, USAF

AFIT/GE/ENG/96D-05

AFIT/GE/ENG/96D-05

PERFORMANCE ANALYSIS OF PREEMPTION ALGORITHMS IN AN IDNX CIRCUIT SWITCH COMMUNICATIONS NETWORK

THESIS

Presented to the faculty of the Graduate School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science (Electrical Engineering)
Eric C. Gumbs, B.S.E.E.
Captain, USAF
December, 1996

Approved for public release, distribution unlimited.

Acknowledgment

I wish to thank the following people who gave of their time and effort to help with this research and thesis preparation: Thesis advisor, Dr. Richard Raines for his valuable input and technical guidance, and Dr. Yupo Chan and Lt. Colonel Gallagher for proofreading this thesis and offering suggestions; and the Department of Defense (DoD) for sponsoring this research. Special thanks to my wife, Mrs. Sharon Gumbs who efficiently kept me on schedule and cheerfully provided excellent typing and moral support.

Table of Contents

List of Figures

List of Tables

Abstract

Access to communication networks is increasing rapidly. The increased access to these networks results in delays and at times loss of data. At times of peak traffic or when trunks or nodes are down, very important customers' communications requirements are not met. One way to combat this problem is to prioritize the network and provide different levels of grade of service (GOS) for each priority. Call preemption provides an effective method of obtaining different levels of GOS. Call preemption terminates lower priority calls to make room for the new high-priority calls. This research seeks to design the best circuit switch communications network preemption model for the DoD by analyzing previously developed preemption algorithms. Four simulation network models are developed. The grades of service per priority are obtained as the network capacity decreases and as the calls generated in node 0 increases. The analysis of preemption network models is based on the grade of service, average number of preemptions, and average network bandwidth. The networks are simulated under the same input parameters. The analysis showed that preemption can significantly lower the grade of service for high priority customers in a congested network. The best configuration preemption model depends on the bandwidth flexibility of the network and the goals of the communications network organization.

Performance Analysis of Preemption Algorithms in an IDNX Circuit Switch Communications Network

1 Introduction

This chapter describes the effort necessary to complete the proposed thesis topic. The chapter begins by presenting some background information and problems that leads to the research effort. Next, current research literature pertinent to the topic is summarized. A list of the assumptions then follows. After the assumptions, the scope of the research is defined. The approach and methodology identifying how the research effort will progress is then given. Finally, a summary of the chapter is presented.

1.1 Background

The rapid advance of technology has enabled more and more users to access communication networks. The increasing access to these networks affects communication systems performance by increasing the average delay encountered by each user, and at times the loss of important data. During peak traffic hours and when some links are down, delays are more prominent. In order to provide better service to their more important customers, some organizations have implemented a system to prioritize access and usage of their networks.

A prioritized network offers different grades of service for its classes of customers. The highest priority customer receives the best grade of service, while the lowest priority customer receives the worst. Grade of service is the percentage of connections, usually measured during busy hours, a network cannot complete, or serve within a certain time period. Networks with priority scheduling can preempt lower priority calls before or while calls are in service. Preemption of lower priority calls also helps to increase the grade of service of high priority calls. The flip side to preemption is the possible loss of data.

The Department of Defense (DoD) has a large communication network. This network has many switching nodes interconnected by various communication links. The links consist of satellites, fiber optics, microwaves and twisted-pair copper wire. This global network enables the DoD to provide timely world-wide transmission of video, facsimile, voice, images and data to its employees and customers.

Over the years, the DoD network has undergone many upgrades designed to satisfy increased user demands. One of these upgrades is the installation of Network Equipment Technologies (N.E.T.) Integrated Digital Network Exchange (IDNX) transmission resource managers. The IDNX is a multiplexer that prioritizes and preempts telecommunications transmissions. Additionally, it routes, switches and performs other network management functions for a wide variety of voice, data, and video applications. Each IDNX has the capability of classifying traffic into 16 different priority levels. Three other features of the IDNX are that it provides bandwidth only on demand, maximizes bandwidth utilization, and automatically optimizes network routing [NeT91].

The DoD needs to know whether they are presently providing the best possible grade of service to their most important customers. They believe that the IDNX's preemption option could be exercised to increase the utilization of the higher priority calls. They also need to know how the use of preemption affects the grade of service of the different priority levels. In addition, it is important to know what preemption algorithm performs the best for the same number of calls generated.

1.2 Problem

The increasing dependence of customers on the DoD's communication networks makes it important to provide not only reliable services, but reasonably good performance under less than optimal conditions. Conceptually, the use of priority and preemption with the IDNX can improve the grade of service to the DoD's most important customers. Additionally, there must be a priority scheme and preemption algorithm that optimizes the efficiency of the network.

A simulation model of an IDNX transmission resource manager using a priority and preemption scheme has yet to be realized. Several factors must be taken into account in developing this model. It is important to use the appropriate call control protocol to efficiently set up and terminate connections. The loading of the network would also affect its efficiency. A lightly loaded network would require less preemption than a heavily loaded network. Finally, the topology of the network will also affect the performance. An originating node close to its destination node would compete more favorably for connection than a node farther away.

This research develops 4 simulation models that analyze the performance of the IDNX transmission resource manager using preemption algorithms. The test-bed for this thesis is a circuit switch communication network comprising 20 nodes and 35 links.

1.3 Summary of Current Knowledge

The transmission resource manager needs to know how to manage calls in a priority and preemption environment. Specific instructions must be given for it to perform its function effectively. For this reason, each node must be equipped with a preemption algorithm. This algorithm is a finite number of well-defined instructions.

Juan A. Garay and Inder S. Gopal of IBM T. J. Watson Research Center developed algorithms for call preemption in communication networks [GaG92]. Garay and Gopal's algorithms preempt a specific number of calls, a specific bandwidth of calls, and a total network bandwidth of calls. The network bandwidth is the number of calls multiplied by the number of hops in the route of the call. Their first algorithm identifies a set of calls and performs a routine that removes these calls from the set of existing calls. Their second algorithm computes the required bandwidth needed to accommodate a higher priority incoming call and removes lower priority calls from the set of existing calls until enough bandwidth is available. Finally, their third algorithm takes into account the complete route of the calls in its decision of what calls to preempt.

Two years after the work done by Garay and Gopal, Mohammad Peyravian attempted to improve on the second algorithm [Pey94]. Mohammad stated that Garay and Gopal's algorithm is suitable for centralized networks but not for decentralized networks. He went on to propose a separate algorithm suitable for both types of networks. Mohammad's algorithm searches for a low priority call to preempt that would satisfy the incoming call's required bandwidth. This algorithm minimizes the number of calls preempted and preempts only the necessary amount of bandwidth.

1.4 Assumptions

The simulation models are accomplished using BONeS DESIGNER. In order to design and simulate the models, the following assumptions are made:

1. The arrivals of calls will be modeled as an exponential distribution.
2. The length of each calls will be exponentially distributed.
3. Each node performs it preemption algorithm independently.
4. The choice of the call control protocol does not significantly affect the utilization of the network.

1.5 Scope

This study seeks to develop four simulation models of a communication network implementing priority and preemption. The network will be composed of 20 nodes and 35 links. Preemption algorithms will be simulated on four circuit switch communication network models. The analysis of simulation results will determine average number of preempted calls, average preempted bandwidth, and average preempted network bandwidth. The simulations also calculate the grade of service per priority level. From the results of the simulations, the best system model is determined given the network operating environment.

1.6 Approach/Methodology

The literature review begins the research effort. A comprehensive literature search of priority/preemption algorithms provides the foundation for this task. Call control protocols relating to the flow and control of traffic across a communication network are explored. An analysis of simulation models and construction of new models on BONeS DESIGNER provides the method to test the performance of the priority/preemption algorithms.

The first step in the methodology is to construct a simulation model of a circuit switch communication network with 20 IDNX nodes and 35 links. Each node utilizes the same priority/preemption algorithm. Each node also has a call control protocol for call routing determination. Each link in the network has a maximum bit rate. Figure 1.1 is a typical network topology comprising ten communication nodes and 14 links. The arrows represent the flow of traffic.

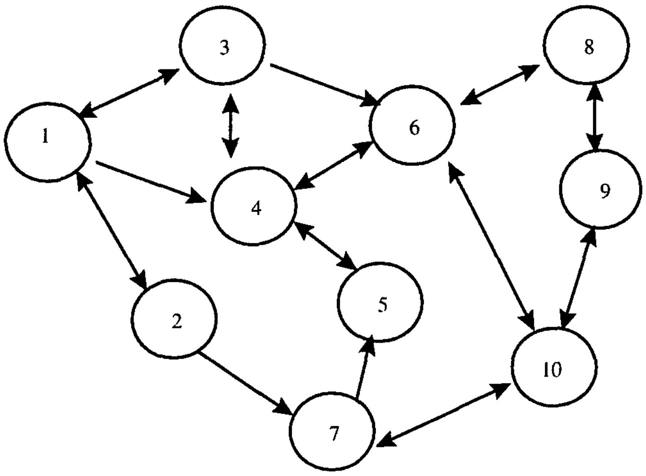

Figure 1.1 Ten-node communication network

Once the communication network model is constructed, pilot simulation trials are performed to verify the model and establish at least a 95% confidence interval. Complete simulations are executed after the model has been verified. Performance analyses are conducted to determine the optimal simulation model configuration that maximizes the grade of service of the high-priority connections.

1.7 Summary

The DoD has experienced a rapid growth in their global communications network. Service to high priority customers needs to get better and the efficiency of the network needs to improve. Preemption of lower priority customers provides a promising way to accomplish this task. The scope of the research focuses on finding the preemption algorithm that provides the best grade of service for the higher priority customers and the biggest increase in network efficiency. The simulation utilizes a circuit switch environment and be accomplished on BONeS DESIGNER network simulator.

2 Literature Review

2.1 Introduction

This chapter reviews reference material pertinent to the performance studies of preemption algorithms over a communication network. Section 2.2 describes preemption and it's importance in a communication network. It also discusses important aspects of a communication network, circuit switched network, and the integrated digital network exchange transmission resource manager. There is also an explanation of how a call is set up using a call control protocol and a short background of performance analysis using simulation models. Finally, Section 2.3 discusses preemption in communications network with emphasis in circuit switched networks using algorithms. This section also reviews the call preemption algorithms in the current literature.

2.2 Preemption

The increasing dependence of customers on communication networks makes it important for network providers to supply reliable services and good performance under less than optimal conditions. Failure of some links and during periods of traffic congestion can reduce the efficiency of a communication network. When this happens, a network may need to guarantee effective and reliable service to its more important customers. An easy way to accomplish this task is to prioritize the network assigning a certain priority to each network user.

Preemption allows prioritized networks to provide a better grade of service to their high priority users. These networks preempt lower priority calls when insufficient resources are available for incoming higher priority calls. Before a lower priority call can be preempted, it must be selected for preemption by an algorithm. An algorithm is a finite number of well-defined steps for the solution of a problem. Once a call is preempted, it may be rerouted if it can preempt a call of lower priority or it may be dropped [GaG92].

2.2.1 Communication Network

A communication network is a series of communication devices, mechanisms, and procedures by which end-user devices attached to the network can exchange meaningful information [SaA94]. To model a communication network, it is necessary to represent all of the network components as links and nodes. Users and switching centers are represented as nodes in the network while the transmission media connecting these nodes are the network links. The source node refers to the node sending the information while the destination node is the node receiving that information. The source and the destination nodes are known as the source-destination pair. Transitory nodes are the nodes in between the source-destination pair. Three parameters describe each network link. The first parameter is its capacity, which is the maximum flow of traffic across the link per unit time. Propagation time is the second parameter and it is the length of time traffic takes to travel across the link. Finally, link reliability is the third parameter. This parameter describes the likelihood that the transmitted information reaches its destination. A path is the name given to the collection of links that forms between a source node and a destination node.

2.2.2 Circuit Switched Network

A circuit switched network is a communication network where a dedicated path between the calling and called stations is established on demand for exclusion use of the circuit until the connection is released. A special signaling message travel across the network and capture channels in the path as it proceeds to set up the circuit [SaA94]. A return signal to the source indicates that transmission of data may proceed. Once the connection is established, the bandwidth allocated is dedicated to that call until the connection is terminated. Circuit switching is not widely used in data networks. However, these networks are used extensively for voice and video traffic.

2.2.3 IDNX

A transmission resource manager that prioritizes and preempts calls is Network Equipment Technology (N.E.T.) Integrated Digital Network Exchange (IDNX). The IDNX performs multiplexing, routing, switching, and network management functions for a wide variety of voice, data, and video transmissions' applications. Three additional features of the IDNX multiplexer are that it provides bandwidth only on demand, maximizes bandwidth utilization, and it automatically optimizes network routing [NeT91].

2.2.4 Call Control Protocol

Before calls can be established, nodes in a network must have a way of talking to each other. Each node must establish a protocol that governs how calls are established, maintained, and disassembled. A typical call control protocol works this way as described by Mohammad Peyravian [Pey94]:

> A call is set up as follows. The origin computes a complete route from the origin to the destination based on its knowledge about the current states and utilization of the links. Then, the origin constructs a call setup request for the call and sends it to all the link managers along the computed route. A link manager along the route accepts the call and returns a positive reply only if it can provide the resources to accommodate the call. Otherwise, it rejects the call and returns a negative reply to the origin. If a link manager accepts a call, it allocates the requested resources for the call. When the origin receives the replies, it determines whether a call setup is successful. The call setup is successful only if all the replies are positive. If the call setup is unsuccessful, the origin computes a new route (which excludes the links that replied unfavorably) and repeats the setup process. When the call setup is unsuccessful, the origin also sends a path takedown request to the link managers along the path of the call that replied favorably. When a link manager revives a path takedown request for a call, it releases the network resources associated with that call.

The effectiveness of the call control protocol greatly affects the effectiveness of network. Calls must be able to get to their destination quickly and efficiently.

2.2.5 Simulation Model

Performance analysis on operational computer systems can be very expensive. For this reason, studies are performed using simulation models. A simulation model provides a relatively low-cost method for predicting the performance of an actual system. A simulation model may be preferred over measurements because it allows the alternatives to be compared under a wider variety of workloads and environments [Jai91]. Even though a simulation model is a good tool for performance studies, care must be taken to ensure the results make sense and the simulation is actively modeling as intended. Each section of the simulation model should be verified individually and different simulations performed with slightly different values to check the consistency of the results. The results should be validated by comparing them with theoretical values, and real-systems measurements.

Selecting a proper language is probably the most important step in the process of developing a simulation model [Jai91]. A simulation language, a general-purpose language, an extension of a general-purpose language, and a simulation package are the choices for the proper language. Simulation languages such as SIMULA and SIMSCRIPT allow the analyst to spend more time on issues specific to the system being modeled and less time on issues that are general to all simulations. General-purpose languages such as Pascal and FORTRAN are usually chosen when the analyst is very familiar with the language. An extension of a general-purpose language such as GASP consists of a collection of routines to handle tasks that are commonly required in simulations. Simulation packages, such as BONeS DESIGNER, have libraries of data structures, algorithms and routines. These packages save the user time when developing simulation models, and thus produce result quickly. On the other hand, a simulation package is not very flexible. Some simulation packages may not have the ability to perform some types of application.

2.3 Preemption in Communications Network

Preemption can be implemented on message, packet, and circuit switching networks. Message switching networks place incoming messages in a queue until a server is available. Preemption of a message occurs when an incoming higher priority message causes the removal of a lower priority message from the server [Tak90]. Preemption in a packet switched network operates similarly to that of the message switched network. Analysis of preemption in a packet switched network has been widely studied [MaT87] [MoL87] [ChG91] [YoC93]. Preemption in a circuit switched network can take place before or after set up of the circuit. Preemption after circuit set up can result in loss of valuable data. In spite of the possible loss of data, preemption after completion of circuit set up will be the focus of this research. In a circuit switched network, preemption can occur in three ways. Preemption can be accomplished based on the total number of calls to be preempted, the total bandwidth of the calls to be preempted, or the total network bandwidth of the calls to be preempted. The network bandwidth is the calls times the number of hops in the route of the call.

2.3.1 Preemption Algorithms

Juan A. Garay and Inder S. Gopal of IBM T. J. Watson Research Center develop an algorithm based on the number of calls to be preempted. This is a simple algorithm that identifies a set of calls to be preempted (P) and performs a routine that removes these calls from the set of existing calls (C). Figure 2.1 shows the first algorithm that Garay and Gopal developed [GaG92]. This algorithm assumes the bandwidth of all the calls is equal.

```
Algorithm H1(new call c);
        {let $r = \{e_1,e_2,\ldots e_n\}$ be the portion of c's route
        without capacity to accommodate c}
        {let $C = \{c_1,c_2,\ldots c_p\}$ be the set of existing calls
        on r; $c_j$'s route is denoted by $r_j$}
        $\pi := r$;
        $P := \phi$; { the set of calls on r to be preempted
        to accommodate c}
        while $\pi \neq \phi$ do
{*}         let $c_i \in C$ be call s.t. $|\pi \cap r_i|$
            $\pi := \pi - (\pi \cap r_i)$;
            $C := C - \{c_i\}$;
            $P := P \cup \{c_i\}$;

        endwhile;

        return (P)
```

Figure 2.1: An algorithm to minimize the number of preempted calls

In Figure 2.1, e_i is the ith link in the route of a call, c_i indicate the ith call arrival, b_i is the bandwidth, and r_i is the route. This algorithm looks for lower priority calls that have route intersection with the incoming higher priority calls. When found, the algorithm deletes the call, and updates C and P. The computer run time for this algorithm is of the order $n^2 \log n$, where n is the number of links in the arriving call's route. The performance of this algorithm is expected to be within a logarithmic factor of optimal. The number of calls to be preempted is of the order $\log n$. Equal bandwidth will restrict the flexibility of the algorithm. This would probably not be used in systems where the traffic is bursty in nature since the bandwidth varies.

Garay and Gopal also developed an algorithm that would preempt calls based on total bandwidth. Their algorithm computes the required bandwidth needed to accommodate an incoming priority call and removes lower priority calls until enough bandwidth is available. Figure 2.2 shows this algorithm [GaG92]. Here w_j is the bandwidth capacity wanted by the arriving higher priority call and p_j is the amount of capacity available on the link. This algorithm computes the capacity wanted by the arriving call and searches for the call with the next higher bandwidth. When found, the algorithm deletes the call, and update C, P, p_j, and, w_j. This algorithm is also designed to achieve a performance within a logarithmic factor of optimal. The amount of bandwidth to be preempted is of the order (β log n), where n is the length of the new call's route and β is the maximal bandwidth of an existing call.

```
Algorithm H3(new call c);
        {let $r = \{e_1, e_2, \ldots e_n\}$ be the portion of c's route without
        enough capacity to accommodate c; $p_j$ is $e_j$'s residual capacity}
        { let $C = \{c_1, c_2, \ldots c_p\}$ be the set of existing calls on r;
        $c_i = < b_i, r_i >$}
        for $e_j \in r$ do
            $w_j := \lceil b - p_j \rceil^+$;
        P := $\phi$; { the set of calls on r to be preempted
        to accommodate c}
        repeat
    {*}     let $c_i \in C$ be call s.t. $\Sigma_{ej \in r} \min(w_k, {}_{bi})$ is maximal;
            $C := C - \{c_i\}$;
            $P := P \cup \{c_i\}$;
            for $e_k \in r_i$ do
                    $p_k := p_k + b_i$;
                    $w_k := \lceil b - p_k \rceil^+$;
            until $p_j \geq b$, $1 \leq j \leq n$;
            return (P)
```

Figure 2.2: An algorithm to preempt minimizing total bandwidth

Mohammad Peyravian decided to improve on this algorithm. He claimed that Garay and Gopal's algorithm is suitable for centralized networks because it considers the complete route of the preempting call, and it shares one or more links with the preempting call [Pey94]. He also claimed that such a scheme cannot fit well into a decentralized network because it considers the complete route of the preempting call. It also considers the calls that share one or more links with the preempting call. As a result, he proposed a separate algorithm that takes care of that problem. Mohammad Peyravian's algorithm picks the lowest priority calls and at the same time minimizes preemption of the number of calls and excess bandwidth. The upper bound on the computational complexity is of the order nk^2 where k is the number of existing calls on a link and n is the number of links the incoming calls occupy. Simulation results comparing Peyravian algorithm to Garay and Gopal algorithm show an improvement in overall performance. Figure 2.3 illustrates his algorithm [Pey94].

Peyravian's algorithm is an improvement over Garay and Gopal's algorithm because it minimizes the amount of preemption. His algorithm searches for a low priority call that would satisfy the incoming calls required bandwidth and preempt it. If one call does not have the necessary bandwidth requirement then the lower priority call with the largest bandwidth would be preempted first. The algorithm does not need to know the complete route of the incoming call before it can preempt an ongoing call as with Garay and Gopal's algorithm.

Peyravian's Algorithm

```
1. $P := \phi$
2. $W_j := B_p - a_j$;   $a_j$ is the available capacity
3. Search for a call $C_i \in C$, where $1 \le i \le k$, with the smallest bandwidth such that $B_i \ge W_j$ and there does not exist a call $C_r \in C$, where $1 \le r \le k$, and $r \ne i$, with $B_r \ge W_j$ and $P_r < P_r$. If such a call (i.e. $C_i$) is found, go to Step 5.
4. Find a call $C_i \in C$, where $1 \le i \le k$, with the largest bandwidth such that there does not exist any other call $C_r \in C$, where $1 \le r \le k$, and $r \ne i$, with $P_r < P_i$.
5. $C := C - \{C_i\}$.
6. $P := P \cup \{C_i\}$.
7. $a_j := a_j + B_i$
8. If $B_p > a_j$, go to step 2.
```

Figure 2.3: An algorithm to preempt minimizing number of calls and total bandwidth

Preemption algorithms based on the total network bandwidth require knowledge of the complete route of the calls. Garay and Gopal also developed an algorithm that takes the total network bandwidth into account. The algorithm is very similar to the one shown in Figure 2.2. In this algorithm, the route of a call helps to determine the set of calls to be preempted. A call with a higher number of hops would be less likely to be preempted than one with a lower number of hops. This method gives long distance calls a lower probability of being preempted than a local call of the same priority.

The algorithms presented by the references cited attempt to provide effective call preemption. These algorithms enable a communication network to support its more valuable customers by increasing the probability of their calls being completed. Garay and Gopal algorithms are based on the total number of calls to be preempted, the total bandwidth of the calls to be preempted, or the total network bandwidth of the calls to be preempted. A communication network can use any of these algorithms. The set of calls to be preempted is based on propositions. No simulations were accomplished to validate these propositions. The algorithm developed by Mohammad Peyravian is an improvement over Garay and Gopal's algorithm that is based on the total bandwidth of the calls to be preempted. His algorithm preempts the lowest priority calls first which eliminate the need for preempted calls to preempt others. This method not only reduces the number of preemption but it also reduces the number of computations. Peyravian algorithm would be effective in any circuit switch network.

2.4 Summary

In the first section, preemption was discussed. The use of this scheme enables a communication network to provide better service to its more important customers. Next, there were discussions on communication network, circuit switched network, and Integrated Digital Network Exchange. It was seen that the communication network is composed of nodes and links. Within each node is a link manager that multiplexes, routes, and switches traffic from one node to another. A call control protocol was reviewed. This procedure used by telecommunication nodes to set up, monitor and terminate calls. A short discussion of simulation was also presented. Finally, the current literature on preemption algorithms was reviewed. The algorithms were based on preempting a number of calls or excess bandwidth in order to accommodate an incoming higher-priority call. The algorithms were developed by Juan A. Garay and Inder S. Gopal, however Mohammed Peyravian modified one of the algorithms.

3 Methodology

3.1 Introduction

This chapter presents a methodology that can be used to analyze the performance of preemption algorithms in an IDNX communication network. First, a brief description of BONeS Designer Network Simulator and its application to computer networks modeling are provided in Section 3.2. Section 3.3 defines the performance metrics. As determined from the literature review, grade of service, and the average number of preemptions are excellent measures to evaluate preemption algorithms. Section 3.4 discusses the development of the preemption model. The network topology and its IDNX nodes were first introduced in this section. Next, there is a discussion of the relevant assumptions, the construction of the network models and the system parameters. Section 3.5 discusses the statistical precision of the simulations. Section 3.6 discusses the steps taken to verify the network model construction and the output results. Finally, Section 3.7 highlights validation techniques.

3.2 BONeS DESIGNER and Computer Network Modeling

In past years, simulations were implemented exclusively using a general-purpose language such as FORTRAN or Pascal. Simulations with a general-purpose language require in-depth knowledge of that language and usually require thousands of lines of source codes. Recently, Alta Group of Cadence Design Systems, Inc developed BONeS DESIGNER to enable the analyst to develop simulation models with greater ease. DESIGNER's graphical environment and collection of model library blocks make it ideally suited for network simulations. The hierarchical structure of DESIGNER makes it simple and easy to learn. For this reason, BONeS DESIGNER was chosen as the tool for use in modeling the preemption algorithms in this IDNX circuit switch network. A brief description of BONeS DESIGNER block oriented network simulator is provided to assist in the understanding of the models presented later in this section.

DESIGNER is an integrated software package for modeling and simulating event-driven data transfer systems such as communications networks, computer architecture and distributed processing systems [AIR94]. DESIGNER is comprised of a core library with over 300 modules and primitives that are written in C/C++. DESIGNER consists of a comprehensive core library and five modules. The library contains models of traffic sources, channels, timers, delays, queues, and arithmetic and logical operators. The modules are Data Structure Editor (DSE), Block Diagram Editor (BDE), Symbol Editor Simulation Manager and Post Processor (PP). The DSE specifies, edits, and stores data structure for the model. The BDE creates, edits, documents, and stores graphical models for the system created. The Symbol Editor is used to create custom symbols or to modify existing symbols from the symbol library. The Simulation Manager is used to input the execution parameters, place data collection probes, automatically generate discrete-event simulations, control simulation execution, and store the simulation results. The Post Processor assists in the analysis of the simulation results, computes statistical and performance measures, and displays the results in graphical plots.

The DESIGNER model of a communication network is specified by a hierarchical block diagram, which represents the flow of packets, messages or transactions. A simple communication network consists of three blocks. These are a traffic source, a transmission system, and a destination node. Each hierarchical block is composed of several component blocks. The traffic source provides the entity that is used to model a call request. Each call request has a set of data structure such as its source, destination, priority and bandwidth. The transmission system block represents the method by which the call request gets from the source node to the destination node. At the destination node, the call request successfully connects and its data are transmitted at the appropriate time.

3.3 Performance Metrics

Previous research suggests that the performance of preemption algorithms can be determined by the number of preemptions and the completion probability [Pey94]. This effort compares the average number of preemptions, grade of service, and average network bandwidth preempted. The grade of service is one minus the completion probability of a call. There is a discussion below on each of the performance metrics.

3.3.1 Number of Preemptions

The number of preemptions is the average number of calls that a call of priority *i* preempts in order to establish a link along its path. The number of preemptions is calculated as follows:

$$\frac{Pi\Pr eemptedCalls}{\Pr eemptingCalls} \qquad i = 0, 1, 2, 3 \qquad (1)$$

where "PiPreemptedCalls" is the number of calls preempted per priority and PreemptingCalls is the number of calls per priority that preempted at least one lower priority call. This value will be greater than or equal to one. This is because each preempting call preempts at least one call.

3.3.2 Grade of Service

The grade of service is the most important performance measure and clearly indicates the importance of call preemption. The grade of service is the probability that a call of some priority will not be successfully completed. The level of service that a communication network provides to it customers is specified by the grade of service. The grade of service is calculated as follows:

$$\frac{PiBlockedCalls + Pi\Pr eemptedCalls}{TotalCalls} \qquad i = 0, 1, 2, 3 \qquad (2)$$

where "PiBlockedCalls" is the number of calls blocked per priority and "TotalCalls" is the total number of call generated per priority. This value will always be less than or equal to one. The lower the grade of service, the better the performance for that priority class.

3.3.3 Network Bandwidth Preempted

The network bandwidth preempted is the average network bandwidth that a call of priority *i* preempts in order to establish a link along its path. Average network bandwidth preempted is calculated as follows:

$$\frac{PiNetworkBandwidth\Pr eempted}{\Pr eemptingCalls} \qquad i = 0, 1, 2, 3 \qquad (3)$$

where "PiNetworkBandwidthPreempted" is the total network bandwidth preempted per priority and

"PreemptingCall" is the total number of calls per priority that preempts at least one lower priority call. Network bandwidth is the calls times the number of hops in the route of the call.

3.4 Preemption Model

The circuit switch preemption network is constructed and modeled in this section. The topology of the network is a 20-node circuit switch network. The nodes are all IDNX transmissions' resource managers. Table 3.1 gives the network nodal configurations. It shows the type and capacity of each IDNX node in the network. The IDNX-20 and IDNX-90 nodes support their maximum bandwidth. The IDNX-70 can support from 1-8 slots of 32Mbps each. The choice of these IDNX nodes gives a good representation of a typical interconnected IDNX network. Similar IDNX networks have been devised but with less nodes [NeT91].

Table 3.1 Network Nodal Configuration

Node	IDNX (type)	Capacity (Mbps)	Node	IDNX (type)	Capacity (Mbps)
0	90	256	10	20	32
1	70	192	11	70	64
2	70	128	12	20	32
3	90	256	13	20	32
4	70	128	14	20	32
5	70	128	15	20	32
6	20	32	16	20	32
7	20	32	17	70	64
8	20	32	18	20	32
9	70	64	19	20	32

Figure 3.1 shows the general graph representation of the network. This 20-node network configuration is similar to the network model used by Mohammad Peyravian [Pey94]. The trunk capacities are the same as those used in similar studies [Pey94]. This network can support voice, data and video. Information is multiplexed and transported along the trunks at or below the maximum specified bandwidth.

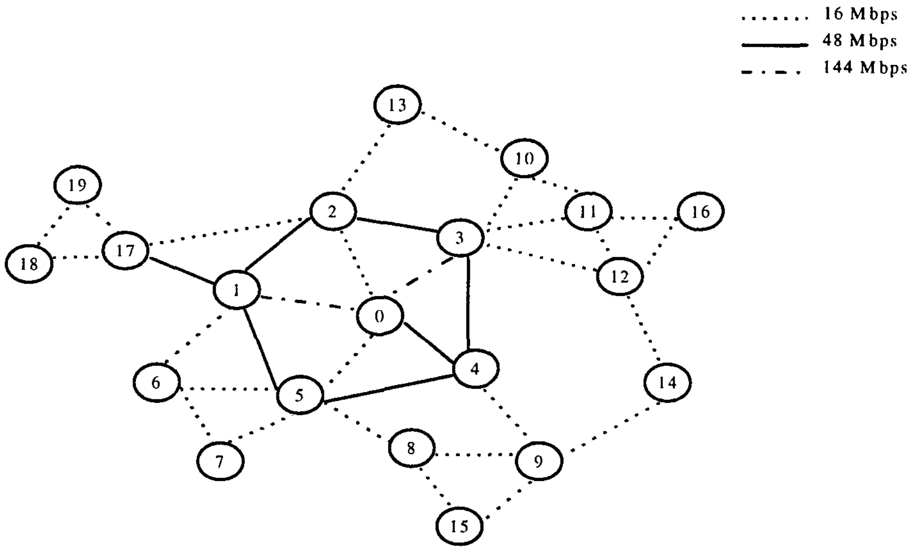

Figure 3.1 20 Node IDNX Communications Network

3.4.1 Objective

The purpose of this experiment is to present analytical data that can be used to compare the performance of preemption algorithms. Three preemption algorithms will be compared in terms of the number of call preempted and the grade of service. The performance parameters are analyzed to determine which algorithm produces the best grade of service for the higher priority calls and at the same time, the least number of preemptions.

3.4.2 Assumptions

In order to accomplish the performance studies on the simulation model, certain assumptions and operating conditions for the network must be established. Some of the operating assumptions were used in previous research [Pey94]. The operating assumptions and conditions for the communication network simulation are as follows:

1. Each source operates independent of each other.
2. Call requests are generated from the source and the inter-arrival times are assumed to be exponentially distributed.
3. The destination of the call request is generated based of a stated probability.

4. Four levels of priority. This number gives a good indication of the network preemption capability.
5. The mean holding time of the connections is fixed at 200 seconds for each priority. This is the same as one used in a similar study [Pey94].
6. The bandwidth of the calls is assumed to be uniformly distributed (except in the constant bandwidth experiments).
7. The average service time of the network is less than the arrival rate of the call requests. This ensures that the network would not always be congested.
8. All nodes and links are 100% reliable.
9. Traffic travels both ways along a link.
10. Errors due to channel noises are negligible.

3.4.3 Network Model

This section describes the construction of the network model on BONeS DESIGNER. Call requests are generated as customers requesting service. The service times correspond to the trunk holding times. Each call request is described by its data structure fields. Discussed below are the call request data structures.

3.4.3.1 Call Request Data Structure

The fields in the call requests describe the attributes of the call. Figure 3.2 shows the fields of the call request data structure. The source is the node where call requests are generated. The destination is the node where the call request will be terminated. The trunk field has multiple roles. Normally, it is the trunk connecting the source to the destination. In the case of call blocking, this trunk connects the source to the alternate node. Call holding time is the session length in seconds for the call request. Time in queue is the time this call request enters its associated queue in its waiting room. Alternate node is the node that the call request uses in case of call blocking. When the trunk connecting the present node and the next node nodes is out of capacity, the call request is rerouted through this node. The alternate trunk is the trunk connecting the alternate node to next node in the path. This trunk is used only in the case of call blocking.

Name: Call Request DS [circuit SW]
Date: Saturday, 11/9/96 08:01:50 pm EST

Name	Type	Subrange	Default Value
Source	INTEGER	[0, +Infinity)	...
Destination	INTEGER	[0, +Infinity)	...
Trunk	INTEGER	[0, +Infinity)	...
Call Holding Time	REAL	[0, +Infinity)	...
Time In Queue	REAL	[0, +Infinity)	...
Alternative Node	INTEGER	[0, +Infinity)	...
Alternative Trunk	INTEGER	[0, +Infinity)	...
Alternative?	INTEGER	[0, 1]	0
Priority	INTEGER	[0, +Infinity)	0
Bandwidth	INTEGER	[0, +Infinity)	...
Next Node	INTEGER	[0, +Infinity)	...
Number of Hops	INTEGER	[0, +Infinity)	0
Queue Number	INTEGER	[0, +Infinity)	...
Node or Trunk Blocked	INTEGER	[0, 1]	0
Previous Node	INTEGER	[0, +Infinity)	...
Trunks in Path	INT-VECTOR	...	...
Nodes in Path	INT-VECTOR	...	...
Source Blocked?	INTEGER	[0, 1]	0
Source or Next Node Delay	INTEGER	[0, 1]	0
Replace Preempted Calls in Queue?	INTEGER	[0, 1]	0
Previous Trunk	INTEGER	[0, +Infinity)	...
Preempt Alternative Trunk?	INTEGER	[0, 1]	0

Figure 3.2 Call Request Data Structure Fields

The field "alternate?" takes on the values zero or one. Zero implies the call request takes the normal route and the one implies the call request takes the alternate route. The priority is an integer from 0 to 3. The 0 represents the lowest priority and the 3 is the highest. The bandwidth is also an integer that represents the capacity of the call. Next node is the node call requests will next seek connection. If this node is the last node in the source-destination path, then it is also the destination node. The number of hops field is the number of trunks a call request traverse. An internal call will not have a zero in that field. The queue number is the number of the queue in the waiting room that connected calls wait. The field "node or trunk blocked" takes on the values zero or one. Zero implies the next node is blocked and the one implies the next trunk is blocked.

The field "previous node" is the node the call request leaves to go to the next node. The previous node and the destination determine the next node in the source-destination path. The trunks in path field is the set of trunks in the source-destination path. The nodes in path field is the set of nodes in the source-destination path. Source blocked? is the field that indicates where a call request was blocked. This field takes on two values zero or one. The zero means that the call is blocked at the next node and the one means that the call is blocked at the source. The field "source or next node delay" determines the path of the call request leaving the node processor. This field also takes on two values zero or one. A zero value implies the call request came directly from the source and the one means the call came from a node besides the source.

Replace preempted call in queue? is one of the fields in the data structure. It determines whether calls preempted should be reconnected. It takes on two values zero or one. The one causes reconnection of the preempted calls while the zero value terminates the preempted calls. Finally, the previous trunk field reinstates the next trunk when an alternative trunk is blocked.

3.4.3.2 Variable Bandwidth Configuration

The first system configuration to be discussed involves preempting calls in a network where the bandwidth is not constant. This configuration occurs in systems where voice, video, and data are integrated. Flexibility of the nodes to accommodate the varying bandwidth is key for this system. The system level for all configurations will be the same. At this level, the network consists of an initialization block, a traffic generator block, an IDNX circuit switch network block, a statistic collection block, and a node & trunk monitor block. The simulation begins at the Initialization block and ends at the Statistic Collection block. Figure 3.3 illustrates the system level block diagram.

The first module in the system level diagram is the Initialization module. The Initialization module reads all input files and places the data in the non local memories. The memories include Trunk Capacity IMatrix, Node Capacity IVector, Next Node Routing IMatrix, Call Generation Factor RVector, Destination Distribution RMatrix, and Alternate Routing IMatrix. Figure 3.4 shows the content of the Initialization module. The Initialization module also modifies the Destination Distribution RMatrix by transposing the input destination probabilities into a cumulative distribution function. The modified destination distribution matrix is later used to generate the destination node for all call requests. Appendix A, Figure A-1 shows the sub-module of the Modify "Destination Distribution RMatrix." The Initialization module simulates the loss of network capacity by reducing the initial capacity of nodes and trunks. This reduction occur in the "Modify Trunk Capacity IMatrix & Node Capacity Ivector block. Appendix A, Figure A-2 shows the sub-module internal representation. Appendix A, Figures A-3 and A-4 show the read input files.

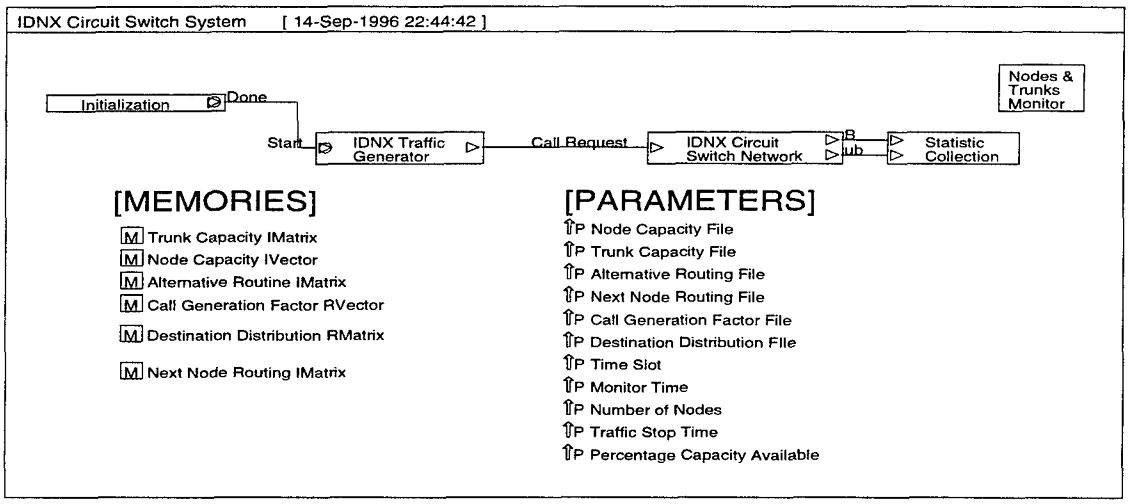

Figure 3.3 IDNX Circuit Switch System

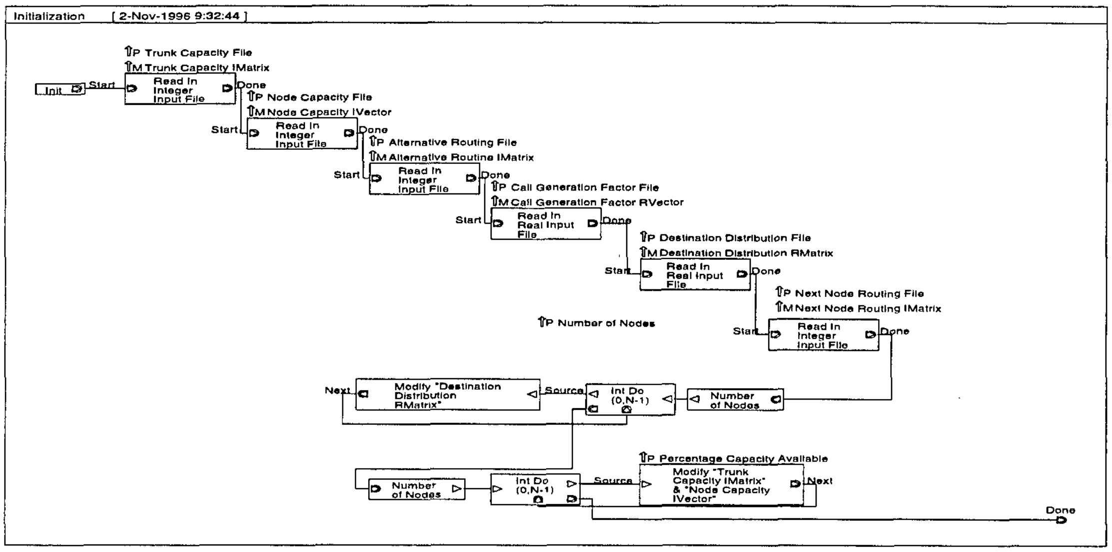

Figure 3.4 Initialization Block

The traffic generator generates call requests for all nodes in the network until the end of the simulation. Figure 3.5 illustrates the traffic generator module. This module contains a source generator that generates exponentially distributed call requests from all 20 nodes. Within the traffic generator, the source node, destination node, priority, bandwidth, next trunk and call holding time are inserted into each call request.

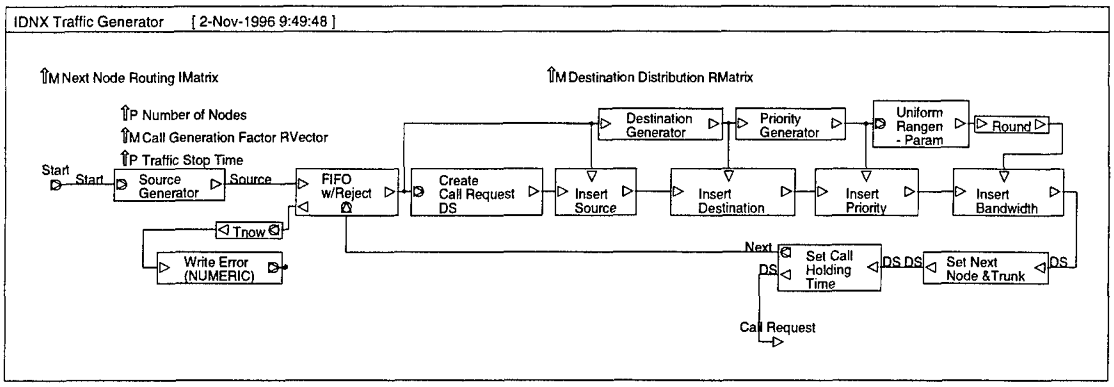

Figure 3.5 Traffic Generation Module

The most important block in the traffic generation module is the source generator block. This source generator produces the call requests for all nodes. The source generator operates by first generating the inter-arrival time of the first call request per node. The requests are then delayed for the appropriate amount of time before leaving the block. Once a call request leaves the source generator, the next inter-arrival for that node is then generated. This procedure continues until the traffic stops time. Figure 3.6 shows the block diagram of the Source Generator.

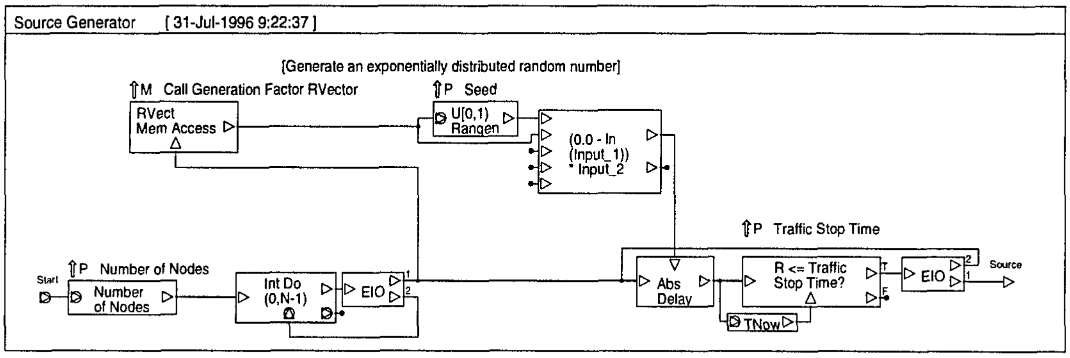

Figure 3.6 The Source Generator Module.

The Destination Generator produces the destination for the source node based on the probability in the Destination Distribution RMatrix file. The priority generator inserts the priority in the call requests. This configuration utilizes a uniformly distributed random generator to generate the bandwidth of each call request. Appendix A, Figure A-5 shows the block diagram of the Destination Generator.

The priority generator separates the call requests into four distinct classes. The probability of each priority based on a fixed parameter and set before simulation. Appendix A, Figure A-6 shows the block diagram of the Priority Generator.

The first section of a call's path and its call length are also set in the traffic generator. The call initial path is set in the block Set Next Node & Trunk. The call length is calculated and set in the Set Call Holding Time block. Appendix A, Figures A-7 and A-8 show these blocks.

The next major module in the IDNX circuit switch system is the circuit switch network module. This module is the heart of the communication network model. It determines if a call request becomes blocked or successfully completes it task. Figure 3.7 shows the IDNX Circuit Switch Network.

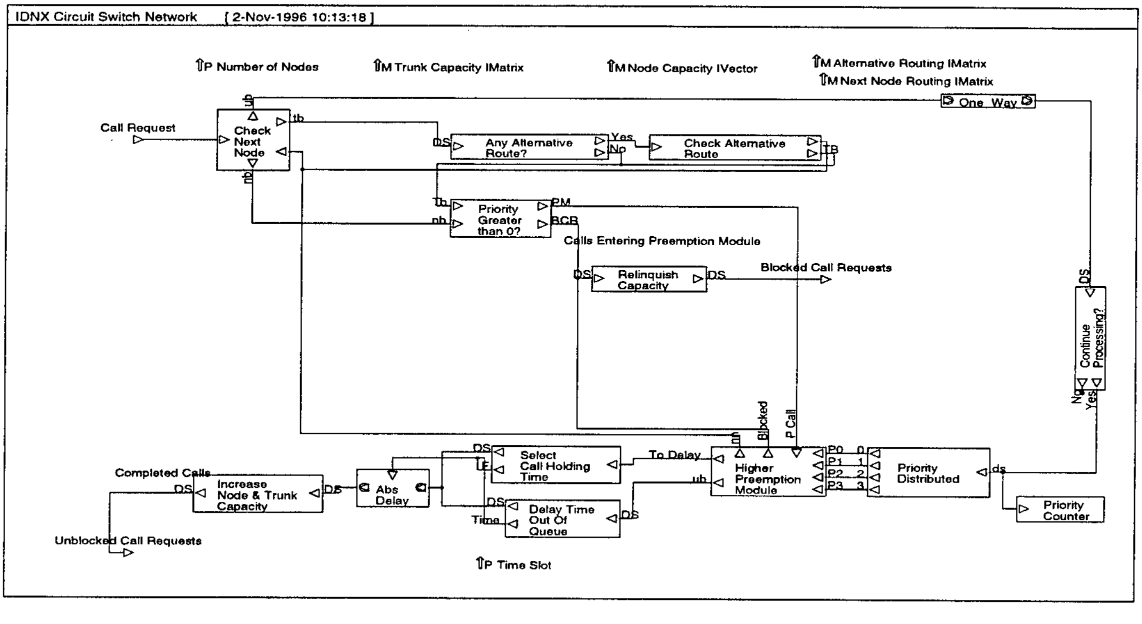

Figure 3.7 IDNX Circuit Switch Network

The major sub-modules in the circuit switch network module are Check Next Node block, Any Alternative Route? block, Check Alternative Route block, Priority Greater than 0? Block, Relinquish Capacity, Higher Preemption block, and Increase Node & Trunk Capacity.

Call requests enter the network module through the Check Next Node block. This block checks the source node, next trunk and next node for blocking. If no blocking occurs, the next node is checked to ascertain if it is also the destination node. If the next node is not the destination node, then the number of hops is incremented and the next node and trunk in the source-destination path is selected. Once again, the selected node and trunk are checked for blocking. This procedure continues until the call request is blocked or connected. Figure 3.8 shows the representation of the Check Next Node module. Appendix A, from Figure A-9 to Figure A-18 show the internal representations of the individual blocks of the Check Next Node module.

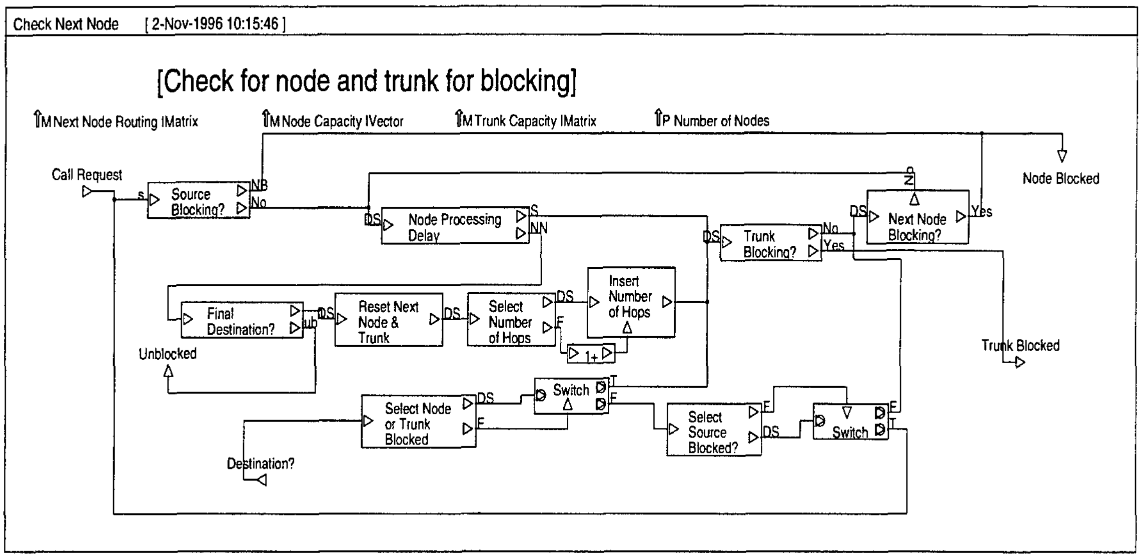

Figure 3.8 Check Next Node Module

If the trunk is blocked, the call request enters the block Any Alternate Route? to determine if there is any alternate route. Appendix A, Figure A-19 shows the internal components of the Any Alternate Route? block. If no alternate route is available, the call is blocked. If an alternate route is available, the call request enters the Check Alternative Route block. This block checks the alternate node and trunk for blocking. If the alternate route is blocked, the call request will be blocked. If not, the call request continues on its source-destination path. Appendix A, Figure A-20 is the diagram of the Check Alternative Route block. If all the routes are blocked and the priority is 0, the call request will be terminated.

All initially blocked call requests enter the Priority Greater than 0? block. In this block, priority-0 call requests are routed automatically to the Relinquish Capacity block, and priority calls greater than 0 are routed through the preemption module. Appendix A, Figure A-21 is the internal representation of the Priority Greater than 0? block.

All blocked call requests enter the Relinquish Capacity block. This block returns the acquired capacity to the respective nodes and trunks. Call requests entering the Relinquish Capacity block are first checked to see if the source and the destination are the same. Figure 3.9 shows the internal representation of the Relinquish Capacity block. If the source and destination are not equal, the call requests enter the Increase Trunk Capacity block to relinquish all trunk capacity. It then enters the Increase Node Capacity to return all received node capacity. Appendix A, Figure A-22, and Figure A-23 show the Increase Trunk Capacity block and the Increase Node Capacity block respectively. If the source and destination are the same, then the call request is checked to see if it was blocked at the source. If it was blocked at the source, no capacity adjustment is necessary. If it was not blocked at the source, then the capacity at the source node only is adjusted.

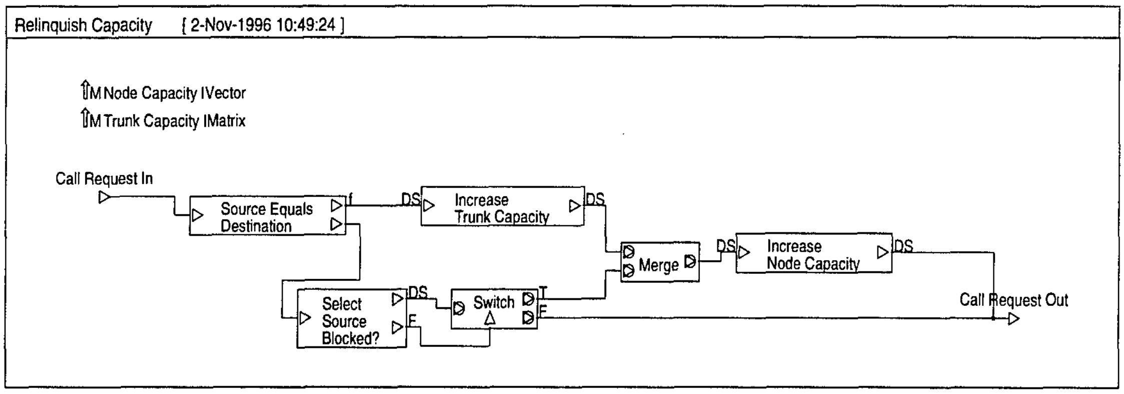

Figure 3.9 Relinquish Capacity

All unblocked call requests are sorted by priority and enter the Higher Preemption module. The Higher Preemption module shown in Figure 3.10 has many levels. The first level groups the four Calls Connection blocks together. Calls entering this module are directed to their respective Calls Connection block. The construction of each Calls Connection is essentially the same.

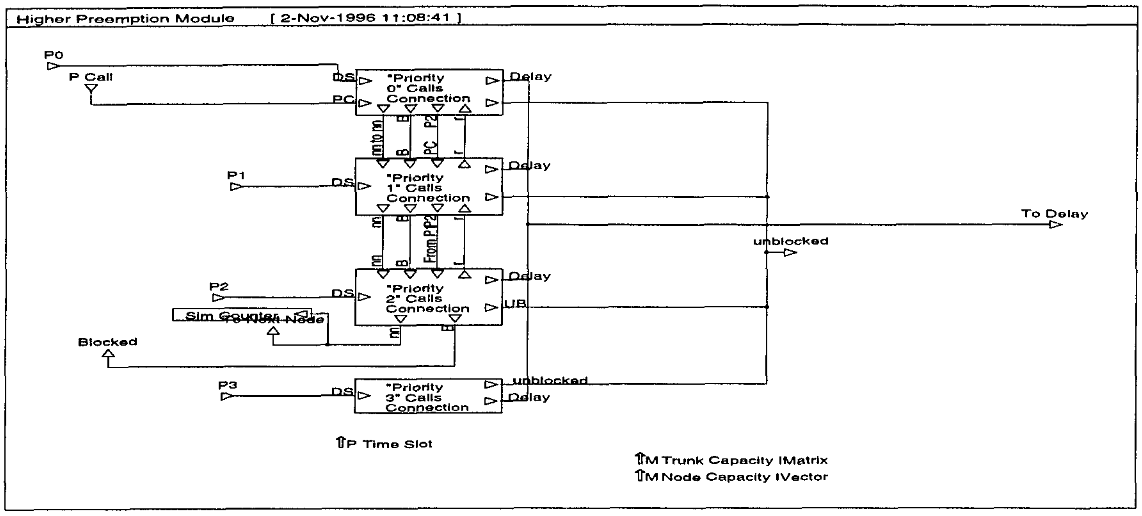

Figure 3.10 Higher Preemption Module

Each calls connection block contains only one priority class. These blocks allow incoming call requests with a call holding time of less than one time slot to go directly to the short term delay block. A time slot is a parameter, set before simulation, which dictates how often a queue in the waiting room will be emptied. It is also the remaining time of a call by which preemption will not take place. A call with a holding time greater than the time slot will be sent to the waiting room. All calls in the waiting room wait until one time slot before going to the delay block. Calls leaving the waiting room enters the Queue Number Correct? block. This block routes calls ready for completion to the short term waiting room and calls considered for preemption to the Call Preemption block. Appendix A, Figure A-24 shows the block diagram of the Queue Number Correct? block. The Calls Connection blocks also contain the appropriate call preemption module. Figure 3.11 is the diagram of the "Priority 0" Calls Connection block. Appendix A, Figures A-25, A-26, and A-27 show the diagrams of Priority 1, 2 and 3 Calls Connection blocks respectively.

Each Call Preemption block allows the preemption of a lower priority call by an incoming higher priority call request. The Call Preemption blocks contain four key lower level modules. These modules are Node or Trunk Blocked, Empty Queues, Preempt If Within Blocked Path, and Next Priority Preemption?. Figure 3.12 shows the block diagram of Priority 0 Call Preemption block. Appendix A, Figures A-28 and A-29 show the block diagrams of Priority 1, and 2 Call Preemption modules respectively.

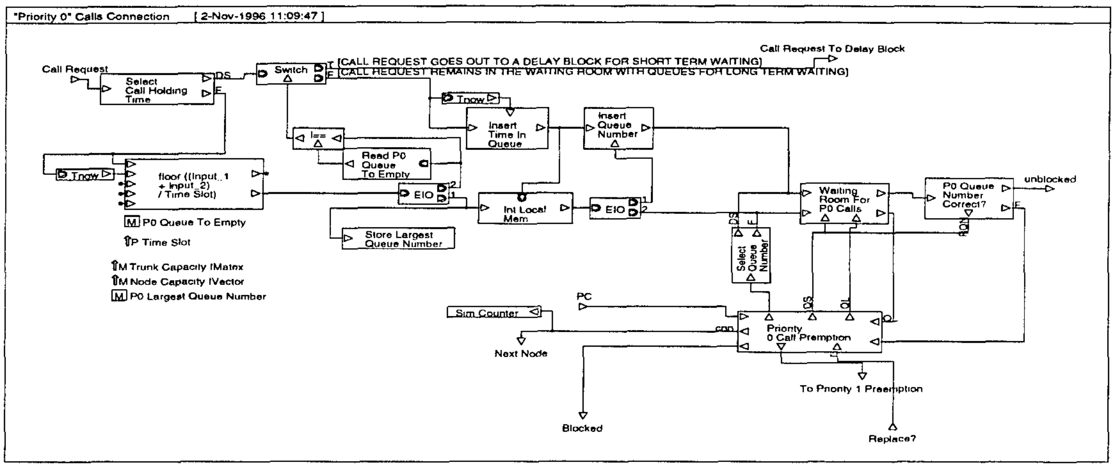

Figure 3.11 "Priority 0" Calls Connection

Before preemption can occur, the preemption module needs to know what segment of the priority calls path is blocked. The determination of the segment blockage is accomplished in the "Node or Trunk Blocked" block. Calls entering this block are first check to see whether the node or trunk is blocked. If trunk blocking, the number of the trunk is stored in memory for future use. If node blocking, the call request is further checked to determine if the next node or the source node was blocked. In any case, the node number is also stored in memory for future use. Appendix A, Figure A-30 shows the diagram of the "Node or Trunk Blocked" block.

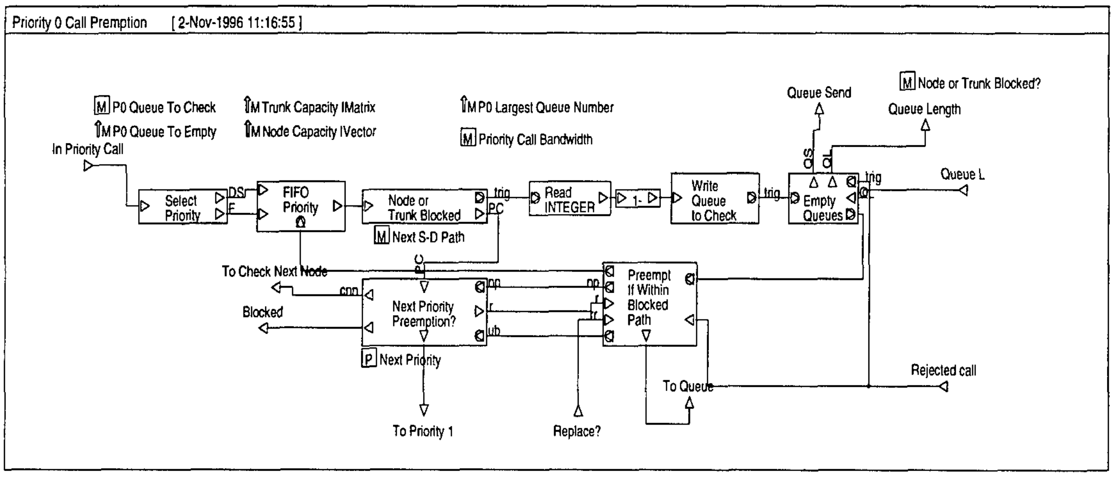

Figure 3.12 Priority 0 Call Preemption

The block Empty Queues does exactly what its name implies. It empties all the queues in the waiting room except for the queue that holds calls within one time slot of being emptied. Figure 3.13 shows the internal representation of Empty Queues block.

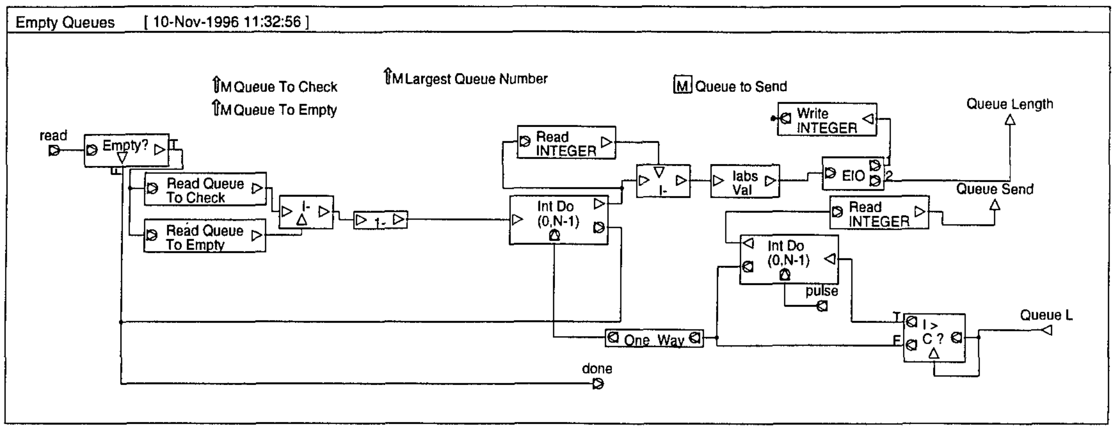

Figure 3.13 Empty Queues

While the queues are being emptied, the Preempt If Within Blocked Path block compares the connected call source-destination for path similarity with that of the blocked path of the higher priority call. Those calls with no path similarity are sent back to their respective queues. Those with path similarity are sent to the Preempt Largest Bandwidth First block. Figure 3.14 shows the block diagram of Preempt If Within Blocked Path block.

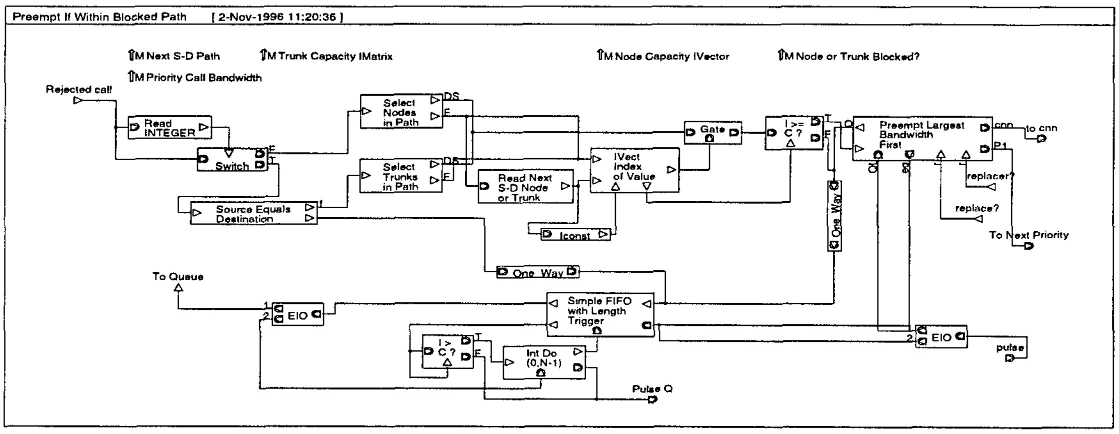

Figure 3.14 Preempt If Within Blocked Path

The Preempt Largest Bandwidth First block arranges the selected calls based on bandwidth in preparation for preemption. Calls entering this block are placed in a queue with the largest bandwidth at the head of the line. When all the connected calls have been checked for path similarity, the calls are emptied from the queue and enter the Preempt This Call Request? block. Figure 3.15 shows the block diagram of Preempt Largest Bandwidth First module.

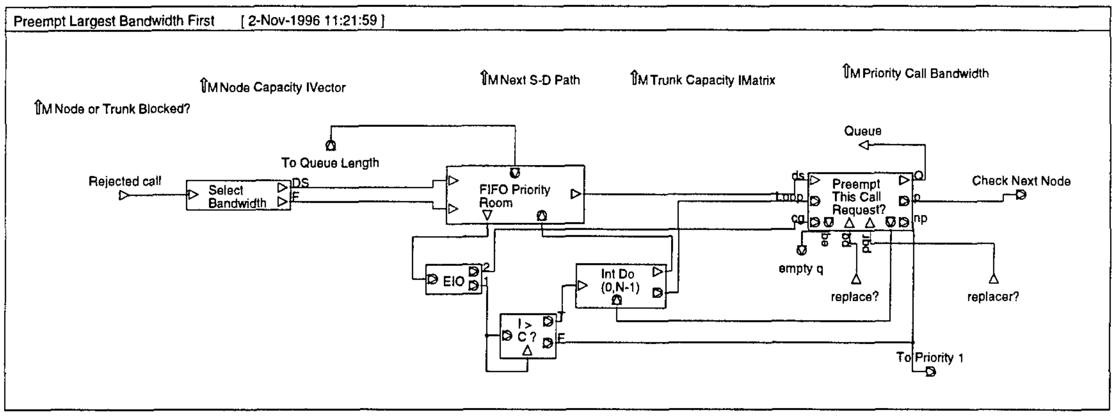

Figure 3.15 Preempt Largest Bandwidth First

The "Preempt This Call Request?" block preempts calls if the path is still blocked, and reconnects calls if enough capacity cannot be released to connect the higher priority call. Figure 3.16 is the representation of the Preempt This Call Request? block. A call enters this block through the Still Blocking? module. At this point, the incoming higher priority call's path is rechecked to see if it is still blocked. Appendix A, Figure A-31 shows the Still Blocking? module. If the path is blocked, the capacity in the selected calls is returned to the appropriate nodes and trunks. Once again, the path is checked for blocking. If there is no blocking, the selected call is preempted and all other selected calls are replaced in their respective queues. If the path is still blocked, the selected call enters a holding area. Calls in the holding area are terminated if the incoming priority call is connected, otherwise the calls are reconnected. Before the calls are reconnected, they enter the Decrease Node & Trunk Capacity block. In this block, the bandwidth capacity is returned to them. Appendix A, Figures A-32, A-33, A-34 shows the diagrams of Decrease Node & Trunk Capacity and its sub-modules. The total numbers of calls, bandwidth, and network bandwidth of the preempted calls are recorded in the Preempted Statistic BW block. Appendix A, Figures A-35 displays this block.

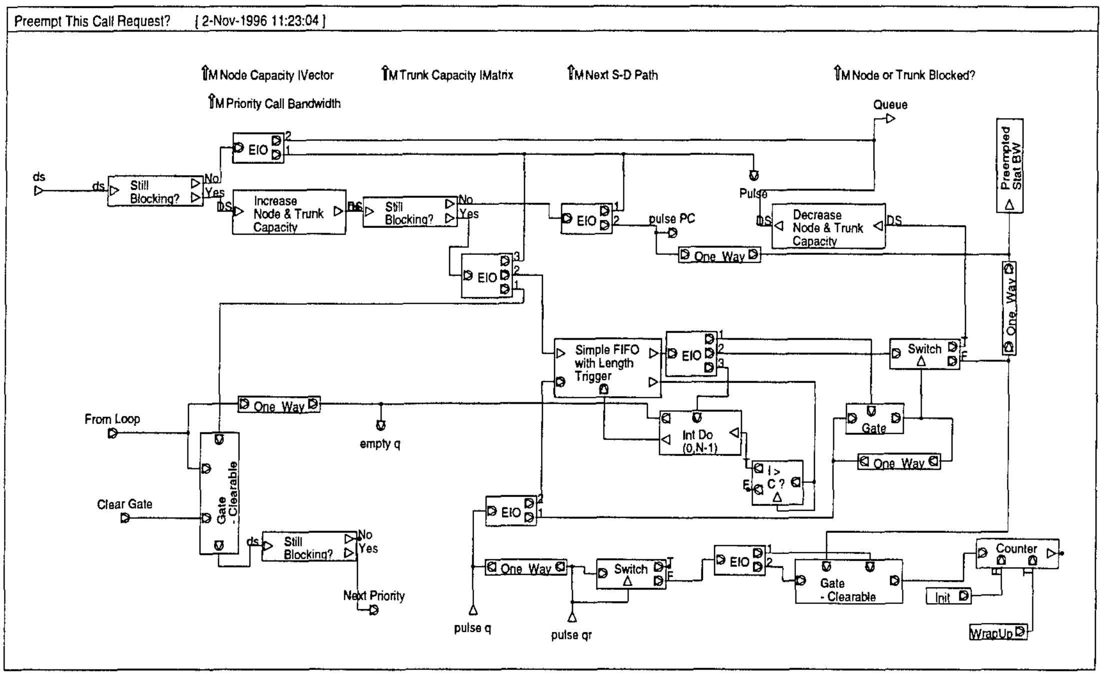

Figure 3.16 Preempt This Call Request?

If no calls or an insufficient number of calls was preempted, the incoming priority call will enter the Next Priority Preemption? block. This block routes the incoming priority call request to the next higher priority preemption or to the blocked port. If the incoming call request priority is greater than the next higher priority preemption, the priority call request will be routed to that preemption module. If not, it will be blocked. Figure 3.17 shows the block diagram of the Next Priority Preemption? block. This block slightly differs for priority-2 preemption because there is no higher priority preemption block. The block is called P2 Next Node or Blocked, and it can be seen in Appendix A, Figure A-36.

Once a successfully completed call leaves the Higher Preemption module, it goes to a delay block. At the delay block, the call waits the remaining time of the call holding time. After the delay, the completed call increases the capacity of the nodes and trunks in its source-destination path.

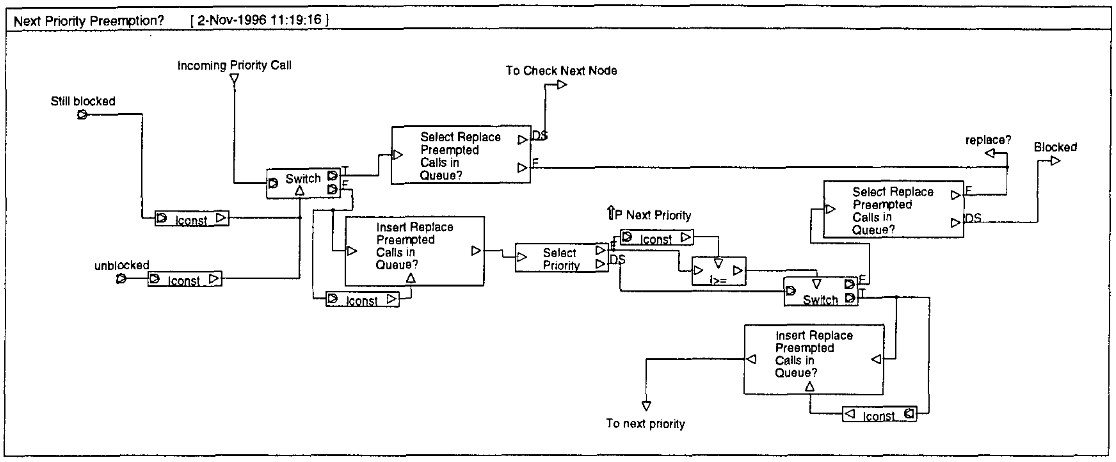

Figure 3.17 Next Priority Preemption?

Once the call finishes updating the nodes and trunks, it enters the Statistic Collection module. The Statistics Collection module counts the total number of blocked calls and the total number of completed calls per priority. Figure 3.18 shows the block diagram of the Statistic Collection block.

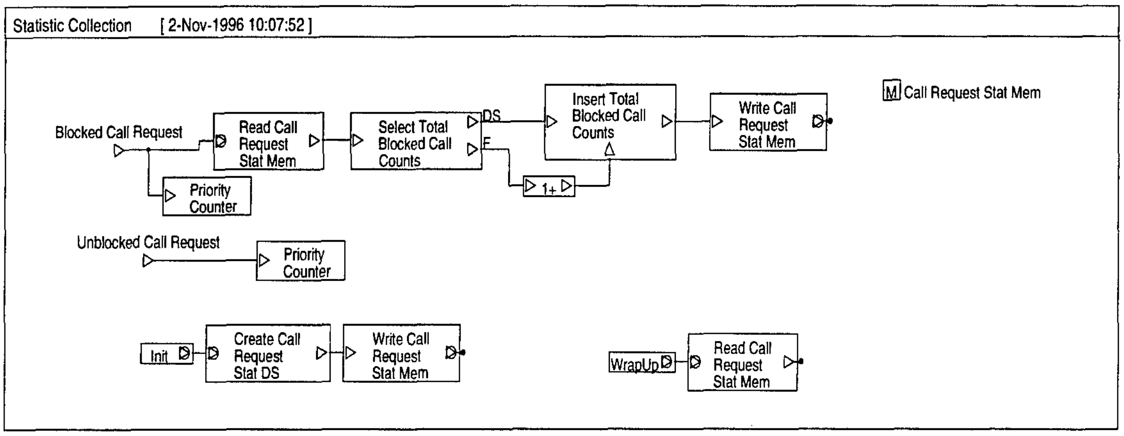

Figure 3.18 Statistic Collection

The final module is the Node & Trunk Monitor. This module records the remaining capacity of the nodes and trunks by accessing the Node Capacity IVector and the Trunk Capacity Imatrix. The Node & Trunk Monitor module helps to determine the steady-state time of the simulations. It also helps to establish the minimum required simulation run time. Figure 3.19 shows the block diagram of the Node & Trunk Monitor.

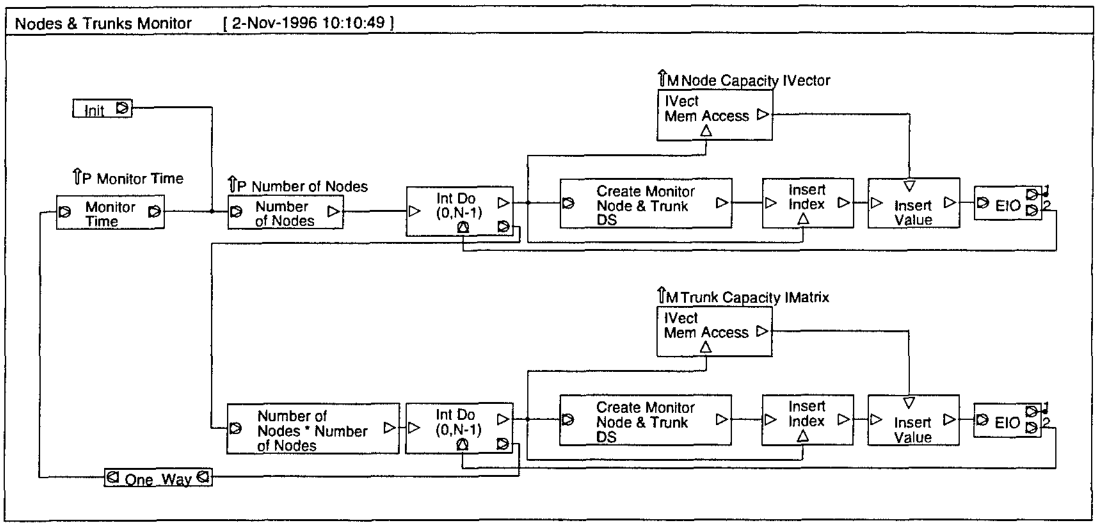

Figure 3.19 Node & Trunk Monitor

3.4.3.3 Constant Bandwidth Configuration

This configuration can occur in a time division multiplexed system where all nodes transmit at a constant bandwidth. The system design is the same as that of above except for the traffic generator block. The traffic generator block in this configuration inserts a constant bandwidth into the call requests data structure. Figure 3.20 shows the traffic generator with the constant bandwidth inserted in the call request data structure. Every other aspect of the model is identical to the first configuration.

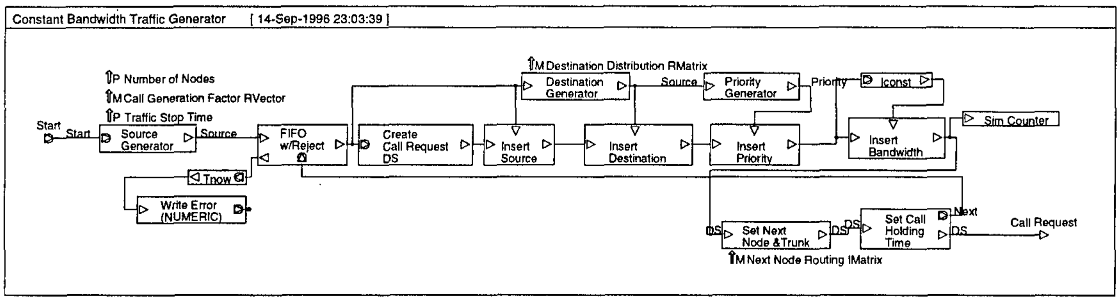

Figure 3.20 Constant Bandwidth Traffic Generator

3.4.3.4 Network Bandwidth Configuration

The call preemption configuration emphasizes network bandwidth. It is concerned with the total usage of bandwidth capacity in the entire network. This configuration seeks to preempt lower priority calls with the least number of hops first. In this way, the network disturbance is reduced. Network bandwidth is the bandwidth of a call multiplied by the number of hops. The network model for this configuration is almost identical to the first model. The differences occur in the lower levels of the Higher Preemption module. The block Preempt If Within Blocked Path is the point where the differences begin. In this block, the module Preempt Least Number of Hops First replaces Preempt Largest Bandwidth First block. Figure 3.21 is a representation of Preempt If Within Blocked Path for the network bandwidth configuration.

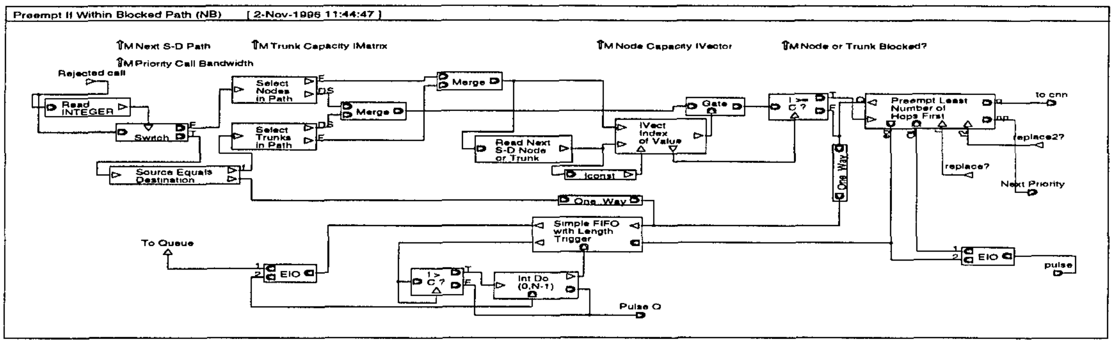

Figure 3.21 Preempt If Within Blocked Path (NB)

The block "Preempt Least Number of Hops First" places the connected calls with path similarity to the higher priority call in a queue. The queue allows the calls with the lowest network bandwidth to go ahead of the line. Calls are then ejected from the queue one at a time. Each ejected call enters the Preempt This Call Request? block. The "Preempt This Call Request?" block ensures the path is blocked before preempting the ejected call. If the path becomes unblocked, the remaining call or calls are sent back to their original queue in the waiting room. Figure 3.22 shows the Preempt Least Number of Hops First block.

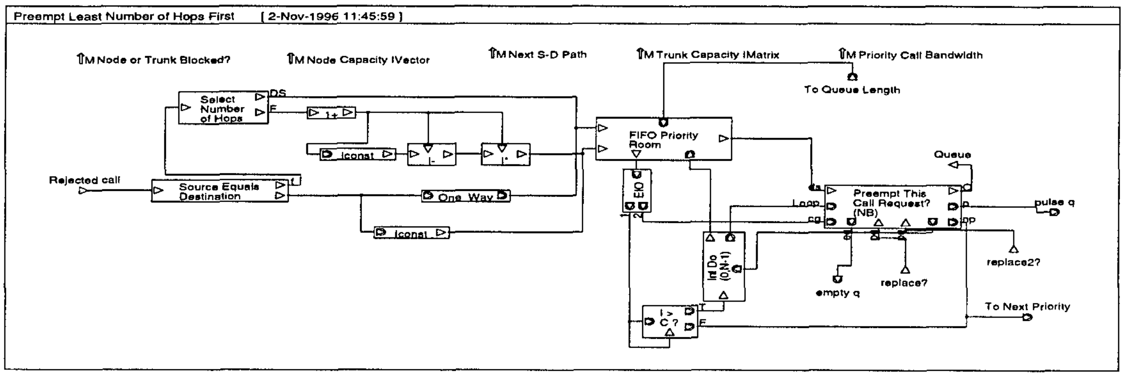

Figure 3.22 Preempt Least Number of Hops First

3.4.3.5 Constant Network Bandwidth Configuration

This configuration is a hybrid between the constant bandwidth and network bandwidth configurations. Once again this is applicable in a time division multiplexed system. The configuration is almost identical to that of call preemption based on network bandwidth except for the constant traffic generator shown in Figure 3.20.

3.5 System Parameters

Several different simulations are required to produce the necessary performance metrics. The parameters for each of the configurations are similar. These parameters are as follows:

1. Node Capacity File
2. Trunk Capacity File
3. Next Node Routing File
4. Alternative Routing File
5. Priority
6. Bandwidth
7. Call Holding Time File
8. Number of Nodes
9. Percentage Capacity Available
10. Time Slot
11. Call Generation Factor File
12. Destination Distribution File

The node capacity file contains the node available capacity for all the nodes in the network. The node capacities chosen represent typical capacities in IDNX-20, IDNX-70, and IDNX-90 nodes. Table 3.2 shows the node capacity for each node. Each unit of capacity in Table 3.2 represents 64 Kbps of bandwidth. The bandwidth generated for each call will be a product of 64,000 in order to reduce the computer processing time.

Table 3.2 Node Capacity

Node	Capacity	Node	Capacity
0	4000	10	500
1	3000	11	1000
2	2000	12	500
3	4000	13	500
4	2000	14	500
5	2000	15	500
6	500	16	500
7	500	17	1000
8	500	18	500
9	1000	19	500

The trunk capacity file contains the trunk available capacity for all the nodes in the network. A zero in this file means no trunk exists between the nodes. The node/trunk capacity file contains the bandwidth capacity of all the nodes and trunks in the network. This file contains 400 values. Table 3.3 lists the node/trunk capacity. The diagonal row in Table 3.3 from top left to bottom right contains the node capacities and the remaining sections of the table reflect the trunk capacities. The trunk capacities used are similar to those used in previous study [Pey94]. Each unit also represents 64 Kbps of bandwidth.

Table 3.3 Node/Trunk Capacity

NODE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	4000	2250	750	2250	750	250	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2250	3000	750	0	0	750	250	0	0	0	0	0	0	0	0	0	0	750	0	0
2	750	750	2000	750	0	0	0	0	0	0	0	0	0	250	0	0	0	250	0	0
3	2250	0	750	4000	750	0	0	0	0	0	250	750	250	0	0	0	0	0	0	0
4	750	0	0	750	2000	750	0	0	0	750	0	0	0	0	0	0	0	0	0	0
5	750	750	0	0	750	2000	250	250	250	0	0	0	0	0	0	0	0	0	0	0
6	0	250	0	0	0	250	500	250	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	250	250	500	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	250	0	0	500	250	0	0	0	0	0	250	0	0	0	0
9	0	0	0	0	750	0	0	0	250	1000	0	0	0	0	250	250	0	0	0	0
10	0	0	0	250	0	0	0	0	0	0	500	250	0	250	0	0	0	0	0	0
11	0	0	0	750	0	0	0	0	0	0	250	1000	250	0	0	0	250	0	0	0
12	0	0	0	250	0	0	0	0	0	0	0	250	500	0	250	0	250	0	0	0
13	0	0	250	0	0	0	0	0	0	0	250	0	0	500	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	250	0	0	250	0	500	0	0	0	0	0
15	0	0	0	0	0	0	0	0	250	250	0	0	0	0	0	500	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	250	250	0	0	0	500	0	0	0
17	0	750	250	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1000	250	250
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	250	500	250
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	250	250	500

The next node routing file contains the next node along the source-destination path for all call requests. This is a static routing system. The path of the call is totally dependent on the routing file which is fixed. Calls with identical source and destination travel along the same path unless, an alternate route is chosen. Table 3.4 shows the next node routing file matrix.

Table 3.4 Next Node Routing File Matrix

Node	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	1	2	3	4	5	1	5	5	4	3	3	3	2	4	4	3	1	1	1
1	0	1	2	0	0	5	6	5	5	5	0	0	0	2	0	5	0	17	17	17
2	0	1	2	3	3	1	1	1	0	0	3	3	3	13	3	1	3	17	17	17
3	0	0	2	3	4	0	0	0	4	4	10	11	11	2	4	4	11	2	2	2
4	0	0	3	3	4	5	5	5	9	9	3	3	3	3	9	9	3	0	0	0
5	0	1	0	0	4	5	6	7	8	4	0	0	0	0	4	4	0	1	1	1
6	1	1	1	1	5	5	6	7	5	5	1	1	1	1	5	5	1	1	1	1
7	5	6	6	5	5	5	6	7	5	5	5	5	5	5	5	5	5	6	6	6
8	9	5	5	9	9	5	5	5	8	9	9	9	9	9	9	15	9	5	5	5
9	4	4	4	4	4	4	4	4	8	9	4	4	14	4	14	15	14	4	4	4
10	3	3	3	3	3	3	3	3	3	3	10	11	3	13	3	3	11	13	13	13
11	3	3	3	3	3	3	3	3	3	3	10	11	12	10	12	3	16	3	3	0
12	11	11	11	11	11	11	11	11	14	14	11	11	12	11	14	14	16	11	11	1
13	2	2	2	10	10	2	2	2	2	2	10	10	2	13	2	2	2	2	2	2
14	9	9	12	12	9	9	9	9	9	9	12	12	12	12	14	9	12	12	12	3
15	9	8	9	9	9	8	8	8	8	9	9	9	9	9	9	15	9	8	8	4
16	11	11	11	11	11	12	12	11	12	12	11	11	12	11	12	12	16	11	11	11
17	1	1	2	1	1	1	1	1	1	1	2	2	2	2	1	1	2	17	18	19
18	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	18	19
19	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	18	19

The alternative routing file contains the possible alternative route available between any pair of nodes in the network. If a call request selected next trunk is blocked, that call then searches the alternative routing file matrix for an alternate route. A negative one (-1) in this file implies that no alternative route is available between that pair of nodes. This file contains 400 values. Table 3.5 lists the available alternative routes between the various nodes.

Four priority classes were chosen for the simulation models. Previous research has demonstrated that the choice of four priorities is sufficient to analyze the performance of a communication network [MaT87] [MoL87] [Pey94]. Each priority has its own probability of occurrence. Table 3.6 shows the probability of the different priority classes.

Table 3.5 Alternative Route

NODE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	-1	5	1	2	3	4	5	1	4	5	2	2	2	3	3	4	4	5	5	5
1	2	-1	0	2	5	0	5	0	0	0	2	2	2	0	5	0	2	2	2	2
2	3	0	-1	0	0	0	0	0	1	3	-1	0	0	-1	0	0	0	1	1	1
3	4	2	0	-1	0	4	2	2	0	0	11	10	10	0	0	0	10	0	0	0
4	3	5	0	0	-1	0	0	0	-1	-1	0	0	0	0	-1	-1	0	5	5	5
5	1	0	1	4	0	-1	7	6	-1	0	4	4	4	4	0	0	1	0	0	0
6	5	5	5	5	1	1	-1	5	7	7	5	5	5	5	7	7	5	5	5	5
7	6	5	5	6	6	6	5	-1	6	6	6	6	6	6	6	6	6	5	5	5
8	15	-1	-1	15	15	-1	-1	-1	-1	15	15	15	15	15	15	9	15	-1	-1	-1
9	-1	-1	-1	-1	-1	-1	-1	-1	15	-1	-1	-1	-1	-1	-1	8	-1	-1	-1	-1
10	11	11	11	11	11	11	11	11	11	11	-1	3	11	-1	11	11	3	-1	-1	-1
11	10	10	10	10	10	10	10	10	10	10	3	-1	16	3	16	16	-1	10	10	10
12	16	16	16	16	16	16	16	16	-1	-1	16	16	-1	16	-1	-1	11	16	16	16
13	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
14	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
15	8	9	8	8	8	9	9	9	9	8	8	8	8	8	8	-1	8	9	9	9
16	12	12	12	12	12	11	11	12	11	11	12	12	11	12	11	11	-1	12	12	12
17	2	2	1	2	2	2	2	2	2	2	1	1	1	1	2	2	1	-1	19	18
18	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	-1	17
19	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	17	-1

The constant bandwidth configurations assume the bandwidth of each call is 12 units. Each unit represents 64 Kbps of bandwidth. This makes the bandwidth of each constant bandwidth 768 Kbps. This bandwidth approximates half the capacity of a T-1 link. A T-1 link is prominent in most IDNX network [NeT91]. The variable bandwidth configurations assume the bandwidth of each call is uniformly distributed between 1 and 24 units. Therefore, the bandwidth of the variable bandwidth configurations lies between 64 Kbps and 3968 Kbps. Table 3.6 also shows constant and variable bandwidth for the different priority classes.

The call holding time is the total connection time of a call. It is exponentially distributed with a mean of 200 seconds. Refer to Table 3.6. The mean call holding time is the same for each priority class. If the average call holding time per priority is kept the same, then any effect of the holding time on the individual priorities is minimal.

Table 3.6 Priority and associated parameters

Priority	Priority Probability	Constant Bandwidth	Uniformly Distributed Bandwidth	Mean Holding Time (Exponential) seconds
0	0.40	12	1 - 62	200
1	0.35	12	1 - 62	200
2	0.20	12	1 - 62	200
3	0.05	12	1 - 62	200

The number of nodes is the total number of nodes in the network. This parameter can be set at the beginning of each simulation. There are 20 nodes in this network. This number of nodes provides the needed platform to analyze preemption in a circuit switch network. A similar research has used 16 nodes [Pey94].

The percentage available capacity is the percentage of the capacities of nodes and trunks available for the simulation iteration. This parameter simulates peak traffic conditions and or loss of network capacity.

The time slot is the time interval when the waiting room of queues empties. It helps to determine the queue number where the call request waits. This number is also the cut-off remaining time by which a call request will not be preempted. The time slot in this simulation was set at 10 seconds as it was done previously [AlR94].

The call generation is the distribution of calls generated by each node. The call generation factor file contains the call rate averages for each node in the network. The calls are exponentially distributed. Exponentially distributed calls are typical in communications network [Pey94]. Table 3.7 shows the average inter-arrival time of the call requests generation.

The destination distribution is the probability that a call generated in one node would go to a particular node. The destination distribution file matrix contains the probabilities used in selecting the destination node for the call request of each node. Table 3.8 shows the Destination Distribution. Generally, nodes in close proximity have a higher probability of connecting than one further away. Also, nodes with high capacity tend to have high destination probabilities. Refer to Figure 3.1 for nodal configuration.

Table 3.7 Call Generation Factors

Node	Factor	Node	Factor
0	0.801	10	4.604
1	1.249	11	3.078
2	1.510	12	5.410
3	0.985	13	5.087
4	1.965	14	4.845
5	2.462	15	5.406
6	4.403	16	5.203
7	5.165	17	3.984
8	5.232	18	5.402
9	3.945	19	5.606

Table 3.8 Destination Distribution

NODE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0.23	0.19	0.05	0.17	0.07	0.05	0.01	0.03	0.03	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
1	0.21	0.20	0.11	0.06	0.06	0.11	0.05	0.02	0.02	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
2	0.14	0.14	0.24	0.16	0.07	0.05	0.03	0.01	0.01	0.01	0.02	0.01	0.02	0.03	0.01	0.01	0.01	0.01	0.01	0.01
3	0.15	0.06	0.12	0.28	0.11	0.04	0.04	0.01	0.01	0.02	0.03	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01
4	0.09	0.11	0.08	0.12	0.21	0.11	0.02	0.02	0.01	0.07	0.03	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01
5	0.10	0.15	0.07	0.06	0.11	0.20	0.05	0.06	0.04	0.02	0.01	0.01	0.01	0.01	0.03	0.02	0.01	0.02	0.01	0.01
6	0.13	0.11	0.09	0.05	0.06	0.10	0.19	0.07	0.03	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.02
7	0.05	0.07	0.05	0.05	0.06	0.11	0.09	0.26	0.06	0.04	0.01	0.01	0.01	0.05	0.01	0.02	0.01	0.02	0.01	0.01
8	0.04	0.09	0.06	0.04	0.07	0.09	0.03	0.05	0.24	0.09	0.01	0.01	0.03	0.01	0.02	0.07	0.01	0.01	0.01	0.02
9	0.10	0.03	0.04	0.09	0.16	0.09	0.02	0.02	0.07	0.18	0.02	0.02	0.02	0.01	0.02	0.05	0.01	0.02	0.02	0.02
10	0.07	0.04	0.05	0.11	0.07	0.04	0.02	0.01	0.01	0.02	0.24	0.08	0.03	0.13	0.01	0.01	0.02	0.01	0.01	0.02
11	0.07	0.05	0.06	0.11	0.03	0.03	0.02	0.02	0.02	0.03	0.07	0.24	0.05	0.05	0.02	0.02	0.06	0.02	0.01	0.02
12	0.06	0.02	0.03	0.09	0.05	0.03	0.01	0.01	0.02	0.06	0.02	0.06	0.33	0.02	0.05	0.01	0.10	0.01	0.01	0.01
13	0.07	0.04	0.16	0.07	0.05	0.05	0.02	0.01	0.02	0.02	0.06	0.04	0.03	0.22	0.02	0.01	0.02	0.05	0.02	0.02
14	0.05	0.06	0.03	0.07	0.04	0.02	0.01	0.01	0.05	0.11	0.02	0.02	0.11	0.01	0.22	0.06	0.06	0.02	0.01	0.02
15	0.06	0.04	0.03	0.05	0.07	0.08	0.02	0.02	0.08	0.10	0.02	0.02	0.03	0.01	0.02	0.31	0.01	0.01	0.01	0.01
16	0.08	0.02	0.03	0.09	0.04	0.03	0.01	0.01	0.02	0.04	0.05	0.04	0.04	0.02	0.02	0.01	0.41	0.02	0.01	0.01
17	0.10	0.14	0.10	0.06	0.06	0.08	0.04	0.02	0.02	0.01	0.01	0.02	0.01	0.05	0.01	0.01	0.01	0.05	0.09	0.11
18	0.06	0.07	0.11	0.05	0.04	0.05	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.03	0.01	0.01	0.01	0.11	0.17	0.15
19	0.04	0.09	0.07	0.03	0.02	0.04	0.03	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.01	0.09	0.16	0.23

3.5.1 Output Metrics

Section 3.3 gives the required metrics. In order to compute the required metrics the following output metrics must be obtained from the simulation.

1. Total Calls Generated

2. Blocked Calls/Priority
3. Completed Calls/Priority
4. Preempted Calls/Priority
5. Preempted Bandwidth/Priority
6. Preempted Network Bandwidth/Priority

3.6 Statistical Precision

Simulation exhibits random variability when random number generators are used to produce the values of the input variables [BaC96]. Necessary steps must be taken to achieve a certain statistical precision. This precision can be achieved by the warm-up period, the length of each simulation run and the number of independent replications. According to Banks and Carson [BaC96], the larger the length of the simulation run, the smaller the initialization bias. Similarly, the bias in the estimator variability can be reduced with a larger number of independent replications.

The warm-up time is required to reduce bias caused by initial transients. The end of the warm-up time is determined by observing the time when the nodes and trunks start to become congested. The length of each simulation run should be at least ten times the warm-up period [BaC96]. Based on the graph of the node monitor in Figure 3.23, the warm-up time is approximately 80 seconds. This means that each simulation should run for at least 800 seconds. Each simulation configuration ran for 1200 seconds with a traffic generation time of 1000 seconds.

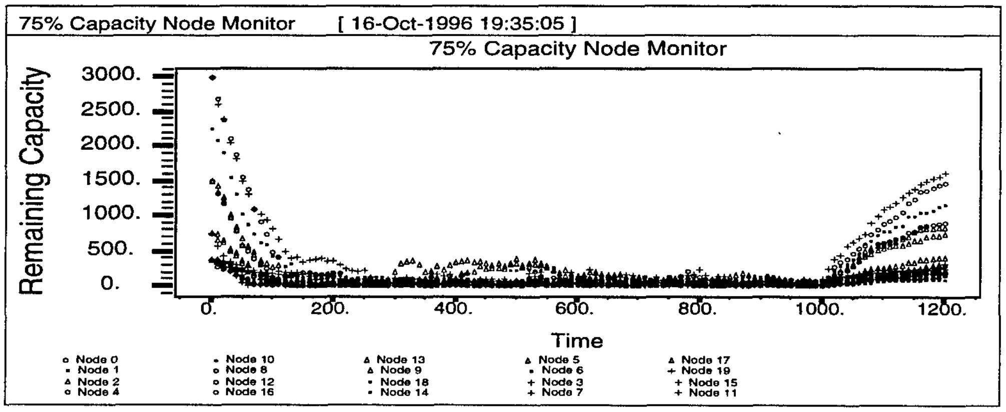

Figure 3.23 Node Monitor

The next decision is to determine the number of replications. The goal is to attain a statistical precision within +/- 0.02 with a 95% confidence. Since the number of replications is small, we can assume that the mean is distributed as a student-t distribution. Therefore 100(1-∝) % confidence interval for the mean is

$$\bar{Y} - t_{\alpha/2}\frac{S}{\sqrt{R}} < \theta < \bar{Y} + t_{\alpha/2}\frac{S}{\sqrt{R}} \tag{4}$$

where $\bar{Y}$ is the sample mean, θ is the true mean, S is the sample standard deviation, R is the number of replications and $t_{\alpha/2}$ is the critical value of the standard distribution with (R-1) degrees of freedom [BaC96]. The number of replications is determined by the equation

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\varepsilon} \right)^2 \tag{5}$$

where $z_{\alpha/2}$ is the half-length cumulative normal distribution, S_0 is the initial sample standard deviation, and ε is the error criterion [BaC96]. Based on the equation R, is calculated to be 30 runs.

3.7 Verification

Verification is concerned with building the model right. The conceptual model must be accurately represented by the computerized system [BaC96]. Verification of the preemption simulation models was accomplished by Designer Module Block Verification, Designer's Interactive run controller, and closely examining the model's output for accuracy under a variety of settings of the input parameters.

3.7.1 Designer Module Block Verification

Each block was built and tested at the lowest level before building upward. Each level of a block was verified as it builds up to the system level. This verification process ensures that all dependencies and block construction were correct. Testing each block was accomplished by placing probes at the output and comparing the output data with the expected output. For example, placing four probes at the output of the priority generator block verified the correct priority distribution was inserted in the call request data structure.

3.7.2 Designer Interactive Run Controller

The simulation was monitored as it progresses using the Interactive Run Controller (IRC). The IRC verified that the correct path was taken within modules and throughout the system. Any warnings or errors encountered were quickly resolved with the IRC's help.

3.7.3 Accuracy of Output Data

No matter how good the design may be, if the output does not make sense, the entire effort will be wasted. For this reason, all output was checked for accuracy. The nodes and trunks were monitored to make sure they become congested so preemption can take place. The total number of preemption per priority was analyzed for correctness. If a higher percentage of priority-2 calls was preempted than priority-0 calls, then the output would not make sense.

3.8 Validation

Validation is concerned with building the right model [BaC96]. Validation tests can be accomplished by comparing the output results with expert intuition, measurements from real systems, and theoretical results [Jai91]. Validation of the models by comparing to real systems cannot be accomplished, since no data is available from a real system with comparable configuration. Similarly, no theoretical results exist that models the performance of these configurations. The validation of the models is determined by using a widely used three-step approach [NaF67]. The three steps are building a model with high face validity, validating model assumptions, and validating the input-output transforms. The first two steps can be performed on these models, but the last is impossible since this system does not exist. Instead, the input parameters were validated and the output results compared with previously accomplished experiments.

3.8.1 Face Validity

Building a model with high face validity involves constructing a model that appears reasonable on its face to model users and others who are knowledgeable about the system being simulated [BaC96]. These simulation models were build with high face validity. The network topology is very similar to that used in previous studies [GaG92, PeM94]. The bandwidth capacities of the nodes used reflect the actual capacities of IDNX transmission resource managers. The trunk capacities were also used in previous research [Pey94]. The choice of four priority classes is typical of real systems. IDNX nodes have a priority service and also provide a preemption option. The choice of an exponential distributed call generation closely reflects call generation in real systems, and also used extensively in previous research.

Sensitivity analysis can also be used to check a model's face validity. Sensitivity analysis involves changing the input variables and observing the changes in the output. Decreasing the inter-arrival time of the call requests resulted in expected increase in the total calls generated, preempted, blocked and connected. An increase in the inter-arrival times produced the opposite effect.

3.8.2 Validating Model Assumptions

Model assumptions encompass structural and data assumptions. Structural assumptions involve questions of how the system operates while data assumptions involve using the correct input data. The model assumptions of the systems being modeled closely resembles those systems found in previously published related works [Hsi90] [GaG92] [LaM94] [Pey94]. These performance studies were based on valid structural and data assumptions.

3.8.3 Validating Input Parameters

The input parameters used in the simulations were consistent with those used in previous studies [LeM94, PeM94, TaH90]. The generation of call requests follows an exponential distribution. The number of nodes was 20. The number is slightly more than the 16 used by Mohammed Pyravian [Pey94]. The use of only four priorities, average call holding time of 200 seconds per priority, and uniformly distributed bandwidth was used before in published work [Pey94]. The topology chosen was a scaled up version of the 16 node model previously mentioned. This topology represents how an IDNX communication network could evolve over time. Node capacities reflect the actual capacities of real IDNX systems [NeT91]. Parameters such as call generation factors, destination distribution, and alternate routing are typical parameters used in BONES Designer manual [ClR94].

3.8.4 Validation of Output Results

The output results were compared with similar results from previous research [Pey94]. The conditions used in the previous research could not be entirely duplicated due to insufficient information. However, the results such as grade of service and the average number of preemption were similar. For example, Mohammed Peyravian's findings of the grade of service for priority 3, 2 and 1 calls were 0.00, 0.034 and 0.288 respectively. The grades of service obtained in this research were 0.00, 0.024 and 0.333 for priorities 3, 2 and 1 calls respectively. Similarly, Mohammed Peyravian's average number of preemption was 1.135 compared to 1.189 for this research.

3.9 Summary

This chapter presented a methodology that can be used to analyze the performance of preemption algorithms in an IDNX communication network. Four separate preemption models were constructed and analyzed. These models were based on uniformly distributed bandwidth, constant bandwidth, network bandwidth and uniformly distributed network bandwidth. Each model was analyzed using a number of measures. Grade of service, number of preemption, and network bandwidth preempted were the measures. The parameters used in the simulation ensured sufficient output data were collected. The network models were designed in modular fashion using DESIGNER. The models were verified and validated.

4 Results and Analysis

4.1 Introduction

This chapter shows the results of the analysis of the IDNX circuit switch communications preemption models. First, there is a presentation on the quantitative description of the call requests generated by the network. Secondly, the network loading capacity is determined for all four configuration models. The capacity of the network and call requests generated in node 0 are allowed to vary to see how they affect the grade of service for each priority class. Plots showing the effect on the grade of service for each configuration are provided. After the network loading is set, the results for each configuration are determined. Tables listing the grade of service, the average number of preemptions, average bandwidth preempted, and average network bandwidth preempted are also provided. The grades of service of the variable bandwidth and constant bandwidth configurations without preemption are determined. Finally, the results of the configuration models are compared with one another.

4.2 Traffic Generation

Before communications begin in a circuit switch communications network, a physical connection between source and destination must be established. Call requests traveling from the source to the destination set up these physical connections. Each call request is generated in the traffic generator of its respective source node. The generated call requests are based on fixed distributions described in Section 3.5. These distributions are the same for all configurations. The bandwidth generator for the constant bandwidth models differs from the variable bandwidth models. The constant bandwidth generator sets the bandwidth of each call request at 768 Kbps. However, the variable bandwidth generator sets the bandwidth of the call requests uniformly between 64 Kbps and 1.536 Mbps. Within the traffic generator, there are five processes of interest. These processes are source, destination, priority, bandwidth, and call holding time generation.

4.2.3 Source Generation

The source generator produced the call requests for all nodes. Call requests are produced exponentially based on each node's inter-arrival time. An increase in the inter-arrival time produces a decrease in the number of calls generated and a decrease produces the opposite effect. Call requests are generated from each of the 20 nodes. The number of calls generated by each source node is the same for all configurations. Figure 4.1 is a histogram showing the total number of call requests generated per node for the length of the simulation. Larger capacity nodes generate more call requests compared to nodes with smaller capacities. Nodes 0 and 3 generated the most call requests while, nodes 7, 13, 19 generated the least call requests.

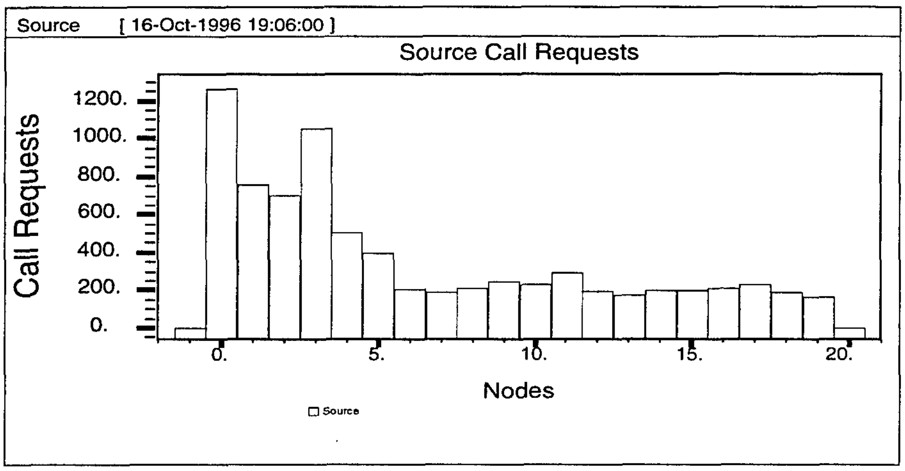

Figure 4.1 Call Requests Leaving Source Node

4.2.2 Destination Generation

The destination generator determines the destination for each call request. The destinations are based on the destination distribution matrix shown in Table 3.5. The destination distribution is the same for all simulation configuration models. Just like the source nodes, larger capacity nodes generally receive more call requests than the smaller capacity nodes. The largest nodes are Node 0 and Node 3. Approximately 1000 calls were destined to these nodes. Figure 4.2 shows the destinations of the call requests per node.

4.2.3 Priority Generation

The priority generator distinguishes the different classes of call requests. There are four separate classes in these simulations. The priority of a call request is based on fixed parameters found in Table 3.6. The total number of call requests generated is 7622. A total of 3433 call requests (40%) had a priority of 0, and 2284 call requests (35%) had a priority of 1. A total of 1523 call request (20%) had a priority of 2, and 382 call requests (5%) had a priority of 3. Figure 4.3 shows the total number of call requests generated per priority.

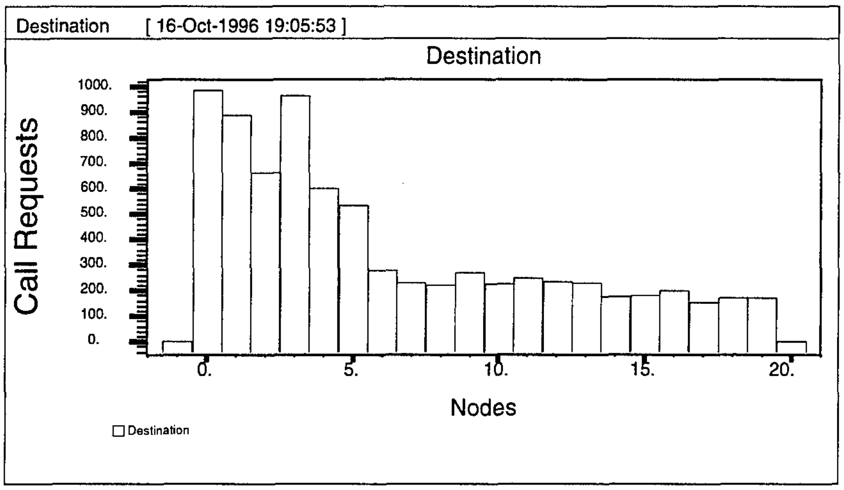

Figure 4.2 Destination of Call Requests

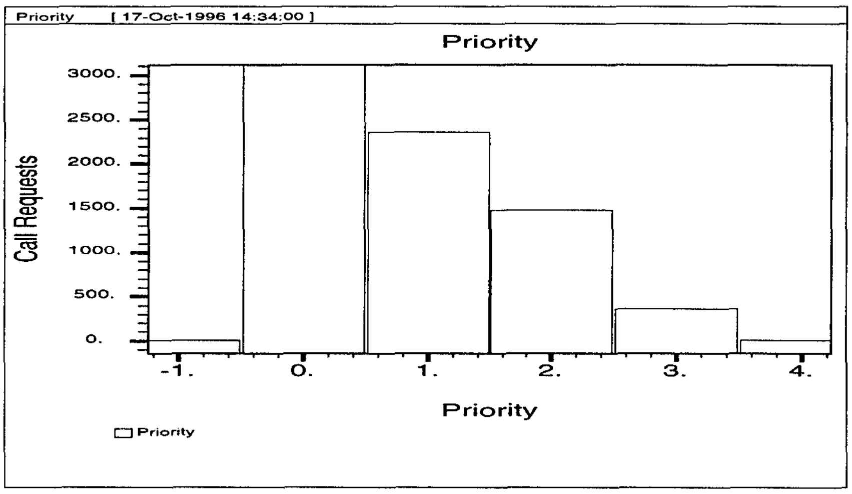

Figure 4.3 Priority Distribution

4.2.4 Bandwidth Generation

The bandwidth of a call represents the capacity of that call. For the constant bandwidth and constant network bandwidth configurations, the bandwidth was fixed at 12 units. For the variable and network bandwidth configurations, the bandwidth for the call requests was generated uniformly between 1 and 24 units. Each unit represents 64 Kbps. Figure 4.4 shows a histogram plot of the bandwidth generated for the uniformly distributed bandwidth configurations. From the figure, approximately the same number of calls was allotted to each bandwidth except for the bandwidth at both ends. Calls with a bandwidth of 1 and 24 received about half the quantity as the others. This reduction at the edges occurs because the uniform number generator is not ideal.

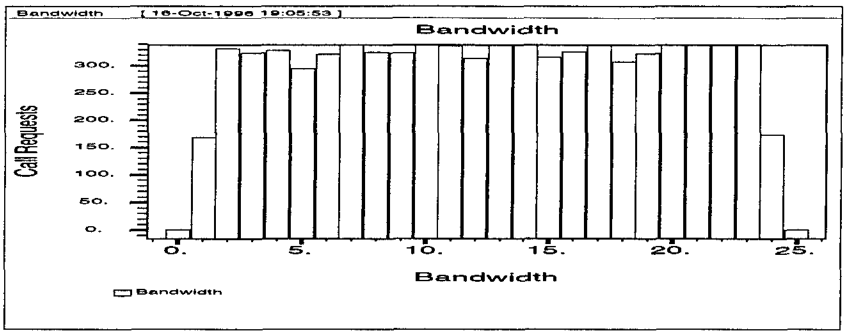

Figure 4.4 Bandwidth for Uniformly Distributed Simulations

4.2.5 Call Holding Time

The holding time for calls is exponentially distributed with a mean of 200 seconds. Figure 4.5 shows the total number of calls in twenty second intervals. As expected, the first interval had the most call requests. It must be noted that some calls were over 1200 seconds long.

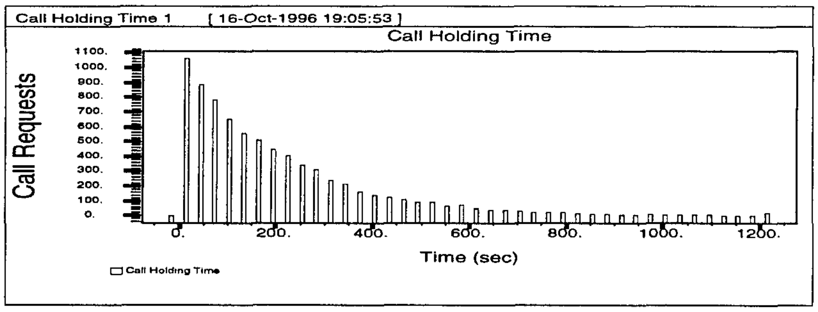

Figure 4.5 Call Holding Time Distribution

4.3 Determination of Network Loading

Under normal conditions, most communications networks have enough capacity to handle customer traffic without the need for preemption. However, in certain peak volume conditions and unexpected loss of network resources, preemption becomes an option. Increase network loading can be simulated by reducing the network capacity or increasing the number of call requests. Preemption only takes place when a node or trunk does not have enough capacity to accommodate a higher priority call. The goal is to reduce the network capacity sufficiently to induce preemption of priority-2 calls but, not too much to cause blockage of priority-3 call requests. Blockage of priority-3 calls prevents the high priority customers from getting the best possible grade of service.

4.3.1 Variable Bandwidth

Network loading for the variable bandwidth configuration is highly dependent on the grade of service of the priority-3 calls. For this reason, the behavior of the grade of service with respect to network capacity must be obtained. The grade of service per priority, for the variable bandwidth configuration, was calculated as the network capacity increases from 0 to 100 percent. The grade of service is the probability that a call will not be successfully completed. It is calculated by dividing the number of generated calls by the number of calls not completed. Figure 4.6 shows the graphs of the grade of service versus percentage network capacity for each priority. Appendix B, Table B-1 shows the same relationship in tabular format.

At 0 percent capacity, the grade of service for each priority is 1.00 or 100 percent. This is expected since all the calls generated are also blocked. As the capacity increases, some calls for each priority become completed and the grade of service for each priority becomes less than 1.00. The rate of decrease in the grade of service is not the same for each priority. Priority-3 calls take advantage of the increase capacity before any other priority. They do so by preempting any other priority call in their source-destination path and utilizing the captured resources. This preemption advantage allows the priority-3 grade of service to reduce drastically before any other priority. The grade of service for priority-3 calls decreases exponentially until it reaches zero at 60 percent network capacity. The grade of service approximately reduces by a factor of 2 for every five percent increase in network capacity. The priority-3 grade of service is dependent only on the network capacity available, not by the number of other priority calls present.

The grade of service for the priority-2 calls also decreases as the capacity increases. The rate of decrease is smaller than that of the priority-3 calls. This is expected since the priority-3 calls preempt the priority-2 calls and the reverse is not true. The priority-2 grade of service depends on the network capacity and the number of priority-3 calls in the system. As the network capacity increases, the priority-2 grade of service steadily decreases. At 100 percent capacity the priority-2 grade of service decreases to 0.005.

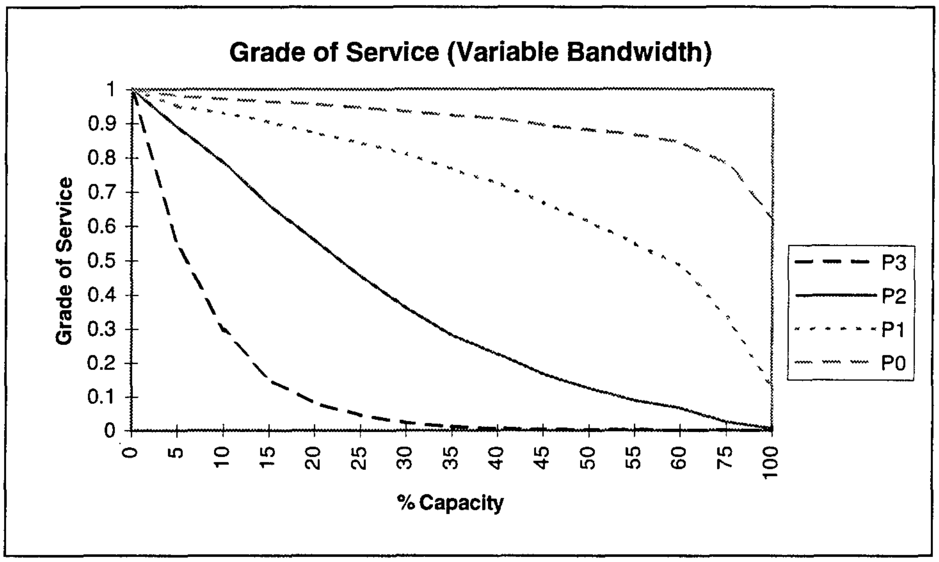

Figure 4.6 Grade of Service per priority(Variable Bandwidth)

Priority-1 and priority-0 grades of service stay high at low network capacity. These priority calls are greatly affected by the number of higher priority calls in the system. As Figure 4.6 shows, only when the grade of service for the higher priorities is very low, a small decrease in network capacity produces a large decrease in the grade of services for the lower priority calls. At 100 percent network capacity, priority-1 grade of service reduces to 0.13 and priority-0 grade of service becomes 0.623.

The graphs in Figure 4.6 indicate that the network capacity must be set at 60 percent or above in order to ensure the highest priority customers receive the best possible communications service. However, if the grade of service for these important customers can be allowed to be as high as 1.1 percent, then the network can meet its goal even if it is degraded down to 35 percent capacity. The network loading for this configuration will be set to 75 percent network capacity. This setting provides the best possible grade of service for the highest priority customers, reflects a realistic network degradation, and provides all the necessary data to compute the desired metrics.

The length of the simulation must be at least 10 times the steady-state time[BaC96]. The steady-state time is the time for the network to be congested. At 75 percent capacity, the network becomes congested at about 90 seconds. The length of the simulation must be at least 900 seconds. Call requests generation lasted for 1000 seconds and the total simulation time was 1200 seconds. Figure 4.7 shows the nodes remaining capacity as the simulation progresses.

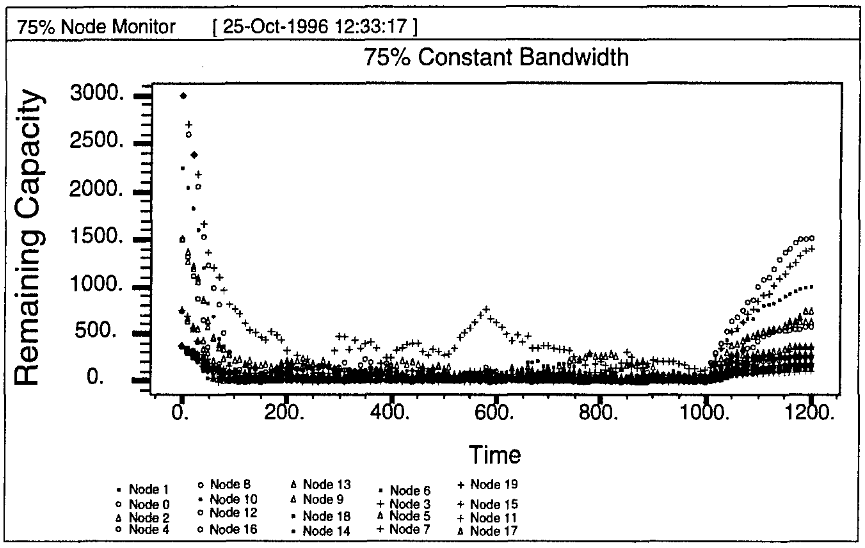

Figure 4.7 Node Monitor (75% Capacity)

In the graphs of Figure 4.6, the number of calls was kept constant. However, what if the calls generated were not constant? How would the grade of service for each priority be affected? What if a crisis has develop and Headquarters (Node 0) needs to increase the amount of traffic to the other nodes? Figure 4.8 shows the grade of service per priority as the amount of traffic generated by node 0 increase from the original setting to 20 times that number. Appendix B, Table B-2 shows the same relationship in tabular format.

When the number of calls in node 0 doubles, there was an immediate increase in the grades of service for the priority-1 and priority-0 calls. At the same time, the grades of service for the two higher priorities were not affected. These changes in the grade of services are expected. The doubling of node 0 traffic increases the number of priority-2 and priority-3 calls in the system thus making it more difficult for the lower priority calls to be completed. As the number of calls in node 0 continues to increase, the rate of increase in the grade of service for the lower priority calls declines and approaches a horizontal line.

The grade of service for the priority-2 calls increases dramatically as the number of calls in node 0 increases past four times its original value. The rate of increase is smaller than the two lower priorities. The priority-2 grade of service approaches the grade of service of the lower priority calls as the number of call increases. If enough calls were generated, it is possible for the three lowest priority grades of service to approach one hundred percent.

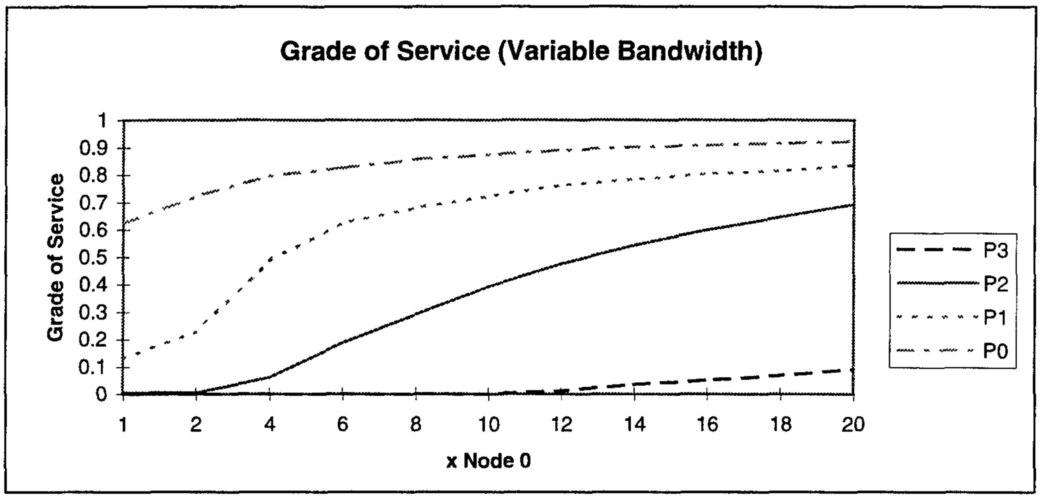

Figure 4.8 Grade of Service per Priority as Node 0 Increases (Variable Bandwidth)

The grade of service for priority-3 calls stayed at 0 up to 10 times the original number of node 0 calls. At 20 times the original number of node 0 calls, the priority-3 grade of service moves up to 0.091. If the network can tolerate a priority-3 grade of service of 5 percent, then the network can accommodate an increase in the node 0 calls of 15 times the normal average number. This way the network still satisfies its most important customers.

4.3.2 Constant Bandwidth

Network loading for the constant bandwidth configuration is also highly dependent on the grade of service of the priority-3 calls. The changes in the grade of service as the network capacity increases also gives insight into the reliability of the network. Figure 4.9 shows the graphs of the grade of service per priority. Appendix B, Table B-3 shows the same relationship in tabular format.

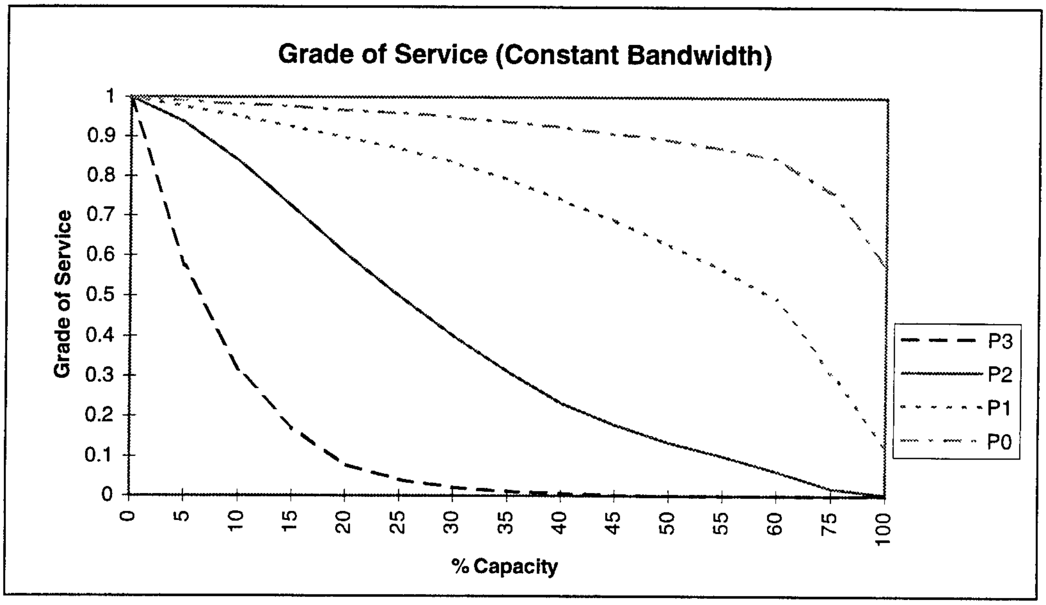

Figure 4.9 Grade of Service per priority (Constant Bandwidth)

Just like the variable bandwidth configuration, the grade of service for each priority is 1.00 at a network capacity of zero. The graphs, in Figure 4.9, look almost identical to those in Figure 4.6. However, the grade of service per priority is slightly smaller when network capacity is high and slightly larger when network capacity is low. For example, the constant bandwidth grade of service for priority 2, 1, and 0 are 0.02, 0.306, and 0.759 compared to 0.024, 0.333, 0.783 for the variable bandwidth at 75% network capacity. However, at 15% network capacity, the constant bandwidth grades of service for priority 3,2,1,0 are 0.173, 0.722, 0.924, and 0.975 compared to 0.154, 0.666, 0.904, and 0.964 for the variable bandwidth configuration. At low network capacity, the remaining capacity in many nodes and trunks are close to zero. If the remaining capacity is less than 12, no constant bandwidth call will be connected without the aid of preemption. Variable bandwidth calls with small bandwidths, on the other hand, are more likely to be connected. At high network capacity, more constant bandwidth calls are likely to be connected than the variable bandwidth calls.

At a network capacity of 55 percent, the grade of service for the priority-3 calls is zero. For the highest priority customers to receive the best possible service, the network capacity must be at 55 percent or above. However, if the grade of service is allowed to be as high as 1.1 percent, then the network can meet its goal even if it is degraded down to 35 percent. The network loading will also be set to 75 percent network capacity. This setting once again provides the best possible grade of service for the highest priority customers, reflects a realistic network degradation, and provides all the necessary data to compute the desired metrics.

At the 75 percent network capacity, the steady-state time is 80 seconds. The length of the simulation must be at least 800 seconds. Just as in the variable bandwidth configuration, the call requests were generated for 1000 seconds and the total simulation lasted for 1200 seconds. Figure 4.10 shows the nodes' remaining capacity for the constant bandwidth configuration as the simulation progresses.

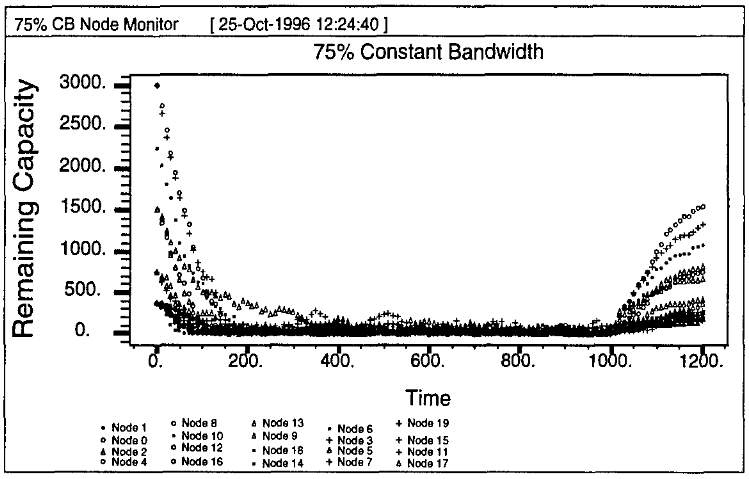

Figure 4.10 Constant Bandwidth Node Monitor (75% Capacity)

A similar scenario of the variable bandwidth configuration was tested. Calls generated at node 0 were allowed to increase and the grade of service per priority was plotted. Figure 4.11 shows the graphs of the grade of service for each priority. Appendix B, Table B-4 shows the same relationship in tabular format.

The graphs of the grade of service per priority are almost identical to that of the variable bandwidth. The grade of service per priority for the constant bandwidth is also slightly smaller when network capacity is high and slightly larger when network capacity is low. Case in point, at the original node 0 number of calls, the constant bandwidth priority-2 grade of service is 0.0034 and the variable bandwidth grade of service is 0.0053. However, at 20 times the original node 0 number of calls priority-2 grade of service is 0.709 while the variable bandwidth grade of service is 0.694. Since the grade of service for priority-2 customers should be very low, this rate of calls generated in node 0 is intolerable.

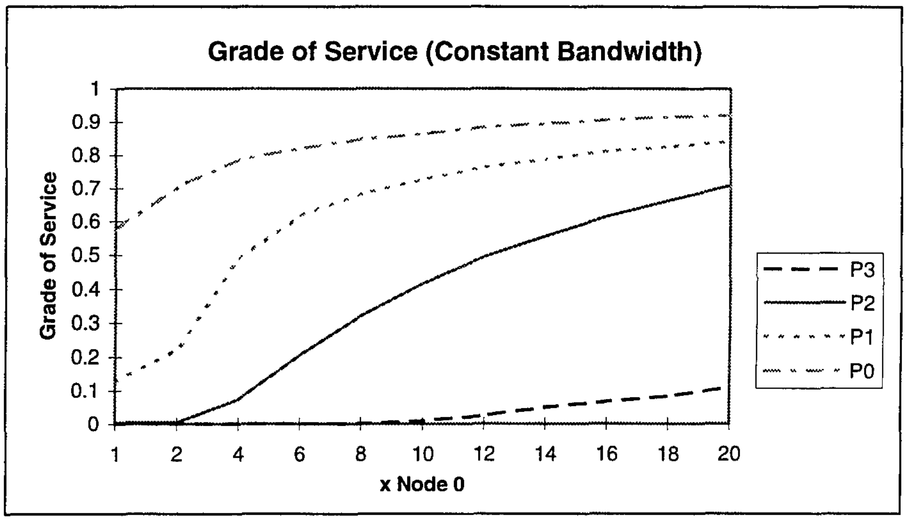

Figure 4.11 Grade of Service per Priority as Node 0 Increases (Constant Bandwidth)

There is no change in the priority-3 grade of service up to an increase of 8 times the original number of calls generated in node 0. If a crisis occurs, the Headquarters can increase its output traffic up to eight times without affecting the highest priority customers' grade of service assuming the traffic at the other nodes remains the same. If a higher priority-3 grade of service is acceptable then even more traffic from node 0 can be tolerated.

4.3.3 Network Bandwidth

As with the other configurations, determination of the network loading requires investigating the grade of service as the network capacity changes. Once again, it is important to have the grade of service of the most important customers be the lowest possible. Some preemption of all lower priorities must occur to insure the proper working of the model. Figure 4.12 shows the graphs of the grade of service for the network bandwidth configuration. Appendix B, Table B-5 shows the same relationship in tabular format.

The graphs depicting the grade of service begin from a grade of service of 1.00 at zero network capacity, and decrease as the capacity increases. The graphs in Figure 4.12 slightly differ from those in Figure 4.6. The similarity is expected, since the only difference in the two configurations is how they preempt lower priority calls. The total number of calls blocked by the network bandwidth configuration and the number of calls blocked by the variable bandwidth configuration is the same. However, the network bandwidth configuration preempted more calls than the variable bandwidth. This explains why the grades of service graphs are slightly higher than those of the variable bandwidth configuration. The network bandwidth configuration preempts 1959 calls and the variable bandwidth configuration preempts only 1826 calls.

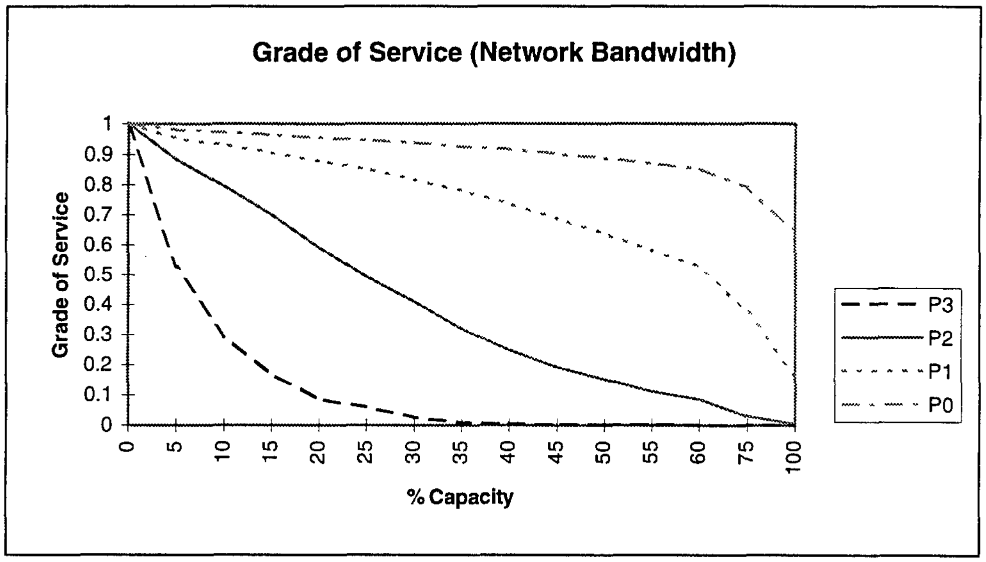

Figure 4.12 Grade of Service per priority (Network Bandwidth)

Just as shown for the variable bandwidth configuration, the priority-3 grade of service is zero at 60 percent capacity. The network loading must also be at 60 percent or above to ensure the priority-3 customers receive the best possible grade of service. However, if a larger grade of service is acceptable, then the network capacity can operate at a lower capacity. The network loading is also the same as the variable bandwidth configuration at 75 percent capacity. The simulations also ran for 1200 seconds.

As the number of calls generated from node 0 increases from its original setting to 20 times that number, the grade of service per priority also behaves like those of the variable bandwidth configuration. Figure 4.13 displays the grade of service per priority for the network bandwidth configuration as node 0 increases. Appendix B, Table B-6 shows the same relationship in tabular format.

The priority-3 grade of service stayed zero only up to 8 times node 0 original calls compared to 10 times for the variable bandwidth configuration. The priority-2 grade of service steadily decreases but, not as fast as the priority-3 grade of service. The grade of service of priority-1 and priority-0 inches downwards until about the 60 percent capacity mark. From this mark, further increase in network capacity produces large decrease in priority 1 and 2 grade of services. If the network can tolerate a 1 percent grade of service for priority-3 customers, the network bandwidth can accept up to 10 times the original calls from node 0.

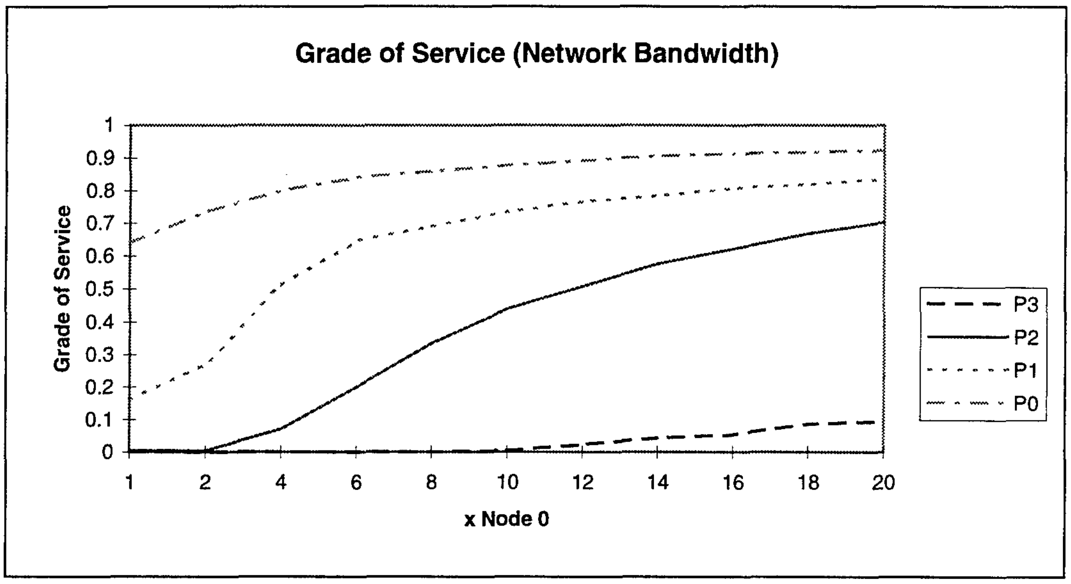

Figure 4.13 Grade of Service per Priority as Node 0 Increases (Network Bandwidth)

4.3.4 Constant Network Bandwidth

The constant network bandwidth configuration network determination is the same as the other configurations. The grade of service for the priority-3 customers must be zero and the some priority-2 call must be preempted. Varying the network capacity while measuring the grade of service for each priority should help to determine a good network loading. Figure 4.14 shows the graphs of the grade of service for the constant network bandwidth configuration. Appendix B, Table B-7 shows the same relationship in tabular format.

The grade of service for the priority-3 calls decreases from 1.00 to 0 at 55 percent network capacity. As the capacity increases, all the grade of service decreases. The higher the priority, the greater the rate of decrease in the grade of service. An examination of the slopes of the grade of service confirms this fact. The rate of decrease in the grade of service depends on the priority of the call and the loading of the network. The priority-3 grade of service graph in Figure 4.14 indicates that the network capacity must be set to 55 percent or above to guarantee the highest possible grade of service for the most important customers. However, if the network can tolerate an 8.5 percent priority-3 grade of service, then the network capacity can drop as low as 35 percent. In keeping with previous decisions, the network capacity was set at 75 percent capacity. The constant bandwidth configuration model steady-state is almost identical to this configuration. Therefore, the simulations also ran for 1200 seconds and the calls were generated for 1000 seconds.

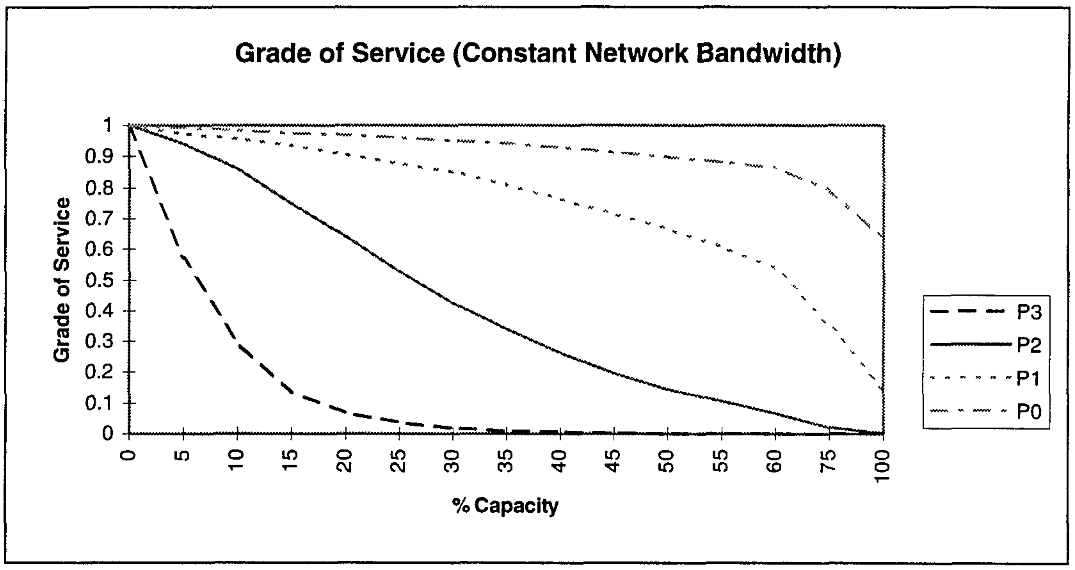

Figure 4.14 Grade of Service per priority (Constant Network Bandwidth)

If the network was at 100 percent capacity and node 0 needed to increase its output traffic as a result of emergency actions, the grade of service of the priority calls will be affected. Figure 4.15 shows the relationship between the grade of service per priority and the increase output of node 0. Appendix B, Table B-8 shows the same relationship in tabular format.

The graphs of the grade of service for each priority behave in a similar manner to all the other configurations. Similar to the other constant bandwidth, the priority-3 grade of service is not affected up to 8 times node 0 original number of calls. At 20 times node 0 calls, priority-3 grade of service becomes 0.1. As the calls in node 0 increases, priority-2 grade of service moves from 0.003 at the original node 0 calls to 0.734 at 20 times the original number of calls. This experiment showed that the network can operate at 8 times the original number of calls generated in node 0 without affecting the highest priority customers' grade of service.

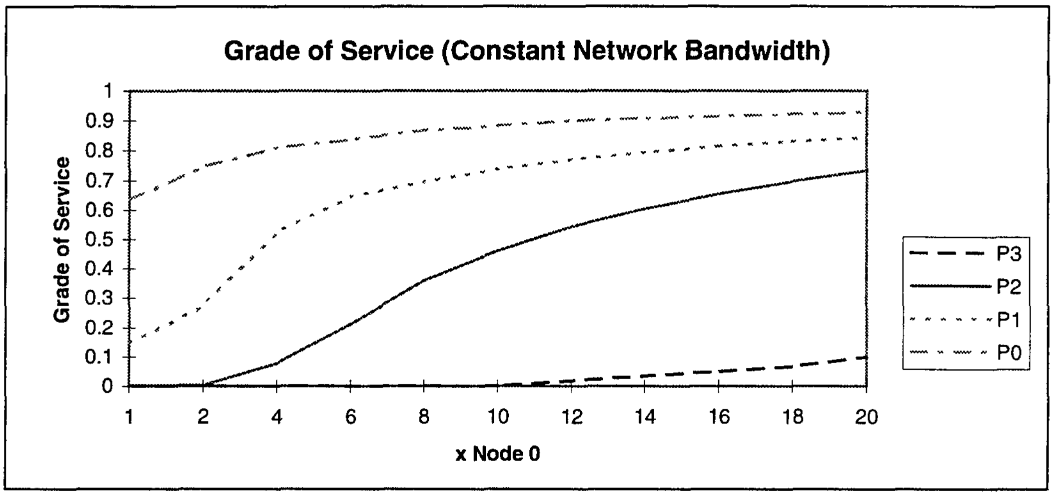

Figure 4.15 Grade of Service per Priority as Node 0 Increases(Constant Network Bandwidth)

4.4 Simulation Results

The results of the simulations are based on the grade of service per priority, average number of preemptions, average bandwidth preempted and average network bandwidth preempted. There is a discussion below of the results for each configuration.

4.4 1 Variable Bandwidth Configuration

The results in this experiment are based on grade of service per priority, average number of preemptions per priority, average bandwidth preempted, and average network bandwidth preempted. Table 4.1 lists the grade of service per priority and the average number of preemptions per priority. The numbers in the table are the average values of 30 different iterations. The calls are rounded to the nearest whole number. The grade of service and average number of preemptions were calculated to be within the specified range with a 95% confidence interval.

Table 4.1 Variable Bandwidth Simulation Results

Priority	Generated Calls	Blocked Calls	Preempted Calls	Preempting Calls	Grade of Service	Average Number of Preemption
0	3433	1242	1444	1202	0.783 ± 0.016	1.202 ± 0.02
1	2284	386	375	328	0.333 ± 0.024	1.145 ± 0.035
2	1523	30	7	7	0.024 ± 0.005	1.014 ± 0.039
3	382	0	0	0	0.00 ± 0	0.00 ± 0
Totals	7622	1658	1826	1537		
Overall Average					0.457	1.189

The grade of service is the probability that a call request would be blocked or preempted. In this simulation, the probability of a call not being completed decreases as the priority increases. This is expected since higher priority calls are allowed to preempt lower priority calls. The grade of service for priority-3 calls is zero. The goal of the simulations was to provide the best possible grade of service to the highest priority customers. The priority-2 calls are almost as important as the priority-3 calls. These calls should also have a very low grade of service. Only priority-3 calls can preempt priority-2 calls, and priority-2 calls can preempt the two lower priorities. For the reasons above, priority-2 grade of service is very low. The grade of service for the priority-2 calls is calculated to be 0.024± 0.005. The grade of service for the priority-1 calls is within the expected range. The priority-1 calls are important but not as important as priority-2 and priority-3 calls. The priority-1 calls can be preempted by up to 25 percent of the total calls generated. However, these calls can preempt up to 45 percent of the total calls. The reasons above explain why the grade of service of the priority-1 calls, lies between the grade of service of the priority-2 calls, and the grade of service of the priority-0 calls. The priority-0 calls are lowest priority calls. They are not allowed to preempt any other calls. If successfully connected, priority-0 calls can be preempted by all other higher priority calls. This means that up to 55 percent of the total calls generated can preempt priority-0 calls. The grade of service for the priority-0 calls is expected to be the highest of all the calls, and the results confirm that fact. The grades of service for priorities 1, 2, and 3 are close to the results obtained by Mohammad Peyravian [Pey94]. The grades of service for priorities 1, 2, and 3 were 0.333, 0.024, and 0.00 versus 0.295, 0.038 and 0.00 obtained by Mohammad Peyravian [Pey94]. The grade of service for priority-0, is almost twice as much as that obtained by Peyravian 0.783 compared to 0.375. The difference in the two results could depend on the network loading and the fact that Peyravian rerouted all preempted calls. Rerouting preempted calls helps to improve the grade of service for all priorities. Since most calls preempted are priority-0, rerouting would produce the biggest effect on priority-0 grade of service.

The average number of preemption is the number of preempted calls divided by the total number of preempting calls. This value will be greater than or equal to one. This is because each preempting call preempts at least one call. This configuration preempts the largest bandwidth first. For that reason, the average number of preemptions is expected to be close to 1. The average number of preemptions is 1.189. This number is very close to previously published work [Pey94]. Mohammad Peyravian's experiments obtained an average number of preemptions of 1.135. Table 4.1 shows the average number of preemptions per priority.

Since this configuration preempts the largest bandwidth first, the average bandwidth preempted is expected to be greater than the average bandwidth of all the call requests. The average bandwidth of the call requests is 12.25, where each point represents 64 Kbps. The average bandwidth preempted is 16.626 and the average network bandwidth preempted is 21.54. The average network bandwidth is higher than the average bandwidth preempted. This is expected, since network bandwidth is the number of hops times the bandwidth and calls with the largest bandwidth were preempted first. Table 4.2 list the average bandwidth preempted and the average network bandwidth preempted.

Table 4.2 Preempted Bandwidth and Network Bandwidth Results (Variable Bandwidth)

Preempted Bandwidth	Preempted Network Bandwidth	Preempting Calls	Average Bandwidth Preempted	Average Network Bandwidth Preempted
22612	29293	1360	16.626 ± 0.8	21.54 ± 0.9

4.4.2 Constant Bandwidth

The results in this experiment are based on grade of service per priority and average network bandwidth preempted. Table 4.3 shows the grade of service per priority. Once again, the numbers in the table are the average values of 30 different iterations. The calls are rounded to the nearest whole number, and the grade of service for each priority is calculated to be within the specified range with a 95% confidence interval.

Table 4.3 Constant Bandwidth Simulation Results

Priority	Generated Calls	Blocked Calls	Preempted Calls	Preempting Calls	Grade of Service	Average Number of Preemption
0	3433	1383	1222	1222	0.759 ±0.018	1.000
1	2284	386	312	312	0.306 ± 0.018	1.000
2	1523	24	7	7	0.020 ± 0.008	1.000
3	382	0	0	0	0.00 ± 0	1.000
Totals	7622	1793	1541	1541		
Overall Average					0.437	1.000

The grade of service per priority is almost the same as those of the variable bandwidth configuration. In fact, the grade of service is slightly smaller in every priority except priority-3, which cannot get any smaller. The difference between the two configurations' grades of service is due to the difference in the average bandwidth and the average number of preemptions. The average bandwidth of the constant bandwidth configuration is 12, where each point represents 64 Kbps. The variable bandwidth configuration is 12.25. The lower average bandwidth effectively lowers the network loading. The lower network loading enables fewer connected calls to be preempted. A higher priority preempting call only needs to preempt one call to free up enough bandwidth for it to gain connection. This one-to-one ratio produces a lower number of preempted calls. Fewer preempted calls help to lower the grade of service per priority. No previously published research with constant bandwidth simulation results was found.

Lower priority calls are preempted on a first-come-first-serve basis. Since the bandwidth is constant, the average bandwidth preempted is the same as the bandwidth of each call. Table 4.4 shows the average network bandwidth preempted for the constant bandwidth configuration. Just as shown for the variable bandwidth configuration, the average network bandwidth preempted is expected to be more than the average bandwidth preempted. The same reason applies as above.

Table 4.4 Preempted Bandwidth and Network Bandwidth Results Constant Bandwidth)

Preempted Bandwidth	Preempted Network Bandwidth	Preempting Calls	Average Bandwidth Preempted	Average Network Bandwidth Preempted
18489	24732	1541	12 ± 0	16.05 ± 0.75

4.4.3 Network Bandwidth

In the network bandwidth configuration, preempting calls first seek to preempt the lowest priority call with the smallest network bandwidth. The results are based on the grade of service per priority, average number of preemptions per priority, average bandwidth preempted, and average network bandwidth preempted. Table 4.5 lists the grade of service per priority and the average number of preemptions per priority. The grade of service and average number of preemptions were calculated to be within the specified range with a 95% confidence interval.

Table 4.5 Variable Network Bandwidth Simulation Results

Priority	Generated Calls	Blocked Calls	Preempted Calls	Preempting Calls	Grade of Service	Average Number of Preemption
0	3433	1233	1485	1171	0.792± 0.012	1.268± 0.013
1	2284	391	462	330	0.373 ± 0.025	1.397 ± 0.026
2	1523	34	12	8	0.030 ± 0.008	1.390 ± 0.184
3	382	0	0	0	0.00 ± 0	0.00 ± 0
Totals	7622	1658	1959	1509		
Overall Average					0.475	1.297

The grade of service is similar to the previous two configurations. The probability of a call not being completed also decreases as the priority increases. By preempting the lowest network bandwidth first, more calls are preempted per priority. The increase in the number of preempted calls causes the grade of service per priority to increase slightly.

This configuration preempts an average of about 1.3 calls per preempting higher priority call. This number is higher than that of the variable bandwidth configuration. This increase in the average number of preemptions stems from the fact that calls with the lowest network bandwidth is preempted first in order to keep network disruption to a minimum. Since lower priority calls with the largest bandwidth are not preempted first, more calls must be preempted to connect the preempting higher priority calls.

Unlike the previous two configurations, the average network bandwidth preempted is smaller than the average bandwidth preempted. Calls with the same source and destination have a network bandwidth of zero. These calls are the first to be preempted and they help to reduce the average network bandwidth. Table 4.6 lists the average bandwidth preempted and the average network bandwidth preempted.

Table 4.6 Preempted Bandwidth and Network Bandwidth Results (Network Bandwidth)

Preempted Bandwidth	Preempted Network Bandwidth	Preempting Calls	Average Bandwidth Preempted	Average Network Bandwidth Preempted
16777	12964	1330	16.22 ± 0.4	9.747 ± 0.3

4.4.4 Constant Network Bandwidth

In the constant network bandwidth configuration, all calls have the same bandwidth. Preempting calls first seek to preempt the lowest priority call with the smallest network bandwidth. The results are based on grade of service per priority and average network bandwidth preempted. Table 4.7 contains the grade of service per priority. Just like the other configurations, the grade of service was calculated to be within the specified range (refer to Table 4.7) with a 95% confidence interval. The grade of service is lower than the network bandwidth configuration. Just as the constant bandwidth configuration, the lower average network loading and the lower number of preempted calls help to reduce the grade of service in comparison to the network bandwidth configuration.

Just like the network bandwidth configuration, the average network bandwidth preempted is smaller than the average bandwidth preempted. Table 4.8 lists the bandwidth and network bandwidth preempted for the constant network bandwidth configuration. The constant bandwidth allows more blocked calls and less preempted calls.

Table 4.7 Constant Network Bandwidth Simulation Results

Priority	Generated Calls	Blocked Calls	Preempted Calls	Preempting Calls	Grade of Service	Average Number of Preemption
0	3433	1466	1257	1257	0.793 ± 0.023	1.000
1	2284	436	364	364	0.350 ± 0.015	1.000
2	1523	25	6	6	0.020 ± 0.004	1.000
3	382	0	0	0	0.00 ± 0	1.000
Totals	7622	1927	1627	1627		
Overall Average					0.466	1.000

Table 4.8 Preempted Bandwidth and Network Bandwidth Results (Constant Network Bandwidth)

Preempted Bandwidth	Preempted Network Bandwidth	Preempting Calls	Average Bandwidth Preempted	Average Network Bandwidth Preempted
19517	12028	1627	12.00 ± 0	7.393 ± 0.16

Preempting the lowest network bandwidth first, allows the model to preempt calls with no network bandwidth first. Calls with no network bandwidth keep the total network bandwidth preempted low and the large number of preempting calls effective reduces the average value. Constant bandwidth configurations are less flexible and produce more blocked calls.

4.4.5 Priority Without Preemption

The variable bandwidth and the constant bandwidth communication models without preemption were simulated to compare their grades of service and average number of preemptions. Table 4.9 shows the results for the variable bandwidth.

The grade of service is statistically the same for each priority. This happens as no preferential treatment occurs in this network. Since no preferential treatment occurs, each priority class is connected with the same probability. The total number of calls generated per node is too low for the nodal queues to have more than one call in queue at any one time. Figure 4.16 is the maximum queue occupancy. This figure confirms the fact that a maximum of one call is in queue at any one time. From Figure 4.16, only nodes 0, 1 and 2 had any calls in their queues.

Table 4.9 Variable Bandwidth (No Preemption)

Priority	Generated Calls	Blocked Calls	Grade of Service
0	3433	1333	0.388 ± 0.008
1	2284	887	0.389 ± 0.013
2	1523	593	0.389 ± 0.02
3	382	149	0.389 ± 0.03
Totals	7622	2962	
Overall Average			0.389

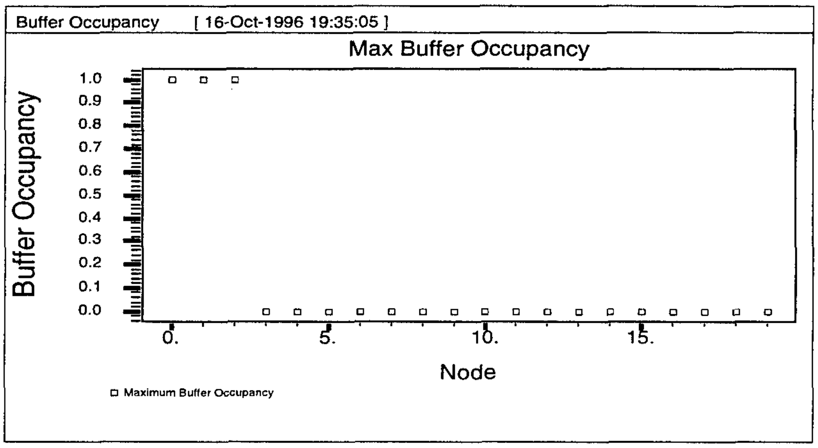

Figure 4.16 Maximum Nodal Queue Occupancy

The average nodal delay also confirms the fact that there is little queuing in the nodes. Nodes 0, 1, and 2 are the only nodes that have any delay. This is due to the fact that they are the only nodes that have any queuing (refer to Figure 4.16). Figure 4.17 shows the average nodal delay for all the nodes. The processing time for a call through each node is normally distributed with a mean of 250 and a variance of 25 microseconds.

Without preemption, preferential treatment can only occur when the different priority calls are competing for connection. Since the time to connect a call is very short, average of 250 microseconds, only in large capacity networks or at low bandwidth, such as a single telephone call, can priority in a circuit switch network provide a very good grade of service for its important customers.

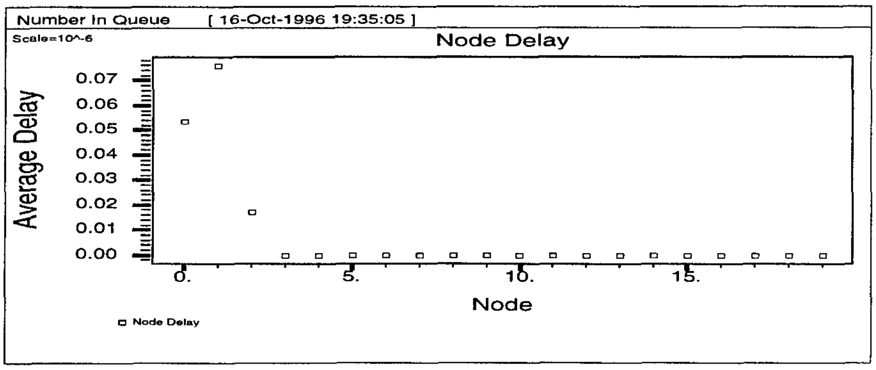

Figure 4.17 Average Nodal Delay

The constant bandwidth model without preemption results in a similar grade of service for each priority. Table 4.10 list the grade of service for the constant bandwidth with no preemption.

Table 4.10 Constant Bandwidth (No Preemption)

Priority	Generated Calls	Blocked Calls	Grade of Service
0	3433	1525	0.444 ± 0.01
1	2284	1011	0.443 ± 0.008
2	1523	679	0.446 ± 0.008
3	382	172	0.449 ± 0.015
Totals	7622	3387	
Overall Average			0.444

The constant bandwidth grade of service was worse than those of variable bandwidth. This happened because the constant bandwidth blocks more calls than the variable bandwidth, and thus a larger grade of service (refer to Table 4.9 and Table 4.10). When the capacity of a node or trunk is less than 12, no calls are connected in this constant bandwidth model. However, calls with a capacity of less than 12 in the variable bandwidth model are connected. This flexibility enables the variable bandwidth configuration to have a lower grade of service. The grade of service is also statistically the same for each priority. Once again, the total number of calls generated per node is too low for the nodal queues to have more than one call in queue at one time (refer to Figures 4.16 and 4.17).

4.5 Configuration Comparisons

This section compares the results of the configurations' models. The comparison is based on the performance metrics highlighted in Section 3.3. These metrics are grade of service, average number of preemption, and average network bandwidth preempted.

4.5.1 Variable Bandwidth Vs Constant Bandwidth

The variable bandwidth simulation model can be viewed as a communications network that accepts traffic with bandwidth uniformly distributed between 64 Kbps and 1.536 Mbps. Similarly, the constant bandwidth simulation model can be viewed as a network where the bandwidth of each call is 768 Kbps. The variable bandwidth and the constant bandwidth configurations are compared using the grade of service and the average number of preemptions.

The grade of service for each priority of the constant bandwidth model is lower than the variable bandwidth model. Table 4.11 shows a comparison of the grade of service and average number of preemptions for both configurations.

Table 4.11 Variable Bandwidth & Constant Bandwidth Metric Comparison

Priority	Variable Bandwidth		Constant Bandwidth	
	Grade of Service	Average Number of Preemption	Grade of Service	Average Number of Preemption
0	0.783 ± 0.016	1.202 ± 0.020	0.759 ±0.018	1.000
1	0.333 ± 0.024	1.145 ± 0.035	0.306 ± 0.018	1.000
2	0.024 ± 0.005	1.014 ± 0.039	0.020 ± 0.008	1.000
3	0.00 ± 0	0.00 ± 0	0.00 ± 0	1.000
Overall Average	0.457	1.189	0.437	1.000

The grade of service per priority was lower for the constant bandwidth model because an extra 81 priority-0, 63 priority-1, and 6 priority-2 more calls were completed. The difference in the grade of service for the two simulation models could be explained by the difference in their average bandwidth. The average bandwidth for the variable bandwidth configuration is 12.25 while that of the constant bandwidth configuration is 12. The loading of the variable bandwidth model is 2.08 percent more than the constant bandwidth configuration. If the loading of the variable bandwidth was the same as the constant bandwidth configuration, the variable bandwidth would have completed about 86 more calls. If the average bandwidth was the same, the overall grade of service would be 0.446 for the variable bandwidth and 0.437 for the constant bandwidth. These values are very close and within a 2% uncertainty.

The average number of preemptions per priority for the constant bandwidth model is 1.00. On the other hand, the average number of preemptions per priority for the variable bandwidth configuration is 1.189. The lower average number of preemptions for the constant bandwidth configuration results in a lower number of calls been preempted. The constant bandwidth configuration disrupts the network less than the variable bandwidth configuration.

4.5.2 Variable Bandwidth Vs Network Bandwidth

Lower priority calls with the largest bandwidth are preempted first in the variable bandwidth model. On the other hand, lower priority calls with the lowest network bandwidth are preempted first in the network bandwidth model. The variable bandwidth and the network bandwidth models are compared using the grade of service, the average number of preemptions, and network bandwidth preempted. Table 4.12 compares the grade of service per priority and average number of preemptions for the two configurations.

The grade of service per priority for the variable bandwidth model is lower than that of the network bandwidth model. The number of blocked calls for both models is approximately the same. However, the number of preemptions each priority is larger with network bandwidth model. Since the network bandwidth model preempts the smallest network bandwidth first, more calls must be preempted in order to accommodate higher priority calls. The larger number of preemptions results in a larger grade of service per priority.

The average number of preemptions for the variable bandwidth model is 1.189 and that of the network bandwidth model is 1.297. The higher average number of preemptions in the network bandwidth model results from a larger number of preemptions compared to the variable bandwidth model. The larger number of preemptions occurs in the network bandwidth model because this configuration first preempts the least network bandwidth. Therefore this configuration would need to preempt more calls in order to connect the higher priority calls.

Table 4.12 Variable Bandwidth & Network Bandwidth Metric Comparison

Priority	Variable Bandwidth		Network Bandwidth	
	Grade of Service	Average Number of Preemption	Grade of Service	Average Number of Preemption
0	0.783 ± 0.016	1.202 ± 0.020	0.792 ±0.012	1.268 ± 0.013
1	0.333 ± 0.024	1.145 ± 0.035	0.373 ± 0.025	1.397 ± 0.026
2	0.024 ± 0.005	1.014 ± 0.039	0.030 ± 0.008	1.390± 0.184
3	0.00 ± 0	0.00 ± 0	0.00 ± 0	0.00 ± 0
Overall Average	0.457	1.189	0.437	1.297

The average preempted network bandwidth for the variable bandwidth model is 21.54 while that of the network bandwidth model is 9.75. Network bandwidth is defined as the bandwidth of a call times the number of hops the call makes. A call with the same source and destination has a zero network bandwidth. If a communications network is interested in minimizing the amount of bandwidth preempted throughout the network, the network bandwidth configuration would better accomplish this task than the variable bandwidth configuration. On the other hand, the variable bandwidth configuration is better if the grade of service and the average number of preemptions are the only determinants.

4.5.3 Variable Bandwidth Vs Constant Network Bandwidth

The constant network bandwidth model is a hybrid of the constant bandwidth and network bandwidth models. The bandwidth throughout the network is constant and the model preempts the lowest network bandwidth first. The variable bandwidth and the constant network bandwidth are compared using grade of service and average network bandwidth preempted.

The overall grade of service for the constant network bandwidth is larger than that of the variable bandwidth. The constant network bandwidth grade of service is 0.466 compared to 0.457 for the variable bandwidth. The grade of service per priority is not consistently higher for any of the two configurations. Table 4.13 compares the grade of service for both configurations. Priority-0 and priority-1 grades of service are higher in the constant network bandwidth model. However, priority-3 grade of service is lower. In the constant network bandwidth model, more of the lower priority calls are preempted due to the preemption method. On the other hand, fewer of the higher priority calls are preempted due to the lower average bandwidth. Since the variable bandwidth configuration is more flexible and produces a better overall grade of service, the better choice is the variable bandwidth configuration model.

Table 4.13 Variable Bandwidth vs Constant Network Bandwidth

Priority	Variable Bandwidth	Constant Network Bandwidth
	Grade of Service	Grade of Service
0	0.783 ± 0.016	0.793 ±0.023
1	0.333 ± 0.024	0.350 ± 0.015
2	0.024 ± 0.005	0.020 ± 0.004
3	0.00 ± 0	0.00 ± 0
Overall Average	0.457	0.466

The average network bandwidth preempted for the variable bandwidth model is 21.54 compared to 7.393 for the constant network bandwidth. The difference between these two values is indicative of the method of preemptions. If a communications network is interested in minimizing the amount of bandwidth preempted throughout the network, the constant network bandwidth configuration would better accomplish this task than the variable bandwidth configuration.

4.5.4 Constant Bandwidth Vs Network Bandwidth

The constant bandwidth preempts calls on a first-come-first-preempted basis. On the other hand, the network bandwidth configuration preempts calls with the lowest network priority. The constant bandwidth and the network bandwidth configurations are compared using the grade of service and network bandwidth preempted.

The grade of service per priority for the constant bandwidth configuration is the lowest of all the configurations. The network bandwidth configuration, on the other hand, has the highest grade of service per priority. Table 4.14 compares the grade of service per priority for the two configurations. The network bandwidth configuration preempts the most calls. This preemption method results in a higher grade of service per priority. If grade of service is the only method of determination, the constant bandwidth configuration is the better option.

Table 4.14 Constant Bandwidth vs Network Bandwidth

Priority	Constant Bandwidth Grade of Service	Network Bandwidth Grade of Service
0	0.759 ±0.018	0.792 ±0.012
1	0.306 ± 0.018	0.373 ± 0.025
2	0.020 ± 0.008	0.030 ± 0.008
3	0.00 ± 0	0.00 ± 0
Overall Average	0.437	0.475

The network bandwidth for the constant bandwidth is 16.05 compared to 9.75 for the network bandwidth configuration. This difference is expected because the method of preemption differs. Since the network bandwidth preempts the lowest bandwidth first, the average network bandwidth preempted would be much lower than the constant bandwidth configuration. If network bandwidth is the only consideration, then the network bandwidth configuration would be better than the variable bandwidth configuration.

4.6 Summary

This chapter showed the results of IDNX communication preemptions models. The distributions of the network traffic with respect to the source, destination, priority, bandwidth and call holding time were determined. For all four configurations, the network loading was determined by varying the network capacity to see how it affected the grade of service per priority. In all configurations, the grade of service for all priorities decreased as the network capacity increased. The plots showed a greater rate of decrease for higher priority calls. With the network capacity set at 100 percent, the rate of calls generated in node 0 was allowed to increase. This was done to determine the effect on the grade of service. It was seen that the grade of service for the lower two priority calls increased immediately. It took at least 8 times the original number of calls in node 0 before there was any change in the priority-3 grade of service. From the grade of service plots, the network loading was set at 75 percent capacity.

The results of the simulations showed that the overall grades of service for all the configurations are within 8% of each other. The variable bandwidth configuration produced a better grade of service and average number of preemptions than the network bandwidth configuration. Only the constant bandwidth produced a better grade of service than the variable bandwidth configurations. Configurations with variable bandwidth initially connected and preempted more calls than the configurations with constant bandwidth. It was also observed that when capacity is low, the variable bandwidth gives a better grade of service and when capacity is high the constant bandwidth configurations produce the better grade of service. The network bandwidth configurations greatly reduce the average network bandwidth preempted but at a price of increasing the grade of service.

The constant bandwidth configuration had the best grade of service per priority and the lowest number of preempted calls. This configuration produced the best performance. However, it is very inflexible and not of much use to the DoD. The variable bandwidth configuration had the second best grade of service and a low average number of preemptions. This configuration can be adapted to any network and the DoD can benefit from it to increase the grade of service for its most important customers. If the DoD is interested in keeping the network disruption to a minimum, then the network bandwidth configurations can help. However, the price would be a lower grade of service.

In the variable bandwidth configuration, the grade of service of priority-3 calls improved from 0.389 without preemption to 0.00 with the aid of preemption. This shows 100% improvement, as all the calls were connected. Similarly, the grade of service of priority-2 calls showed a 93.8% improvement and a 1.4 % improvement for priority-1 calls. These improvements in the grades of service of the higher priorities came at the expense of priority-0 calls which deteriorated by 101%.

In the constant bandwidth configuration, the grade of service of priority-3 calls improved from 0.449 without preemption to 0.00 with the aid of preemption. Once again, there was a 100% improvement for the priority-3 grade of service. The grade of service of priority-2 calls showed a 95.5% improvement and a 31% improvement for priority-1 calls. Once again, these improvements came at the expense of priority-0 calls which declined by 71%.

Clearly, preemption provides better access to the network for high priority customers. The DoD can benefit tremendously by these preemption models to insure network resources are given to the right people.

5 Conclusion and Future Recommendations

5.1 Introduction

The objective of this research was to develop the best preemption configuration model for the DoD. The development of this model predicated on the performance of preemption algorithms previously developed. This chapter concludes the research effort. An overview of the research effort is first presented. This overview summarizes the major topics in each of the previous chapters. Next, the conclusions are presented in two parts. The conclusions directly comparing specific configurations and an overall conclusion are presented. Following the conclusions, recommendations for future research are presented. Finally, an overall summary is provided.

5.2 Overview

Chapter 1 began with a short background of priority and preemption in a circuit switch communications network. Next, there were a definition of the problem and a presentation on the plan of attack. The DoD wanted to know if the use of preemption in their IDNX switches would improve the grade of service for their most important customers. They also wanted to know what preemption model would produce the best performance. The scope of the research was narrowed to the performance analysis of existing call preemption algorithms.

The review of call preemption algorithms was accomplished in Chapter 2. The literature review found preemption algorithms developed by Garay and Gopal [GaG92]. These algorithms differ from each other based on the distribution of the bandwidth and the method of preemption. The literature search also found a comparison analysis of one of Garay and Gopal's algorithm with a modified version of itself by Mohammed Peyravian [Pey94].

Chapter 3 presented the methodology used in obtaining the required metrics to perform the analysis. It also stated the required performance metrics. The metrics used were grade of service, average number of preemptions, and average network bandwidth preempted. Four configuration models were constructed. Each model represented a single algorithm. These models were based on variable bandwidth, constant bandwidth, network bandwidth and constant network bandwidth. The input parameters were specified. These parameters were fixed parameters and must be set before the simulations begin. The input traffic was exponentially generated from each node based on the nodes' inter-arrival time. Steady-state analysis followed. The simulation warm-up time and the run time were determined. Finally, the communication network models were validated and verified.

The results were presented in Chapter 4. With the traffic load constant, the grades of service per priority were plotted as the network capacity increases. Also with the network capacity set at 100 percent; the grade of service per priority was investigated, as the traffic generated in node 0 increases 20 folds. The operating network loading for each configuration was determined and the required metrics were obtained. Simulations of the variable and constant bandwidth communications models without preemption were accomplished for comparison purpose. Finally, relevant configuration models results were compared.

5.3 Conclusion

Each preemption configuration model can add significance to a communication network. The results showed that preemption increases the grade of service for high priority users significantly over lower priority users. There are also some distinct differences in the results of the various configurations. Below are the conclusions directly comparing specific configurations and an overall conclusion.

5.3.1 Variable Bandwidth Versus Network Bandwidth Conclusion

The variable bandwidth model differs from the network bandwidth model only in the method of preemption. While the variable bandwidth preempts the call with the largest bandwidth first, the network bandwidth model preempts the call with the lowest network bandwidth first. This contrast in the method of preemption produced differences in the various output metrics.

The results showed that the variable bandwidth preemption model produced a better grade of service per priority and average number of preemptions. For the same number of calls generated, the grade of service per priority was consistently lower in the variable bandwidth model. Priority-2 grade of service was 25% lower, priority-1 grade of service was 12% lower and priority-0 grade of service was 1% lower. At the same time, the network bandwidth produced a much lower (55%) average preempted network bandwidth. In terms of preempted calls, the variable bandwidth had a 9.1% lower average number of preemptions. If the goal is to provide the best possible grade of service for the high priority customers or preempt the least number of calls, then the variable bandwidth model is the better choice. On the other hand, if limiting network disruption is more important, then the network bandwidth model is the better choice. This comparison gives the DoD a choice based on the goals of the communications network organization.

5.3.2 Variable Bandwidth Versus Constant Bandwidth

The variable bandwidth model differs from the constant bandwidth only in its bandwidth distribution. The variable bandwidth configuration model is uniformly distributed while the bandwidths are the same for the constant bandwidth model. This difference produced dissimilarities in the results.

The results showed that the constant bandwidth model performed better than the variable bandwidth model. The constant bandwidth model yielded a 3% lower grade of service for priority-0, 8% lower for priority-1 and 17% lower for priority-2 calls. Therefore, the constant bandwidth model is better based solely on the grade of service.

5.3.3 Variable Bandwidth Versus Constant Network Bandwidth

The variable bandwidth and the constant network bandwidth models differ in two ways. Firstly, the variable bandwidth model call requests are uniformly distributed while the bandwidths are the same for the constant bandwidth model. Secondly, the variable bandwidth model preempts the largest bandwidth first, while the constant network bandwidth model preempts the lowest network bandwidth first. The differences in the models produce dissimilar results.

The results showed advantages for both preemption models. The variable bandwidth model had a 1% lower grade of service for priority-0 and 5% lower for priority-1 calls. However, the constant network bandwidth had a 17% lower grade of service for the priority-2 calls and a 66% lower overall preempted network bandwidth. Since the goal of preemption is to produce a lower grade of service for the highest priority customers, the constant network bandwidth model performed better than the variable configuration model.

It was also found that the constant network bandwidth model preempted the lowest average network bandwidth of all the configuration models. Thus, if keeping the network disruption to a minimum is the most important factor, the constant network bandwidth model gives the best results. The constant bandwidth model traded flexibility for improved performance.

5.3.4 Constant Bandwidth Versus Network Bandwidth Conclusion

The constant bandwidth and the network bandwidth also differ in two ways. The bandwidth of the call requests is fixed for the constant bandwidth model and uniformly distributed for the network bandwidth model. Additionally, the constant bandwidth preempts the first call while the network bandwidth model preempts the call with the lowest network bandwidth first. The output metrics produced by each configuration also differ.

The results showed that the constant bandwidth preemption model produced a 4% lower grade of service for priority-0, 18% lower for priority-1 and 33% for priority-2 calls. However, the network bandwidth model produced a 39% lower average preempted network bandwidth. The constant bandwidth model produced the best grade of service per priority than all other configuration models. It is also the least flexible of all configuration models. If limiting network disruption is more important, then the network bandwidth model is the better choice If grade of service is the only option, the constant bandwidth is the best preemption model.

5.3.5 Overall Conclusion

Each preemption configuration model has its advantages and disadvantages. The best configuration model depends on the need of the communications network organization. In a network where the bandwidth is not constant, the variable bandwidth configuration produced the best grade of service and average number of preemptions. In that same network, the network bandwidth preemption model best limits the amount of network disruption. In a network where the bandwidth is constant, the constant bandwidth preemption model produced best grade of service, and the constant network bandwidth preemption model best limits the amount of network disruption. Each preemption model significantly increases the grade of service over a network without preemption. The simulations showed that preemption improves the grade of service of the higher priority calls. The grade of service per priority is not significantly better for any one priority and the choice of configuration depends on the need of the communications network organization.

5.4 Future Recommendations

This effort analyzed the performance of preemption algorithms in a circuit switched IDNX communications network that implements a static table-driven routing procedure. Other communications network models can also be used to analyze the preemption algorithms. Some possible areas of research are as follows:

1. Develop models based on an actual communications networks. Implement the network routing based on Dijkstra's algorithm. This algorithm finds the shortest path between nodes. The performance obtain in this network could be compared with the results obtained using a static table-driven routing procedure.

2. Implement a network using Dynamic Non-hierarchical Routing (DNHR). This routing strategy is a hybrid of time-variable routing and simple implementation of adaptive routing, combined with sophisticated off-line optimization of trunking and routing pattern design [Ash90]. Gerald Ash described the implementation of a DNHR switch network [Ash90]. The results can be compared with the results obtained in this research.

3. Incorporate periodic failures of nodes and trunks instead of decreasing the network capacity to simulate network degradation. Compare the performance of the preemptions algorithms with the performance of results obtained in this research.

5.5 Summary

This chapter concludes the research effort. An overview of the entire thesis was provided and all conclusions were given. The analysis showed that preemption can significantly lower the grade of service for high priority customers in a congested network. It was also shown that each configuration model has its advantages and disadvantages. The best configuration preemption model depends on the bandwidth flexibility of the network and the goals of the communications network organization. The constant bandwidth model produced the lowest grade of service. The variable bandwidth model produced the lowest average number of preemptions for uniformly distributed bandwidths. The constant network bandwidth model preempted the lowest average network bandwidth. Based on the present DoD network and their need to improve the grade of service of their high priority customers, the variable network preemption model would best fit their need. However, if their communications network or needs' changes then, the other preemption models might best serve them.

Appendix A

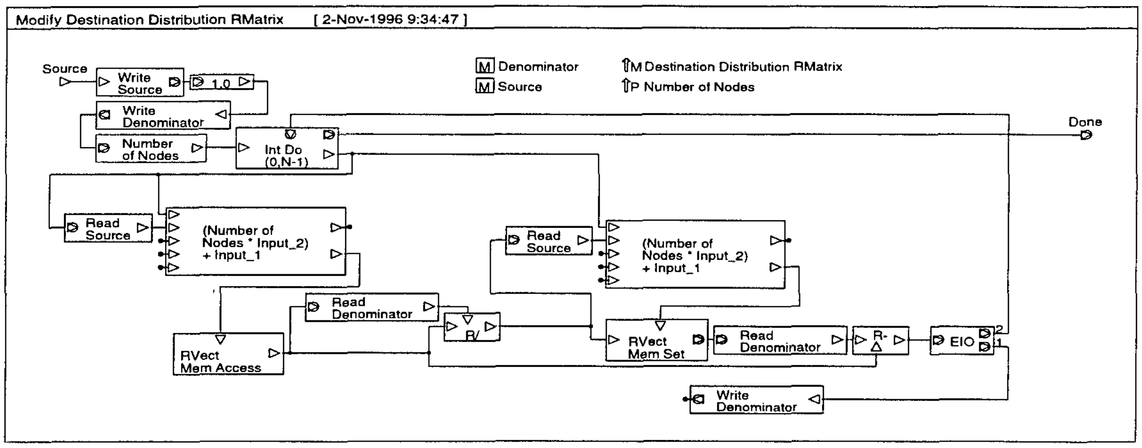

Figure A-1 Modify Destination Distribution RMatrix

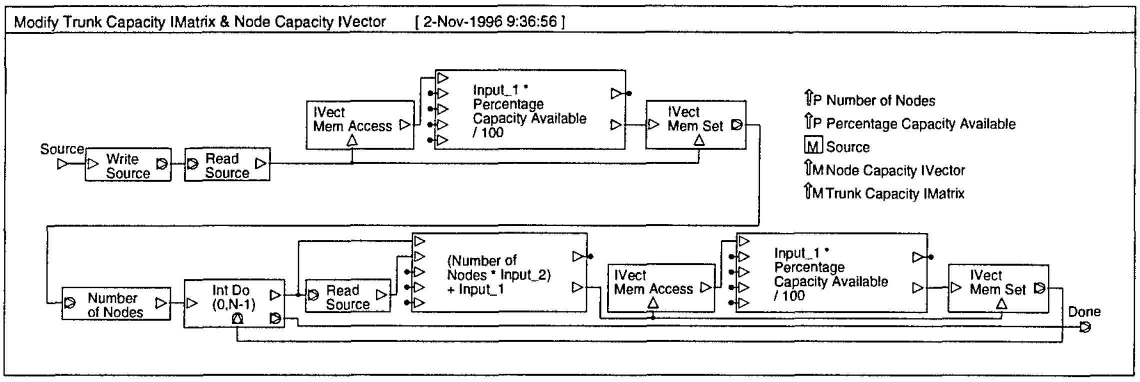

Figure A-2 Modify Trunk Capacity IMatrix & Node Capacity

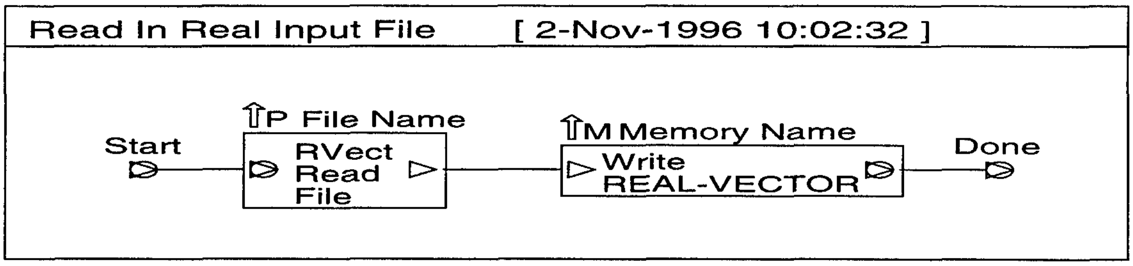

Figure A-3 Read in Real Input File

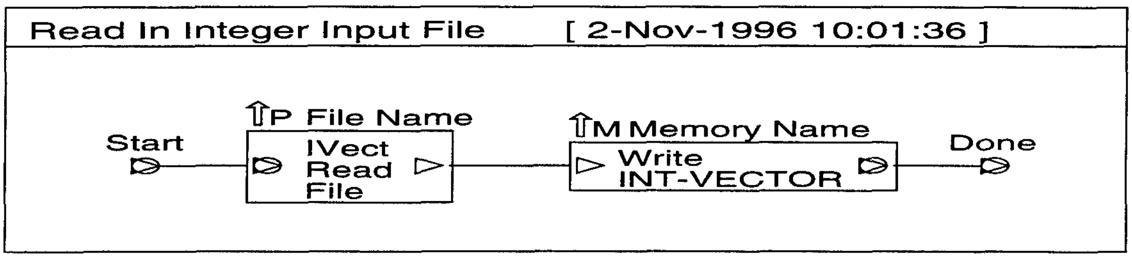

Figure A-4 Read in Integer Input File

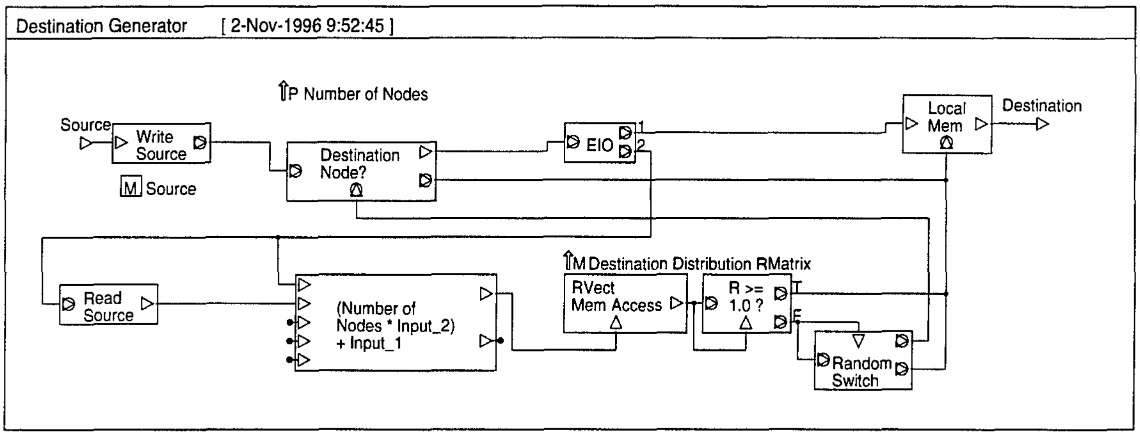

Figure A-5 Destination Generator

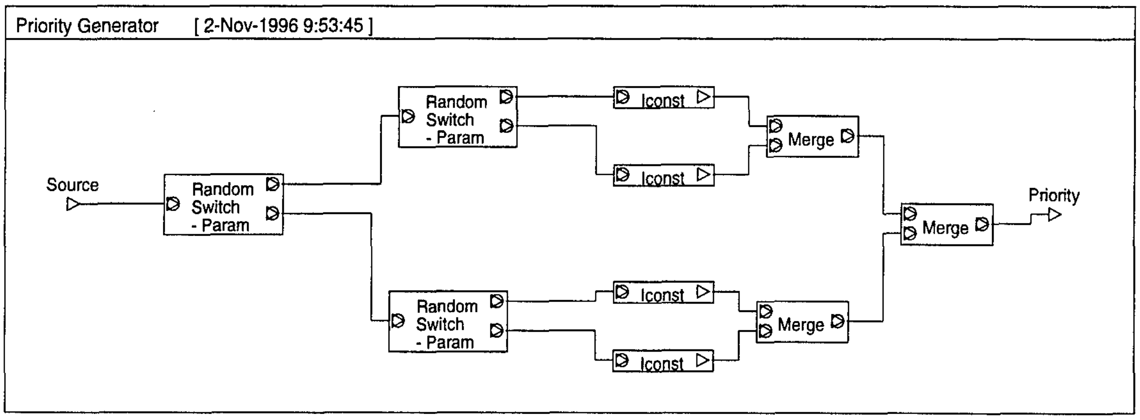

Figure A-6 Priority Generator

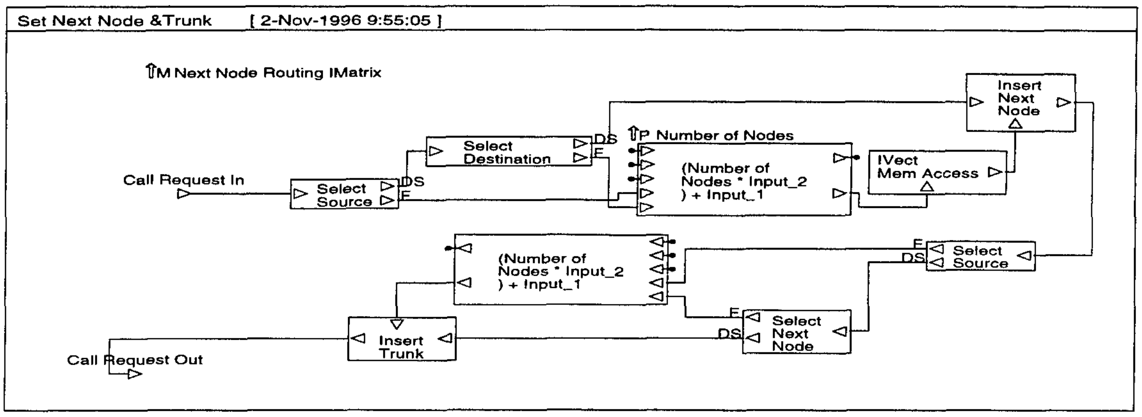

Figure A-7 Set Next Node & Trunk

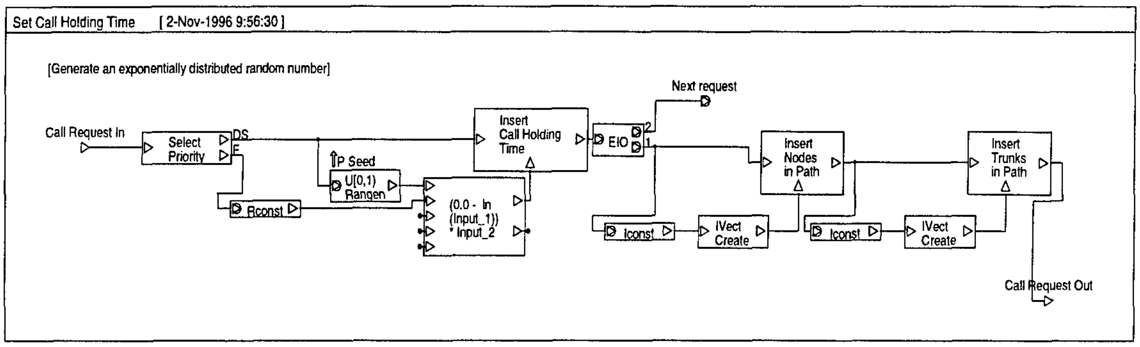

Figure A-8 Set Call Holding Time

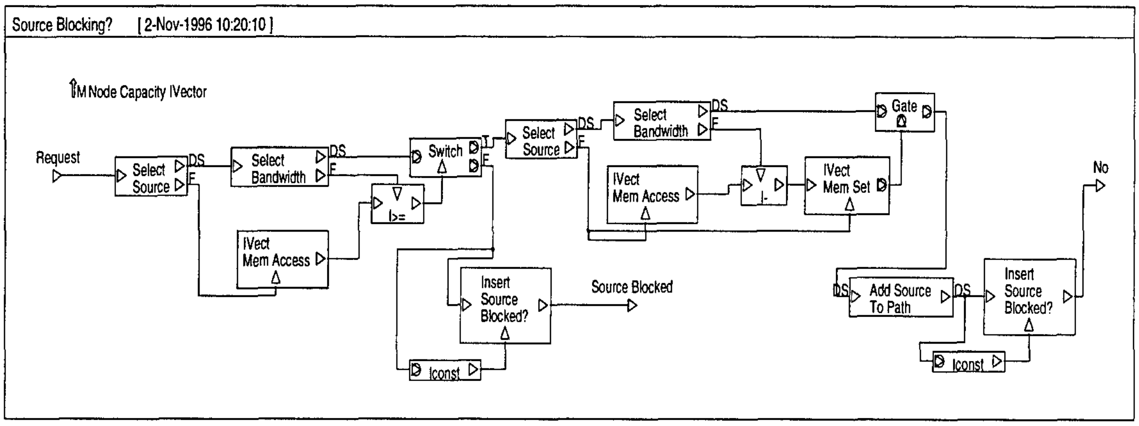

Figure A-9 Source Blocking?

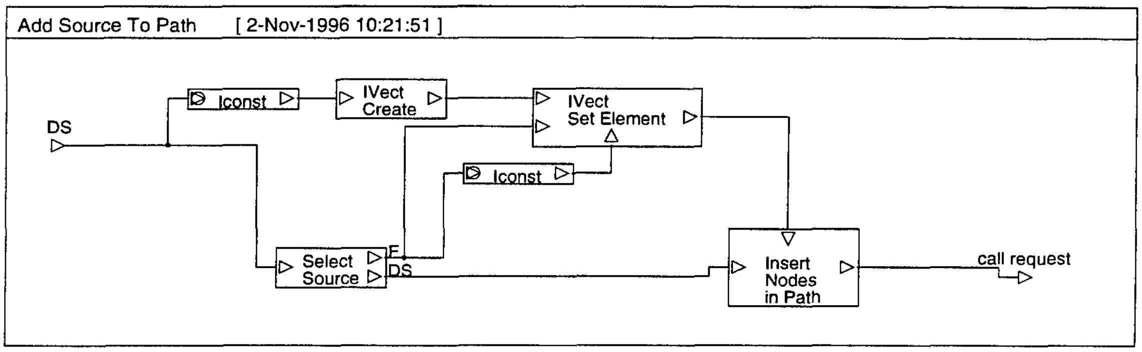

Figure A-10 Add Source to Path

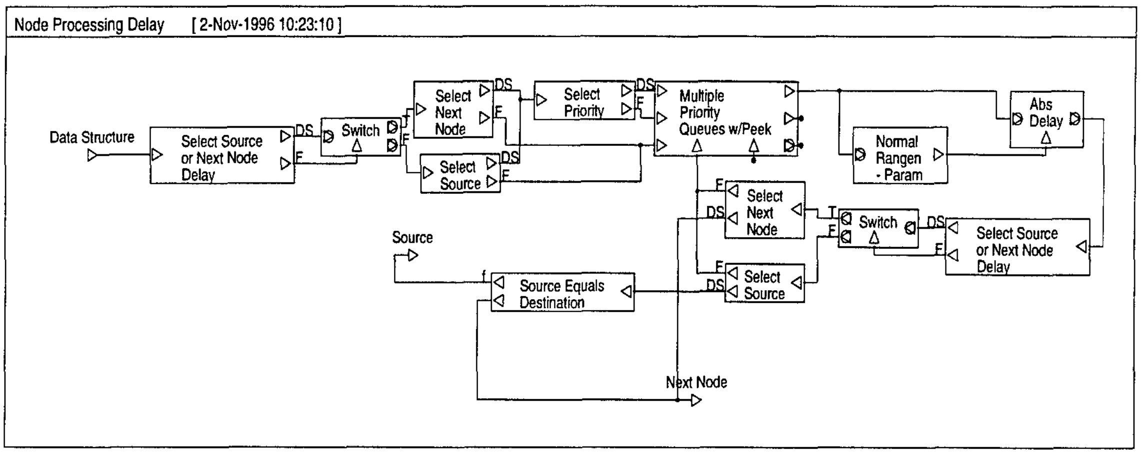

Figure A-11 Node Processing Delay

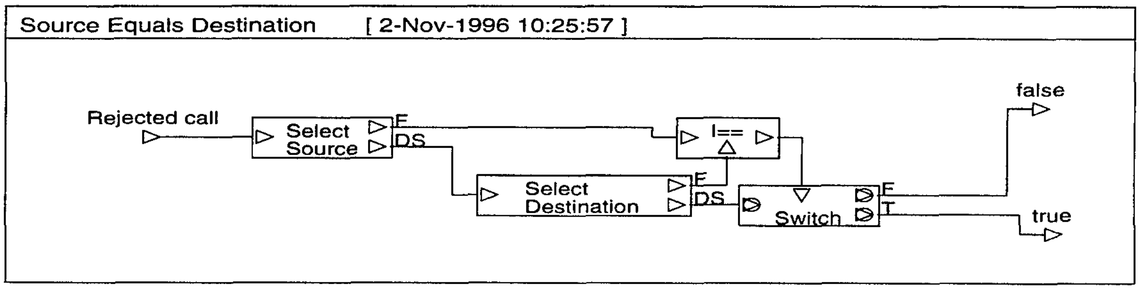

Figure A-12 Source Equals Destination

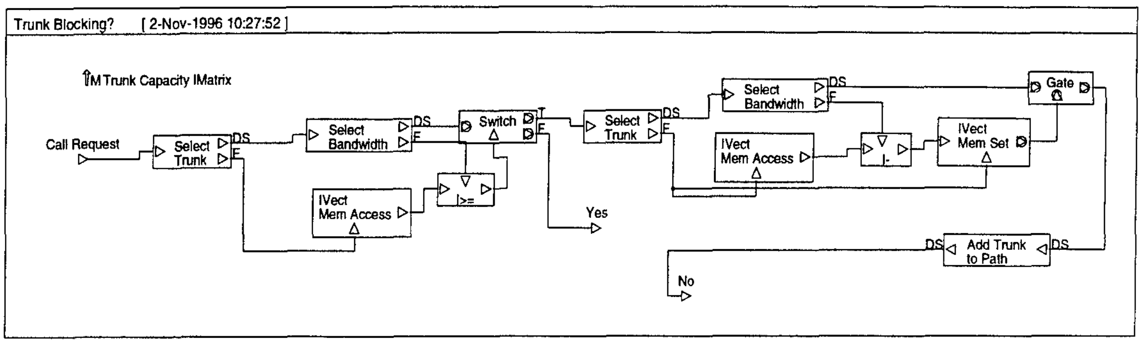

Figure A-13 Trunk Blocking?

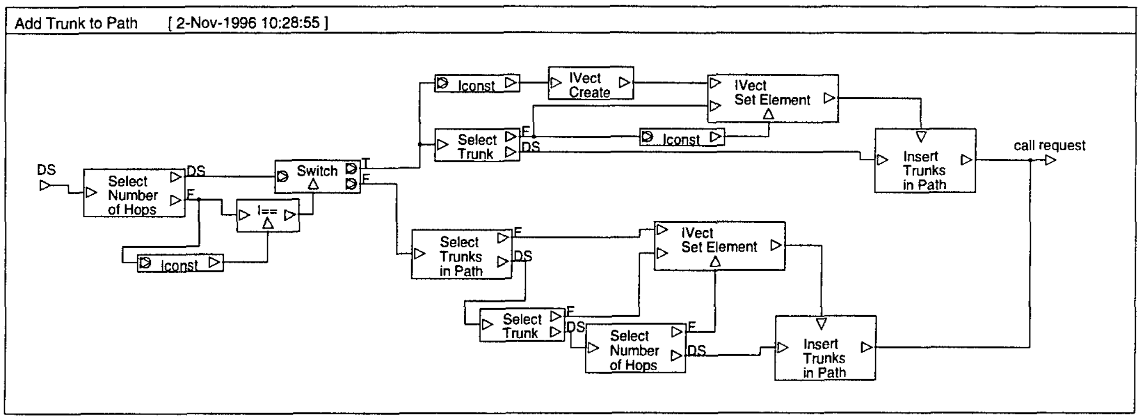

Figure A-14 Add Trunk to Path

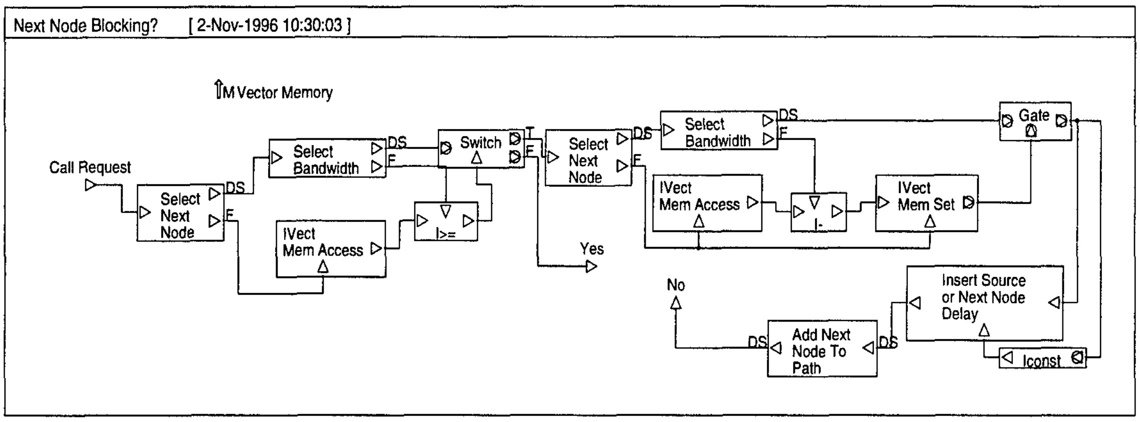

Figure A-15 Next Node Blocking?

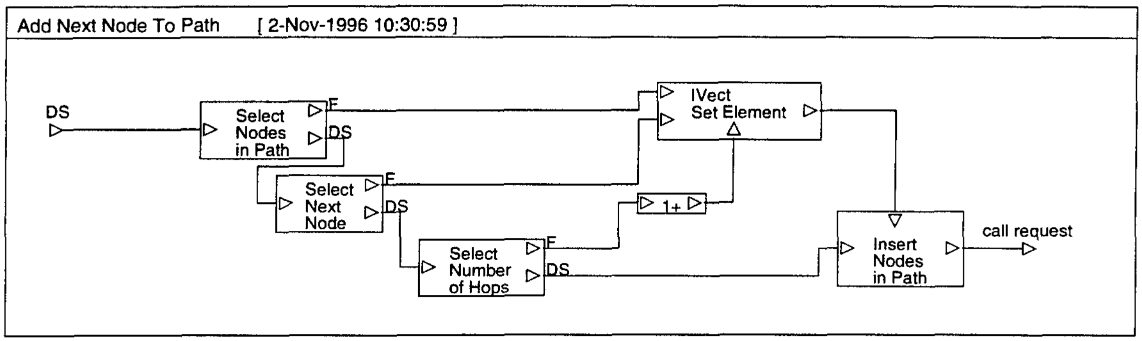

Figure A-16 Add Next Node to Path

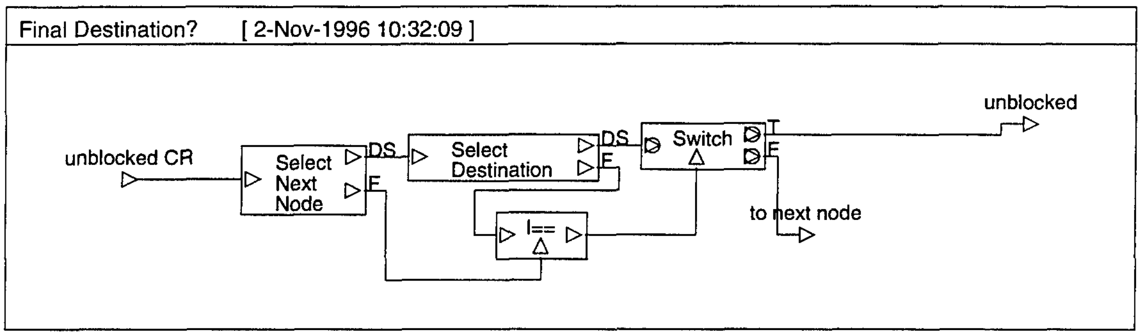

Figure A-17 Final Destination?

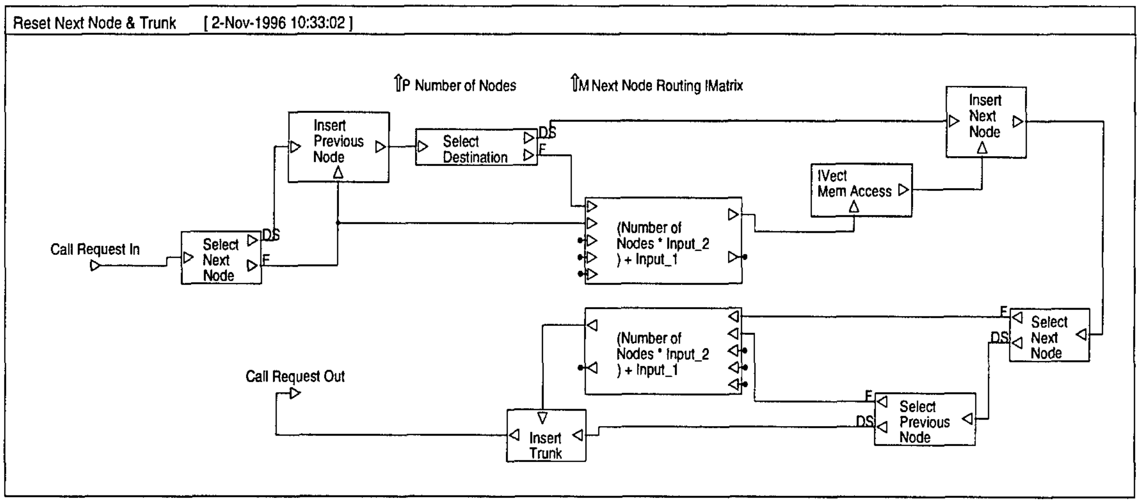

Figure A-18 Reset Next Node & Trunk

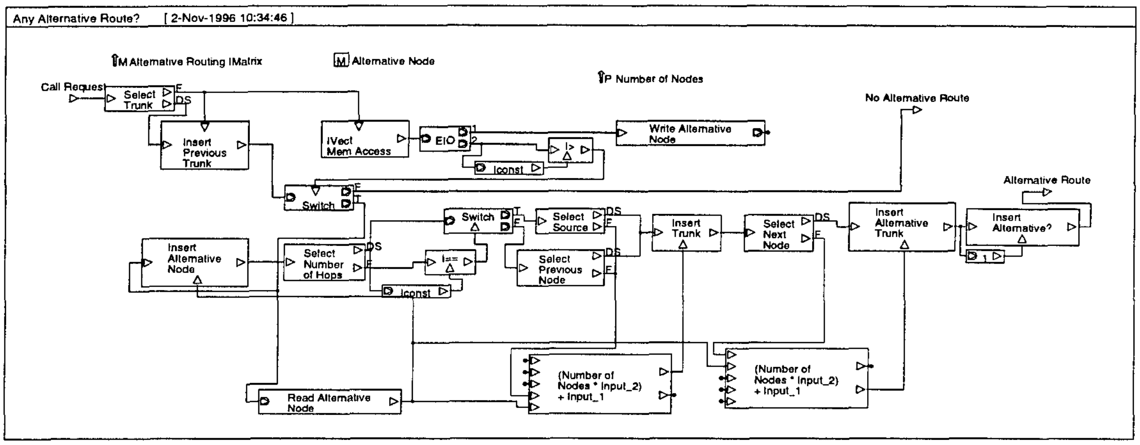

Figure A-19 Any Alternate Route?

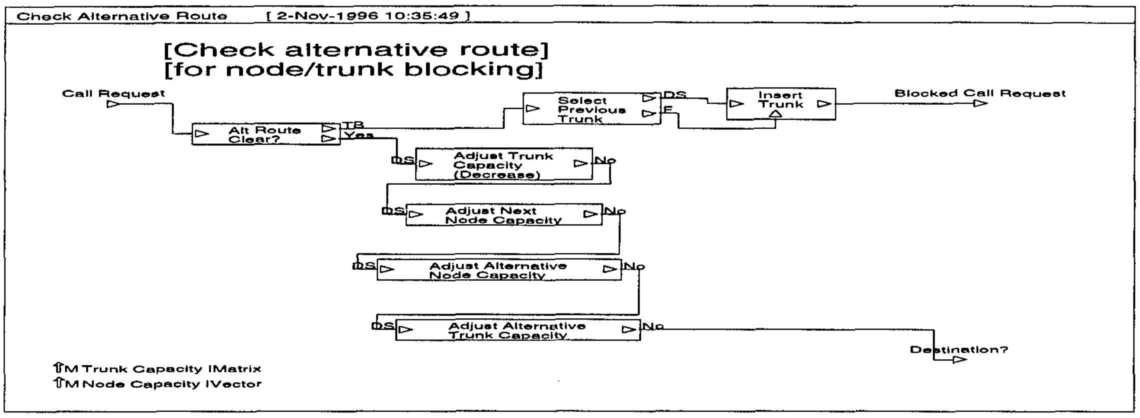

Figure A-20 Check Alternative Route block

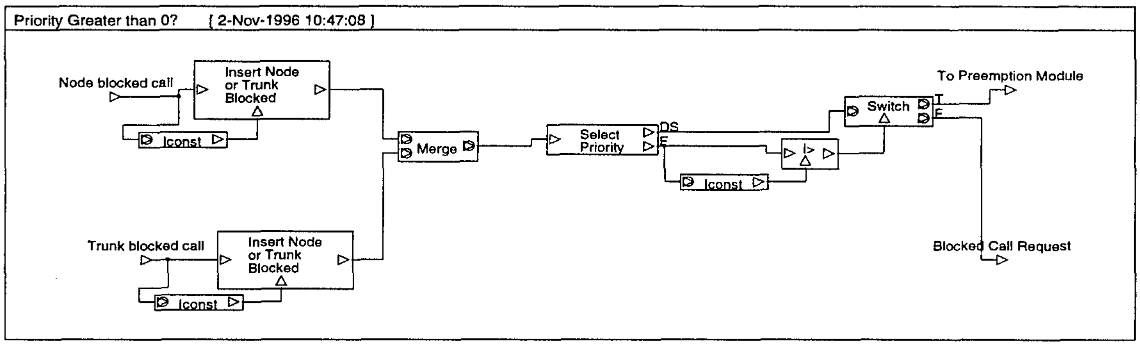

Figure A-21 Priority Greater than 0?

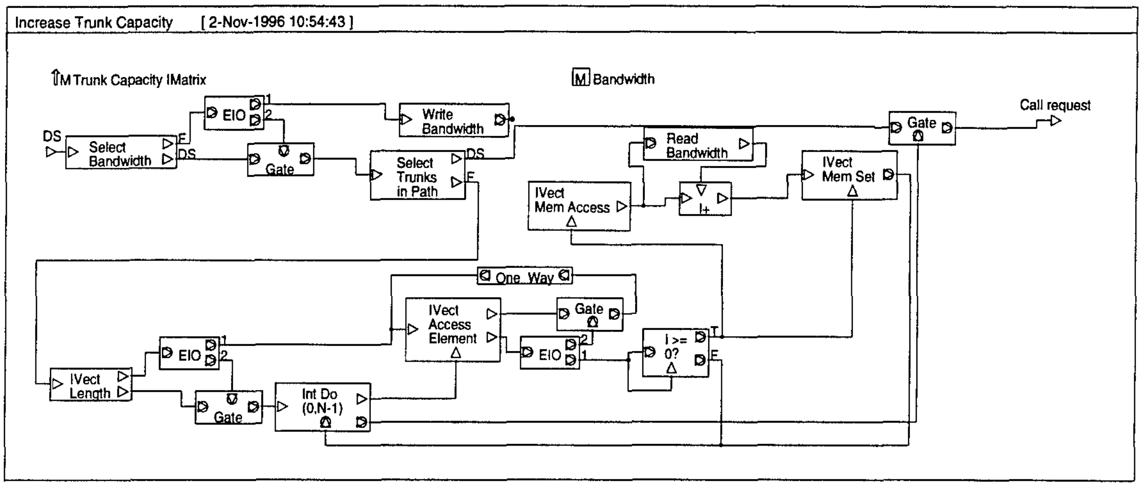

Figure A-22 Increase Trunk Capacity

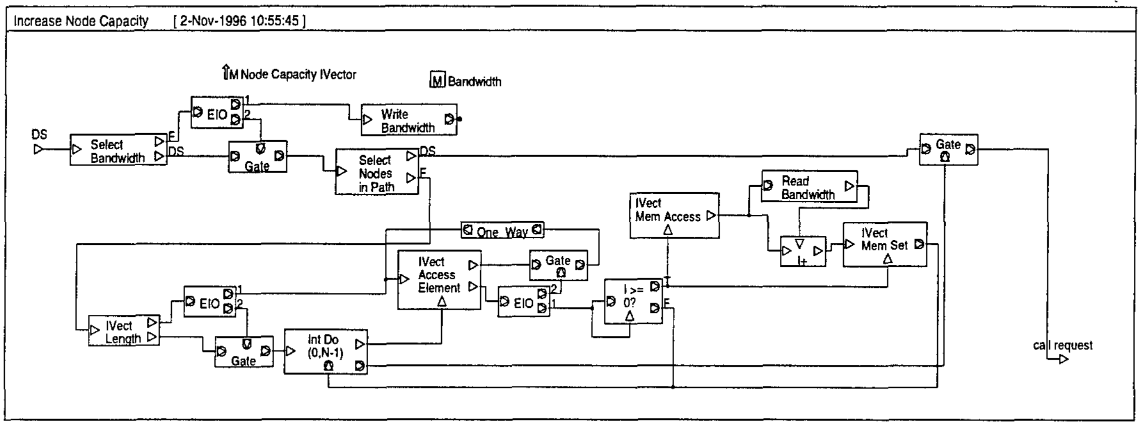

Figure A-23 Increase Node Capacity

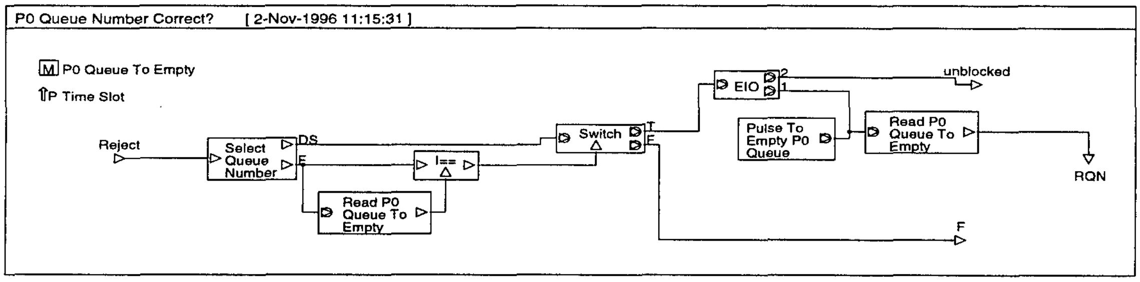

Figure A-24 P0 Queue Number Correct?

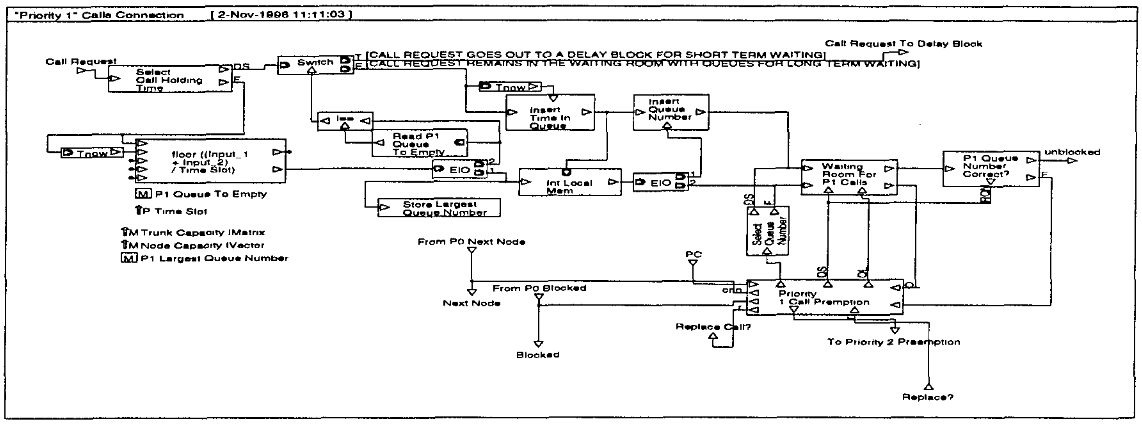

Figure A-25 Priority 1 Calls Connection

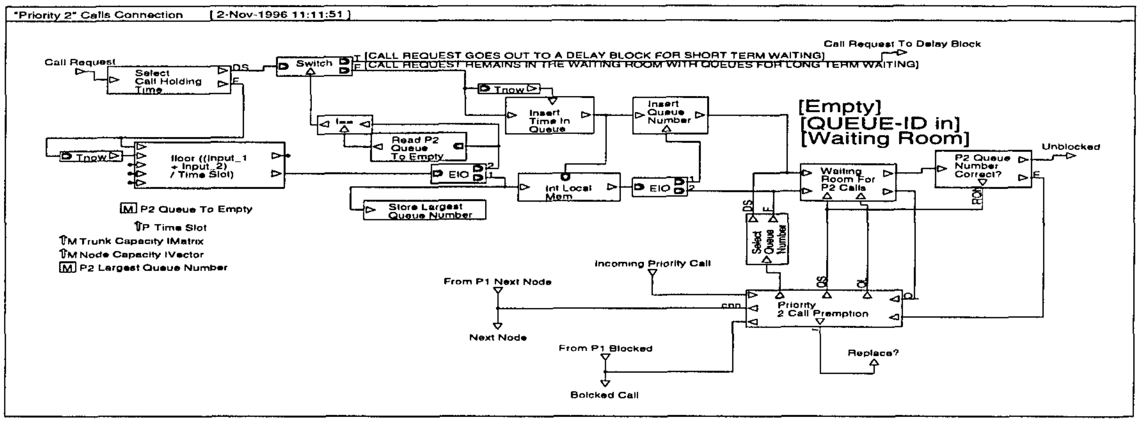

Figure A-26 Priority 2 Calls Connection

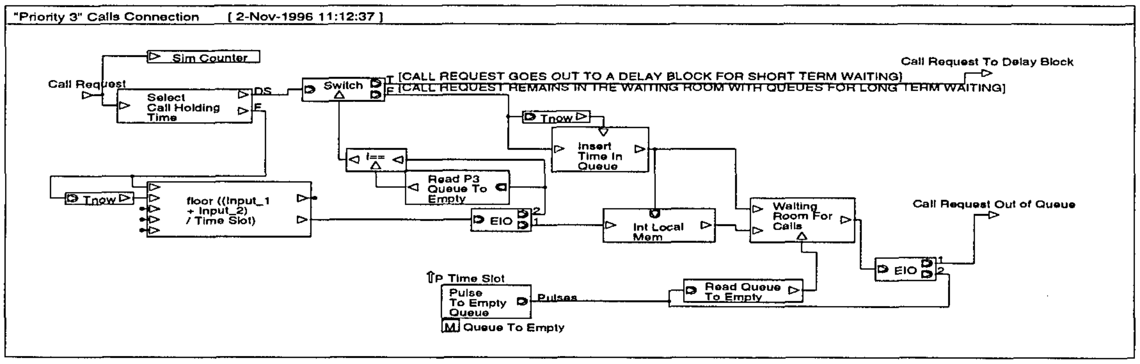

Figure A-27 Priority 3 Calls Connection

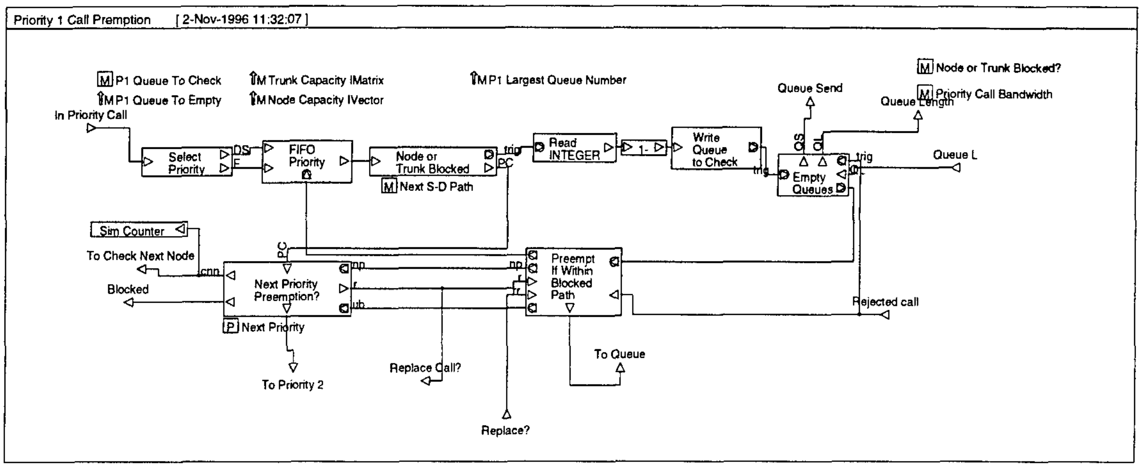

Figure A-28 Priority 1 Call Preemption

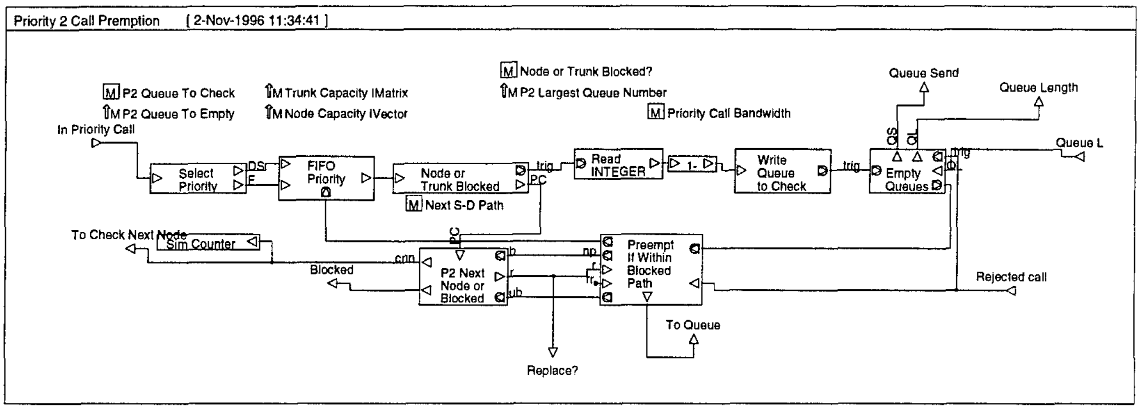

Figure A-29 Priority 2 Call Preemption

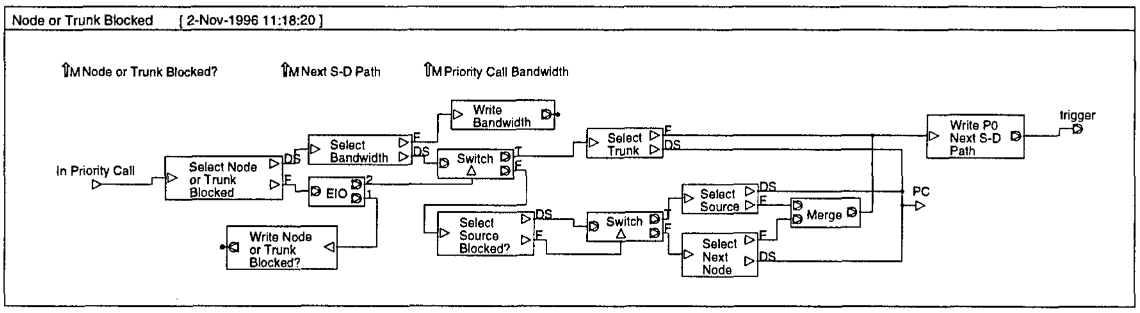

Figure A-30 Node or Trunk Blocked

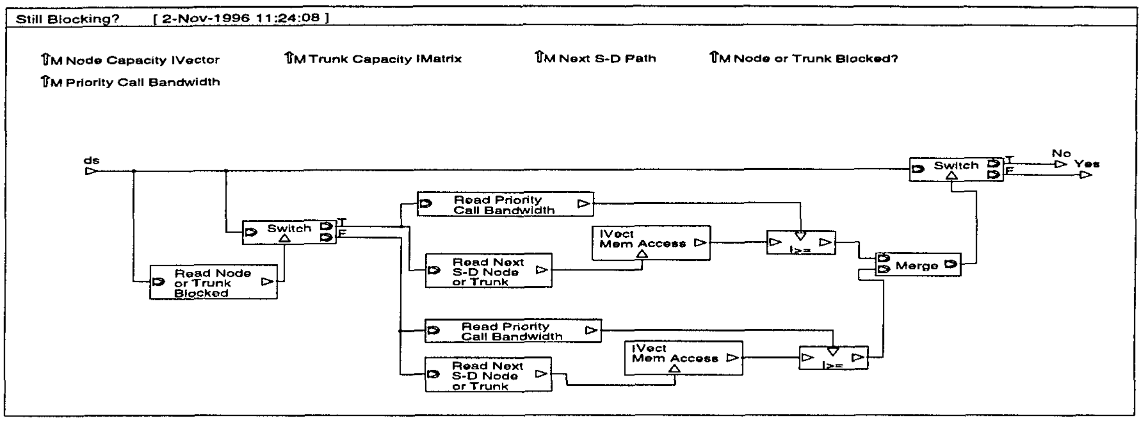

Figure A-31 Still Blocking?

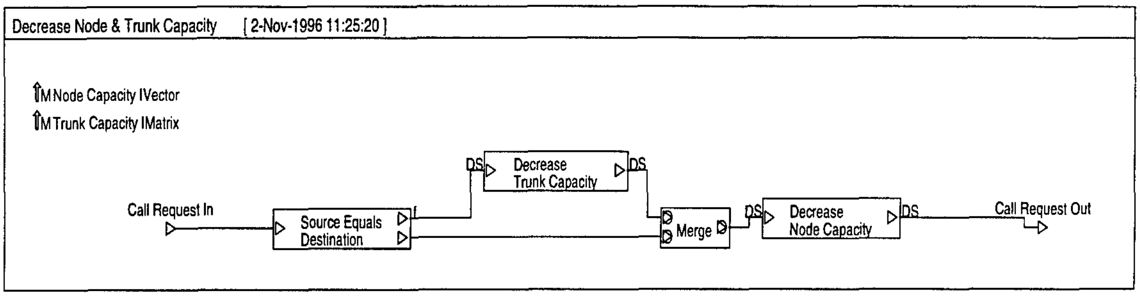

Figure A-32 Decrease Node & Trunk Capacity

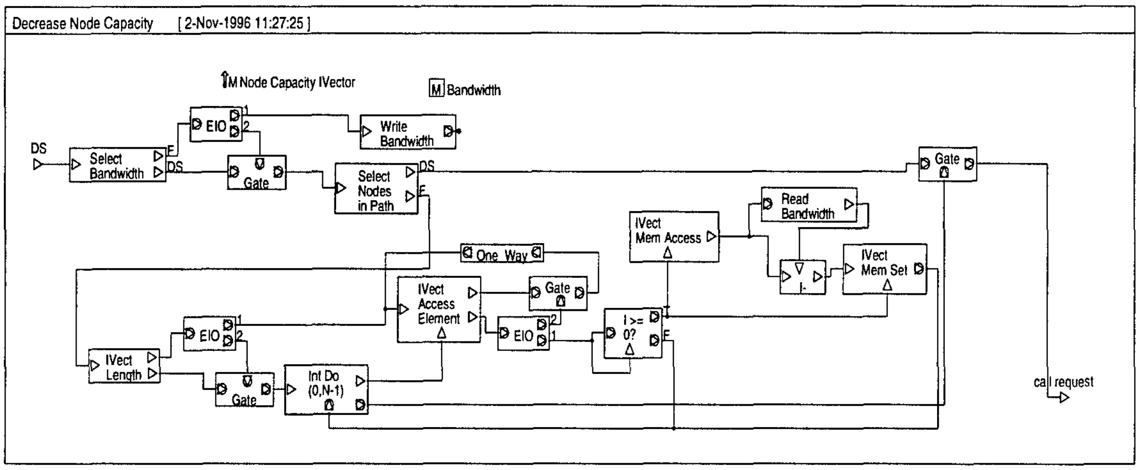

Figure A-33 Decrease Node Capacity

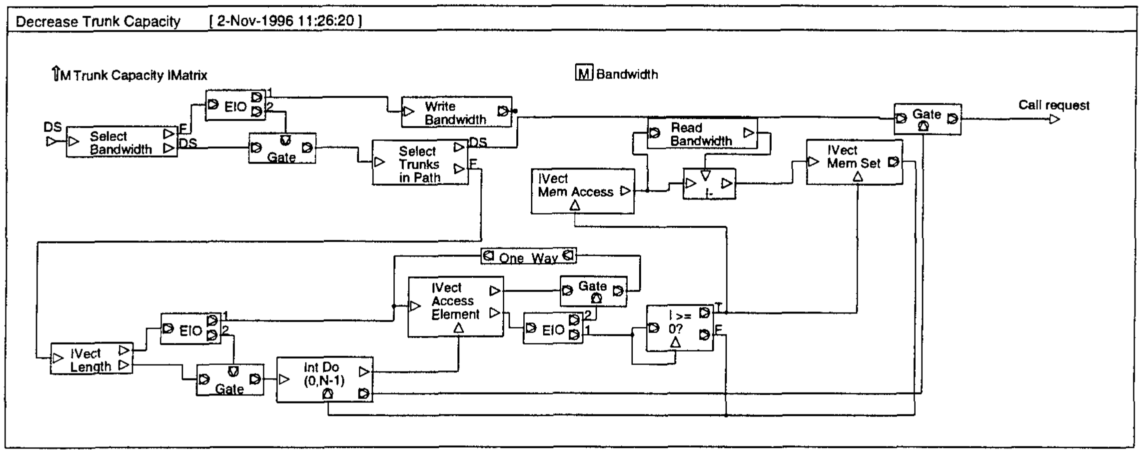

Figure A-34 Decrease Trunk Capacity

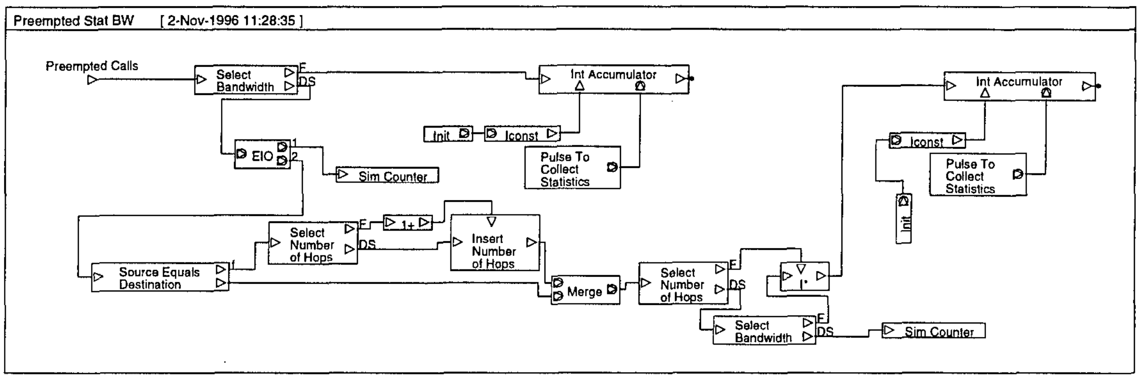

Figure A-35 Preempted Statistics

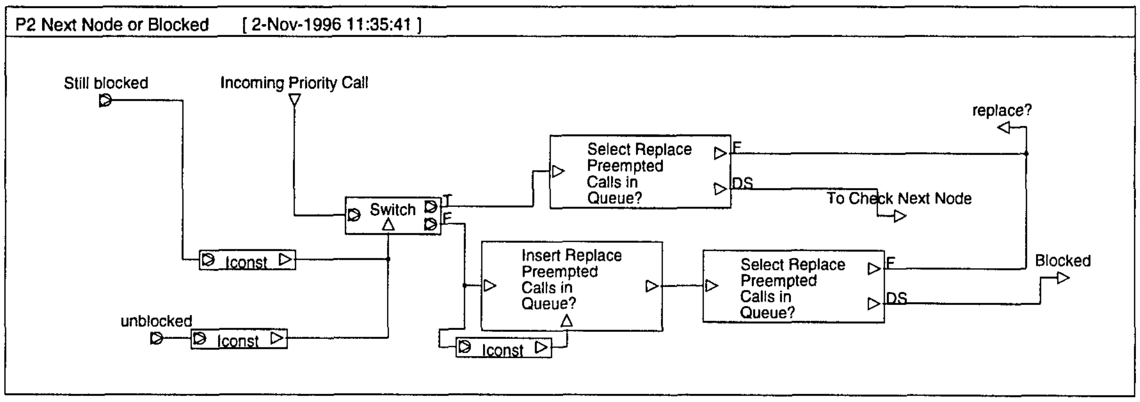

Figure A-36 P2 Next Node or Blocked

Appendix B

Table B-1 Grade of Service per priority (Variable Bandwidth)

Grade of Service				
Capacity	P3	P2	P1	P0
0	1	1	1	1
5	0.546	0.888	0.952	0.98
10	0.297	0.785	0.93	0.972
15	0.154	0.66	0.904	0.964
20	0.08	0.559	0.872	0.957
25	0.045	0.454	0.843	0.947
30	0.022	0.36	0.809	0.935
35	0.011	0.282	0.766	0.922
40	0.006	0.226	0.725	0.913
45	0.003	0.168	0.667	0.895
50	0.002	0.125	0.612	0.882
55	0.001	0.089	0.55	0.866
60	0	0.066	0.486	0.845
75	0	0.024	0.333	0.783
100	0	0.005	0.13	0.623

Table B-2 Grade of Service per Priority as Node 0 Increases (Variable Bandwidth)

Grade of Service				
x Node 0	P3	P2	P1	P0
1	0	0.005326	0.129055	0.623162
2	0	0.005375	0.228541	0.721201
4	0	0.061887	0.486849	0.796119
6	0	0.188029	0.625615	0.826055
8	0	0.292386	0.680121	0.855851
10	0	0.392038	0.723035	0.873507
12	0.010666	0.475531	0.762597	0.89051
14	0.034725	0.544234	0.783903	0.90206
16	0.051473	0.600348	0.803829	0.908769
18	0.068194	0.647289	0.814794	0.91593
20	0.091442	0.694227	0.834119	0.921339

Table B-3 Grade of Service per priority (Constant Bandwidth)

Grade of Service				
Capacity	P3	P2	P1	P0
0	1	1	1	1
5	0.574	0.936	0.976	0.99
10	0.313	0.84	0.951	0.983
15	0.173	0.722	0.924	0.975
20	0.079	0.604	0.897	0.966
25	0.041	0.498	0.87	0.959
30	0.021	0.399	0.835	0.949
35	0.011	0.31	0.792	0.937
40	0.006	0.232	0.739	0.925
45	0.002	0.179	0.686	0.909
50	0.001	0.135	0.627	0.892
55	0	0.102	0.565	0.87
60	0	0.063	0.492	0.845
75	0	0.02	0.306	0.759
100	0	0.003	0.124	0.575

Table B-4 Grade of Service per Priority as Node 0 Increases (Constant Bandwidth)

Grade of Service				
x Node 0	P3	P2	P1	P0
1	0	0.003377	0.124153	0.574899
2	0	0.004579	0.221823	0.705267
4	0	0.072447	0.488158	0.786676
6	0	0.206081	0.617406	0.819083
8	0	0.320767	0.68301	0.847605
10	0.009694	0.416433	0.726838	0.863681
12	0.024786	0.494868	0.763444	0.884482
14	0.05078	0.557492	0.789089	0.894356
16	0.066516	0.615647	0.811343	0.906323
18	0.082148	0.663325	0.825565	0.914905
20	0.109565	0.709398	0.842638	0.919821

Table B-5 Grade of Service per priority (Network Bandwidth)

	Grade of Service			
Capacity	P3	P2	P1	P0
0	1	1	1	1
5	0.524483	0.884656	0.954095	0.979825
10	0.295974	0.796296	0.931361	0.972193
15	0.167573	0.699206	0.90617	0.963318
20	0.084875	0.592063	0.879575	0.9561
25	0.059848	0.493386	0.854033	0.94616
30	0.025027	0.408333	0.816905	0.938587
35	0.009249	0.319709	0.780567	0.924506
40	0.003264	0.249339	0.736856	0.916933
45	0.002176	0.191799	0.686562	0.902142
50	0.002176	0.150397	0.638111	0.888061
55	0.001088	0.110847	0.579917	0.87114
60	0	0.085053	0.527254	0.853331
75	0	0.03	0.373	0.792
100	0	0.004	0.16	0.64

Table B-6 Grade of Service per Priority as Node 0 Increases (Network Bandwidth)

	Grade of Service			
x Node 0	P3	P2	P1	P0
1	0	0.004497	0.160186	0.639924
2	0	0.003074	0.270957	0.735973
4	0	0.071258	0.513461	0.799559
6	0	0.198989	0.647624	0.84195
8	0	0.334242	0.691368	0.859021
10	0.004323	0.439491	0.735359	0.876165
12	0.023336	0.509448	0.766689	0.891498
14	0.042637	0.576802	0.785188	0.905509
16	0.052484	0.621928	0.807512	0.914253
18	0.086556	0.668749	0.820859	0.91831
20	0.09669	0.70534	0.837908	0.924498

Table B-7 Grade of Service per priority (Constant Network Bandwidth)

	Grade of Service			
Capacity	P3	P2	P1	P0
0	1	1	1	1
5	0.572327	0.942406	0.974091	0.991344
10	0.295073	0.86052	0.955523	0.983777
15	0.137317	0.750163	0.935746	0.976152
20	0.068658	0.642027	0.907332	0.96956
25	0.036164	0.527752	0.877969	0.960158
30	0.018344	0.424971	0.849987	0.950183
35	0.00886	0.338514	0.809655	0.941584
40	0.005241	0.26368	0.763192	0.928629
45	0.002621	0.195899	0.714915	0.913896
50	0.000524	0.144835	0.666854	0.897673
55	0	0.106439	0.607306	0.882022
60	0	0.066736	0.539166	0.864882
75	0	0.02	0.35	0.793
100	0	0.003	0.145	0.635

Table B-8 Grade of Service per Priority as Node 0 Increases(Constant Network Bandwidth)

		Grade of Service		
x Node 0	P3	P2	P1	P0
1	0	0.002873	0.145349	0.634805
2	0	0.00202	0.274204	0.747709
4	0	0.078813	0.523735	0.810094
6	0	0.211649	0.645673	0.838388
8	0	0.358649	0.694843	0.868024
10	0.000856	0.460533	0.738391	0.882756
12	0.017398	0.543761	0.771597	0.899422
14	0.034693	0.603262	0.794926	0.905906
16	0.049168	0.65459	0.816146	0.915897
18	0.067596	0.697247	0.83217	0.921757
20	0.099871	0.734111	0.845141	0.928792

Bibliography

[AlR94] Alta Group. BONeS DESIGNER Application Library Reference. Product Number: DES 8018. City: Cadence Design Systems, Inc., December 1994

[Ash90] Ash, Gerald R. "Design and Control of Networks with Dynamic Nonhierarchical Routing." IEEE Communications Magazine, October 1990, Vol.46 , pp. 34-40

[BaC96] Banks, Jerry. Carson, John S. Nelson, Barry L. Discrete - Event System Simulation. New Jersey: Prentice Hall, 1996.

[ClR94] Alta Group. BONeS DESIGNER Core Library Reference. Product Number: DES8004. City: Cadence Design Systems, Inc., December 1994

[ChG91] Chen, J.S.-C. Guerin, R. "Performance Study of an Input Queueing Packet Switch with Two Priority Classes," IEEE Transactions on Communication, vol. 39, pp. 117-126, Jan. 1991

[Jai91] Jain, Raj. The Art of Computer Systems Performance Analysis. New York: John Wiley & Sons, 1991.

[GaG92] Garay, Juan A. Gopal, Inder S. "Call Preemption in Communication Networks," Proceedings of INFOCOM '92, 1043-1050. New York: IEEE Press, 1992.

[Hsi90] Hsing Deh-phone K. "Simulation and Performance Evaluation of an ATM Multiplexer Using Priority Scheduling," Proceedings of Globecom'90. 1946-1952. New York: IEEE Press, 1990.

[LaM96] Law, Averill M. McComas, Michael G. "Simulations in Communications Networks," Proceedings of the 1994 Winter Simulation Conference, December 1994, pp 166 - 169.

[MaT87] Marie, Raymond. Trivedi, Kishor S. "A Note on the Effects of Preemptive Policies on the Stability of a Priority Queue," Information Processing Letters. Vol. 24, No. 6 pp 397-401.

[MoN94] Mooney, Francis W. Newport, Kris T. Schroder, Michael A. "Video Teleconferencing (VTC) in an Integrated Digital Network Exchange (IDNX) based Communication System," Proceedings of Milcom'94. 102-106. New York: IEEE Press, 1994.

[NaF67] Naylor, T. H. Finger J. M. "Verification of Computer Simulation Models," Management Science, Vol. 2 pp. B92 - B101.

[NeT91] Network Equipment Technologies. IDNX Integrated Digital Network Exchange Redwood City: Network Equipment Technologies Inc., 1991.

[Pey94] Peyravian, Mohammad. "Providing Different Levels of Network Availability in High-Speed Networks," Proceedings of Globecom '94. 941-945. New York: IEEE Press, 1994.

[SaA94] Saadawi, Tarid N. Ammar, Mostafa H. with Hakeem, Ahmed El. Fundamentals of Telecommunication Networks. John Wiley and Sons, Inc.1994.

[Tak90] Takagi, Hideaki. "Priority Queues with Setup Times." Operations Research, Vol 38, No. 4, July-August 1990, pp 667-677.

[YoC93] You, Ze Cho. Chong, Kwan Un. "Analysis of the M/G/1 Queue under a Combined Preemptive/Nonpreemption Priority Discipline", IEEE Transactions on Communications, Vol. 41, No. 1, January 1993, pp 132-141.

Vita

Captain Eric C. Gumbs [redacted]s. He graduated from Basseterre High School in 1976. He graduated from Texas A&M University with a Bachelor of Science degree in Electrical Engineering in August 1988. He received his commission on 22 November 1988 upon graduation from Officer Training School. In May 1995, he entered the School of Engineering, Air Force Institute of Technology.

[redacted]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 1996	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE
PERFORMANCE ANALYSIS OF PREEMPTION ALGORITHMS IN AN IDNX CIRCUIT SWITCH COMMUNICATIONS NETWORK

5. FUNDING NUMBERS

6. AUTHOR(S)
Eric C. Gumbs, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
2750 P Street
WPAFB, OH 45433-7126

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GE/ENG/96D-05

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Department of Defense (R55)
9800 Savage Road, Suite 6550
Fort Meade, MD 20755-6000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Access to communication networks is increasing rapidly. The increase access to these networks results in delays and at times loss of data. At times of peak traffic or when trunks or nodes are down, very important customers' communications requirements are not met. One way to combat this problem is to prioritize the network and provide different levels of grade of service (GoS) for each priority. Call preemption provides an effective method of obtaining different levels of GoS. This research seeks to design the best circuit switch communications network preemption model for the DoD by analyzing previously developed preemption algorithms. Four simulation network models are developed. The grades of service per priority are obtained as the network capacity decreases and as the calls generated in node 0 increases. The analysis of preemption network models is based on the grade of service, average number of preemptions, and average network bandwidth. The networks are simulated under the same input parameters. The analysis showed that preemption can significantly lower the grade of service for high priority customers in a congested network. The best configuration preemption model depends on the bandwidth flexibility of the network and the goals of the communications network organization.

14. SUBJECT TERMS
Call Preemption, Priority, Communication Networks, BONeS DESIGNER, IDNX
Grade of Service (GoS), Preemption Algorithms, Simulation, Topology

15. NUMBER OF PAGES
114

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract
G - Grant
PE - Program Element
PR - Project
TA - Task
WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

★U.S.GPO:1993-0-336-043