10-31-1992

Adaptive map configuration and dynamic routing to optimize the performance of a satellite communication network

Santhalingam Balasekar
New Jersey Institute of Technology

ABSTRACT

ADAPTIVE MAP CONFIGURATION AND DYNAMIC ROUTING TO OPTIMIZE THE PERFORMANCE OF A SATELLITE COMMUNICATION NETWORK

by
Santhalingam Balasekar

Availability of alternate routes are shown to enhance the performance of non-hierarchical circuit switched networks at moderate load conditions on which the networks typically operate. But alternate routes introduce instability at heavy and overloaded conditions and at these load conditions the network performance is found to deteriorate. To alleviate this problem, one of the control mechanism used is to reserve a fraction of the capacity of each link for direct routed calls. In this work, the feasibility of changing the facilities of a satellite network is taken advantage in optimizing the performance of such a network. A mesh connected, circuit switched satellite communication network is optimized by reconfiguring the network with proper allocation of link capacities and placing an optimal reservation scheme. The network load is measured and the network is continually adapted by reconfiguring the map to suit the current traffic conditions. The routing is performed dynamically. The results from the simulation study shows this method of traffic management performs better than the pure dynamic routing with fixed configuration.

ADAPTIVE MAP CONFIGURATION AND DYNAMIC ROUTING TO OPTIMIZE THE PERFORMANCE OF A SATELLITE COMMUNICATION NETWORK

by
Santhalingam Balasekar

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering
Department of Electrical and Computer Engineering
October, 1992

APPROVAL PAGE

ADAPTIVE MAP CONFIGURATION AND DYNAMIC ROUTING
TO OPTIMIZE THE PERFORMANCE
OF A SATELLITE COMMUNICATION NETWORK

by

Santhalingam Balasekar

7/22/92

Dr. Nirwan Ansari, Thesis Advisor
Assistant Professor of Electrical and Computer Engineering
New Jersey Institute of Technology

7/22/92

Dr. Joseph Frank, Committee Member
Associate Professor of Electrical and Computer Engineering
New Jersey Institute of Technology

7/22/92

Dr. Zoran Siveski, Committee Member
Assistant Professor of Electrical and Computer Engineering
New Jersey Institute of Technology

BIOGRAPHICAL SKETCH

Author : Santhalingam Balasekar

Degree : Master of Science in Electrical Engineering

Date : October, 1992

Date of Birth :

Place of Birth :

Undergraduate and Graduate Education :

- Master of Science in Electrical Engeneering,
  New Jersey Institute of Technology,
  Newark, New Jersey, 1992

- Bachelor of Science in Electrical Engineering
  Rutgers, The State University of New Jersey,
  New Brunswick, New Jersey, 1990

Major: Electrical Engineering

Professional Position Held:

Research Assistant, (Jan. 1992 - Aug. 1992)
Department of Electrical and Computer Engineering,
New Jersey Institute of Technology, Newark, New Jersey

This thesis is dedicated to
my parents

ACKNOWLEDGMENT

The author wishes to express his sincere gratitude to his supervisor, Professor Nirwan Ansari, for his guidance, friendship, and support in completing this thesis successfully. The author is grateful to Prof. Ansari for making him available for consultation whenever I needed his help.

Author extends his thanks to Dr. Joseph Frank and Dr. Zoran Siveski at NJIT for serving as members of the committee.

Author is grateful to the members of his family and his good friends for their continuous encouragement and support throughout this research work.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The introduction of satellite communications came as the solution when researchers in the field of communications were looking for ways to achieve greater coverage and capacity while the cost is not prohibitive.

With the advent of data communication, during the last two decades, there has been an explosion of growth in the need to communicate. Consequently the communication systems are also evolving to meet the demand. The satellite communication system plays an important role in both military and the civilian sector. The properties that are intrinsic to the satellite communication systems, such as broadcasting property and geographical flexibility explain the use of this system mainly in **W**ide **A**rea **N**etworks (**WAN**). With the commercial sector expanding especially with number of locations and distances between them becoming widespread, there is potential for growth in satellite communication networks to take advantage of their relatively constant cost (Rees 1990).

The introduction of optical communication medium has resulted in many fold increase in the speed with which processes communicate with each other. Optical medium is said to have given cable based long haul terrastial communication systems a second life which was being replaced by the satellite communication systems. Battle for the supremacy here is not over yet. In recent years space communications that employs optical communication links have received considerable attention (Kiasaleh 1991).

Even though optical medium gives enormous bandwidth, the power and signal processing aspects of a communication system places limits on such systems. Numerous schemes have been proposed to improve the efficiency of satellite communication networks. Among others, related works include channel allocation and scheduling. Many studies carried out for circuit switched telecommunication networks are also useful here since a wide spectrum of the applications in satellite networks are circuit switched networks. This is due to the large propagation delay of satellite networks.

The advantages of nonhierarchical networks over hierarchical networks are well known (Schwartz 1987). The **D**ynamic **N**on-**H**ierarchical **R**outing(**DNHR**) of AT&T (Ash 1990) and an adaptive routing algorithm known as Dynamic Control Routing (DCR) that is being considered by Telecomm Canada (Gerard and Bell 1989) are two of the most common examples for dynamic nonhierarchical networks. Several schemes have been proposed and analyzed by many in this field (Gersht and Shulman 1989; Kaufman 1981; Aein and Kosonych 1977; Akinpela 1984). From those studies, it is understood that, for circuit- switched networks, nonhierarchical dynamic routing performs better than the hierarchical, static routing. It is also shown that allowing alternate paths in nonhierarchical networks results in improved performance at low and moderate loads but suffer from instabilities at high and overload conditions. Some control mechanisms have been proposed and shown to overcome this instability problem. In one of these methods a call that is blocked on a direct path is tried on an alternate path with a probability. Another control scheme is to reserve a fraction of the capacity of a link for first routed traffic. In a study done in Mitra, Gibbens and Huang (1991), optimal trunk reservation is found for a fully connected, completely symmetric network. Work done on this thesis greatly benefited from the work of Mitra *et al.* (Mitra, Gibbens and Huang (1991).

A traffic management scheme is proposed here for improving the efficiency of circuit switched satellite communication networks of geostationary orbital type.

The proposed scheme incorporates the idea of dynamically adapting the network as well as dynamically routing each arrival. The proposed scheme allows the network to change according to the input pattern thus improving the grade of service (e.g. blocking rate) while maximizing the network utilization. The system is modeled such that it continuously organizes itself to minimize the cost for varying traffic conditions.

In the following chapter the proposed traffic management scheme is described in detail. Analytical model for the scheme is given in Chapter 3. Chapter 4 gives a brief introduction to the neural network based optimization technique - simulated annealing. The implementation of this technique in the proposed traffic management scheme is also included in Chapter 4. Analysis of the simulation as well as the analytical results are given in Chapter 5. Conclusions from this thesis work is in Chapter 6.

CHAPTER 2
NETWORK MODEL

The objective here is to design a network such that the network configuration will adapt itself towards minimizing the cost. In this work, for the purpose of illustration the overall network blocking rate is used as the cost of the network. Feasibility of this scheme is mainly due to the flexibility of satellite networks in allocating channels (trunks) to different links. The constraint here is that the total capacity of the network is finite and fixed. Thus the capacity of the links are constrained to a finite number of circuits in a given configuration.

The satellite network can be modeled as a mesh connected topology of a communication network as in Figure 2.1. The specific features of the satellite network are transparent.

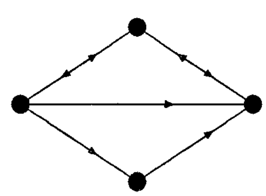

Figure 2.1: A Mesh Connected Network with 4 Nodes and 7 Links

Each node represents either a geostationary satellite or an earth station. The connections between nodes in the direction shown by the arrows denote the links between stations. All the connections and therefore the set of possible alternate routes for each origin-destination (O-D) pair are predefined. The proposed scheme can be explained with the help of the block diagram in Figure 2.2.

Functionally, the proposed traffic management scheme is divided into four modules. Based on the configuration given, which was the output of the Map generator module, Routing module dynamically route each call. The Controller performs the decision making process as to when to request for a new map. The arbitrator will decide whether to actually implement the change in configuration that is to be used by the routing module. The functions of these modules are further described in the subsequent sections.

2.1 Routing

The routing is performed dynamically for every call arriving at the network as follows.

1. An arriving call is first tried on the direct link
2. If direct link has no idle circuits then an alternate route is tried.
3. The choice of the alternate route is random.
4. If none of the alternate routes can accommodate this call it is then blocked and the call is lost from the network.

Here the routing module performs a simple routing function without much

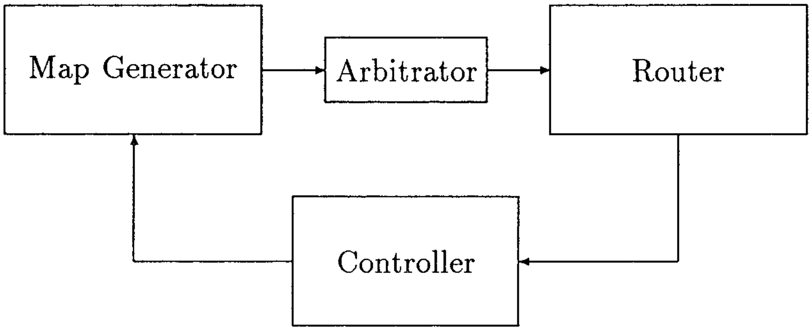

Figure 2.2: System Model

computation involved. This helps to reduce the processing delay of each call. The minimization of blocking rate comes mainly from having proper map in this module. There are other schemes which try to dynamically compute the alternate route that minimize the blocking rate. For instance, the least busy alternative (LBA) model in Mitra, Gibbens and Huang (1991). Here in this work the proper choice of map eliminates the need for computing the least busy alternate route. Also, such computations, as it will become evident in later chapters, will reduce the efficiency of the map generation process.

2.2 Map Generator

Average arrival rates for each origin-destination pair and the total capacity of the network along with the current status of the network are given as input to this module. Based on this information, using simulated annealing optimization technique, a map is produced as output. That is, simulated annealing technique is used to search for the map that best suits the current network conditions and constraints. Simulated annealing and its implementation in this work are illustrated in chapter 4. A map differs from another by two parameters namely $\vec{c}$ and $\vec{r}$. Vector $\vec{c}$ denotes the link capacities of the network and elements of vector $\vec{r}$ denotes the number of circuits that can be used by the alternately routed calls. That is, $c - r$ circuits of a particular link are reserved for the direct calls on that link. This parameter $\vec{r}$ is also referred to as *reservation parameter* in the text to follow.

2.3 Controller

The Controller keeps track of the network status and the performance. Network status is updated at regular intervals. Controller uses this information to decide whether a new map is needed. As mentioned before, link capacities and the reservation parameter are the variables in a map. Apart from the reservation scheme, previous studies have shown that proper assignment of link capacities improve the network performance. The optimal performance is obtained when the capacity assignment is such that all links saturate simultaneously (Kleinrock 1975). In this ideal case the flow to capacity ratio of all the links shall be equal to the average flow to capacity ratio of the network. Also, it is known that circuit reservation for first-routed calls are effective as control mechanism against instabilities at high load conditions (Akinpela 1984). Based on these observations, it is clear that a map should be configured by selecting appropriate values for link capacities and the reservation parameter.

We can reasonably expect to find a map with optimally minimum blocking rate by allowing both $\vec{c}$ and $\vec{r}$ to vary simultaneously. But this grows as a more complex combinatorial problem with the network size and also with the total capacity of the network. Other possibilities are changing either of the two parameters $\vec{c}$ and $\vec{r}$ while holding the other at a constant value. In this work, all three possibilities are considered. Further details and their relative merits are discussed in subsequent Chapters. The basis on which the Controller decides to call for a new map is described in detail in Chapter 5.

The function of the arbitrator is deferred until the end of Chapter 4. In the following chapter, the mathematical model is developed and this is used in optimizing the network by simulated annealing as explained in that chapter.

CHAPTER 3
ANALYTICAL MODEL

A queuing model is used to analyze the network. New arrivals to any O-D pair follow Poisson process. Holding times of calls are exponentially distributed. Each link is represented by an $M/M/m/m$ queuing model where m is the number of circuits in that link. To simplify the analysis, the average holding time of calls is assumed to be one time unit. The following assumptions are made.

1. All arrivals (including overflow calls) to any link form a Poisson process and are independent.
2. Link blocking probabilities are independent.
3. Processing and propagation delays are negligible.

The last of these three assumptions permit the study of the performance of the network solely based on the proposed scheme. Furthermore, this is close to real situation in circuit switched networks due to long average holding time. Even though calls arriving at an alternate route is clearly not Poisson, this is found to be a reasonable assumption especially when each link derives traffic from many end users. Similarly, link blocking independence assumption is found to be reasonable and this greatly reduced the complexity of the analysis. Detail discussion of the validity of these assumptions can be found in Gerard and Bell (1989). However, the first two assumption are not used in the simulations to be discussed in Chapter 5. The validity of these assumptions in this work are briefly discussed when analyzing the simulation results.

Some of the commonly used notations are listed below.

(i,j)	Link from node i to node j
$(i{-}j)$	Call originating from node i and destined for node j, also denotes the O-D pair from node i to node j
λ_{i-j}	External (new) arrival rate of $(i{-}j)$
γ_{ij}	Total arrival (new & overflow) rate to (i,j)
c_{ij}	Capacity (in number of circuits) of (i,j)
r_{ij}	Number of circuits in (i,j) that allow alternately routed calls
M_{i-j}	Set of tandem nodes that forms the alternate routes for $(i{-}j)$
R^k_{i-j}	An alternate route for $(i{-}j)$ with node k as the tandem node
B_{i-j}	Probability that call $(i{-}j)$ is blocked from the network
B_{ij}	Probability that any call is blocked in (i,j)
B^R_{ij}	Probability that an alternately routed call is blocked in (i,j)

Suppose the average network blocking probability is denoted by $\bar{B}$. This is obtained by summing all the call blocking probabilities and normalizing it by the total arrivals to the network.

$$\bar{B} = \frac{1}{\Lambda} \sum_{(i-j)} \lambda_{i-j} B_{i-j} \tag{3.1}$$

where Λ, the rate of the total input traffic to the network and is given by

$$\Lambda = \sum_{(i-j)} \lambda_{i-j} \tag{3.2}$$

Arrival rates λ_{i-j} are known quantities. Under the previously stated assumption of independent link blocking probabilities, B_{i-j} can be expressed in terms of B_{ij} and B_{ij}^R. A call $(i-j)$ is blocked from the network only when all available routes are occupied. Further, each alternate route is busy when either or both of the links constituting that route are busy *i.e.* they cannot take any more alternately routed calls. Hence the call blocking probability B_{i-j} is,

$$B_{i-j} = B_{ij} \prod_{m \in M_{i-j}} \left[1 - (1 - B_{im}^R)(1 - B_{mj}^R)\right] \tag{3.3}$$

Therefore, if the link blocking probabilities B_{ij} and B_{ij}^R are known, the average network blocking rate can be found by substituting (3.3) and (3.2) in (3.1). These probabilities are derived from the birth-death process of an $M/M/m/m$ queuing model (Kleinrock 1975).

Let p_i denote the steady state probability of having j circuits occupied in a given link. Let λ and v denote the arrival rates of new and overflow calls respectively. The $M/M/m/m$ queuing model can be described by the following set of equations. If the transition rate from state i_1 to state i_2 is denoted by $p(i_1, i_2)$ then,

$$p(i, i+1) = \begin{cases} \lambda & (c > i \geq r) \\ \lambda + v & (r > i \geq 0) \end{cases} \tag{3.4}$$

$$p(i, i-1) = \; i\mu \quad (c \geq i > 0) \tag{3.5}$$

With $\mu = 1$ and defining,

$$\gamma \doteq \lambda + v$$

$$\alpha \doteq \frac{\lambda}{\gamma}$$

following steady state probabilities are obtained from the balance equations for the above queuing model. Detail derivation is found in Kleinrock (1975).

$$p_i = \begin{cases} \frac{\gamma^i}{i!} p_0 & (0 \leq i \leq r) \\ \frac{\gamma^i (1-\alpha)^{i-r}}{i!} p_0 & (r < i \leq c) \end{cases} \tag{3.6}$$

where p_0 is found using the flow conservation equation, $\sum_{i=0}^{c} p_i = 1$ and is given below.

$$p_0 = \left[\sum_{i=0}^{r} \frac{\gamma^i}{i!} + \sum_{i=r+1}^{c} \frac{\gamma^i (1-\alpha)^{i-r}}{i!} \right]^{-1} \tag{3.7}$$

Substituting (3.7) in (3.6), p_i can be solved in terms of γ, α and c. The link blocking probabilities B and B^R are then related to these parameters via (3.8) and (3.9) below.

$$B = p_c \tag{3.8}$$

$$B^R = \sum_{i=r+1}^{c} p_i \tag{3.9}$$

Substituting (3.8) and (3.9) in (3.3), and from there $\bar{B}$ is computed from (3.1). But to solve for the link blocking probabilities B and B^R one needs to relate the unknown link arrivals γ and the parameter α to the known node-to-node arrivals λ and the reservation parameter r of all the links in the network. This is done by finding the flow f_{ij} on each link in two different methods. Knowing the link blocking probabilities, first flow is found by using γ and then by using λ and the overflow rate from other O-D pairs.

First, for a given map, i.e. with c and r for each link are specified, the flows f_{ij} of all links (i,j) can be computed from,

$$f_{ij} = \gamma_{ij}(1 - B_{ij}) \tag{3.10}$$

Alternatively, flows in each link can also be given in terms of the node-to-node traffic λ and the link blocking probabilities. Suppose, for link (i,j), the number of accepted direct calls are denoted by u_{ij} and the number of accepted alternately routed calls are denoted by y_{ij} then the flow on link (i,j) is,

$$f_{ij} = u_{ij} + y_{ij} \tag{3.11}$$

where

$$u_{ij} = \lambda_{i-j}(1 - B_{ij}) \tag{3.12}$$

With $f^k_{i'-j'}$ denoting the accepted calls (i',j') on alternate route $R^k_{i'-j'}$, y_{ij} in the above equation is given by,

$$y_{ij} = \left[\sum_{k:(i,j)\in R^k_{i'-j'}} f^k_{i'-j'} \right] \tag{3.13}$$

and

$$f^k_{i'-j'} = \lambda_{i'-j'} B_{i'j'} Q^k_{i'-j'} \tag{3.14}$$

where $\lambda_{i'-j'} B_{i'j'}$ is the rate of overflow of calls $(i'-j')$ from the direct link (i',j') and $Q^k_{i'-j'}$ is the probability that call $(i'-j')$ is not blocked on the alternate route

$R^k_{i'-j'}$ while it is blocked on all other alternate routes available for that O-D pair. This is related to the link blocking probabilities through the following equation.

$$Q^k_{i'-j'} = \left[\prod_{\substack{m \neq k \\ m \in M_{i'-j'}}} B(R^m_{i'-j'}) - \prod_{m \in M_{i'-j'}} B(R^m_{i'-j'}) \right] \tag{3.15}$$

where $B(R^m_{i'-j'})$ is the blocking probability of the alternate route $R^m_{i'-j'}$ and this can be expressed in terms of the link blocking probabilities for alternately routed calls of the links that form this alternate route as given below,

$$B(R^m_{i'-j'}) = \left[1 - (1 - B^R_{i'm})(1 - B^R_{mj'})\right] \tag{3.16}$$

Using the initial estimate of the quantities α, link arrivals γ are obtained from external arrival rates λ. Then using (3.4) through (3.9) the link blocking probabilities are computed. These probabilities are used to compute the flow by using (3.11) through (3.16). For these flows the new link arrivals can be computed from (3.10) and then a new set of α for each link from,

$$\alpha_{ij} = \frac{\lambda_{ij}}{\gamma_{ij}} \tag{3.17}$$

Above steps are repeated until the flows found in successive iterations converge. Finally, from the resulting blocking probabilities the desired quantity $\bar{B}$ is obtained from equations (3.1) through (3.3). This $\bar{B}$, the network blocking rate is used as the cost function in determining the best map in the Map Generator module as explained in the next chapter.

CHAPTER 4
NEW MAP GENERATION THROUGH SIMULATED ANNEALING

As mentioned in the previous chapter, a map that would minimize the cost function which is the blocking rate for the network, should be chosen. Since this quantity, $\bar{B}$, depends on number of circuits per link and on number of circuits reserved per link and these are independent variables the solution space for the problem of finding the best map is very large. Hence, the problem is a large combinatorial problem. This emphasizes the need of a powerful optimization technique. In this thesis a neural network based optimization technique called *simulated annealing* is used. Simulated annealing is described in the following section. The subsequent section gives more detail on generating and replacing a network configuration using simulated annealing.

4.1 Simulated Annealing

Simulated annealing is a stochastic computational technique to solve complex problems. This is first proposed by Kirkpatrik (1983). Since then many applications have been found in areas such as VLSI, image reconstruction, *etc.*

Suppose one needs to find the minimum of a function described by many variables. The basic idea here in reaching the minimum is to compute the value, call it the cost of the function, for different input and keep updating the current minimum value of the function till the optimally minimum value of the function is reached. But, suppose if the function has more than one minimum and not all of them are optimally minimum (they are referred to as local minima) then this simple method might as well reach a local minimum depending on the starting point chosen. So, there should be a way to prevent from getting stuck in a local minimum. To overcome this an occasional movement in the opposite direction is permitted. That is, when updating the current minimum with the new cost, a solution that is not better than the previous one is also accepted but with a probability. This shall provide a way out of a local minimum and eventually the procedure will result in the desired global minimum. This is the basis of what is called *stochastic relaxation method* developed by Metropolis, Rosenbluth A., Rosenbluth M., Teller A. and Teller E. (1953).

4.1.1 Analogy to Statistical Mechanics

Statistical mechanics is the study of a very large complex system with many interacting components at thermal equilibrium. In analogy to this, a large combinatorial problem is also a large complex system described by the objective or cost function with many parameters. In statistical mechanics, probability of finding the system in a particular state s, $P(s)$, is proportional to the Boltzman probability factor:

$$P(s) \;\; \alpha \;\; e^{-E(s)/kT} \tag{4.18}$$

where $E(s)$ is the energy of the system when at state s, k is the Boltzman constant and T denotes the temperature at which thermal equilibrium is maintained. Then p is defined as the ratio between the probabilities of finding the system in two states s_1 and s_2 is as following.

$$p = \frac{P(s_2)}{P(s_1)} = e^{-\frac{[E(s_2)-E(s_1)]}{kT}} \tag{4.19}$$

For a fixed quantity of kT which is always positive, if $E(s_2) < E(s_1)$, then we have $p > 1$, else $0 < p \leq 1$. Also, if the difference $\mid E(s_1) - E(s_2) \mid$ is large compared to kT then p is very small. Alternately, if the difference in energy is held constant, higher the value of T, the smaller the value of p which can be seen as the "relative probability" of finding the system between two states. Therefore, higher energy states are more probable at higher temperature than at lower temperature. Hence the temperature is gradually reduced lower energy states become more likely and the high energy states become much less likely to be found. This behavior of a physical system is what simulated in the simulated annealing. There, a control parameter is used as "temperature" to gradually increase the probability of finding the solution for the problem described by the cost function.

4.1.2 The "Simulated" Annealing Process

As mentioned in the preceding discussion, in simulated annealing the behavior of the system under consideration is simulated by stochastic relaxation method in analogous to that of an annealing process of statistical mechanics. Suppose the system is in state s_1 at time t. Next, another possible state s_2 is generated randomly. If the energy, or the cost, of the second state $E(s_2)$ is less than that of the first, $E(s_1)$, then the state s_2 is accepted as the state at time $(t + 1)$ with probability one. But, if $E(s_2) \geq E(s_1)$, then the new state s_2 is accepted with probability p as in (4.19). This provides a way out of local minimum. Note that the larger the difference between the energies of the states lower the probability of being accepted. It has been shown that as $t \rightarrow \infty$, the probability of finding the system in a particular state converges to Boltzman distribution. Detailed discussions on the convergence of simulated annealing type algorithms and other similar algorithms can be found in (Gemen and Gemen 1984; Gidas 1985; Gelfand 1987).

In thermal annealing process the temperature is elevated to a higher level and then gradually reduced so that the system attains the lowest energy level in the steady state. It is important to spend sufficient time at each temperature level to attain the optimally low energy state. In analogous, proper choice of the control parameter (corresponding to the temperature) and the "annealing" schedule is vital in optimizing the system as given by the objective function using simulated annealing.

4.2 Map Generation

This section concerns how the simulated annealing method is implemented in the proposed traffic management scheme. The cost function, the solution space and the annealing schedule must be defined when using the simulated annealing. Each of these aspects are described in the following subsections.

4.2.1 Block Rate as Cost Function

In reality there are many parameters, such as power, available bandwidth of the individual links and also the block rate of the network, that need to be taken into account when optimizing a satellite network. In this work, the blocking probability of the network is chosen to illustrate how this method can be implemented to optimize the network. Since the simulated annealing, as described above, only requires the proper cost function and the the careful choice of an annealing schedule, by choosing an appropriate cost function any parameter that affects the network performance can be included in optimizing the network performance. With the assumptions stated in the beginning of Chapter 3 are holding, the cost function

$E(s)$ is then as given by eqn.(3.1). Rewriting this here,

$$E(s) = \bar{B}(\vec{c}, \vec{r}) \tag{4.20}$$

where

$$\vec{c} = \{c_{ij} : (i,j) \in all\ links\}$$

and

$$\vec{r} = \{r_{ij} : (i,j) \in all\ links\}$$

4.2.2 Solution Space

Another important aspect is the solution space that is defined for this problem. The network performance, measured by the network throughput, is attempted to optimize by having the best map for the Routing module. Given the network is mesh connected and the rates with which calls arrive at the nodes are non-symmetric, the two parameters $\vec{c}$ and $\vec{r}$ can vary arbitrarily from one state of the network to another. Therefore, theoretically any combination of $\vec{c}$ and $\vec{r}$ is a state in the solution space. But, including all these possible states of the network in the solution space will not be effective in real time application of the simulated annealing process. Some measures are taken to limit the solution space to a reasonable subset of the entire state space. Implementation of this is discussed in the next chapter.

Three different methods are experimented here in generating states for the simulated annealing. The first method allows only the link capacities $\vec{c}$ varied from state to state. In other words, $\vec{c}$ is the independent variable in the cost function of (4.20). In the second method, the $\vec{r}$ is the independent variable. The third is the combination of the above two. That is, both $\vec{c}$ and $\vec{r}$ are varied simultaneously and independently. Therefore, third method will result in a larger solution space while the length of the vectors $\vec{c}$ and $\vec{r}$ remains the same, which is equal to the number of links in the network. Relative efficiency of these methods are discussed in the next chapter.

4.2.3 Annealing Schedule

Once the cost function and the solution space is defined the next task is to choose proper annealing schedule so that the simulated annealing algorithm will quickly converge to the desired solution. For simplicity let k in (4.19) be equal to 1. When choosing values for the control parameter that corresponds to the temperature as discussed in the previous section two things need to be given close attention. One, the distance between two successive temperature levels and the other is the value of this parameter in relative to the difference in the the possible range of the costs of two states in the solution space, $\Delta(E) = E(s_2) - E(s_1)$. Since the size of the network affects $\Delta(E)$, and the constraint set forth in the beginning is the total capacity of the network, the control parameter T is related to C, the total capacity of the network. Also, in order to have more temperature levels as the annealing nears a solution and larger steps at the beginning where most of the states are accepted as a possible solution the parameter T is expressed as a nonlinear function given below.

$$kT = T = aC/ln(j) \tag{4.21}$$

where a is a constant and j is the iteration index, which is incremented linearly to produce the desired temperature schedule. For different experiments and for various networks, a is adjusted to obtain the proper annealing schedule.

CHAPTER 5
NETWORK SIMULATION AND DISCUSSION OF RESULTS

The implementation of the proposed traffic management scheme is simulated as described in this chapter. Then the results from the simulation as well as the analytical results are analyzed and discussed in detail.

5.1 Network Simulation

A reasonably sized network that is chosen as an example for the analysis here is shown in Figure 5.1. This mesh connected network is non-symmetric in terms of link capacities as well as the arrival rates into the nodes. There are eleven nodes and 47 links in it, and also 47 origin-destination pairs in this example. Each of 47 origin destination pairs have predefined alternate routes to them. Number of alternate routes is different from one O-D pair to the other, ranging from 0 (no alternate routes) to 4 in the experimental network. Since the alternate routes are predefined O-D pairs that use a particular link for alternate routing are also predefined. These information can be derived from Figure 5.2 where an "1" indicates the existence of a link from the node that is given by the row to the node that is given by the column.

The total number of circuits for the network is varied in different experiments. These are stated more specifically when describing the simulation results.

External arrivals to any O-D pair are simulated to follow Poisson distribution and the service time of the calls are simulated to follow exponential distribution with mean 1.0 time unit per arrival. Arrival rates to the O-D pairs are varied in different experiments.

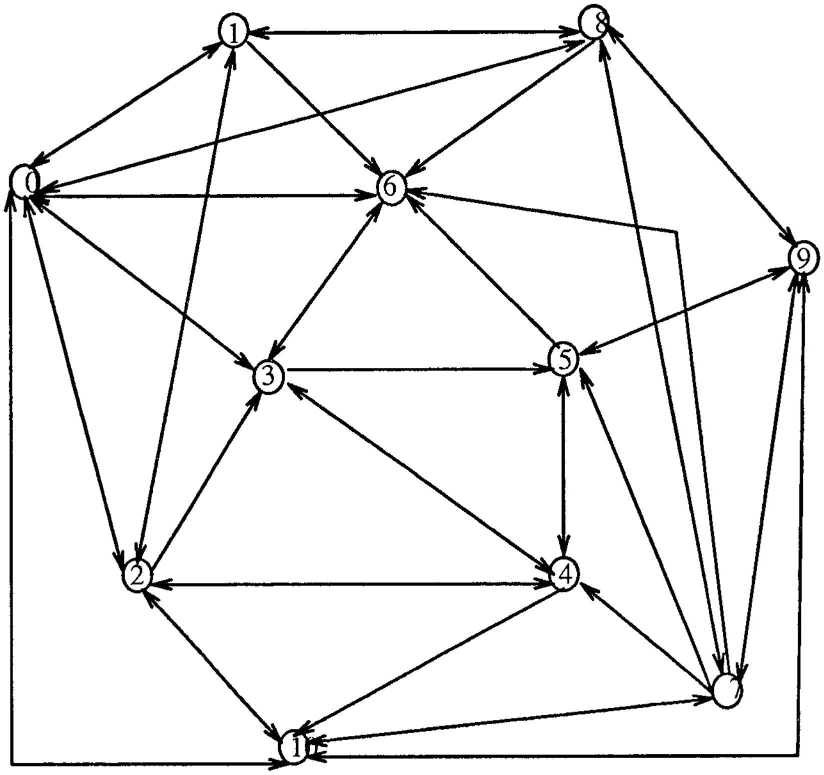

Figure 5.1: Network used in simulation

Node j

Node i	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1	1	0	0	1	0	1	0	1
2	1	0	1	0	0	0	1	0	1	0	0
3	1	1	0	1	1	0	0	0	0	0	1
4	1	0	0	0	1	1	1	0	0	0	0
5	0	0	1	1	0	1	0	0	0	0	1
6	0	0	0	0	1	0	1	0	0	1	0
7	1	0	0	1	0	0	0	1	0	0	0
8	0	0	0	0	1	1	0	0	1	1	1
9	1	1	0	0	0	0	1	1	0	1	0
10	0	0	0	0	0	1	0	1	1	0	1
11	1	0	1	0	0	0	0	1	0	1	0

Figure 5.2: Connections Between Links of the Network

5.1.1 Function of the Controller

Controller of Figure 2.2 is fed with information of the current status of the network at regular time intervals. Also at regular intervals this module updates some measurements of the network performance and the network status. One of them is the rate of arrivals to each O-D pair in the network. Another is the load balance of the network. Let Δ denotes the ratio of the network's total flow to the total capacity and let δ_{ij} denotes the ratio of flow to capacity of link (i, j). At each update a measure of the network's load imbalance, denoted by d, is computed from these ratios using sum-square-error method as given below.

$$\Delta = F/C$$

$$\delta_{ij} = f_{ij}/c_{ij}$$

$$d = \frac{1}{C}\sum_{(i,j)}(\Delta - \delta_{ij})^2 \qquad (5.20)$$

where C and F are total capacity and the total flow of the network respectively. Therefore the parameter d indicates the amount by which the network's current load balance deviates from that of the fully balanced network. This measure of disparity in the network load is taken as the indication of potential premature saturation of the network. There is a threshold value, d_t, defined with respect to the network load balance Δ. When the measure of network's load imbalance d is larger than the threshold value d_t then the network is considered to be operating in an inefficient state for the current traffic condition. In the simulations done here the threshold value is defined as follows.

$$d_t = (0.1) \times \Delta$$

Once the value of d gets larger than the threshold value d_t the Controller module calls the Map Generator module to come up with a better map for the current traffic condition. When d_t is small and close to the origin then even a small deviation of network load balance from the ideal condition is not tolerated. In this situation Map Generator is called very frequently. But too frequent changes in the configuration may not be cost effective when considering other costs associated with changing the configuration of a satellite network. The traffic condition used for generating the map is derived from the recent past experience of the network. In the simulations here network status is updated at the end of every time unit and the parameter d is computed after 10 updates. Up to 100 time units are used in measuring the traffic pattern.

5.1.2 Arbitrator

The function of the arbitrator of Figure 2.2 is simple. This is basically a cost saving measure. The routing of calls must be continued uninterrupted and also the optimization by simulated annealing has to be done in real time. Hence there may be some instances where the map configured from the recent past experience of the network may not actually reflect the optimal performance for the present traffic conditions if the traffic pattern changes too quickly. Therefore,the Arbitrator's function is to get the map that is generated and actually change the current map of the network only if it will result in better performance. The above situation can in fact be eliminated to a certain extent by properly choosing the length of the duration for which the network is monitored for the purpose of measuring the network traffic condition and also by properly choosing the tolerance level i.e by reducing the threshold value δ_t.

5.2 Discussion of Results

Simulation 1

First simulation is intended to show that allowing alternate routes in a non hierarchical network results in better performance of the network at moderate load conditions. For this the network in Figure 5.1 is simulated with each link having 20 circuits. The arrival rates into the O-D pairs are equal. The reason is to study the effect of the alternate route alone. The network is first simulated with no alternate routes allowed and then calls any O-D pair were allowed to use the predefined alternate routes if the direct link is found busy, *i.e.* there are no idle circuits on the direct link. As evident from the simulation results depicted in Figure 5.3, the network throughput has improved when alternate routes were made available. All the measurements in this and the subsequent plots are normalized to the total capacity C of the network. All measurements of arrival rates and the throughput have *number of calls per one time unit* as their units. Notice that the gain in throughput at the low load (near 0.55 calls per time unit) is not significant. And, when the normalized arrival rate goes into the high load region again the difference in the throughput becomes very small. The cases of high and over loaded conditions are treated later.

Simulation 2

Here the network is simulated under similar conditions as described above except each link now has 100 circuits. Results are given in Figure 5.4. Note the improvement in throughput in the moderate load condition when alternate routes are allowed. In both of these simulations the alternate routes were not subject to any reservation policy. Since typically networks operate at moderate load conditions allowing alternate routes are clearly advantageous.

Simulation 3

For the same network as used in Simulation 1, the measured throughput at high and overloaded condition is depicted in Figure 5.5. It shows that not only the advantage of alternate routes is lost at these load conditions but also the alternate routes deteriorates the performance and network becomes more and more instable. The explanation for this behavior is, when the offered load goes into the heavy load region, many calls get blocked on direct links and subsequently routed via an alternate route. Every alternately routed call occupies two links in the network thus further increasing the probability of the future calls being blocked from the network and as a consequence the network soon becomes saturated.

Simulation 4

Similar results are found for the case of Simulation 2 at high and overloaded conditions. These are given in Figure 5.6. And this shows that the network saturate prematurely because of the alternate routes. Next, reservation scheme is used with alternate routes and is shown to be effective in preventing the instability that is experienced at higher load.

Simulation 5

In order to alleviate the instability and to improve the performance of the network with alternate routes at high and over loaded conditions reservation scheme as explained in previous chapters is introduced. The network is simulated as described in Simulation 2 above but now with some portion of the capacity of each link reserved for direct calls. With each link in the network having 100 circuits the reservation placed on all the links are also equal. Of course there may be a better solution by having different amount of reservation for each link. These cases are treated later in this section. But, even in this case there is significant improvement in the performance of the network with alternate routes. In addition to the two cases of no alternate routes and with alternate routes but nothing reserved, two more cases, one with 10% of the link capacities and the other with 50% of the link capacities are reserved are plotted in Figure 5.7. The results for the case where 50% of the circuits are reserved ae similar to that of no alternate routes at all. This is expected since for the throughput above 0.7 calls per time unit, on the average links have more than 70 calls (average holding time is 1.0 time unit) and therefore having 50 circuits reserved does not affect the performance significantly. When 10% of the link capacities are reserved this shows better performance observed at the moderate load region and it performs as good as the one with no alternate routes in the high and overload regions. But, notice that network is more stable for this case than any of the other cases plotted in Figure 5.7.

Simulation 6

Results similar to the one described in simulation 5 were obtained when the network is simulated with 20 circuits per link instead of 100 circuits. Simulation results are given in Figure 5.8.

Comparing the best cases of the last two simulations, *i.e.*when 10% of the circuits are reserved for direct calls, having more number of circuits per each link improves the normalized throughput as seen in Figure 5.9. This is because of the arrival process and the service time of calls are not deterministic. For instance, with less number of capacities on each link a call with longer holding time make the network's blocking probability higher than it would in a network whose link capacities are higher. This is seen better in Figure 5.10 where the block rate of the network is plotted against the arrival rate both normalized to the total capacity of the network. For the case of 100 circuits per link the block rate stays close to zero until the normalized arrival rate is increased to 0.8 calls per time unit.

For all of the above stated simulations analytical results were also obtained.

The analytical results agree in most cases very closely with the simulation results. But when permitting all the circuits in a link for alternately routed calls then the analytical results do not agree with the simulation results. The throughput of the network of 100 circuits per link is plotted in Figure 5.11 for three different amount of reservation of the link capacities for direct calls. As seen in the figure when there is no reservation, the analytical results do not agree with the simulation results. The explanation for this is when there is no reservation scheme, at higher offered load the network will have many calls that are alternately routed. Each alternately routed call occupies two links of the network. And with no reservation scheme in place many of the calls found in the network at heavy load will be alternately routed calls. The arrival process of alternately routed calls into a node is in fact not Poisson. Also the link blocking probabilities for alternately routed calls are not independent since there is correlation between the blocking on a direct link and the blocking on an alternate route. Since the analytical results are under these assumptions that all arrivals form Poisson process and all link blocking probabilities are independent, and these assumptions are not used in the simulations, these assumptions become invalid when the amount of alternately routed calls in the network is more than the number of direct routed calls. Therefore in the heavy load situation the result for the case where no reservation scheme is used, simulation results do not agree with the analytical results.

In all the simulations above the network used here has equal amount of circuits in all the links. And for this case the 10% reservation throughout the links are found to perform better than any other cases considered. Even better performance can be expected by "tuning" the parameters $\vec{r}$ and $\vec{c}$ for the traffic condition. But since the network is mesh connected and with number of alternate routes available for each O-D pair are not equal, finding the point where the performance is optimal is not simple. In the next simulations simulated annealing is used and attempted to find a map that will result in better performance than what is achieved here.

Simulation 7

The network is simulated as described in Simulation 2 (20 circuits per link). Now the threshold value d_t is set to $(0.1 \times \Delta)$ as described in the beginning of this chapter. Initially, the routing was performed with 10% of the link capacities reserved for direct calls. The simulated annealing was performed with $\vec{r}$ being the independent variable in the cost function of (4.18). Since the arrivals to O-D pairs are equal and the number of circuits per link is equal for all the links in this case $\vec{r}$ is expected to be more "sensitive" parameter than $\vec{c}$. The results are plotted in Figure 5.12. The annealing is done for four different arrival rates between 0.8 and 1.1 calls/time unit (normalized). For these arrival rates the network is simulated and the normalized throughput is again measured when the routing is done with the map that was selected by the Map Generator module. In all four cases the throughput showed improvement as can be seen from the plot. The time spent at each temperature during annealing is 2350 trials. The generation of new states is stopped and the next temperature level is tried when either the number of accepts at a temperature level reaches 470 or the total trials (generation of new states) reaches 2350. Table (5.1) gives the number of accepts at each temperature cycle. Values here are obtained when annealing is done for the normalized arrival rate equals to 0.8 calls.

Simulation 8

Next the network is simulated with arbitrary arrival rates into the nodes, ranging from 0.3 to 1.4 calls per time unit, normalized to the total capacity of the network. Initially the routing was done with the network having 20 circuits on each link and

Table 5.1: Annealing Schedule 1

Temp. cycle	No. of accepts
1	470
2	108
3	116
4	14
5	15
6	2
7	0
8	0
9	0

10% of this is reserved for the direct calls. Then with d_t having the same value as in the previous simulation, the network is simulated with only the $\vec{c}$ of the cost function of (4.18)) is varied. Number of circuits reserved in each link was held constant at 10% for this simulation. The measured throughput for the network for the initial map and for the new map after the annealing are plotted in Figure 5.13. The map that is chosen by the simulated annealing method shows better performance for all four arrival rates for which the simulation is done. As far as the simulation is concerned, the time that had to be spent in a given temperature level was double that of the previous case, 4700 trials compared to 2350 trials, to get the algorithm to converge at a good solution. In this simulation, the next temperature cycle is begun when the number of accepts at a temperature cycle reaches 940 or when number of total trials reaches 4700. Table (5.2) gives the number of accepts at each temperature cycle for the case where the normalized arrival rate of the network is 0.754341 calls.

Table 5.2: Annealing Schedule 2

Temp. cycle	No. of accepts
1	940
2	940
3	940
4	831
5	376
6	111
7	33
8	0
9	0
10	0

Simulation 9

Here the third method of annealing is tried on the network that was used for the first method of annealing. In this simulation, number of circuits per link and the reservation parameter were allowed to vary simultaneously and independent from each other. The annealing was performed starting with all links having 20 circuits and all of them were being allowed for alternately routed calls. Even after spending much more time at each temperature, 9400 states generated in contrast to 2350 and 4700 in the first and the second methods respectively, the algorithm did not converge to a good solution. This indicates that the annealing needs to be done for longer time since we know there are better solutions at least for the four cases as plotted in Figure 5.12. The validity of the assumptions as stated in Chapter 3 and described in this chapter may also have affected the results of this simulation since the network was begun with all the circuits being allowed for alternately routed calls. Therefore, in comparing the previous two methods this method is not as good as the other ones at least in the time it takes to converge to a good solution. This could also be due to the large solution space.

Simulation 10

In this simulation, the network is simulated as described in Simulation 10 except the network is initially assigned with link capacities that are proportional to the arrival rates to that particular O-D pair. The throughput after annealing here is compared with the the throughput after annealing as described in Simulation 8. This is plotted in Figure 5.14. There is a slight increase in the throughput. Also the time needed at each temperature was half of the one in Simulation 10. The number of accepts at each temperature cycle for the case with normalized arrival rate equal to 0.754341 calls is given in Table (5.3). Fast annealing is observed here.

Table 5.3: Annealing Schedule 3

Temp. cycle	No. of accepts
1	470
1	470
3	205
4	43
5	7
6	0
7	0
8	0
9	0

From the last four simulation results it is understood that adapting the network for the current traffic condition results in increased throughput. Also, it has been shown that the optimization technique, simulated annealing, can be used to adaptively change the network configuration as the input pattern for the network is varied. From the results obtained in this work it is found that when configuring a new map it is efficient to start the annealing algorithm with the map having link capacities that are proportional to the respective arrival rates into the nodes. Then changing the value of the vector $\vec{r}$ to find the optimal map that will result in a configuration that best suits the traffic condition.

In the simulations, choosing the proper annealing schedule was important in arriving at a good solution. In all the simulations, when annealing was performed, the annealing temperature had to be varied.

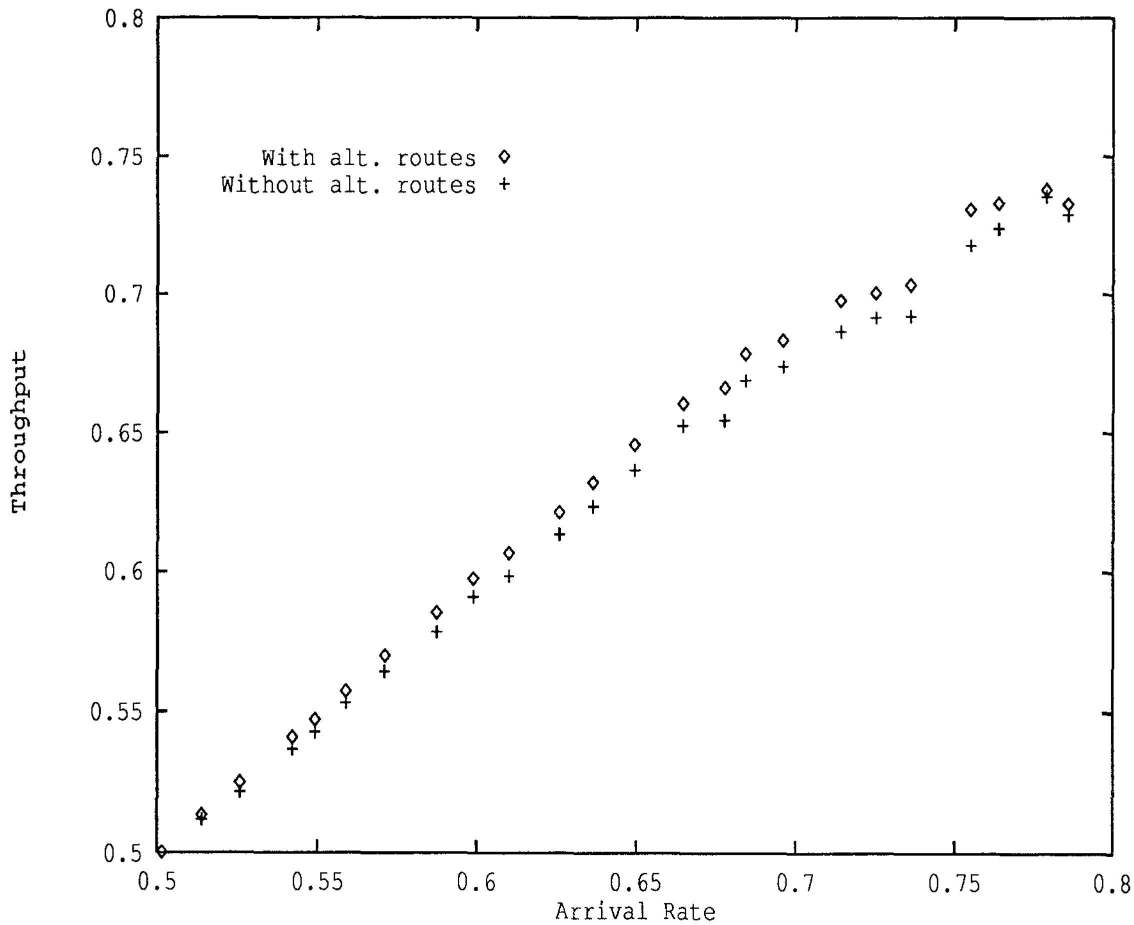

Figure 5.3: Network Performance at Moderate Load, 20 circuits/link.

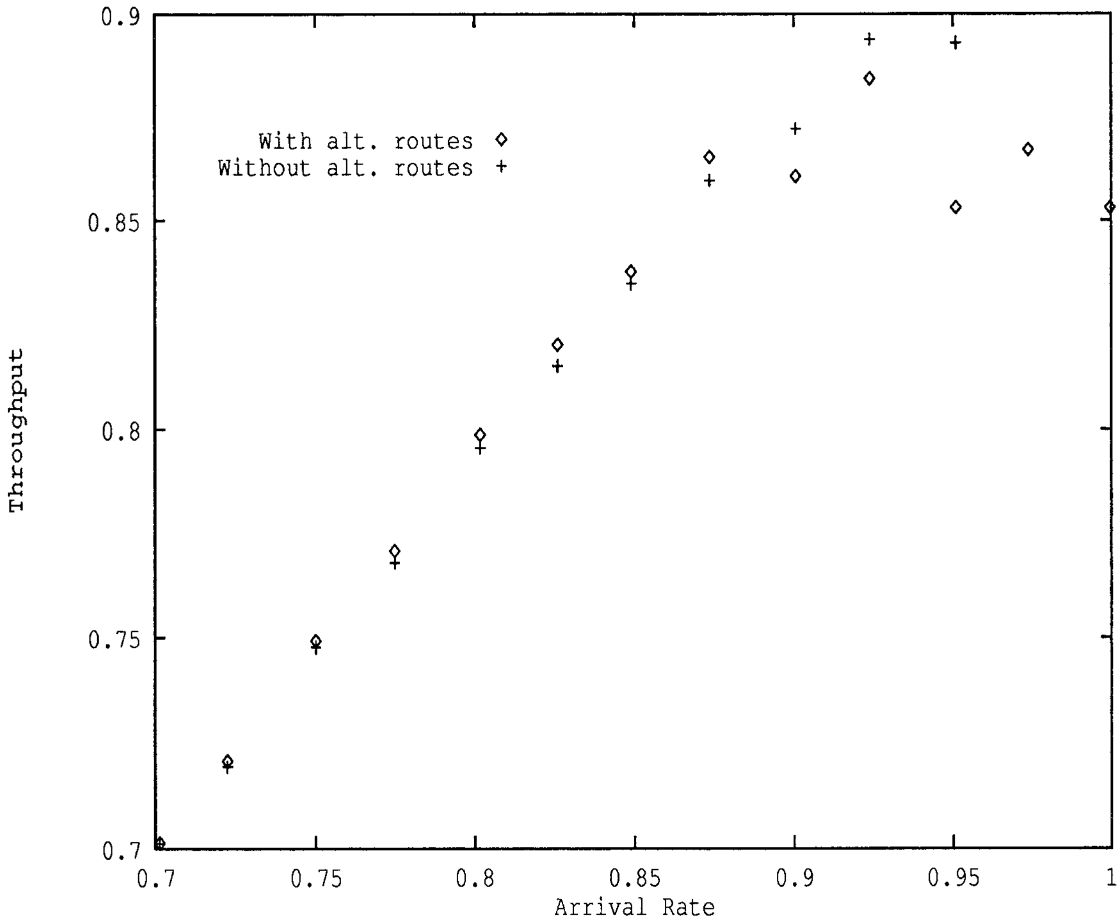

Figure 5.4: Network Performance at Moderate Load, 100 circuits/link.

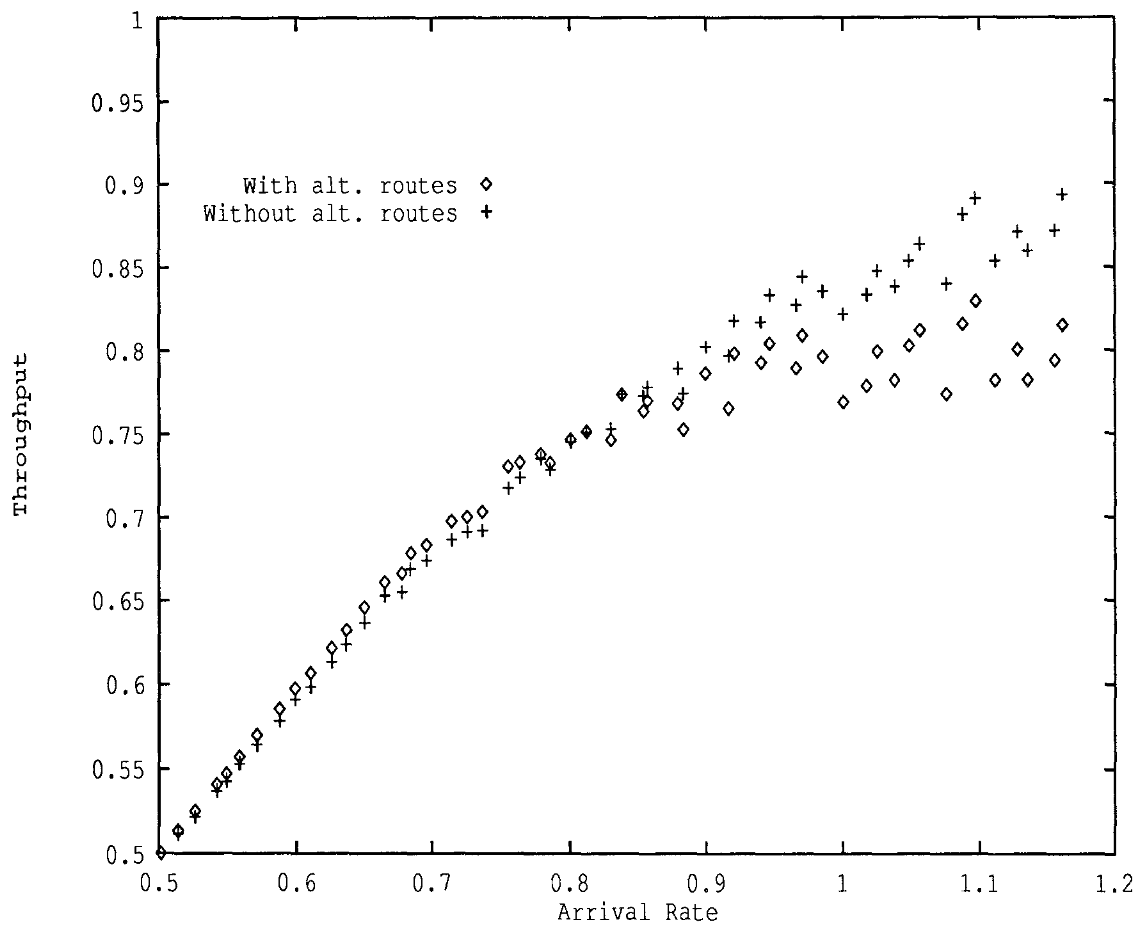

Figure 5.5: Network Performance at Overload, 20 circuits/link.

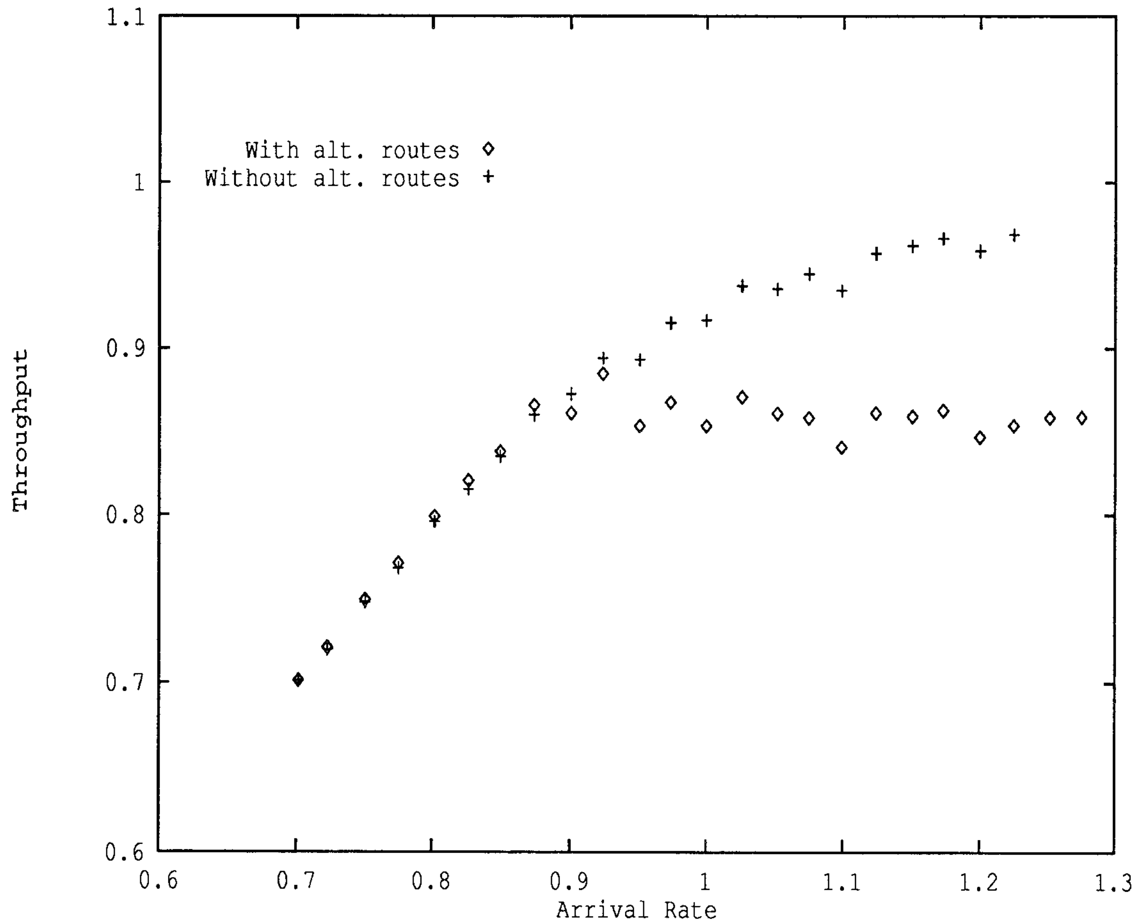

Figure 5.6: Network Performance at overload, 100 circuit/link.

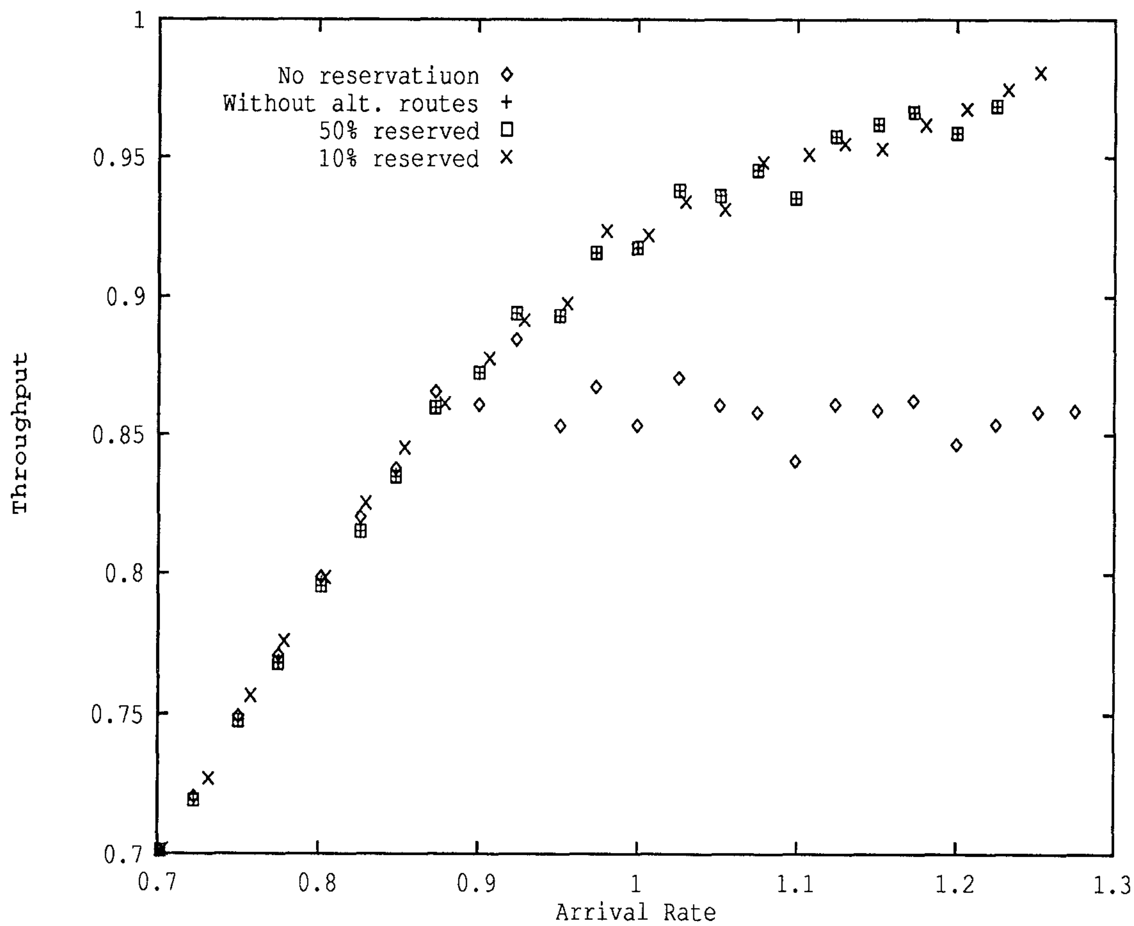

Figure 5.7: Network Performance with Reservation Scheme, 100 circuits/link

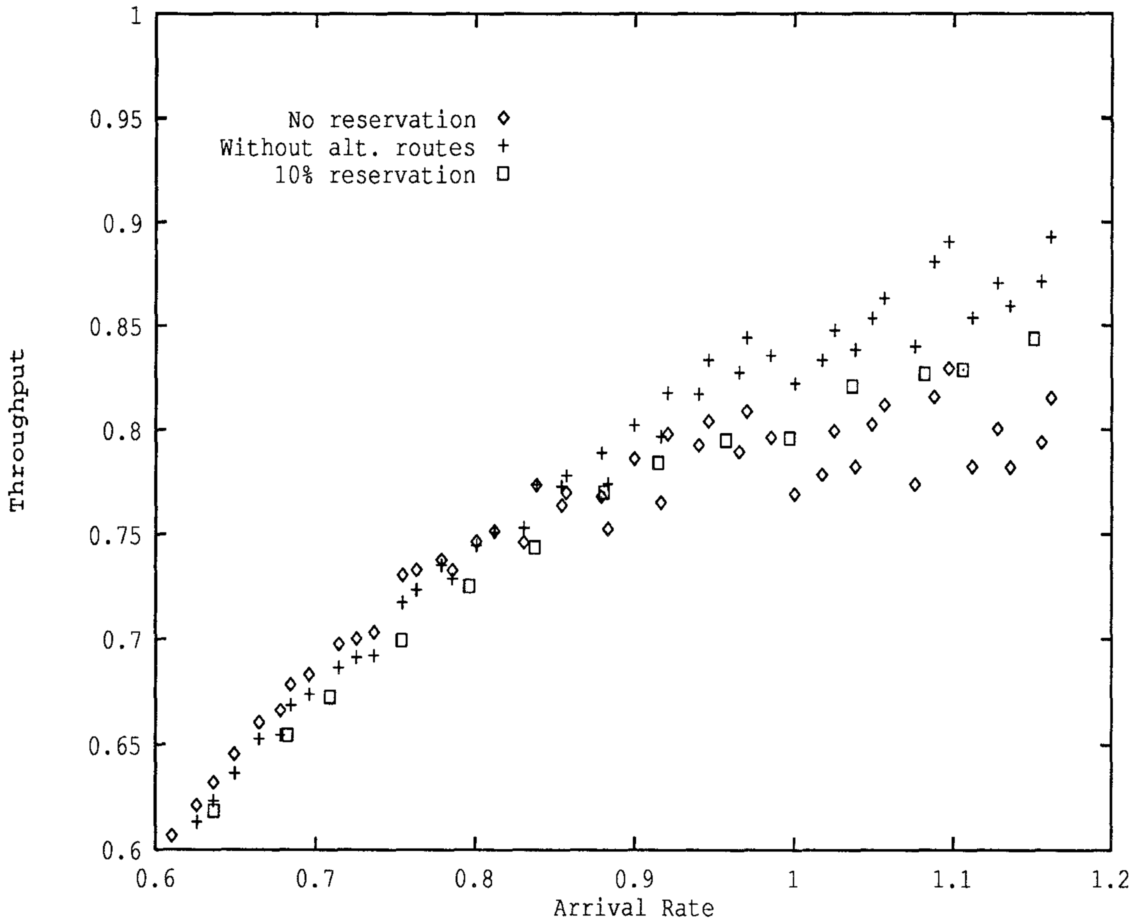

Figure 5.8: Network Performance with Reservation Scheme, 20 circuits/link

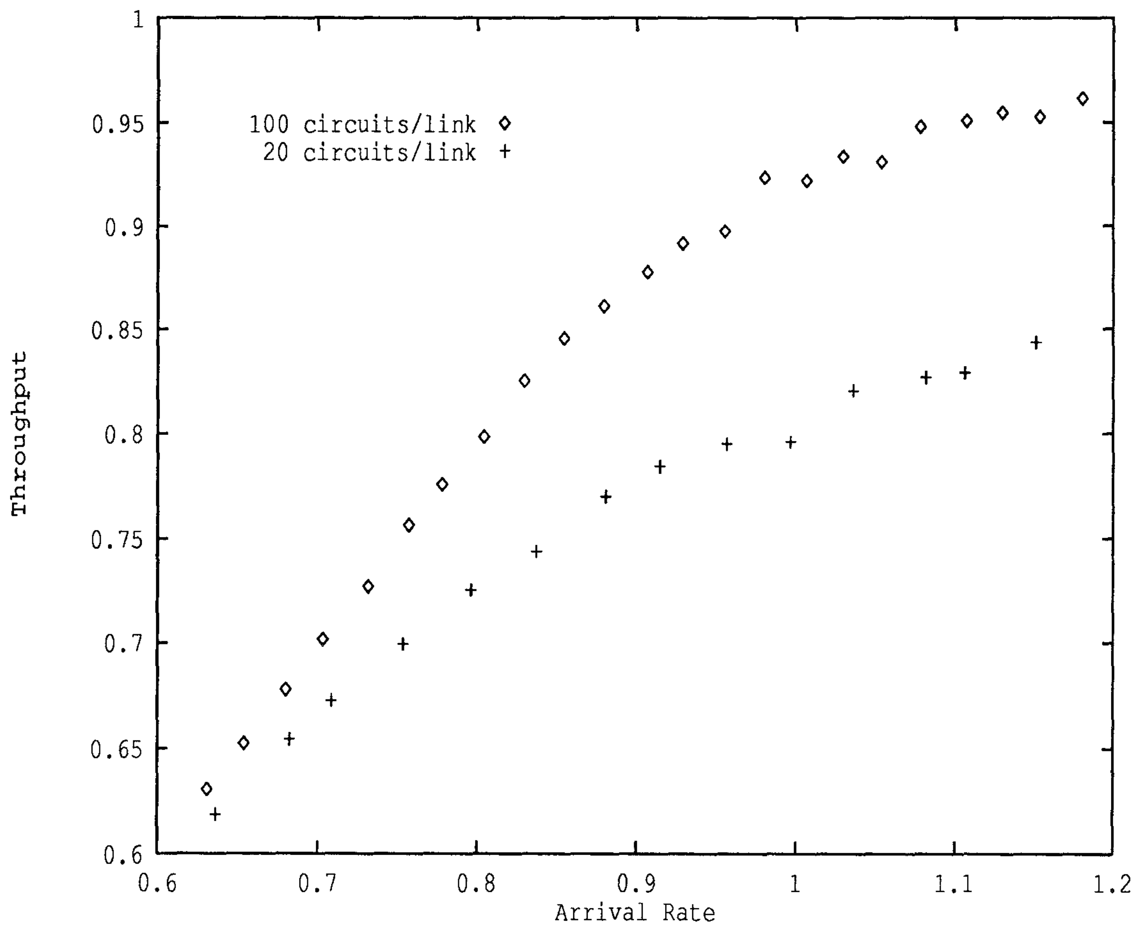

Figure 5.9: Effect of Network Size on Throughput

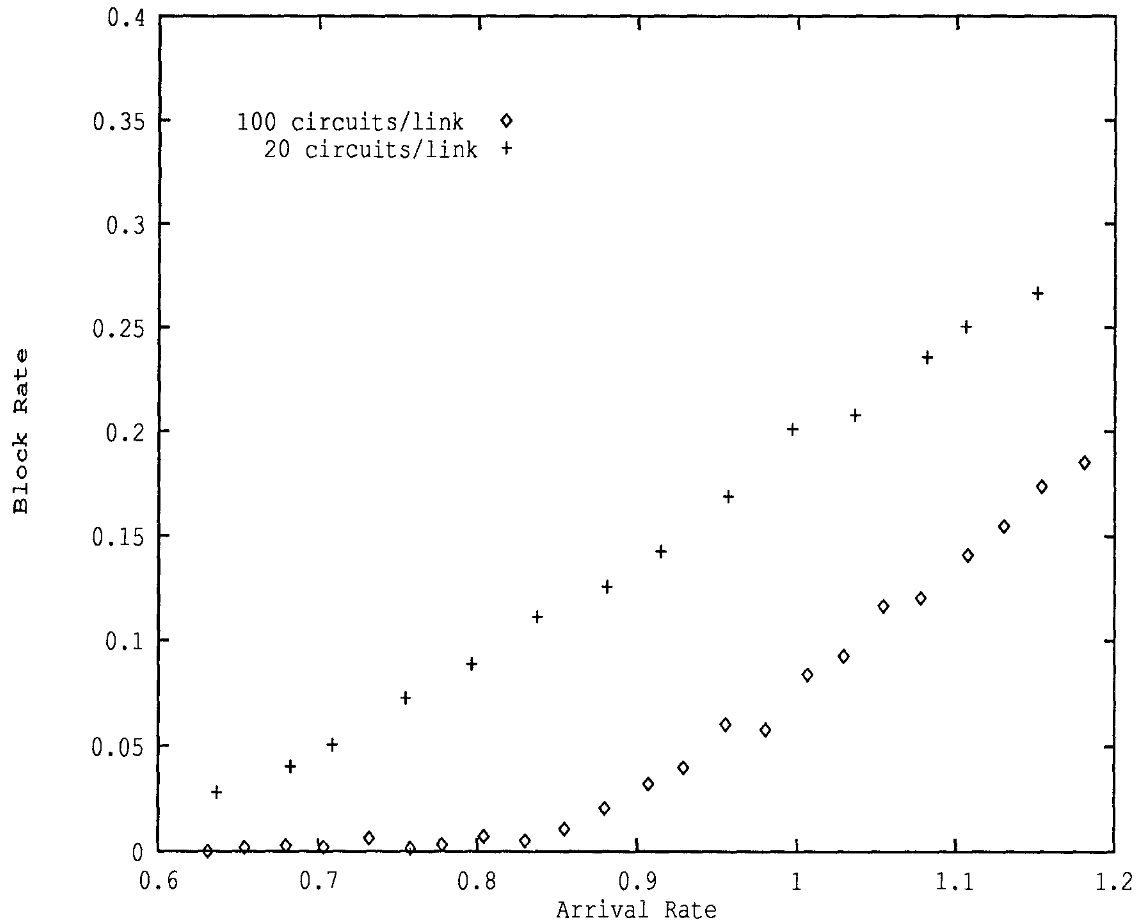

Figure 5.10: Effect of Network Size on Bolck Rate

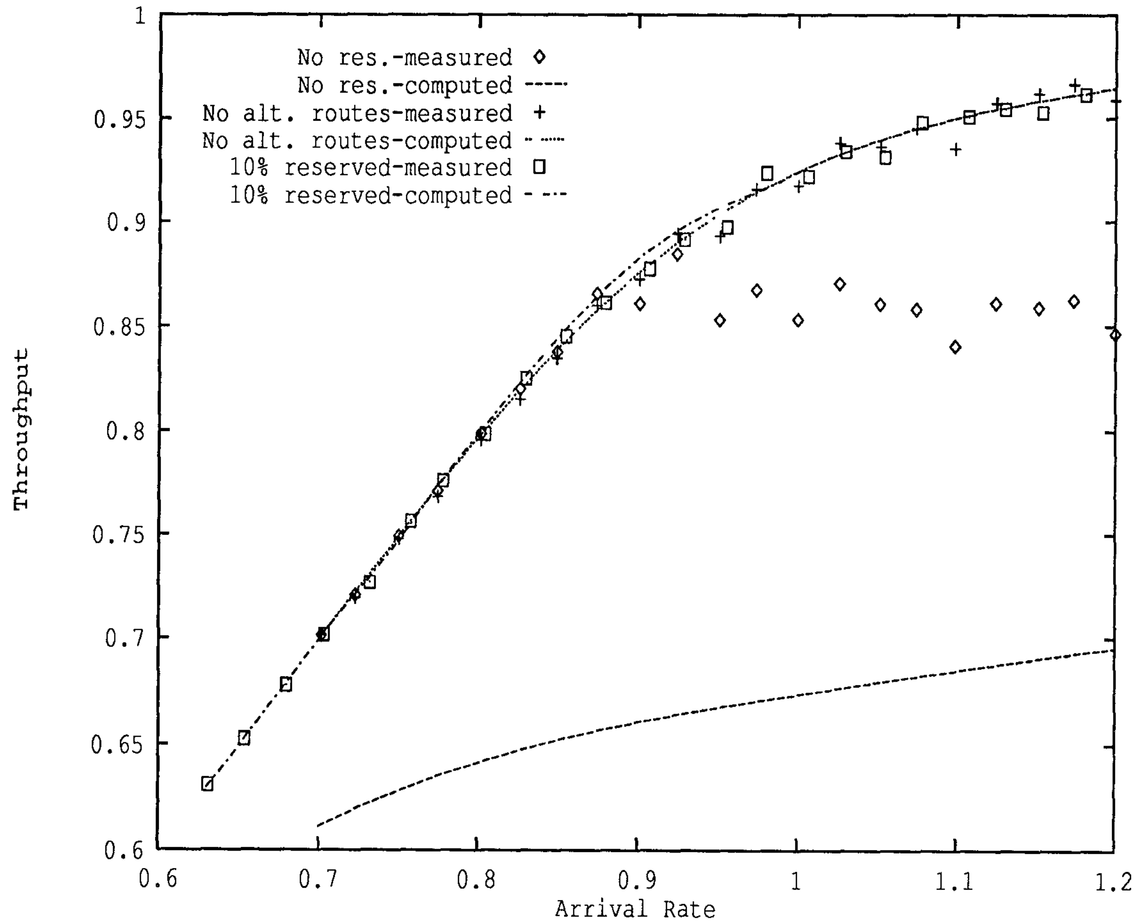

Figure 5.11: Analytical Results Compared to Simulation Results.

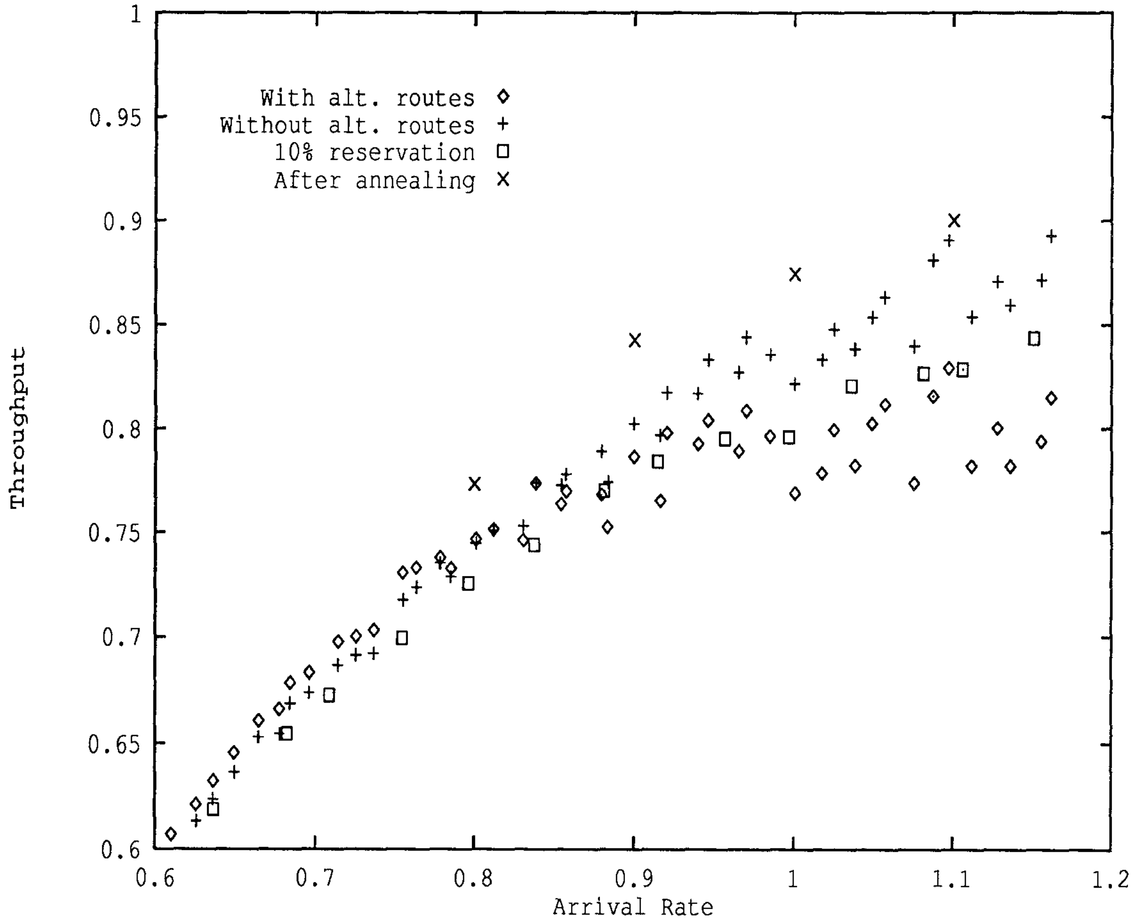

Figure 5.12: Performance After Annealing with $\vec{r}$ as Variable.

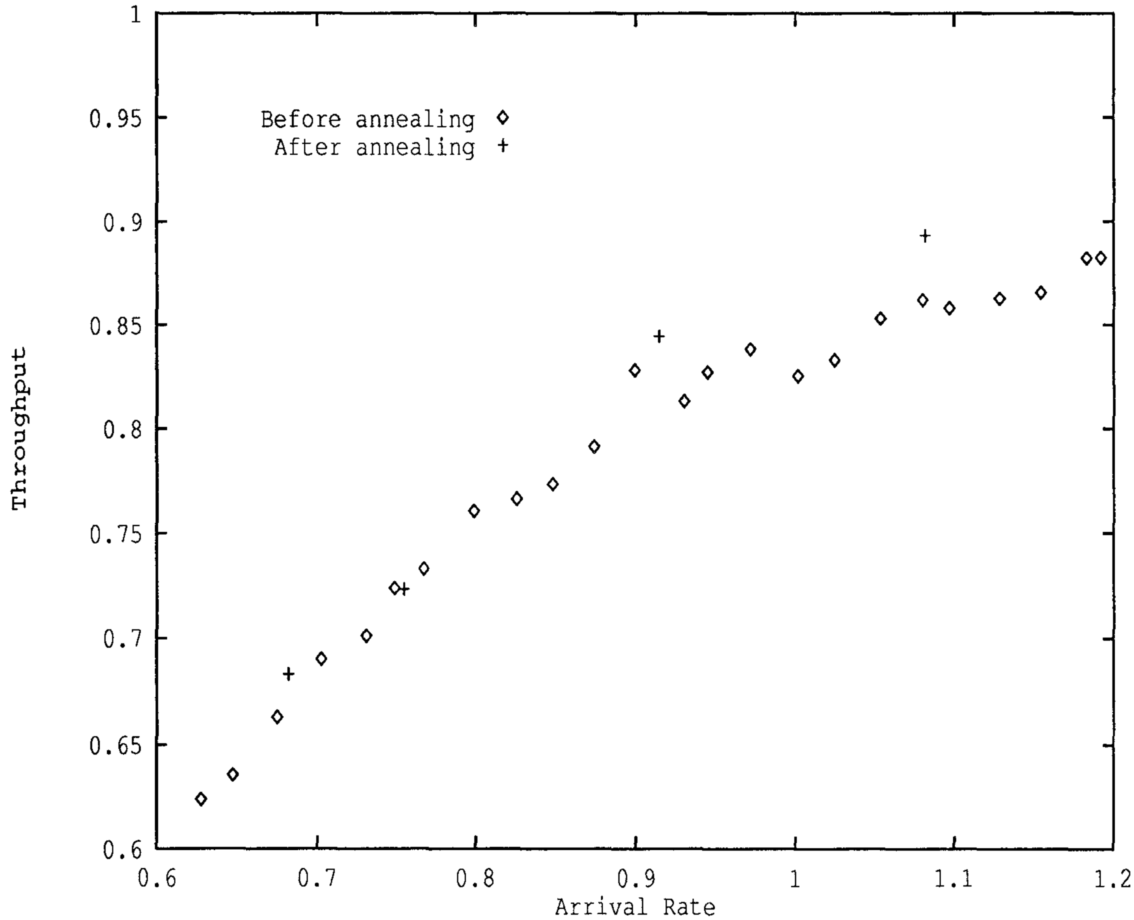

Figure 5.13: Performance After Annealing with $\vec{c}$ as variable.

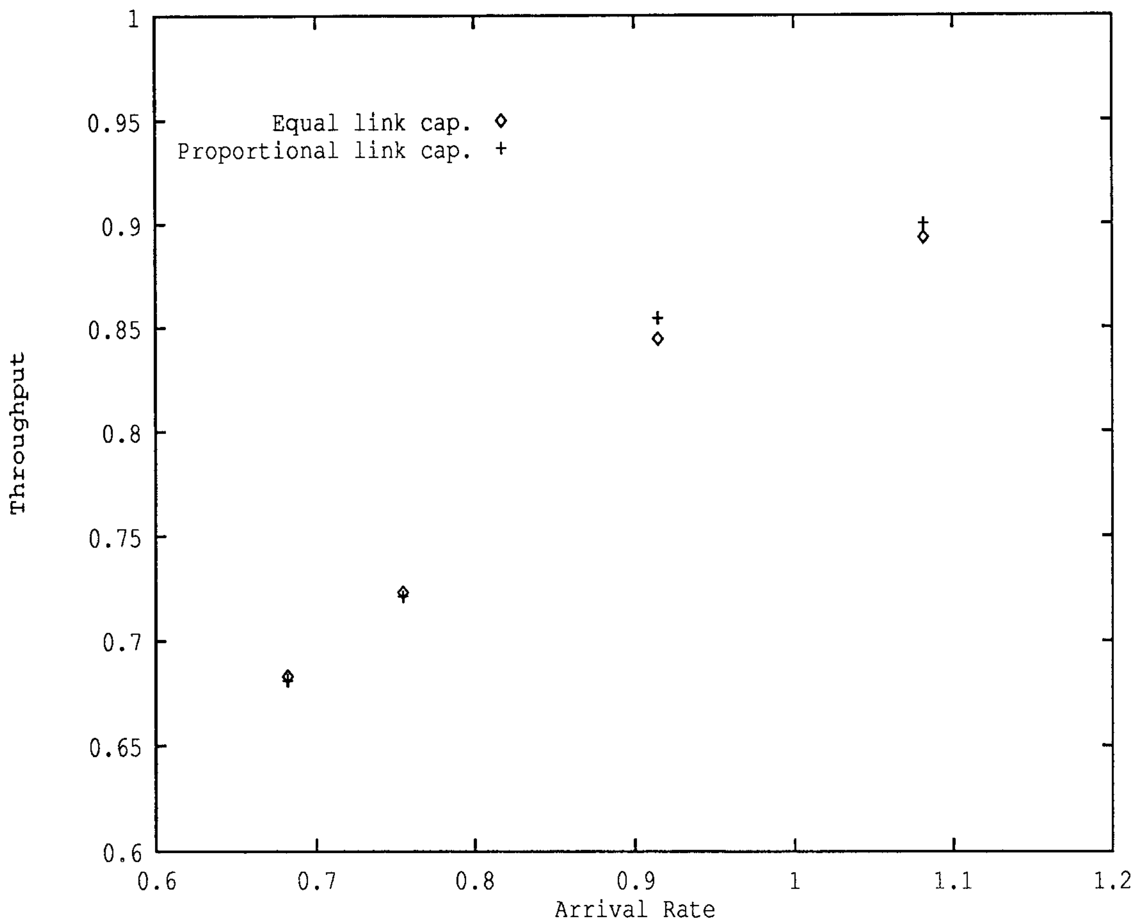

Figure 5.14: Effectiveness of $\vec{c}$ and $\vec{r}$ in Simulated Annealing.

CHAPTER 6
CONCLUSION AND SUGGESTIONS FOR FURTHER STUDY

As in other circuit switched telecommunication networks, having alternate routes showed improvement in the performance of a satellite network at moderate load conditions. But when having alternate routes, the network becomes unstable as the offered load is increased to heavy load region and after a critical point the performance deteriorates. In order to prevent this undesirable effect a control scheme is introduced where some portions of the link capacities is reserved for routing direct calls. Results obtained here shows that this is an effective control mechanism in avoiding the above stated instability. Also use of this scheme resulted in better throughput. Largely due to the feasibility of changing the link capacities in a satellite communication network with ease compared to terrestrial networks, this network was able to adapt to the current traffic pattern while routing each call dynamically. For the traffic pattern measured by the network, the performance of the network was optimized by changing the configuration adaptively. The optimization technique *simulated annealing* is used to find the configuration that best suits the current traffic needs. Results indicate the effective use of simulated annealing in optimizing the performance of the satellite network. After the network is assigned with link capacities that are in proportion with the arrival rates of the corresponding O-D pairs, the optimal reservation parameter is found by simulated annealing to "fine tune" the configuration.

One of the important features of the proposed traffic management scheme is the flexibility with which the factors that affect the network performance can be included. This is done effectively by properly defining the cost function for the simulated annealing. An added feature of the scheme in improving the throughput of the network here is that only the external arrival rates were used and this was easily measured from the network.

Further studies may be done in this field to optimize the network faster using other optimization techniques. Other types of communication networks, cellular communication networks for example, where the facilities can be changed without increasing the cost significantly can be tried to optimize their performance using the scheme that is proposed in this work.

REFERENCES

Rees, D.W.E. *Satellite Communications: The First Quarter Century of Service.* John Wiley & Sons, Inc. New York, 1990.

Kiasaleh, K. "PN Code-Aided Ranging for Optical Satellite Communication Systems ." *IEEE Transactions on Communications*, vol. 39, no. 2, 1991.

Schwartz, M. *Telecommunication Networks: Protocols, Modelling and Analysis.* Addison–Wesley Publishing Co., 1987.

Ash, G. R. " Design and Control of Networks with Dynamic Non-Hierarchical Routing. " *IEEE Communications Magazine*, vol. 28, no. 10, 1990, pp. 34–40.

Gerard, A. and M. Bell. " Blocking Evaluation for Networks with Residual Capacity Adaptive Routing. " *IEEE Transactions on Communications*, vol.37, no. 12, 1989, pp. 1372–1380.

Gersht, A. and A. Shulman. " Optimal Routing in Circuit Switched Communication Networks. " *IEEE Transactions on Communications*, vol. 37, no. 11, 1989, pp. 1203–1211.

Kaufman, J.S. " Blocking in Shared Resource Environment." *IEEE Transactions on Communications*, vol.29, no.10, 1981, pp. 1474–1481.

Aein, J.M. and O.S. Kosonych. " Satellite Capacity Allocation." *Pro ceedings of IEEE*, vol. 65, no. 3, 1977, pp. 332–342.

Akinpela, J.M. "Over-Load Performance of Engineering Networks with Non-Hierarchical and Hierarchical Routing." *AT&T Bell Labs Tech Journal.*, vol. 63, no. 7, 1984, pp. 1261–1281.

Mitra, D., R.J. Gibbens and B.D.Huang. " State-Dependent Routing on Symmetric Loss Networks with Trunk Reservations I." *AT&T Bell Laboratories*, Murray Hill, NJ.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller. "Equations of State Calculations by Fast Computing Machines." *J. of Chemical Physics*, no. 21, pp. 1087-1091, 1953.

Kleinrock, L. *Queing Systems.* vol. 1, John Wiley & Sons, NY.

Kirkpatrick, S., C.D. Gelatt Jr., and M.P. Vecchi. "Optimization by Simulated Annealing." *Science* 220: 671-680.

Gemen, S. and D. Gemen. "Stochastic Relaxation, Gibbs Distribu tions and Baysian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, 1984, pp. 721–741.

Gidas, B. "Non-stationary Markov Chains and Convergence of the Annealing Algorithm." *J. of Statistical Physics*, no. 39, 1985, pp. 73–131.

Fujioka, M. "Universal Service Creation and Provision Enviornment for Intelligent Network." *IEEE Communications Magazine*, vol. 29, no. 1, 1991, pp. 44–51.

Jabbari, B. Goode and R. Fike. "Intelligent Networks: Toward Implementation in Fixed and Mobile Networks." *IEEE Communications Magazine*, vol. 1, no. 2, 1992, pp. 9–10.

Cameron, H., P. Galloy and W.J. Graham. "Report on the Toronto Advanced Routing Concept Trial." *Network Planning Symposium*, 1979, pp. 171–177.

Tanenbaum, A.S. *Computer Networks*, Prentice–Hall Inc., 1988.

Kummerle, K. and H.Rudin. "Packet and Circuit Switching: Cost/Performance Boundaries." *Computer Networks*, vol. 2, 1978.

Kleinrock, L. *Queuing Systems*, vol. 2, John Wiley & Sons, NY.

Gelfand, S.B. "Analysis of Simulated Annealing Type Algorithms." *PhD Thesis*, Massachusetts Institute of Technology, 1987.