

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

**A Bell & Howell Information Company
300 North Zeeb Road Ann Arbor MI 48106-1346 USA
313 761-4700 800 521-0600**

**PERFORMANCE FEATURES OF OPTIMALLY CONFIGURED GLOBAL
NETWORKS**

by
Ahmet Mete K k

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1996

UMI Number: 9618080

**Copyright 1996 by
Kok, Ahmet Mete**

All rights reserved.

**UMI Microform 9618080
Copyright 1996, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© 1996
AHMET METE KOK
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

12/15/94

Date

SAHamed

Chair of Examining Committee

8/31/95

Date

Shant, 2/1/95

Executive Officer

Professor Tom Wesselkamper

Professor George S. Moschytz

Dr. Victor B. Lawrence

Dr. Jean-Jacques Werner

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract**PERFORMANCE CHARACTERISTICS OF OPTIMALLY CONFIGURED GLOBAL NETWORKS**

by

Ahmet Mete Kk

Adviser: Professor Syed V. Ahamed

Global telecommunication networks are becoming reality as we approach twenty-first century. These networks currently carry information in the form of voice and data. As the usage of computers in the world increases, there is a greater demand for faster, more efficient and sophisticated digital communication networks which transport data, audio, video and other forms of information. These networks differ from the traditional analog networks both in functionality and operation. This thesis identifies the fundamental elements of a network, such as link characteristics, traffic patterns and topology that are crucial to a well functioning, efficient and scalable global networks which can handle the demands of the future applications. It also investigates the theoretical foundations for a framework of general guidelines for future designers of global networks.

TABLE OF CONTENTS

I. INTRODUCTION	1
II. CONCEPTS AND METHODOLOGY	13
2.1 Network Concepts	13
2.1.1 Topology	13
2.1.1.2 Classification of Network Topologies	13
2.1.1.3 Fully Connected Topology	14
2.1.1.4 Hierarchical (Tree) Topology	15
2.1.1.5 Ring Topology	16
2.1.1.6 Bus Topology	17
2.1.2 Traffic Flow	18
2.1.2.1 Concepts of Traffic Flows and Network Control	18
2.1.2.2 Traffic Entities	18
2.1.2.3 Traffic Attributes	22
2.1.2.4 Network Connectivity	23
2.1.2.5 Traffic Paths and Routing	25
2.1.2.6 Traffic Control	26
2.1.3 Transmission and Signalling	29
2.1.3.1 Transmission	29
2.1.3.2 Signalling	43
2.1.4 Network System Performance	46
2.2 Research Methodology	48
2.2.1 Network Analysis	49
2.2.2 Modelling and Simulation	49
2.2.3 Evaluation	50
III. NETWORK ANALYSIS	51
3.1 Analytical Techniques	51
3.1.1 Queuing Theory	52
3.1.2 The Poisson Process	55
3.1.3 Network Attributes	58
3.1.3.1 Network Connectivity	58
3.1.3.2 Network Flow	59
3.1.3.3 Network Capacity	59
3.2 Network Analysis	60
3.2.1 Packet Switched Networks	60
3.2.1.1 Link utilization	60
3.2.1.2 Packet and Message Delay	61
3.2.1.3 Transmission cost	62
3.2.1.4 Routing	62

3.2.1.5 Traffic	62
3.2.2 Circuit Switched Networks	63
3.2.2.1 Call connect time	63
3.2.2.2 Blocking Probability	64
3.3 Node Analysis	64
3.3.1 Critical Resources	65
3.3.2 Other Resources	67
IV. MODELLING AND SIMULATION METHODOLOGY	68
4.1 Statistical Distributions	69
4.1.1 Arrival rate distribution	69
4.1.2 Service time distribution	69
4.2 Simulation model elements	70
4.2.1 Global Network Model Characteristics	73
4.2.2 Topology of Simulated Networks	75
4.2.3 Simulation Traffic	78
4.2.4 Nodes	84
4.2.5 Links	87
4.2.6 Routing	93
4.3 Network Performance Measures	93
V. SIMULATION RESULTS.	95
5.1 Categories of results	95
5.2 Simulation Results	96
5.2.1 First Simulation Model Results.	96
5.2.1.1 Full Duplex Links	97
5.2.1.1.1 9.6 kb/s Links	97
5.2.1.1.2 56 kb/s Links	100
5.2.1.1.3 380 kb/s Links	103
5.2.1.1.4 1544 kb/s Links	106
5.2.1.2 Half Duplex Links	109
5.2.1.2.1 9.6 kb/s Links	109
5.2.1.2.2 56 kb/s Links	112
5.2.1.2.3 380 kb/s Links	115
5.2.1.2.4 1544 kb/s Links	118
5.2.1.3 Link Failure	121
5.2.1.3.1 9.6 kb/s Links	121
5.2.1.3.2 56 kb/s Links	124
5.2.1.3.3 380 kb/s Links	127
5.2.1.3.4 1544 kb/s Links	130
5.2.2 Second Simulation Model Results	133
5.2.2.1 Full Duplex Links	133
5.2.2.1.1 9.6 kb/s Links	133
5.2.2.1.2 56 kb/s Links	136

5.2.2.1.4 1544 kb/s Links	140
5.2.2.2 Link Failure	143
5.2.2.2.1 9.6 kb/s Links	143
5.2.2.2.2 56 kb/s Links	146
5.2.2.2.3 380 kb/s Links	149
5.2.2.2.1 1544 kb/s Links	152
VI. DISCUSSION OF RESULTS	155
6.1 Network Results for the First Simulation Model	155
6.1.1 Throughput	155
6.1.2 Message Delay Times	166
6.1.2.1 Average Message Delay	166
6.1.2.2 Maximum Message Delay	177
6.1.3 Percent of Messages over Three Seconds	185
6.2 Network Results for the Second Simulation Model	192
6.2.1 Throughput	192
6.2.2 Message Delay	200
6.2.2.1 Average Message Delay	200
6.2.2.2 Maximum Message Delay	208
6.2.3 Percent of Messages over Three Seconds	215
6.3 Further Analysis and Findings	222
6.3.1 Throughput Separation	222
6.3.2 Throughput Saturation	245
6.3.2 Other Factors That Affect Performance	259
VII. CONCLUSIONS	261
7.1 Summary of Findings	261
7.2 Further Research Areas	263
APPENDIX	264
REFERENCES	267

LIST OF TABLES

Table 1.1	BITNET/NORTHNET/EARN/JNET Connectivity	8
Table 1.2	Telecommunication services present and future	9
Table 6.1	Point of Separation occurrences for full duplex links	204
Table 6.2	Points of separation for Model One	207
Table 6.3	Points of separation for Model Two	207
Table 6.4	Traffic intensity coefficients of throughput for Model One	209
Table 6.5	Traffic intensity coefficients of throughput for Model Two	209
Table 6.6	Link speed coefficients of throughput for Model One	212
Table 6.7	Link speed coefficients of throughput for Model Two	212
Table 6.8	Traffic Intensity coefficients at point of separation	214
Table 6.9	Traffic Intensity coefficients at point of separation	215
Table 6.10	Point of saturation for full duplex links (Model One)	219
Table 6.11	Points of saturation for Model One	220
Table 6.12	Points of saturation for Model Two	220
Table 6.13	Throughput coefficients for point of saturation for Model One	227
Table 6.14	Rate of change of p coefficients for Model One	228
Table 6.15	Throughput coefficients for point of saturation for Model Two	229
Table 6.16	Rate of change of p coefficients for Model Two	229

LIST OF FIGURES

Figure 1. 1: The Intelligent Network Components	3
Figure 1. 2: Number of computer installations in the world	5
Figure 1. 3: Percentage of world main lines	6
Figure 1. 4: Main lines per 100 population	7
Figure 2. 1: Fully Connected Topology	14
Figure 2. 2: Hierarchial Topology	15
Figure 2. 3: Ring Topology	16
Figure 2. 4: Bus Topology	17
Figure 2. 5: Frame Types	21
Figure 2. 6: Connection Matrices for different topologies	24
Figure 3. 1: M/M/1 Queueing Model	59
Figure 4. 1: Simulation Block Diagram	80
Figure 4. 2: Global BITNET Connectivity	83
Figure 4. 3: Topology for First Simulation Model	85
Figure 4. 4: Topology for Second Simulation Model	86
Figure 4. 5: Traffic activity for May 1993 at CUNY gateway	88
Figure 4. 6: CUNY Gateway Traffic Activity for May 6, 1993	89
Figure 4. 7: CUNY Gateway Traffic Activity for May 6, 1993	90
Figure 4. 8: Packet Switch Node Organization	93
Figure 5. 1: Full Duplex with 9.6 kb/s links	108
Figure 5. 2: Full Duplex with 9.6 kb/s links	109
Figure 5. 3: Full Duplex with 9.6 kb/s links	110
Figure 5. 4: Full Duplex with 56 kb/s links	111
Figure 5. 5: Full Duplex with 56 kb/s links	112
Figure 5. 6: Full Duplex with 56 kb/s links	113
Figure 5. 7: Full Duplex with 380 kb/s links	114
Figure 5. 8: Full Duplex with 380 kb/s links	115
Figure 5. 9: Full Duplex with 380 kb/s links	116
Figure 5.10: Full Duplex with 1544 kb/s links	117
Figure 5.11: Full Duplex with 1544 kb/s links	118
Figure 5.12: Full Duplex with 1544 kb/s links	119
Figure 5.13: Half Duplex with 9.6 kb/s links	120
Figure 5.14: Half Duplex with 9.6 kb/s links	121
Figure 5.15: Half Duplex with 9.6 kb/s links	122
Figure 5.16: Half Duplex with 56 kb/s links	123
Figure 5.17: Half Duplex with 56 kb/s links	124
Figure 5.18: Half Duplex with 56 kb/s links	125
Figure 5.19: Half Duplex with 380 kb/s links	126
Figure 5.20: Half Duplex with 380 kb/s links	127
Figure 5.21: Half Duplex with 380 kb/s links	128
Figure 5.22: Half Duplex with 1544 kb/s links	129

Figure 5.23: Half Duplex with 1544 kb/s links	130
Figure 5.24: Half Duplex with 1544 kb/s links	131
Figure 5.25: Link Failure with 9.6 kb/s links	132
Figure 5.26: Link Failure with 9.6 kb/s links	133
Figure 5.27: Link Failure with 9.6 kb/s links	134
Figure 5.28: Link Failure with 56 kb/s links	135
Figure 5.29: Link Failure with 56 kb/s links	136
Figure 5.30: Link Failure with 56 kb/s links	137
Figure 5.31: Link Failure with 380 kb/s links	138
Figure 5.32: Link Failure with 380 kb/s links	139
Figure 5.33: Link Failure with 380 kb/s links	140
Figure 5.34: Link Failure with 1544 kb/s links	141
Figure 5.35: Link Failure with 1544 kb/s links	142
Figure 5.36: Link Failure with 1544 kb/s links	143
Figure 5.37: Full Duplex with 9.6 kb/s links	144
Figure 5.39: Full Duplex with 9.6 kb/s links	146
Figure 5.40: Full Duplex with 56 kb/s links	147
Figure 5.41: Full Duplex with 56 kb/s links	148
Figure 5.42: Full Duplex with 56 kb/s links	149
Figure 5.43: Full Duplex with 380 kb/s links	150
Figure 5.44: Full Duplex with 380 kb/s links	151
Figure 5.45: Full Duplex with 380 kb/s links	152
Figure 5.46: Full Duplex with 1544 kb/s links	153
Figure 5.47: Full Duplex with 1544 kb/s links	154
Figure 5.48: Full Duplex with 1544 kb/s links	155
Figure 5.49: Link Failure with 9.6 kb/s links	156
Figure 5.50: Link Failure with 9.6 kb/s links	157
Figure 5.51: Link Failure with 9.6 kb/s links	158
Figure 5.52: Link Failure with 56 kb/s links	159
Figure 5.53: Link Failure with 56 kb/s links	160
Figure 5.54: Link Failure with 56 kb/s links	161
Figure 5.55: Link Failure with 380 kb/s links	162
Figure 5.56: Link Failure with 380 kb/s links	163
Figure 5.57: Link Failure with 380 kb/s links	164
Figure 5.58: Link Failure with 1544 kb/s links	165
Figure 5.59: Link Failure with 1544 kb/s links	166
Figure 5.60: Link Failure with 1544 kb/s links	167
Figure 6.1: Throughput for Full Duplex links	170
Figure 6.2: Throughput for Half Duplex links	171
Figure 6.3: Throughput for Link Failure	171
Figure 6.4: Throughput for 9.6 kb/s links	174
Figure 6.5: Throughput for 56 kb/s links	174
Figure 6.6: Throughput for 380 kb/s links	175
Figure 6.7: Throughput for 1544 kb/s links	175

Figure 6.8: Average Message Delay for Full Duplex links	177
Figure 6.9: Average Message Delay for Half Duplex links	177
Figure 6.10: Average Message Delay for Link Failure	178
Figure 6.11: Average Message Delay for 9.6 kb/s links	182
Figure 6.12: Average Message Delay for 56 kb/s links	182
Figure 6.13: Average Message Delay for 380 kb/s links	183
Figure 6.14: Average Message Delay for 1544 kb/s links	183
Figure 6.15: Maximum Message Delay for Full Duplex links	185
Figure 6.16: Maximum Message Delay for Half Duplex links	186
Figure 6.17: Maximum Message Delay for Link Failure	186
Figure 6.18: Maximum Message Delay for 9.6 kb/s links	187
Figure 6.19: Maximum Message Delay for 56 kb/s links	187
Figure 6.20: Maximum Message Delay for 380 kb/s links	188
Figure 6.21: Maximum Message Delay for 1544 kb/s links	188
Figure 6.22: Percent of Messages over 3 sec for Full Duplex links	189
Figure 6.23: Percent of Messages over 3 sec for Half Duplex links	190
Figure 6.24: Percent of Messages over 3 sec for Link Failure	190
Figure 6.25: Percent of Messages over 3 sec for 9.6 kb/s links	192
Figure 6.26: Percent of Messages over 3 sec for 56 kb/s links	192
Figure 6.27: Percent of Messages over 3 sec for 380 kb/s links	193
Figure 6.28: Percent of Messages over 3 sec for 1544 kb/s links	193
Figure 6.29: Throughput for Full Duplex links	196
Figure 6.30: Throughput for Link Failure	196
Figure 6.31: Throughput for 9.6 kb/s links	197
Figure 6.32: Throughput for 56 kb/s links	197
Figure 6.33: Throughput for 380 kb/s links	198
Figure 6.34: Throughput for 1544 kb/s links	198
Figure 6.35: Average Message Delay for Full Duplex links	201
Figure 6.36: Average Message Delay for Link Failure	201
Figure 6.37: Average Message Delay for 9.6 kb/s links	202
Figure 6.38: Average Message Delay for 56 kb/s links	202
Figure 6.39: Average Message Delay for 380 kb/s links	203
Figure 6.40: Average Message Delay for 1544 kb/s links	203
Figure 6.41: Maximum Message Delay for Full Duplex links	205
Figure 6.42: Maximum Message Delay for Link Failure	205
Figure 6.43: Maximum Message Delay for 9.6 kb/s links	206
Figure 6.44: Maximum Message Delay for 56 kb/s links	206
Figure 6.45: Maximum Message Delay for 380 kb/s links	207
Figure 6.46: Maximum Message Delay for 1544 kb/s links	207
Figure 6.47: Percent of Messages over 3 sec for Full Duplex links	209
Figure 6.48: Percent of Messages over 3 sec for Link Failure	209
Figure 6.49: Percent of Messages over 3 sec for 9.6 kb/s links	210
Figure 6.50: Percent of Messages over 3 sec for 56 kb/s links	210
Figure 6.51: Percent of Messages over 3 sec for 380 kb/s links	211

Figure 6.52: Percent of Messages over 3 sec for 1544 kb/s link	211
Figure 6.53: Rate of change of throughput (Model 1 - 9.6 kb/s)	199
Figure 6.54: Rate of change of delay times (Model 1 - 9.6 kb/s)	199
Figure 6.55: Rate of change of throughput (Model 1 - 56 kb/s)	200
Figure 6.56: Rate of change of delay times (Model 1 - 56 kb/s)	200
Figure 6.57: Rate of change of throughput (Model 1 - 380 kb/s)	201
Figure 6.58: Rate of change of delay times (Model 1 - 380 kb/s)	201
Figure 6.59: Rate of change of throughput (Model 1 - 1544 kb/s)	202
Figure 6.60: Rate of change of delay times (Model 1 - 1544 kb/s)	202
Figure 6.61: Rate of change of throughput (Model 1 - full duplex)	204
Figure 6.62: Rate of change of throughput (Model 1 - half duplex)	205
Figure 6.63: Rate of change of throughput (Model 1 - link failure)	205
Figure 6.64: Rate of change of throughput (Model 2 - full duplex)	206
Figure 6.65: Rate of change of throughput (Model 2 - link failure)	206
Figure 6.66: Throughput Interpolation for 56kb/s full duplex links	221
Figure 6.67: Throughput Approximation for 56kb/s full duplex links	221
Figure 6.68: Throughput Interpolation for 380 kb/s full duplex links	223
Figure 6.69: Throughput Approximation for 380kb/s full duplex links	224
Figure 6.70: Throughput Interpolation for 1544 kb/s full duplex links	225
Figure 6.71: Throughput Approximation for 1544 kb/s full duplex links	226
Figure 6.72: Throughput Approximation Curve for 56 kb/s with link failure	232

I. INTRODUCTION

As the twentieth century draws to a close, the telecommunications industry, utilizing the latest computer and telecommunications technologies, provides an ever increasing array of computer and telecommunications services in all areas of life. Initially restricted to plain old telephone service (POTS), the telecommunications industry was revolutionized in the 1960s with the use of computers in transmitting data. This meant that not only computer data can be exchanged over telecommunications lines, but also the data can be manipulated by computers. Initially used to improve telephone service, in the three decades since computers are now being utilized in all types of telecommunications applications. This in turn has made controlling and managing telecommunication network easier, faster and more reliable making these networks more flexible, and enabling them to offer new and expanded services to their subscribers.

One such example is AT&T's Intelligent Network (IN). The marriage of computer and telecommunications technologies, offer the subscribers a variety of services that go beyond regular telephone service. Intelligent Networks are defined as collections of "intelligent" processors connected in a distributed computing environment, designed to carry information from one point to another. The intelligence of these processors is embedded in their hardware software and firmware. It is microcoded as firmware in the memories of monitoring and signalling components. The firmware consist of software utilities and information modules. The incorporation of these mechanisms allows the

modification of system functions in real time and provides the capability for the network to adapt to changes quickly. The network functions are dependent on the different components which include equipment used to switch information from one point to another, transmission facilities used to transport the information, and databases that reside at different points in the network containing local and global intelligence. Service subscribers to a network can change their own environmental variables, such as routing intelligence, as often as they like, thus controlling their own network resources. In general, an Intelligent Network is required to be able to adapt in real time to dynamic changes of the environment. These changes include network overload, switching delays, and equipment failure.

The Intelligent Network can exist as an overlay network or an integrated network within the existing Public Switched Telephone Network (PSTN). Intelligent Networks are digital networks but may contain some analog components. By having an open architecture, one can set the stage to implement rules that ensure effective communication regardless of the environment in which the network is deployed. Currently a common set of protocols supported by an open architecture known as Common Channel Signalling System No. 7 provide for an efficient communication between equipment manufactured by various vendors.

The switches used in the Intelligent Network environments are stored program controlled and are called Service Switching Points (SSP). They are access points into the network.

for Intelligent Network services. When a request arrives to a SSP, it may either process it or query other Intelligent Network components for additional intelligence to complete the service requests. A Service Control Point (SCP) is a centralized database that contains call processing information for service subscribers. It may be updated to change the type of services available to the subscribers. The Service Management System (SMS) allows subscribers to control their own resources in the Intelligent Network. Subscribers can, for example, design their own routing plans to maximize the flow of traffic to their locations. An Intelligent Peripheral (IP) provides enhanced services controlled by the SCP or the SSP. Typical IP functions are announcements, speech synthesizing, voice messaging and speech recognition. Finally, the Operations, Administration and Management System (OA&M) provides the functions necessary to manage and control the various network

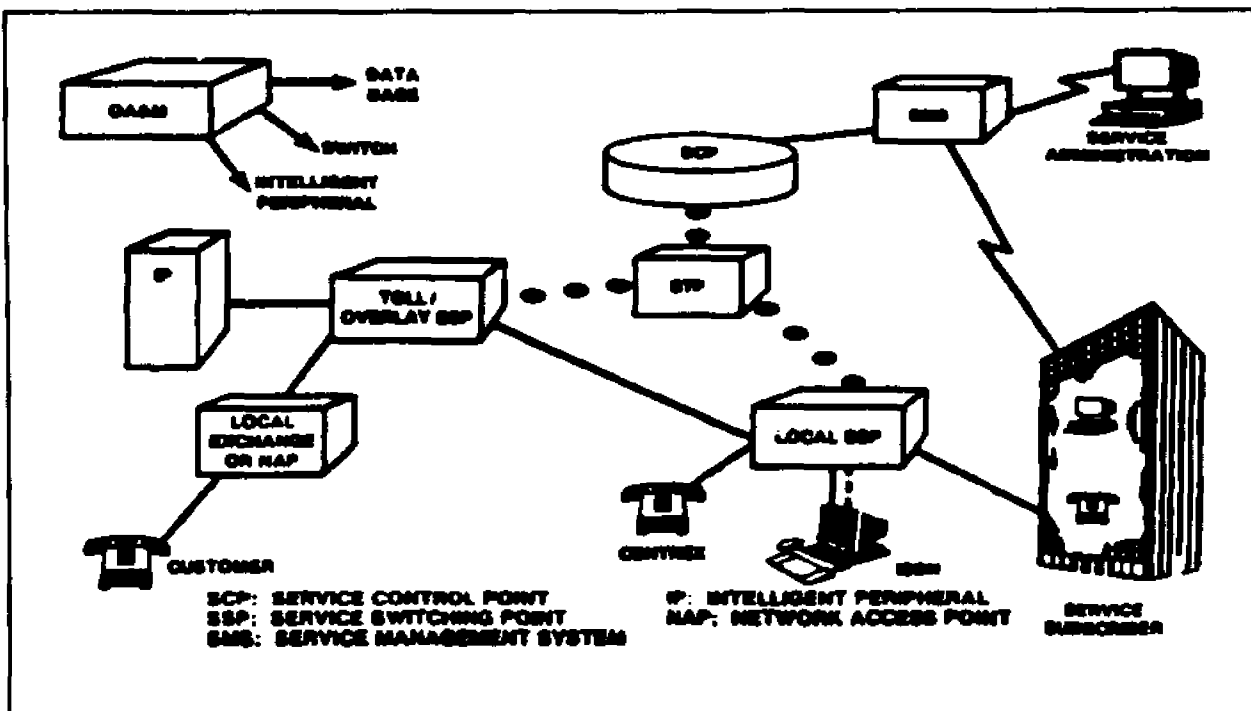


Figure 1-1 The Intelligent Network Components

resources. It provides maintenance functions for the network elements, traffic data collection, and network traffic management. Some of the services offered by the Intelligent Network are the 800 services, 911 public emergency service, Private Virtual Networks (PVN), Alternate Billing Service (ABS) and 900 services. With the advances in IN technology, Value Added Networks (VAN), Integrated Services Digital Network (ISDN) and broadband ISDN which offer large number of diverse services within single access line, will be a common service offered to a large number of subscribers.

As the number of computers and telecommunications services offered to subscribers increased, the large users of these networks, such as airlines, banks and governments started to need these services in a multinational setting. The growth of international data transmission came in the 1970s. Since then it has played an increasingly important role in the global economic and social structure. It provides valuable economic gains for multinational companies both by the international coordination of production, distribution and activities, and through the sharing of expensive computer and communications resources. It facilitates inherently international activities, such as international banking, airlines, and world meteorological cooperation. It aids international cooperation such as pooling of international resources and ideas for scientific and technical research and police activities. In addition, it contributes to greater communications between individuals thus contributing to the better understanding and greater democracy and equality between the nations of the world.

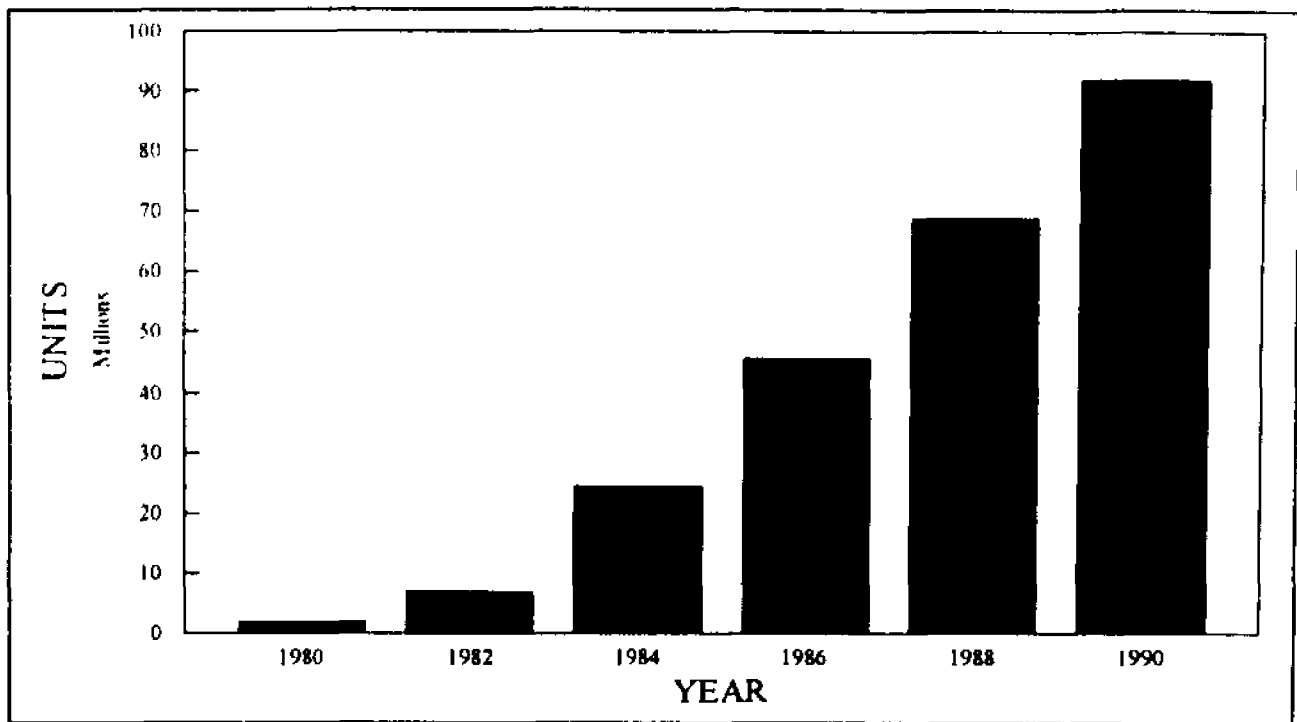


Figure 1. 2: Number of computer installations in the world

The reasons for international data communications are numerous. The most important is the fact that a large volumes of data need to be transmitted and that voice or postal communication may be unsuitable, slow or expensive. The data to be transmitted need to be stored and processed by a distant computer and should be in a form which allows it be entered directly into the computer. The transmission needs to be accurate, and validation of the data before and after transmission may be necessary to ensure its integrity and accuracy. Data also may need to be encrypted for security, so that only the authorized persons can access it.

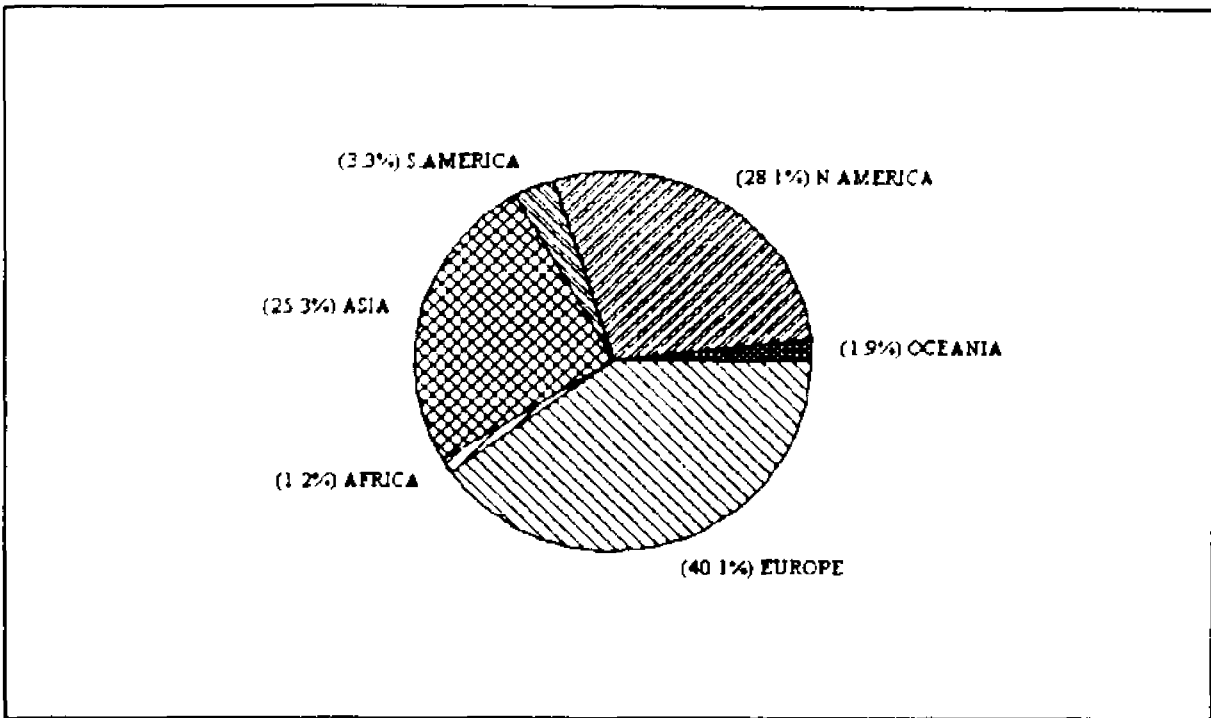


Figure 1. 3: Percentage of world main lines

Many corporations or organizations have set up their own data communication networks to serve their own purposes. Therefore there are many international networks that span the same region(s) resulting in some duplication of communication facilities. Physical duplication is usually avoided by sharing of physical transmission facilities by more than one network through public or private leased lines.

The characteristics of these networks not only depend on the type of data that is being transmitted, but also on the communication facilities that are available in each of the networked countries. Countries may have different levels of telecommunication saturation

depending on their economic status. Countries that are developed have the most advanced facilities, whereas developing and underdeveloped countries have less advanced facilities. In the 1980s with the digitalization of communication links, a phenomenon known as "leap frogging" emerged wherein the less developed countries started installing new digital equipment, bypassing - technologically - some of the developed countries that had not replaced all their non-digital equipment because of economic infeasibility.

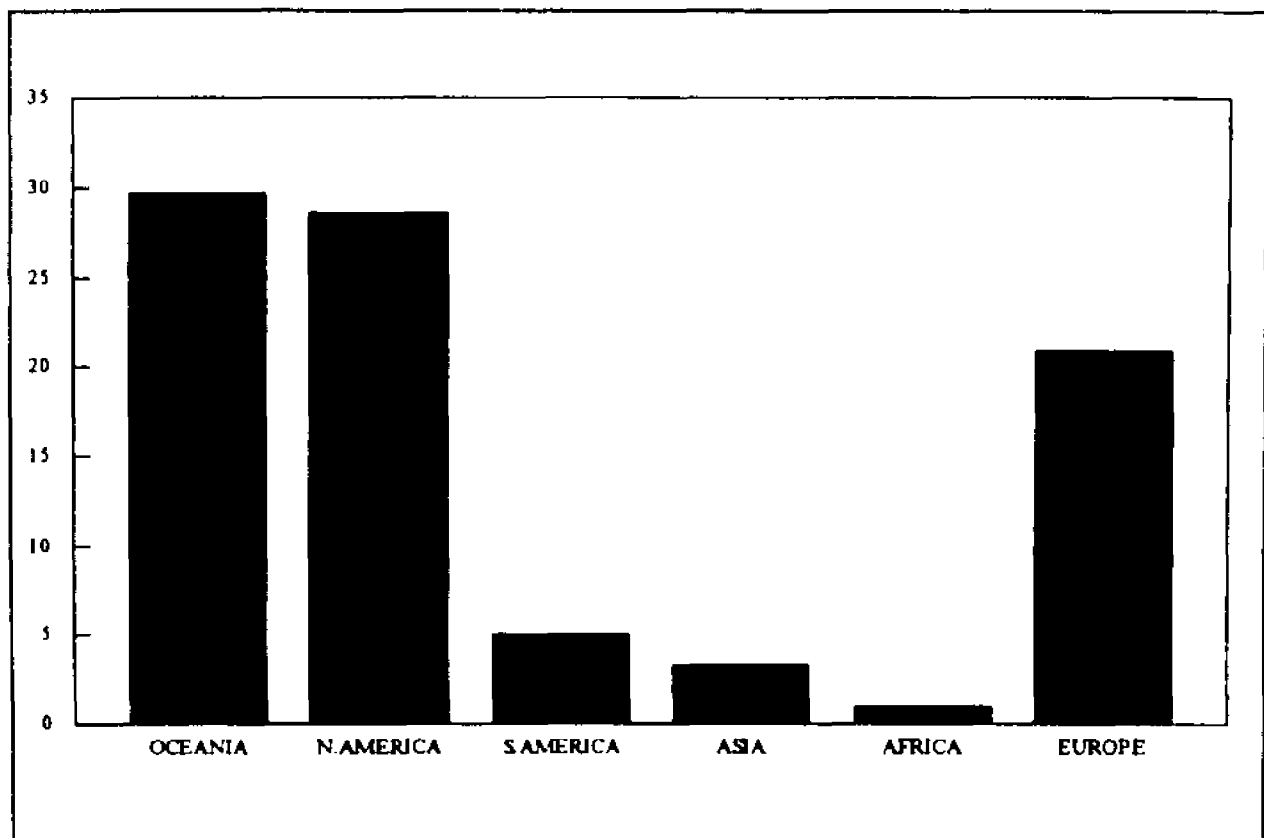


Figure 1. 4: Main lines per 100 population

One of the most common types of international networks is the research and development network. These networks are usually regional such as EARN in Europe which connects universities and research laboratories in Europe, BITNET in United States of America and Canada and JNET in East Asia which connects major university centers and research laboratories. These academic and research networks were loosely connected to each other with low speed links as an afterthought. Recently, as many public and private networks are moving toward global interconnection, the task of designing and developing a blueprint for modification and expansion of existing regional networks has taken the center stage. Currently the most prominent, but not the only, global network is the Internet. This network, with an estimated 15 million users, contains both public and private institutions as members and is constantly evolving and growing. Its services, however, are limited. Aside from electronic mail and file transfer, it offers electronic bulletin board, remote logon as standard services. Certain sites offer other enhanced services such as video conferencing, but that is by no means an established service. Despite this, the number of people using the Internet is growing rapidly.

Connected countries	36
Connected sites	1181
Connected nodes	3271
Connected gateways	36

Table 1. 1: BITNET/NORTHNET/EARN/JNET Connectivity

As we look to the future, the combination of freer access and higher transmission demands of newer services will inevitably stretch the limitations of the current global networks. On the other hand advances in the telecommunications technology such as use of fiber optics and efficient compression algorithms will ease those growing pains. Any additions or modifications to be made to these network must be thought out carefully, paying close attention to the factors that may impact the performance of the network or the delivery of the services. It is important to have a set of performance guidelines that guide the future designers of the networks to make intelligent and cost effective decisions.

TELECOMMUNICATIONS SERVICES		
OLD	NEW	FUTURE
Telephone	Digital telephone	Video telephone
Telex, teletex	Textfax	Fast facsimile
Low speed data	Video conferencing	Bulk document transfer
Mobile phone	Audio teleconference	High speed data
Low speed facsimile	Electronic Mail	High quality videotex
	Wide area mobile telephone	Online graphic design
	High resolution videotex	Remote printing/publishing
	High definition television	Dynamic computer load
		Burst mode transfer
		Distance learning
		High definition television

Table 1. 2 Telecommunication services present and future

Performance issues are considered and investigated in this research. Since the number and type of all networks and all the services are too numerous to consider, some simplifying assumptions are made. The performance of a network is evaluated as the demand on its resources increases. The demand is represented by the traffic intensity, which represents the amount of data requested to be transferred in a unit of time within the network elements. The measure of performance is throughput, delay and Quality of Service (QoS). Throughput is the amount of data actually transferred in a unit of time. Delay is the measure of elapsed time of transmission of data from its source to destination. Quality of Service is a metric that may be defined differently depending on a criteria such as cost, availability, or overall delay time that the network experiences.

The switching that is implemented in a network is very important and has a great impact on its performance. Most computer data, and recently many voice and video networks are packet switched, since it is more efficient than circuit switching. For services that are connection oriented, such as video conferencing, virtual circuit packet switching is used. For connectionless services, such as credit card approvals, datagram packet switching is used. A new switching technology, Asynchronous Transmission Mode (A.T.M.), which provides high data rates, uses packets of small size called cells. A.T.M. provides support for a large number of applications including computer data, audio and video transmission and utilizes virtual circuit switching to accomplish those tasks. Since A.T.M. is quickly becoming the preferred switching technique, the body of this work only deals with Virtual

Circuit Packet Switched Networks, further delimiting the area of research. A virtual circuit switched network can deliver services that require circuit switching as well as services that require datagram switching, therefore the general results of the research can be applied to those types of networks as well

In this dissertation it is shown that the topology as well as characteristics of nodes and links play an important role in determining a network's performance. The work discussed in this dissertation has practical applications in future designs of global networks. The adaptations of the findings lead to reduce cost and time to design and install a network.

Chapter 2 contains an overview of network concepts and research methodology. Concepts of topology, traffic and transmission and their functions and importance are discussed to help the reader gain background information for subsequent chapters. In addition the research methodology is summarized.

Chapter 3 outlines the concepts and methods of network analysis. It introduces the basic concepts of queueing theory and the Poisson process that are instrumental in telecommunications systems analysis. It then enumerates the elements in a network that must be considered when performing the analysis.

Chapter 4 describes the modelling and simulation methodology used to conduct the simulations. The need for simulations is discussed as time and cost saving mechanisms.

The organization of the simulator as well as the nature of the simulation model used in the simulators is discussed.

Chapter 5 enumerates the results from the simulation studies as well as drawing graphs for each set of simulations

Chapter 6 discusses the results of the simulations. These results are categorized according to topology, link properties (full duplex, half duplex and link failure) and performance criteria (throughput, delay times and Quality of Service). It enumerates the parameters that control network performance and places boundaries depending on topology and traffic.

Chapter 7 reviews and summarizes the conclusions and enumerates the factors that affect the performance of a network, so that consideration of these factors avoids the pitfalls of network design. The contributions of this work are summarized with recommended future studies. This work is of importance to those dedicated to the successful design and implementation of global networks

II. CONCEPTS AND METHODOLOGY

2.1 Network Concepts

2.1.1 Topology

Network topology deals with the manner in which the network elements, nodes, and links, are interconnected to achieve the desired flows of information. Diverse network applications may dictate different topologies. With the advent of new hardware and software for telecommunications, new topologies may emerge in coming years, providing improved and different services. This section classifies some well-known topologies.

2.1.1.2 Classification of Network Topologies

As the number of network elements increases, the ways in which one can interconnect the various elements grows rapidly. Consequently, it becomes very difficult to classify possible topologies for a network. After observing various networks there emerge four general network topologies: fully connected, hierarchical, bus, and ring.

Large and more complex networks may consist of one or more of the above topologies, depending on the uses of the network. For example, the bus and ring topologies are multiple access topologies and are used mostly in local networks. Hierarchical and fully connected topologies are point to point topologies, used by wide area networks. These network topologies are discussed in the following sections.

2.1.1.3 Fully Connected Topology

A fully connected topology provides a direct connection between every pair of nodes in the network. Figure 2.1 illustrates this topology. Such a topology is economically feasible only when there are few nodes in the network. The number of links required for realizing such a structure is $\frac{n(n-1)}{2}$, where n is the number of nodes in the network. A fully connected network, despite its relatively high cost, provides better response time and throughput than those for other topologies. This topology is especially suitable for a local network where the need for low delay times and high throughput is important.

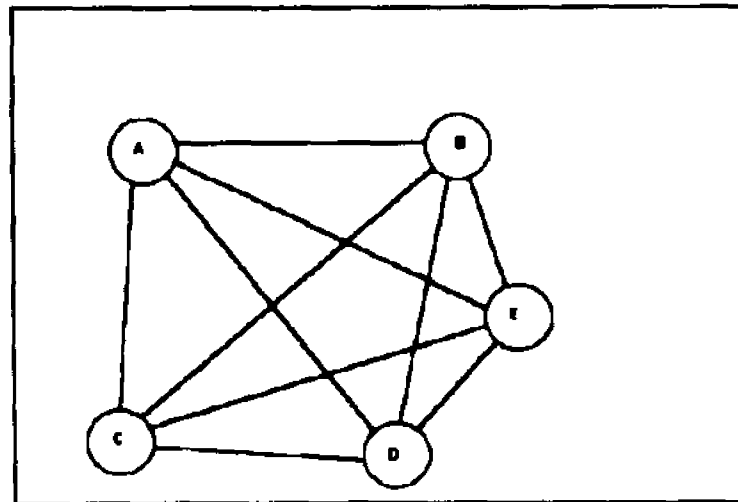


Figure 2.1 Fully Connected Topology

2.1.1.4 Hierarchical (Tree) Topology

The hierarchical topology is a multilevel structure that allows only a single path between any two nodes in the network. Figure 2.2 illustrates a three-level hierarchical topology, which becomes economically advantageous when there are a large number of nodes in the network. The number of links is $n-1$, where n represents the number of nodes. This topology is used mostly in commercial environments where there is a clear division of labor between the departments corresponding to the different levels of the hierarchy.

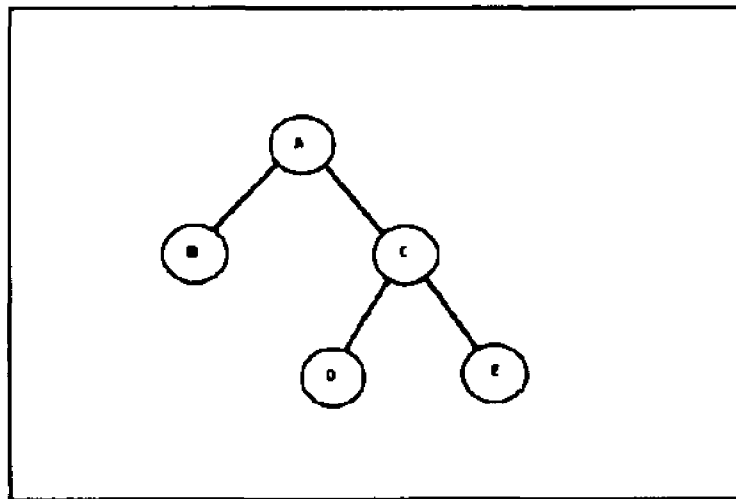


Figure 2. 2: Hierarchical Topology

2.1.1.5 Ring Topology

In a ring topology, illustrated in Fig. 2.3, each node is connected to two other nodes in a ring configuration. Each node in the ring is responsible for forwarding messages to the next node unless a message is destined for itself. The traffic in the loop is usually unidirectional. To avoid simultaneous multiple transmissions, a token schema is implemented allowing only the node with the token to transmit.

The number of links on the network is equal to the number of nodes and grows linearly with the number of nodes, as each node can be connected with only two other nodes. This type of topology is used mostly for local networks due to the complexities introduced by token maintenance and management.

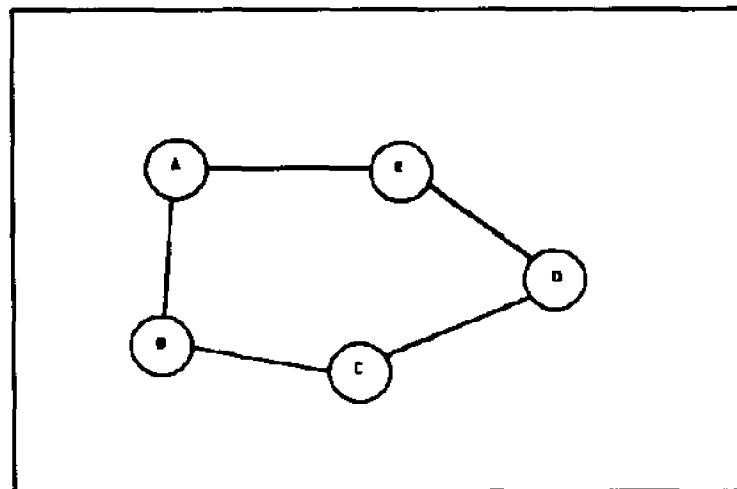


Figure 2.3. Ring Topology

2.1.1.6 Bus Topology

The bus topology, illustrated in Fig. 2.4, employs a shared broadcast transmission facility ("bus") for information exchange between several nodes. The bus access management can be accomplished in two ways. In the first method all nodes are identical and must communicate among themselves for access to the bus. In the second method a single control node manages the traffic flows between nodes. Although the structure of the bus topology is simple, techniques such as token-passing and CSMA/CD must be used to eliminate contention for the use of the bus. These techniques usually result in some waste of bus bandwidth. The bus topology is usually employed in local networks, such as LANs or multiprocessor systems, where traffic is at most moderate.

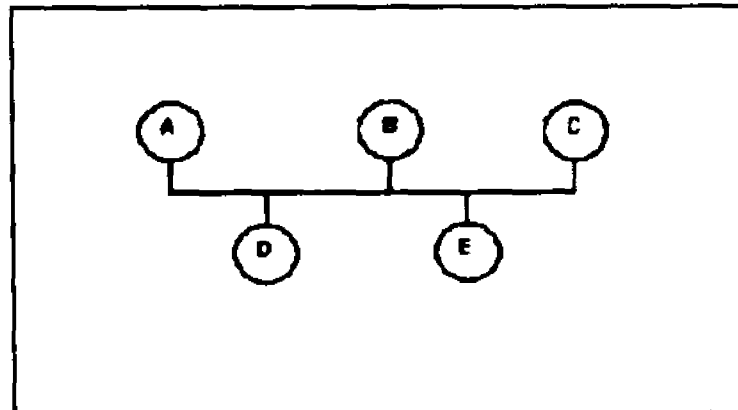


Figure 2.4 Bus Topology

2.1.2 Traffic Flow

2.1.2.1 Concepts of Traffic Flows and Network Control

In order for a network to provide reliable and robust services to its users, it must be able to control traffic between nodes. This is done through identification of the traffic entities that flow within the network and the traffic patterns that occur within the network. In the case of a global network, knowledge of traffic flow distribution for a twenty four hour period becomes essential. To control a network and manage it efficiently, a network control mechanism must exist that takes into account the traffic entities and the flow of information. Based on this information, the network control center can dynamically reroute traffic to prevent congestion, and can monitor the network for faulty nodes and links or improper usage.

2.1.2.2 Traffic Entities

All communication networks are characterized by flows of discrete traffic units from one node to another via the available links. A discrete traffic unit may be defined as the most elementary entity of data that is exchanged between nodes having the shortest fixed length. A typical traffic unit may be a bit for a synchronous link or a byte for an asynchronous link. These traffic units are transmitted by the Physical Layer of the Open System Interconnect reference model designed and approved by International Organization

for Standardization (ISO)¹ and International Telephone and Telegraph Consultative Committee (CCITT)² in 1983.

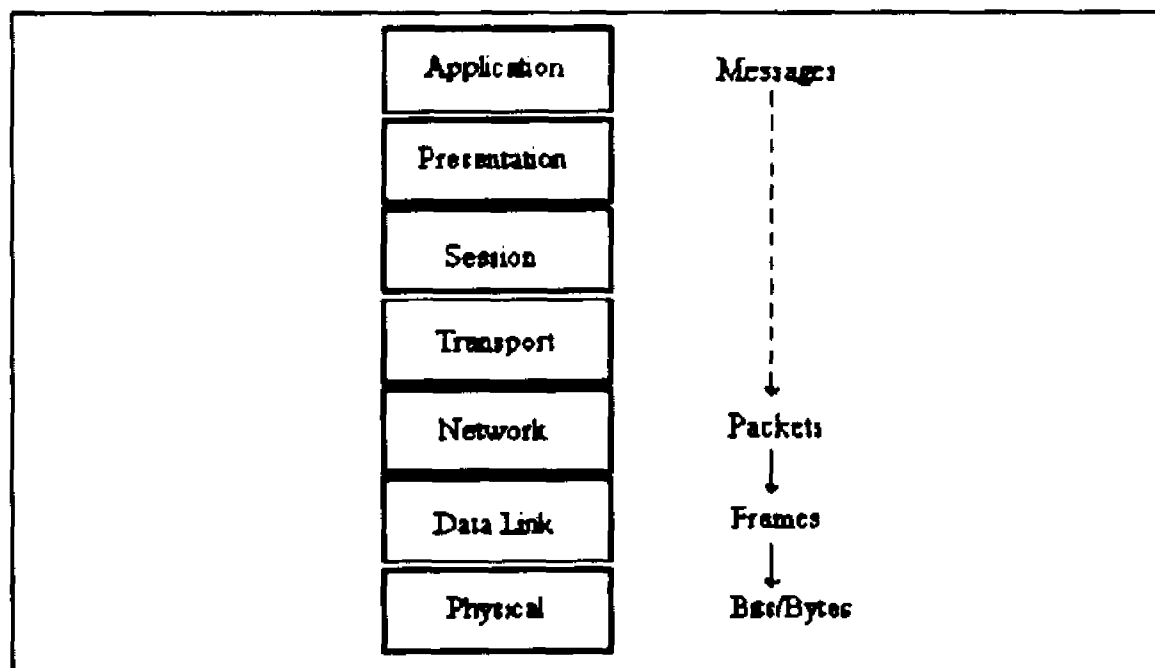


Figure 2. 5: Open System Interconnect Reference Layers

Frames

A frame is the unit of traffic at the Data Link Layer of the Reference Model and it consists of a number of discrete traffic units. The data, along with cyclic redundancy check and flow control information, is encapsulated in a frame to ensure reliable transmission. Figure 2.5 displays some common frame types used in various networks.

¹ISO International Standard 7498, *Information Processing Systems - Open Systems Interconnection - Basic Reference Model*, Geneva, Oct 1983.

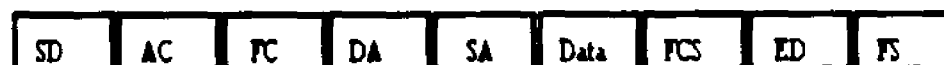
²CCITT Draft Recommendation X.200, *Reference Model of Open Systems Interconnection for CCITT Applications*, Geneva, June 1983.



a. High level Data Link Control (HDLC) Frame Format



b. Collision Sense Multiple Access/Collision Detection (CSMA/CD) Frame Format



c. Token Ring Frame Format



d. Link Access Protocol - D channel (LAP-D) Frame Format



e. SNA data encapsulation format

AC= Access Control
 DA= Destination Address
 ED= Ending Delimiter
 FC= Frame Control

FCS=Frame Check Sequence
 FS= Frame Status
 LH= Link Header
 RH=Req/resp Header

SA= Source Address
 SAPI=Service Access Point Identifier
 SD= Starting Delimiter
 SFD=Starting Frame Delimiter
 TEI= Terminal Endpoint Identifier
 TH=Transmission Header

Figure 2. 6. Frame Types

Packets

A packet is a discrete entity of data that is exchanged between a source and a destination node at the Network Layer. It may contain data or control information. Routing information such as the destination node address is also included within each packet to ensure delivery to its destination. A packet usually does not exceed a certain length in order to optimize the transmission line utilization.

Messages

A message is a logical unit of data transfer at the application layer level. These units may, in turn, be made up of subunits, depending upon the maximum size of the data units allowed at other lower layer protocols.

Call

In connection-oriented services, a call is a session where two or more nodes exchange messages. For each call there is an initiation process (call setup) where an available path(s) between the calling and the called node(s) is found, and a termination process where the path(s) is freed. Between these two processes any number of messages may be exchanged by the nodes.

2.1.2.3 Traffic Attributes

The amount of data that passes through a network is a function of rate of arrival and length of messages arriving at the network nodes. The combined traffic handled by all network nodes is known as traffic intensity. The traffic intensity is sum of the products of message arrival rate and message lengths at each node. These factors affect the amount of data the network must handle. Applications using a network may have different properties and, in turn, different traffic characteristics.

Message Arrival Rate

Message arrival rate is the measure of how frequently messages arrive at any network node during a unit of time, commonly a second. We can also express this rate in terms of message interarrival time which measures the amount of time elapsed between the arrival of two consecutive messages. In practice, the calling rate varies significantly over different periods of time, such as the hours of the day, days of the month, or months of the year. Furthermore, the calling rate distribution may also be influenced by the number of samples or the periods of observation. The calling rate is generally denoted by the symbol λ , and it is generally determined for the average second of a typical peak hour.

Average Message Length

The average message length is the average of the lengths of all messages arriving at a node over a period of time. In packet switched networks it is measured in terms of a

some examples of connectivity matrices for different topologies shown in Figures 2.1 - 2.4.

2.1.2.5 Traffic Paths and Routing

In order to send a message from one node to another in a network, the network management must first select a path along which the message is transmitted. The selection criteria may depend upon factors such as cost, congestion, or delay. The available routes from the source to the destination can be derived from the connection matrix by finding all the possible routes a message can take. These routes can be represented by directed vectors, whose elements are the nodes in the route.

The selected route may vary for different situations. If a message needs to be delivered urgently, the selection criterion is shortest time for delivery. However, if economy is the goal, the criterion is lowest cost. In many instances such criteria are in conflict with each other.

Each node keeps a table of values for all the available routing criteria. This table contains the costs, link delay data, etc., between each pair of connected nodes. This becomes, in effect, a weighted connection matrix for a particular criteria. These matrices are updated regularly to reflect the changing conditions of the network.

2.1.2.6 Traffic Control

During periods of heavy traffic congestion may appear at certain network elements, giving rise to a degradation in network performance. In some cases, either the request for service or the traffic data gets lost due to the unavailability of some critical nodal resources, such as incoming traffic buffers. Consequently, some traffic flow control techniques are employed to maximize the system throughput and minimize delays due to network congestion.

Reasons for network congestion which degrade performance include the peaking of traffic intensities during busy hours, the degradation of service in some parts of the network, or complete failure of one or more parts of the network system. Several techniques are used to control the traffic so that the network performs satisfactorily. Two major techniques commonly employed are flow control and routing.

Flow Control

Flow control restricts traffic within the network when the flow becomes so intense that it begins to overwhelm the network resources. Flow control is usually implemented at the frame level. Different techniques are used to control the traffic flow, depending on the network type.

Stop-and-wait flow control means the destination node, after receiving a frame of data, indicates its willingness to receive another frame of data by sending an acknowledgement.

to the source node. The source node waits for an acknowledgement before sending another frame. This technique is acceptable if the frame sizes are large, so that a message is not broken into many frames, thereby necessitating many acknowledgements.

The sliding window flow control technique directs the source to send a predetermined number of frames to the destination without any acknowledgement. After that, it continues only if it gets the acknowledgements for those frames already sent.

Routing

In order to deliver a message from the origin to the destination node, a route must be selected. More than one route generally exists between the two nodes. The choice of the route usually depends upon factors dictated by the type of service that the network provides. This routing decision depends on the performance criteria, decision time, and the overall routing strategy.

The selection of a route is generally based on some performance criteria. The simplest criteria is the shortest route - that traverses the least number of nodes resulting in the least number of hops - between the source and destination nodes. If, however, costs are associated with each link, then a route with the lowest cost is selected. A route may also be selected in order to maximize the network throughput.

In addition to route selection, routing also deals with timing. This is dictated by the type of switching the network is using. In circuit switched and virtual circuit packet switched systems the route selection occurs at the beginning of a session. In datagram switching it occurs when a packet is ready to be sent.

The location at which the routing decision is made may also vary. The decision could be made at the originating node, at any node along the route, or at a central node not involved in the transmission.

Other routing considerations include the location of network information source and the routing strategy. Network information is needed by all nodes to make the optimal routing decision under changing network conditions. This information includes downed nodes, congested links, etc., which affect the routing decision. The location of the source of the information may be the sending node, or adjacent nodes, or nodes along the way of arriving packets, or all the nodes in the network.

Lastly, the routing strategy deployed affects the route selection. If the strategy used is flooding, there is no need for route selection. However, if it is a random strategy, the selection depends upon a random route selection from all possible choices. If it is adaptive strategy, the network information is considered to select the optimal route between source and destination.

2.1.3 Transmission and Signalling

In order to have a complete understanding of the operation of a network, we need to analyze the underlying transmission techniques and signalling methods that control the operations of the network. These aspects control most of the physical operations of the network, such as bandwidth, error rate, and the data transmission.

2.1.3.1 Transmission

Transmission is the ability to transfer a piece of data from one point to another. It is a physical layer concept and is accomplished by sending a signal through a physical medium. Reliable data transmission, on the other hand, depends on the combination of the characteristics of the transmission medium and the quality of the signal, as well as error checking performed by the data link layer. The transmitted signals can be either analog or digital, depending upon the medium. In addition, the data transmitted can be either analog and digital, depending upon the applications. The combination of transmission media and data encoding gives a network its transmission characteristics.

Transmission Media

The transmission media commonly used in communications networks can be broken down into two major classes: conducted (guided) and radiated (unguided). Conducted media include twisted pair, coaxial cable and fiber optics. Radiated media include terrestrial microwave, satellite microwave and radio. There are many factors that make each medium

desirable or undesirable in different situations, including speed, security, distance, susceptibility to error, and cost.

Twisted Pair.

Twisted pair consists of two insulated copper wires arranged in a regular spiral pattern. A wire pair acts as a single communications link. The twisting of the individual pairs minimizes electromagnetic interference between the pairs. Twisted pair can be used for analog and digital data. The most common use for it is in "local loops" in the telephone system, in intrabuilding communications such as a private branch exchange (PBX), and in microcomputer local area networks.

Twisted pair may be used to transmit both analog and digital signals. Compared to other transmission media, twisted pair is limited in distance, bandwidth, and data rate. The attenuation for twisted pair is a very strong function of frequency. Other impairments are also severe for twisted pair. The medium is quite susceptible to interference and noise because of its easy coupling with electromagnetic fields. Impulse noise also easily intrudes into twisted pair. Several measures are taken to reduce impairments. Shielding the wire with metallic braid or sheathing reduces interference. The twisting of the wire reduces low-frequency interference, and the use of different twist lengths in adjacent pairs reduces crosstalk. Another technique is the use of a balanced transmission line. With an unbalanced line, one wire is at ground potential; with balanced transmission, both wires are above ground potential, carrying signals with equal amplitude but opposite phase. For

point-to-point analog signaling, a bandwidth of up to about 250 KHz is possible. For voice transmission, a 6-km section of line represents an upper limit on the distance that can be covered¹. For digital point-to-point lines, data rates of up to a few Mbps are possible.

Coaxial Cable

Coaxial cable, like twisted pair, consists of two conductors, but it is constructed differently to permit it to operate over a wider range of frequencies. It consists of a hollow outer cylindrical conductor which surrounds a single inner wire conductor. The inner conductor can be either solid or stranded, the outer conductor can be either solid or braided. The inner conductor is held in place by either regularly spaced insulating rings or a solid dielectric material. The outer conductor is covered with a jacket or shield. A single coaxial cable has a diameter of from 0.4 in. to about 1 in.

Coaxial cable is utilized in a variety of applications that include long distance telephone and television transmission, television distribution (cable TV), local area networks, and short run system links such as high speed I/O channels on computer systems.

Coaxial cable is used to transmit both analog and digital signals. Long-distance systems are either analog or digital. Cable TV is analog, and both analog and digital techniques have been used for local networks. Coaxial cable has superior frequency characteristics.

¹ William Stallings, Data and Computer Communications, New York: Macmillan Publishing Company, 1991 p. 71

to twisted pair, and so can be used effectively at higher frequencies and data rates. Because of its shielded, concentric construction, coaxial cable is much less susceptible to interference and crosstalk than twisted pair. The principal constraints on performance are attenuation, thermal noise, and inter-modulation noise. The latter is present only when several channels or frequency bandwidths are in use on the same cable.

For long-distance transmission of analog signals, amplifiers are needed every few kilometers, with closer spacing required if higher frequencies are used. The usable spectrum for analog signaling extends to about 400 Mhz. For digital signaling, repeaters are needed every kilometer or so, with closer spacing needed for higher data rates⁴.

Fiber Optics

Fiber optic cable is a thin flexible medium capable of conducting an optical ray. Various glass and plastic materials are used to make optical fibers. An optical fiber has a cylindrical shape and consists of three concentric sections: the core, the cladding, and the jacket. The core is the innermost section and consists of one or more very thin strands, or fibers. Each fiber is surrounded by its cladding, a glass or plastic coating that has optical properties different than those of the core. The outermost layer, surrounding the cladded fiber(s) is called the jacket. The jacket is composed of plastic and other materials layered to protect against moisture, crushing or other environmental dangers.

⁴Stallings p. 75

One of the most significant technological breakthroughs in data transmission has been the development of practical fiber optic communications systems. Its greater bandwidth, smaller size and lighter weight, low attenuation and electromagnetic isolation, and greater repeater spacing distinguish optical fiber from twisted pair and coaxial cable.

Data rates of 2 Gbps over tens of kilometers have been demonstrated for optical fiber. Optical fibers are very thin and the advantage of small size is considerable. The corresponding reduction in weight reduces structural support requirements.

Attenuation is significantly lower for optical fiber than for coaxial cable or twisted pair and is constant over a wide range. Optical fiber systems are also not affected by external electromagnetic fields. Thus the system is not vulnerable to interference, impulse noise, or crosstalk. By the same token, fibers do not radiate energy, causing little interference with other equipment and providing a high degree of security from eavesdropping. In addition, fiber is inherently difficult to tap.

The advantages of optical fiber over twisted pair and coaxial cable become more compelling as the demand for all types of information (voice, data, image, video) increases.

Optical fiber transmits a signal-encoded beam of light by means of total internal reflection. Total internal reflection can occur in any transparent medium that has a higher index of refraction than the surrounding medium. In effect, the optical fiber acts as a wave guide for frequencies in the range 10^{14} to 10^{15} Hz, which covers the visible spectrum and part of the infrared spectrum. Light from a source enters the cylindrical glass or plastic core. Rays at shallow angles are reflected and propagated along the fiber, other rays are absorbed by the surrounding material. This form of propagation is called multi-mode, referring to the variety of angles that reflect. When a fiber core radius is reduced, fewer angles reflect. By reducing the radius of the core to the order of a wavelength, only a single angle or mode can pass the axial ray. This single mode propagation provides superior performance for the following reason. With multi-mode transmission, multiple propagation paths exist, each with a different path length and hence time to traverse the fiber. This causes signal elements to spread out in time and limits the rate at which data can be accurately received. Since there is a single transmission path with single-mode transmission, such distortion cannot occur. Finally, by varying the index of refraction of the core, a third type of transmission, known as multi-mode graded index, is possible. This type is intermediate between the other two in characteristics. The variable refraction has the effect of focusing the rays more efficiently than ordinary multi-mode, also known as multi-mode step index.

Two different types of light source are used in fiber optic systems: the light emitting diode (LED) and the injection laser diode (ILD). Both are semiconductor devices that emit

a beam of light when a voltage is applied. The LED is less costly, operates over a greater temperature range, and has a longer operational life. The ILD is more efficient and can sustain greater data rates.

The amplitude-shift keying technique is commonly used to transmit digital data over optical fiber; in this context, it is known as intensity modulation. For LED transmitters, binary one is represented by a short pulse of light and binary zero by the absence of light. Laser transmitters normally have a fixed "bias" current that causes the device to emit a low light level. This low level represents binary zero, while a higher-amplitude light wave represents binary one.

Microwave

The most common type of microwave antenna is the parabolic "dish." A typical size is about 10 ft in diameter. The antenna is fixed rigidly and focuses a narrow beam to achieve line-of-sight transmission to the receiving antenna. Microwave antennas are usually located at substantial heights above ground level in order to extend the range between antennas and to be able to transmit over intervening obstacles. With no intervening obstacles, the maximum distance between antennae conforms to⁵

$$d = 7.14\sqrt{Kh}$$

⁵ Stallings, p. 80

where d is the distance between antennas in kilometers, h is the antenna height in meters, and K is an adjustment factor to account for the fact that microwaves are bent or refracted with the curvature of the earth and will hence propagate farther than the optical line of sight.

The primary use for terrestrial microwave systems is in long-haul telecommunications service, as an alternative to coaxial cable for transmitting television and voice. Like coaxial cable, microwave can support high data rates over long distances. The microwave facility requires far fewer amplifiers or repeaters than coaxial cable for the same distance, but requires line-of-sight transmission. Microwave is commonly used for both voice and television transmission.

Another increasingly common use of microwave is for short point-to-point links between buildings. This can be used for closed-circuit TV or as a data link between local networks. Finally, a potential use for terrestrial microwave is for providing digital data transmission in small regions.

As with any transmission system, a main source of loss for microwave is attenuation. For microwave, the loss of signal varies as the square of the distance with attenuation increasing with rainfall.

Another source of impairment for microwave is interference. With the growing popularity of microwave, transmission areas overlap and interference is always a danger.

Satellite

A communication satellite is another type of microwave relay station. It is used to link two or more ground-based microwave transmitter/receivers, known as earth stations or ground stations. The satellite receives transmissions on one frequency band (up link), amplifies (analog transmission) or repeats (digital transmission) the signal, and transmits it on another frequency (down link). A single orbiting satellite operates on a number of frequency bands, called transponder channels, or simply transponders.

There are two common uses of communications satellites. Firstly, a satellite is used to provide a point-to-point link between two distant ground-based antennas. Secondly, a satellite provides communication between one ground-based transmitter and a number of ground-based receivers.

For a communication satellite to function effectively, it is generally required that it remain stationary with respect to its position over the earth. To remain stationary, the satellite must have an altitude of 35,784 km. Two satellites using the same frequency band, if close enough together, interfere with each other. To avoid this, current standards require a 4° spacing (angular displacement as measured from the earth).

Among the most important applications for satellites are television distribution, long-distance telephone transmission and private business networks. Because of their broadcast nature, satellites are well suited to television distribution and are being used extensively for this purpose.

Because of the long distance involved, there is a propagation delay of about 240 to 300 ms from transmission from one earth station to reception by another earth station. This delay is noticeable in time critical applications. It also introduces problems in the areas of error control and flow control. Second, satellite microwave is inherently a broadcast medium. Many stations can transmit to the satellite, and a transmission from a satellite can be received by many stations thereby increasing the security risk of eavesdropping.

Radio

The principal difference between radio and microwave is that radio is omnidirectional and microwave is focused. Thus radio does not require dish-shaped antennas, and the antennas need not be rigidly mounted to a precise alignment.

Radio covers FM radio and UHF and VHF television. In addition to these traditional uses, new applications have grown up in this band such as packet radio. A packet radio system uses ground-based antennas to link multiple sites in a data transmission network.

Radio waves are less sensitive to attenuation from rainfall. For digital data communications, achievable data rates are lower than that of microwave and satellite

As a line-of-sight propagation technique, the amount of attenuation of radio waves is due to distance. However, because of the longer wavelength, radio waves suffer relatively less attenuation. A prime source of impairment for radio waves is multipath interference. Reflection from land, water, and natural or human-made objects can create multiple paths between antennas.

The first radio data communication system was the ALOHA system in Hawaii. Two frequency bands were used, one for transmission from users' terminals to a central controller and another for transmission in the opposite direction. Transmission is in the form of short bursts of data called packets. The point-to-point range is about 30 km, repeaters were used to extend the system to a radius of about 500 km.

Data Encoding

The data that is transmitted can be either analog (voice) or digital (computer data) in nature. Depending on the type of media, data may be encoded before transmission. Analog transmission takes the form of a continuous sine wave, varying the amplitude, frequency and the phase of a wave to represent data. Digital transmission, on the other hand, is a discrete square wave in which the voltage is varied to represent the data.

When digital data is transmitted over media using analog transmission, it is first encoded (converted) into analog data before transmission at the source node, then decoded back into digital after transmission by the receiving node. This encoding/decoding process is called modulation/demodulation and accomplished by a modem (modulator/demodulator). To represent the data, modems use amplitude modulation (or amplitude shift keying - ASK) or frequency modulation (or frequency shift keying - FSK) or phase shift modulation (or phase shift keying - PSK) or a combination of the three.

In Amplitude Shift Keying, the amplitude of the sine wave is varied to represent the data as shown in Figure 2.8

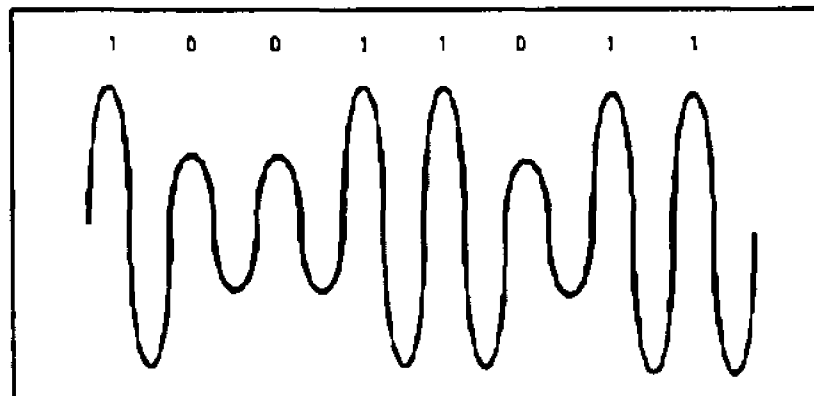


Figure 2.8. Amplitude Shift Keying

In Frequency Shift Keying, the frequency of the sine wave is varied to represent the data as shown in Figure 2.9.

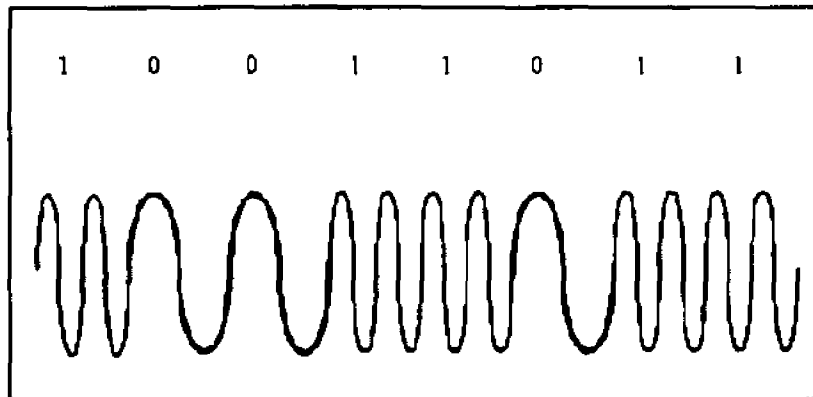


Figure 2.9. Frequency Shift Keying

In Phase Shift Keying, the phase of the wave is shifted to represent the data as shown in Figure 2.10.

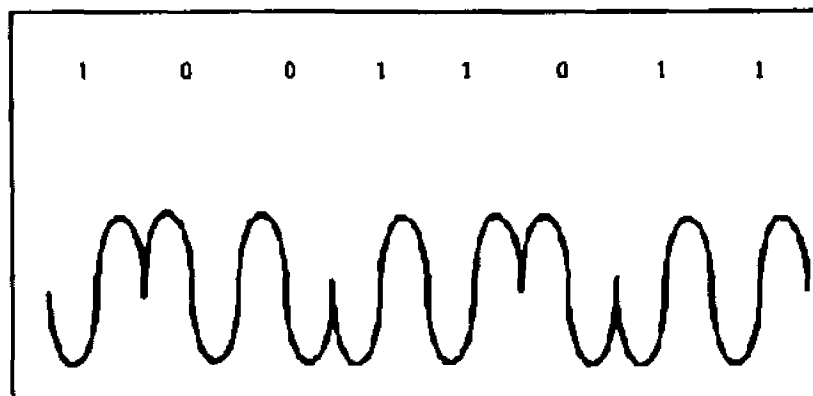


Figure 2.10. Phase Shift Keying

For higher transmission speeds, the encoding techniques may be combined. One example is Quadrature Amplitude Modulation (QAM), which combines eight phase changes with two amplitudes for a total of 16 different signals, yielding four bits for each cycle of the wave.

For analog data, that is transmitted digitally, a similar procedure of conversion must take place. A codec (coder/decoder) converts the analog data into digital signals at the source node and back to analog data at the destination node. Pulse code modulation (PCM) and Delta modulation are some of the most commonly utilized techniques for encoding analog data into digital signals. In PCM, the analog data is sampled 8000 times per second and each sample is quantized into a n-bit value and then transmitted. For voice data, 128 different levels are satisfactory, resulting in a 7 bit value. Delta modulation examines the change in the amplitude of the analog data and generates a 1 if there is an increase or a 0 otherwise.

Both analog and digital transmission are subject to impairments. The errors resulting from these impairments include thermal noise, impulse noise, echo, and attenuation. In addition, analog transmission is susceptible to crosstalk and phase jitter. Errors may be prevented by a variety of methods that include shielding of the media, utilizing an echo suppressor, and repeaters. Digital transmission is superior to analog transmission because it has lower error rates and can achieve higher transmission rates. The discrete nature of the digital

transmission enable the signals to be regenerated, overcoming the effects of any attenuation and noise thus resulting in a more reliable type of transmission.

2.1.3.2 Signalling

Signalling is a process of transferring information over a distance to control the setup of calls or sessions, to maintain connection, to gather billing or cost allocation data, and to terminate the calls or sessions. It involves the ability to send signals over the network that control the operation of the network rather than the transmission of data

There are four major categories of signals: supervisory, address, information, and network management. Supervisory signals are used to initiate requests for service, to initiate a call, to hold or release an established call, to request assistance from a network operator, to initiate and terminate billing, and to alert users. Address signals convey information regarding the called and the calling nodes and all other network elements involved during the connection within the network. Information signals are used to alert the users about the call progress or inform the user about any difficulties experienced by the system in completing the call, or provide advice to users about making requests for services. Network management signals control and manage the network by means of traffic routing, fault isolation and reporting, node recovery, and traffic protection.

The signals must be transmitted concurrently with the data without any interference. In order to achieve this, different signalling methods have been developed. They are in-band, out-of-band and common channel signalling

In-band Signalling

In-band signalling transmits control signals through the same band of frequencies used by data signals. This technique is the simplest of the three, and is quite flexible because signalling can take place anywhere data can be transferred. However, since control and data signals share the same band of frequencies, signalling can take place only when there are no data signals on the circuit. Another drawback is that the control signals are subject to the delays of the network, which may prove detrimental when trying to establish or terminate connections.

Out-of-band Signalling

Out-of-band signalling transmits control signals using the same facilities as data signals, but at a different part of the frequency band. This method is advantageous because it provides continuous supervision during the life of a connection. It is also more complex than in-band signalling, and therefore needs extra electronics to handle the signals.

Common Channel Signalling

In common channel signalling, the control signals are transmitted over links that are dedicated to control signals which are common to a number of data channels. This

signalling system was developed by CCITT and is called Common Channel Signalling (CCS) System No. 7. It is the successor to CCITT signalling system No. 6, which is also known as Common Channel Interoffice Signalling (CCIS). System No. 7 has been approved by ANSI and ISO and is rapidly becoming an international standard. Signal Transfer Points (STP) provide the generation and the transfer of the signalling messages over the dedicated links. It uses packet switching to transmit control signals and has two operational modes: associated and non-associated.

In the associated mode, the control signals closely follow the path of the actual data signals. In the non-associated mode, the control signals may take separate paths than the data signals. This enables the signalling systems to become more flexible as the network conditions change. The non-associated mode of signalling is recommended for signalling in Integrated Services Digital Networks (ISDN). In ISDN, a protocol has been defined for control signalling (I451/Q.931) that closely follows the specifications of System No. 7 non-associated mode. The control signalling is provided by the D channel which may also provide some low speed packet switching services such as videotex or teletex provided that there is available bandwidth. The D channel controls the B channels which provide digital voice, high speed data, and other services such as facsimile and video.

Common channel signalling reduces the transmission times of the control signals, and is flexible enough that it may be adapted to other functional needs. The disadvantage of this technique is in its complexity. Essentially, a signalling subnetwork must be established

in order to implement it, and this may prove costly. However, the recent decrease in costs of the networking hardware renders it a cost effective alternative to other methods of switching.

2.1.4 Network System Performance

In order to evaluate the effectiveness of a network, one must be able to measure the system performance utilizing statistics derived from the observation of the network's operation. There are many statistics which may be extracted from a network, but throughput, delay times, and cost are among the most important. In addition, the concepts of Grade of Service (GoS) and Quality of Service (QoS) help us measure the effectiveness of a network.

The cost associated with the performance of a network may be divided into two categories. The first represents the cost that is associated with the hardware and software needed to actually put a network into operation, and to make further improvements as the marketplace dictates. This is a capital investment on the part of the company or government that is implementing the network, and is very costly.

The second cost represents the expenses that a subscriber incurs in order to access the network. These may be in the form of connection charges or equipment procurement expenses. This cost is relatively small, and is designed to be kept as low as possible.

because higher cost would deter users from using the network, thereby underutilizing the resources that were acquired with great expense.

Throughput

Throughput is the amount of data that is transferred within a network in a unit of time. It is usually measured in bytes per second. In a circuit switched network, throughput is defined as the maximum number of arriving calls that can be handled during an average second of a busy hour, and the maximum number of calls that can be handled concurrently by the system during a busy hour. For a packet switched network, it is defined as the maximum number of input packets that can be handled during an average second of a peak hour. The higher the throughput of a network, the better the utilization of the network resources.

Quality of Service (QoS)

The performance of a network may also be affected by the quality of transmission and the performance of other physical components. The distortion of transmission due to crosstalk, attenuation, and noise all affect the quality of service. Random component failures, which may result in loss or degradation of network services, also affect the QoS. The mean time between failure (MTBF) for network components must be specified to assess the effect of component failures on the QoS of a network.

Grade of Service (GoS)

Grade of service indicates the effectiveness of the network performance by reflecting the degradation in service caused by the contention for the critical network resources. In circuit switched networks the GoS may be specified as the combination of the time between the request for a circuit and the availability of the circuit, the call setup time and the amount of call loss due to unavailability of system resources.

For packet switched systems, the GoS represents the time it takes for a message to reach its destination, or the call setup time for virtual circuit packet switched systems. For both types of switching, the smaller the number of calls/messages that exceeds a generally accepted time frame, the better the system's Grade of Service would be.

2.2 Research Methodology

The research methodology for the thesis is the following:

- a. The analysis of a global network and its components.
- b. The creation of a model based on this analysis, and simulation of that model under various conditions.
- c. The gathering of simulation results and the evaluation of those results.

2.2.1 Network Analysis

The network analysis entails examining existing global networks and their functionality, topology, switching, and traffic characteristics. The characteristics of the hardware utilized by these networks are not analyzed as this is beyond the scope of the thesis.

2.2.2 Modelling and Simulation

A model of a global network is built that incorporates the characteristics of an existing network. It is very rare that a network is built from scratch. Rather, new networks are extensions of old ones and have newer equipment that allows for other applications. This way the capital expense initially incurred when setting up a network is reduced, and the older equipment is phased out rather than thrown out. When modelling a network, the general characteristics of topology, switching, routing, and traffic are studied, then specific characteristics of node and link characteristics are analyzed. The approach to modelling and simulations is top down and multi-level iterative. In this approach a simple simulation model of a network is built and complexity is added until the desired level of modeling is achieved.

This body of work only considers Virtual Circuit Packet Switched networks and their performance.

2.2.3 Evaluation

The evaluation of the models and their simulation results comprise the final section of the thesis. The results are analyzed according to the evaluation criteria of throughput, message delay and Quality of Service. Cost, while significant, is not considered, as pricing and tariffs are beyond the scope of the thesis.

III. NETWORK ANALYSIS

3.1 Analytical Techniques

Before a network is modeled, the attributes critical to the proper operation of the network must be evaluated, so predetermined performance criteria become necessary. These criteria include cost, throughput, delay times, and others which are dictated by the requirements of the network users. This evaluation may be performed on two levels: the network level and the nodal level. The network level of the evaluation entails studying the general characteristics of the network, seen as collection of nodes and connecting links. The nodal level focuses on the attributes of a single node and its characteristics.

The network analysis involves two types of variables: dynamic and static. The dynamic variables, such as traffic characteristics, are expressed in terms of statistical probabilities and probability distribution functions. For traffic attributes, which are a time dependent component of the analysis, we rely upon queuing theory concepts so that we may perform an efficient analysis of the effect of different traffic patterns on a network. The static variables, such as characteristics of nodes and links, account for the major components of the network and must be enumerated for proper analysis.

3.1.1 Queuing Theory

The Queuing theory plays a crucial role in the analysis of networks regardless of the switching method. It enables us to identify probability distributions for events under varying conditions. In the queuing process, information units (packets or calls) requiring service are generated by an input source. These units enter the queuing system and join a queue. At certain time intervals, a member of the queue is selected for service by the server(s), according to some rules known as service discipline. After being serviced, the information unit leaves the system.

A queue is a waiting line in which the incoming information units wait for servicing. The size of the queue may be either finite or infinite. We only consider infinite size queues in our research. The service discipline determines the order in which these units may be serviced. First-in-first-out (FIFO) and priority are the most commonly used service discipline schemes utilized when modelling network traffic activity. The servers are those units within the queuing system which service (process) the informational units. There may be one or more servers working in parallel. Each server completes the service of a unit before starting to serve another. The time spent on service by each server is called the service time.

The input source has two characteristics: population size and distribution function. The population size refers to the number of information units that may need servicing at any time. This number may be finite or infinite. For a large model it is safe to assume infinite population size. The distribution function defines the way the information units arrive at the queuing system. For completeness, we specify a distribution function for the service time for each server. Often, the distribution function for each server is assumed to be identical to the others.

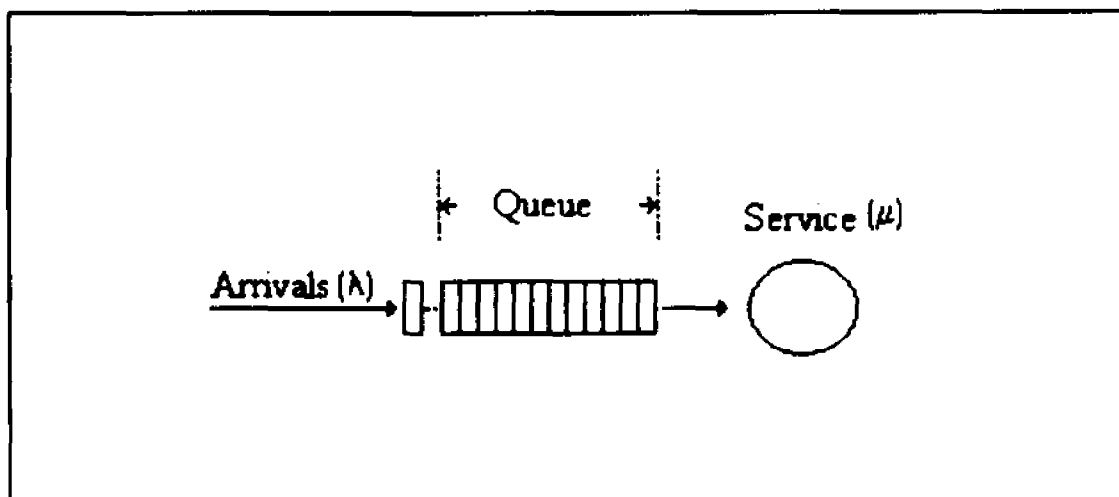


Figure 3. 1 M/M/1 Queueing Model

A shorthand notation, called Kendall notation, after David Kendall, has been developed to describe queueing systems and has the form $A/B/c/K/m/Z$. Here A describes the interarrival time distribution, B the service time distribution, c the number of servers, K the system capacity (the maximum number of requests allowed in the queue), m the number in the source and Z the queue discipline. Usually the shorter notation $A/B/c$ is used and it is assumed that there is no limit to the queue size, the message source is infinite and the queue discipline is FIFO. Some of the symbols traditionally used for A and B are:

- GI general independent service time;
- G general service time distribution;
- M exponential interarrival or service time;
- D deterministic (constant) interarrival or service time distribution;

When modelling telecommunications systems, it is customary to utilize $M/M/n$ queueing system since these systems closely resemble a Poisson process.

3.1.2 The Poisson Process

In a telecommunications network, the number and the size of messages arriving at a node may vary with time resulting in different probability distributions at different times. This type of statistical behavior is called a stochastic process. A type of stochastic process that is especially useful when modelling networks is the Poisson process. A process is a Poissonian with a rate λ if the following conditions hold.

- i. The process has independent increments. (Events occurring in non-overlapping intervals of time are independent of each other)
- ii. The increments of the process are stationary (The distribution of the number of events in any interval of time depends only upon the length of the interval and not when the interval begins)
- iii. The probability that exactly one event occurs in any time interval of length h is $\lambda h + o(h)$.
- iv. The probability that more than one event occurs in any time interval of length h is $o(h)$.⁶

⁶A.O. Allen, *Probability, Statistics and Queueing Theory with Computer Science Applications*. Academic Press, New York, 1978.

The above definition implies that the Poisson process has the Markov property which is sometimes called the memoryless property. That is to say that the probability that an event takes place does not depend on the events that have already taken place. Instead this probability depends on the time interval between the two events. Consequently, if a process is a Poisson process with rate λ , then it has a Poisson distribution which has the probability of k events (arrivals) taking place in any time interval of length $t > 0$ as

$$e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k=0,1,2,\dots$$

with the mean and variance each equal to λ . The random variables associated with a Poisson distribution are discrete random variables. For modelling and analysis of mathematical models it is more helpful to work with continuous random variables. Fortunately, if we consider the time between two consecutive arrivals (interarrival time) rather than number of arrivals within a time interval, we find that it has continuous random variables with a negative exponential distribution. This distribution has the probability function

$$\lambda e^{-\lambda x}$$

for $x > 0$, where x is the waiting time between two arrivals, with the mean equal to $1/\lambda$ and variance equal to $1/\lambda^2$. The exponential function also preserves the Markov property, that is to say that the distribution of the wait time does not depend on the amount of time that has already passed.

The Markov property of the Poisson process is very useful when modelling arrivals to a queue at a node. The arriving messages may come from many different nodes, therefore are independent of each other. Hence the number of arrivals within a time interval is independent of other arrivals, depending only on the length of the time interval

As with all statistical models, it must be pointed out that the empirical data obtained from observations may not exactly follow the Poisson process. However, the nice mathematical properties of the Poisson process enables one to build and analyze a queueing system model easily. Fortunately, the assumption of a Poisson process can be justified on the grounds other than those of analytical convenience. The Poisson distribution, like the normal distribution, turns up as the limiting distribution of large sums.⁷

The service time of a message depends on its size. If the length of a message has Poisson distribution, then the time it takes to service that message has exponential distribution. Therefore, we can use the exponential distribution to model service times by the servers, implying that the probability of completing the service of a request is independent of how long that request has been serviced. Consequently, if the mean service time is μ per request, then the average service rate per server would be $1/\mu$

⁷P. Bratley, B.L. Fox and L.E. Schrage, *A Guide to Simulation*, Springer-Verlag, New York, 1983.

Hence, when studying the behavior of telecommunications networks, M/M/1 queues are utilized, implying that these queues have exponential interarrival and service times with a single server. The queue size is unbounded, with an infinite number of messages in the system and the service discipline is First Come First Serve.

3.1.3 Network Attributes

The network attributes that impact network operation and therefore must be considered during network analysis, are connectivity, traffic flow, and capacity.

3.1.3.1 Network Connectivity

Network connectivity of an n node network can be expressed in terms of a connection matrix $C(n)$ as shown below:

$$C(n) = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}$$

where $c_{ij}=1$ represents a direct connection between node i and node j , and $c_{ij}=0$ implies an absence of a direct connection between node i and node j . A measure of connectivity can be defined as:

$$K = \sum_{i=1}^n \sum_{j=1}^n c_{ij} \leq n^2$$

which is the number of links connecting the network nodes.

3.1.3.2 Network Flow

In order to study the network flows, we first enumerate the paths between two nodes. A path between nodes i and j can be represented by a directed vector whose elements are the links encountered en route:

$$P(i,j) = (c_{i_1}, \dots, c_{i_j})$$

With the aid of a connectivity matrix, one can enumerate many paths between two nodes representing the network flows. These paths may be of different lengths, depending on the topology of the network. Once these paths are enumerated, we can assign traffic loads to each of the branches of a selected path to assist us in computing the load of a network.

3.1.3.3 Network Capacity

The capacity of a network can be defined as the total number of calls or packets that can be sustained during an average second of the busy hour. The busy hour is that 60 minute period during the day when the traffic is the highest. The capacity of the network is directly related to the capacity of each node and link to handle traffic, as well as the length of the chosen path. The capacity for a packet switched network is expressed in terms of kilobits per second (kb/s).

3.2 Network Analysis

In order to understand how a network behaves under changing conditions it is important to analyze those parameters that affect the network's performance. These parameters differ depending on the type of switching utilized by the network.

3.2.1 Packet Switched Networks

For Packet Switched networks, some of the parameters that have an impact upon the performance of a network are link utilization, packet and message delays within the network, transmission costs, traffic intensity, and routing.

3.2.1.1 Link utilization

A close relationship exists between link utilization, frame length, link capacity, and propagation delay on the link, bit error rate on the link and the data link control procedure utilized. One can develop mathematical models for a link utilization between two nodes and since most systems have similar links and physical properties, we can extend these results to all the links within the network. We can define link utilization as

$$U = \frac{T_f}{T_t}$$

where T_f is time for transmitter to emit a single frame and T_t is total time that the line is engaged in the transmission of a single frame. If there are any errors, then the utilization

becomes:

$$U = \frac{T_f}{N_r T_r}$$

where N_r is the number of retransmissions of a frame

3.2.1.2 Packet and Message Delay

In order to analyze delay times within the network, we first need to enumerate all the paths within the network. Then the delay in each link of a typical path can be computed, using the queueing model discussed above. If C is the capacity of the link (bits/sec), λ is the arrival rate per second, and B is the average packet or message length in bits, then we can express average delay time as:

$$W = \frac{1}{\frac{C}{B} - \lambda}$$

with standard deviation represented by the same expression. Therefore by knowing the actual arrival rate and capacity of each link, one can compute the average delay of each link in the path. In addition to these delays we must consider those introduced in each node in the path.

3.2.1.3 Transmission cost

Transmission cost analysis is difficult, due to the varying nature of traffic demands and tariff structures. These costs can be fixed or variable. The fixed costs include the costs related to the hardware or software acquisition. The variable costs depend on the amount of traffic that a network handles as well as the timing of the traffic. Most routing strategies consider cost to be one of the criteria upon which routing decisions are based.

3.2.1.4 Routing

Routing schemes influence the operation of a network. The optimal routing scheme depends first on the performance goals, usually maximum throughput and minimum delays, then on the connectivity and the transmission cost structure of a network. Hierarchical routing can be efficient for large networks composed of regional clusters by minimizing the size of routing tables, and reducing nodal complexity. A dynamic routing scheme may be more appropriate for smaller networks.

3.2.1.5 Traffic

The analysis of nature and traffic patterns helps us understand and streamline the operations of a network. Different types of traffic may have different requirements. While simplex links may be adequate for one directional traffic, it is essential to utilize full duplex links for interactive or bidirectional traffic. In any case, both half duplex and full duplex traffic have the same mathematical modelling properties of the Poisson process.

In addition to the amount of data carried by the traffic, the traffic patterns in a 24 hour period dictate the characteristics of a network link or node. If there is a very high volume of interactive traffic going through a particular node at a certain time, that node needs extra resources to be allocated to it.

3.2.2 Circuit Switched Networks

Since circuit switched networks provide connection oriented services, it is appropriate to analyze the end-to-end performance of the network. In order to accomplish this, we need to examine the call connect time and the blocking probability of the call within the network.

3.2.2.1 Call connect time

Call connect time is measured as the time elapsed between the call request and call connect. The total connect time can be expressed as the sum of the processing time delays experienced by each network node and the transmission delays experienced by signaling data on each link of the path. Because of the Poissonian nature of the incoming calls at each node, the connect time can be represented by random variables with negative exponential probability distribution.

During call connect time, an available path between the source (caller) and the destination (receiving) nodes is found. During peak usage hours, due to limited availability, it might

take longer to locate an available circuit. Once an available path is found, the connection is made and the data transfer resumes.

3.2.2.2 Blocking Probability

The blocking probability is the probability that a call request is not satisfied due to unavailability of resources. The blocked calls are either cleared or postponed by placing them in a queue. The amount of traffic in the network depends on the average holding time and the number of calls or carried traffic. Offered traffic is the demand that users place on the network. Both carried and offered traffic are measured by erlangs which is call arrival rate (per hour) times average holding time (in hours). The difference between the offered and carried traffic determines the blocking probability. Given that the blocked calls are cleared, the Erlang-B formula calculates the probability that a call is lost by:

$$P_B(s,E) = \frac{\frac{E^s}{s!}}{\sum_{k=0}^s \frac{E^k}{k!}}$$

where s is the number of resources (circuits) and E is the offered traffic in erlangs

3.3 Node Analysis

A network node performs a number of functions, of which switching is the most important. Switching may provide temporary storage for messages in transit, may maintain

and update routing information and may perform functions which are occasionally necessary, such as assuming the role of the network master node.

3.3.1 Critical Resources

In order to analyze and measure the performance of a network node, a set of resources and their impact on the node must be identified. In reality there are many resources that impact the performance of a node. The magnitude of this impact differs, depending on the type of switching used by the network. The initial task is to identify and analyze those resources that have the greatest impact upon node performance in a virtual circuit packet switched network node.

3.3.1.1 Packet Buffer Size

Packet buffer size is the amount of memory in the node available to store incoming packets until they are processed and sent to the next node via links. There are two types of buffers: incoming packet buffer and outgoing packet buffer.

The incoming packet buffer holds those packets that have arrived from the incoming links and are ready to be processed by the processing unit. If the size of the incoming packet buffer is too small, the buffer fills up quickly and some of the incoming packets are blocked resulting in the loss of information and forcing retransmission. If the buffer is too large, the result is the waste of buffer space.

The outgoing packet buffer contains those packets that have been processed and are ready to continue on their trip toward the destination node. Number of outgoing packets varies depending on the number of outgoing links originating from the node, and the different priorities each of those links has. There is at least one outgoing packet buffer per outgoing link on the node

3.3.1.2 Processor

The nodal processor is responsible for processing each incoming packet and routing it to the appropriate outgoing packet buffers. Depending on the overall routing scheme, the processor performs a varying number of tasks with different complexities. Each of these tasks usually takes a small amount of time to perform. As the number of packets in the system grows, the processing time also increases, thereby contributing to the overall packet and message delay times.

In our studies, when measuring total delay times the processor delay times are not considered, as this time is constant for each packet and is minimized by either utilizing a very powerful processor or by utilizing an array of processors.

3.3.2 Other Resources

The other resources in the node include hardware and software that perform the switching functions. Main memory affects the amount of information handled by the processor per unit time. It contains buffers and tables (e.g. virtual circuit tables) which contain data needed by the processor. The secondary memory dictates the amount of information that can be stored at that node. Software resources also have an impact upon the performance of a node.

Buffering software, scheduling software, the node operating system, and routing algorithms are among software resources that contribute to the overall efficiency of the node.

IV. MODELLING AND SIMULATION METHODOLOGY

A model is an imitation or an approximate representation of a concept, an object, a system, or a process which is referred to as a prototype. Models are used to study, plan, and design prototypes. In most cases modelling reduces cost and risk. Frequently, modelling is the only way to accomplish complex tasks because the actual system may be unavailable or unusable. In predicting network behavior, several types of modelling are used. Analytical modelling is a technique that uses a model of a system that is mathematically solvable. All questions relating to the system's behavior can be answered easily as long as the system is small and well understood. For large and complex systems, however, analytical modelling proves to be inadequate. As the complexity increases, the solvability of the mathematical equations decreases. The alternative to analytical modelling is simulation. Usually it is possible to simulate each component of a system in detail, resulting in an emulation of the system. In addition, simulation of the statistical behavior of a component or a group of components result in statistical simulation. When modelling with simulation, both statistical simulation and system emulation is used to produce logical and thorough results.

While modelling a network, there are many parameters that need to be carefully set up in order for the simulation to produce meaningful results. Once these parameters are in place, it becomes possible to vary some of them using multilevel iterative modelling to investigate the performance of the network under varying operating conditions and complexity. These parameters fall into two categories: the statistical distributions which

control the frequency, and quantity of the traffic and characteristics of the network elements through which the traffic is handled.

4.1 Statistical Distributions

The statistical distributions used for the simulation models determine the type and the amount of traffic going through the network. These distributions control the packet arrival rate, packet service time, and the size of a message. As discussed in section 3.1.2, the message arrival and service times are part of a Poisson process, which imply that the interarrival and service times have negative exponential distribution. The queues implemented in the simulation are unbounded (no limit to the amount of buffer space) with a single server. Using the notation introduced in section 3.1.1, the queueing model is $M/M/1$.

4.1.1 Arrival rate distribution

The message arrival rate determines the frequency of arriving messages from a link into the incoming data buffer at a node. This, in turn, affects the amount of traffic that is carried by the network. The arrival rate is also significant in determining the size and utilization of the incoming link queues.

4.1.2 Service time distribution

The service time distribution is the amount of time that it takes for a message to be processed and sent out on an outgoing link. The total service time depends on the size of

the message being processed. If the size of the incoming messages changes depending on the time of day, then the message length must have Poisson distribution. Consequently, the service time for those message can be modelled by a negative exponential distribution

4.2 Simulation model elements

The simulation model consists of many elements that work together to make up a complete network. Since the network is a global one there are certain restrictions placed on the topology of a model by the physical constraints of the actual network. The other components of the network are the characteristics of the nodes and links, as well as the traffic patterns and routing information. All of these components have profound effects upon the performance of the network model. The simulation software needs to include all these criteria in order to produce results that are as close as possible to the real operation of a network.

The simulation package used in this research is a commercially available capacity planning and network design aid called COMNET II.5 (Release 5) produced by CACI Products Company of La Jolla, California. The general network input parameters into the simulator are topology, traffic data, and routing. Other input parameters are node and link characteristics that are specific to a node or a link and the characteristics for each virtual circuit. The simulation package produces a main report as its output. This report includes statistics for links, node utilization, message delay, and virtual circuit calls, as well as the overall network statistics.

Link statistics are average bandwidth utilization, percent of capacity utilization, the number of frames sent and resent, the number of packets sent, and the average and maximum packet queue time.

Node utilization statistics represent the average and maximum node buffer utilization, the number of packets processed, the node processor utilization, and the average and the maximum packet switch wait time. For each virtual circuit, message delay statistics include the number of messages sent and received, the number of packets processed, average size of messages, the average and maximum message delay and the percentage of message delays above three seconds. This time interval is set as the threshold for many connection oriented services above which the end-to-end message delay becomes unacceptable. The virtual circuit call statistics represent the number of calls, and the average and maximum call setup delay. The simulation package also produces link and buffer utilization graphs. These graphs can either be viewed by the package's graphical interface or can be exported in the ASCII format for further processing.

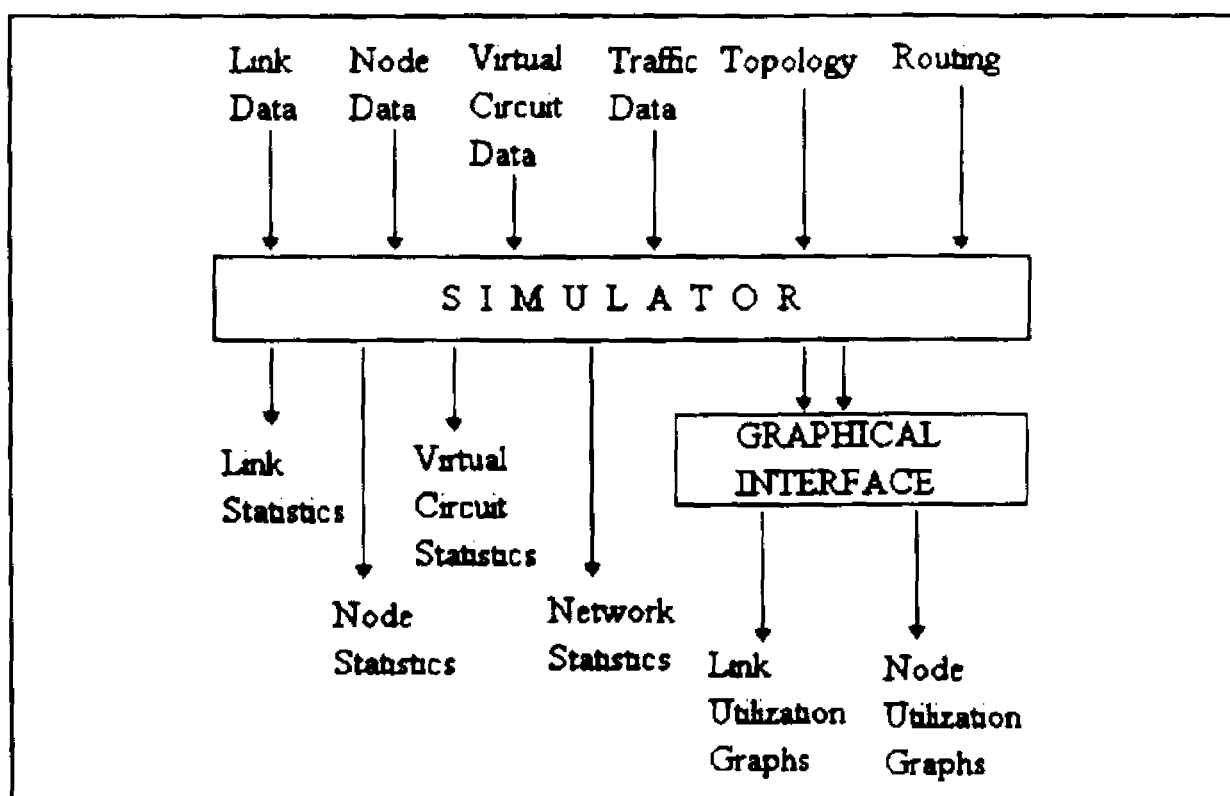


Figure 4.1 Simulation Block Diagram

4.2.1 Global Network Model Characteristics

The global network models that are used in the simulations have been derived from an actual global network, that is, from a collection of many national and regional networks. This international network, made up mostly of universities and corporate research centers, is called BITNET. It contains the regional networks of BITNETII in U.S.A., Netnorth in Canada, JANET in Japan and the Far East, and European Academic Research Network (EARN) as well as other smaller regional networks.

Historically, BITNET was the first of several transparently connected cooperating networks which now constitute the international BITNET network. This network connects more than 1,400 organizations in 49 countries for the electronic exchange of non-commercial information in support of research and education. BITNET users share information via electronic mail to individuals and shared-interest groups, transfer documents, programs, and data, access BITNET server machines and associated data services, and exchange brief, nearly-interactive messages. Nearly 3,000 discussion groups on BITNET cover most topics of academic interest, and may have from five to several thousand participants. Gateways allow the exchange of electronic mail among BITNET, the Internet, and other networks.

BITNET is a "store-and-forward" network based on IBM's NJE communications protocol, information originating at a given BITNET-connected computer (node) is received by

intermediate nodes and forwarded to its destination. Computers in BITNET are interconnected by leased lines or other permanent links. BITNET nodes may be IBM VM/CMS or MVS systems, or non-IBM computers that can communicate using the NJE or TCP/IP protocols running under VAX/VMS, Unix, CDC, and other systems.

Because this network is a hybrid of other networks, the connections between the regional networks are few and are through high bandwidth links. These links are connected to gateways, which accept and forward a large amount of the incoming message traffic to other nodes within the regional networks. The crucial nodes of the global network make up the gateway nodes, and the crucial links of the network are the links that connect the gateway nodes.

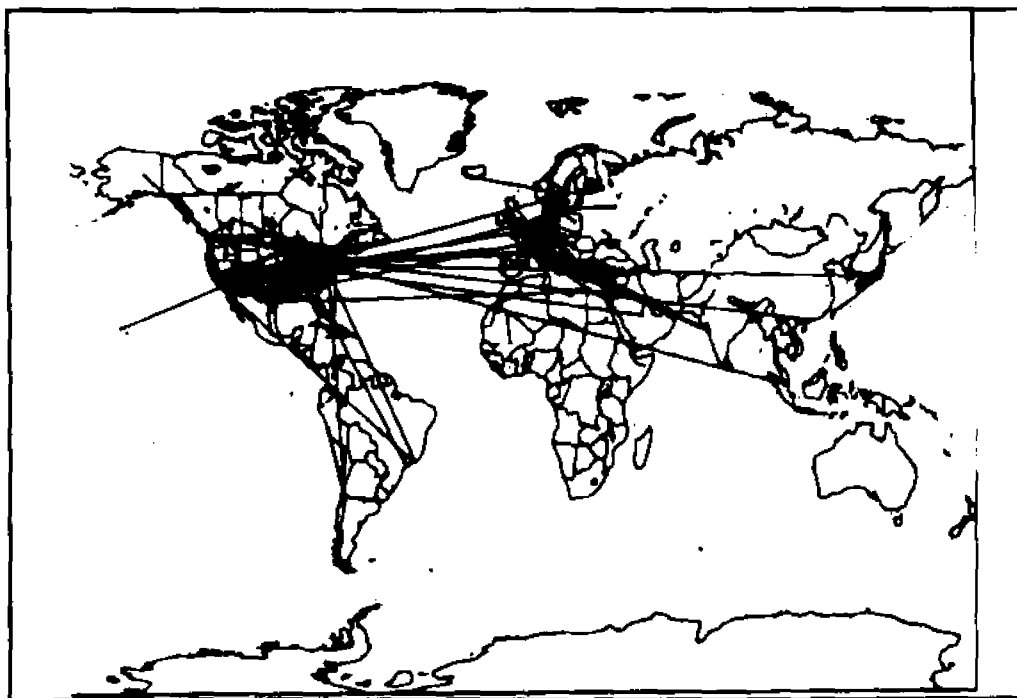


Figure 4. 2 Global BITNET Connectivity

4.2.2 Topology of Simulated Networks

The topologies of the simulation networks investigated in this thesis are derived from the actual topology of the global network of BITNET. The simulation nodes represent the major nodes of the global network and are connected with links that exist between those nodes. Each node in the model represents a gateway node that connects a regional network to others. For the first iterative simulation model, the topology contains two intercontinental links as well as other short links within the regional networks. It is designed so that if one intercontinental link fails the other one can carry the extra load and deliver the data to its destination.

There are six nodes and six links in this model. In addition there are 30 virtual circuits between the nodes. The nodes are clustered in continents of North America and Europe, where each cluster represents a regional network. The nodes in North American continent are in New York City (CUNY), New York, U.S.A. and Princeton, New Jersey, U.S.A., which are major gateway nodes to other regional networks in North America and other continents. The nodes in Europe are found in Montpellier, France, Milan, Italy; Bern, Switzerland, and Bonn, Germany. Each of these nodes is a major gateway and plays an important role in connecting other European, Middle Eastern and North African countries to the global network.

The links in the model are high bandwidth links connecting major gateways. Two links - New York, U.S.A. to Montpellier, France, and Princeton, U.S.A. to Bern, Switzerland - are transatlantic links with high bandwidth. Other links connect gateway nodes of regional networks to each other

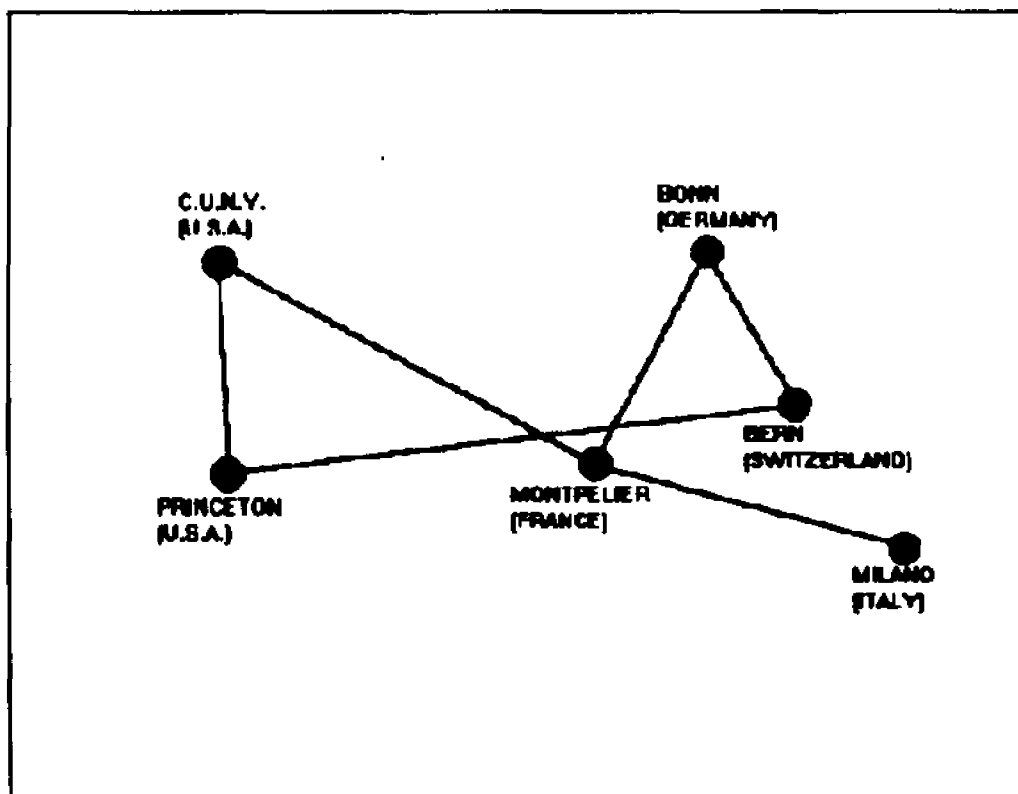


Figure 4. 3. Topology for First Simulation Model

The second iterative model topology is more complex than the previous model. In addition to all the first model's nodes and links, this topology contains 6 other gateway nodes, connecting networks in the Far East, Middle East, South America and North America, including Canada for a total of 12 nodes. There are 36 links, 30 more than the first iterative model, resulting in more paths between any two nodes. There is a total of 132 virtual circuits between the nodes. The net effect is a more robust topology and a more reliable network.

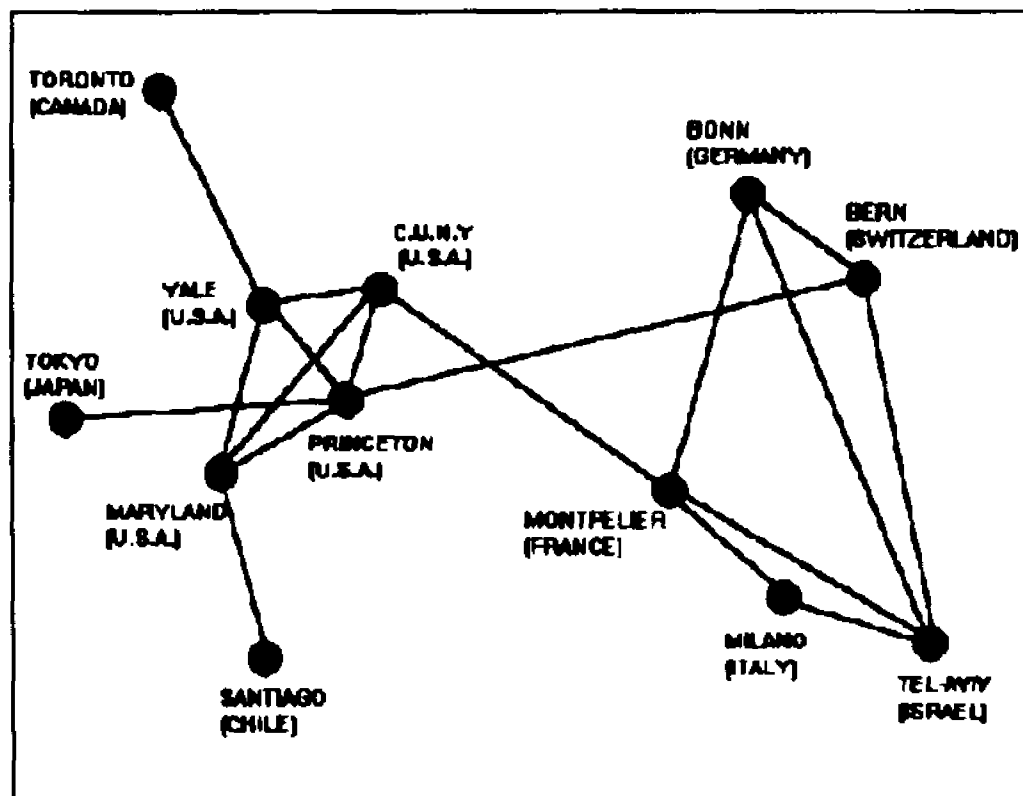


Figure 4. 4 Topology for Second Simulation Model

4.2.3 Simulation Traffic

One of the important parameters of the simulation studies is the amount of traffic generated by a node and destined for other nodes. Since each node in the model represents a regional network, the traffic emanating from this node represents the sum of the traffic originating from the individual nodes within the regional network

The traffic of the model is derived from the traffic data obtained from the studies of BITNET. Traffic records for all messages processed by the CUNY Gateway node was obtained for the month of May, 1993. These records indicate the time (hh:mm:ss), date, source and destination nodes and the size of all messages that were routed through the node during that month. The CUNY gateway is a gateway node for regional as well as distant nodes. The messages passing through the gateway node are categorized into: (a) messages that originate from a regional node and are destined to another regional node, (b) messages that originate from a regional node and are destined to a distant node, (c) messages that originated from a distant node and are destined to a regional node and (d) messages that originated from a distant node and are destined to another distant node (i.e. passing through.)

In order to obtain an accurate number for traffic generated from the C.U.N.Y. region to other gateways, only the messages originating within the region and destined to a distant

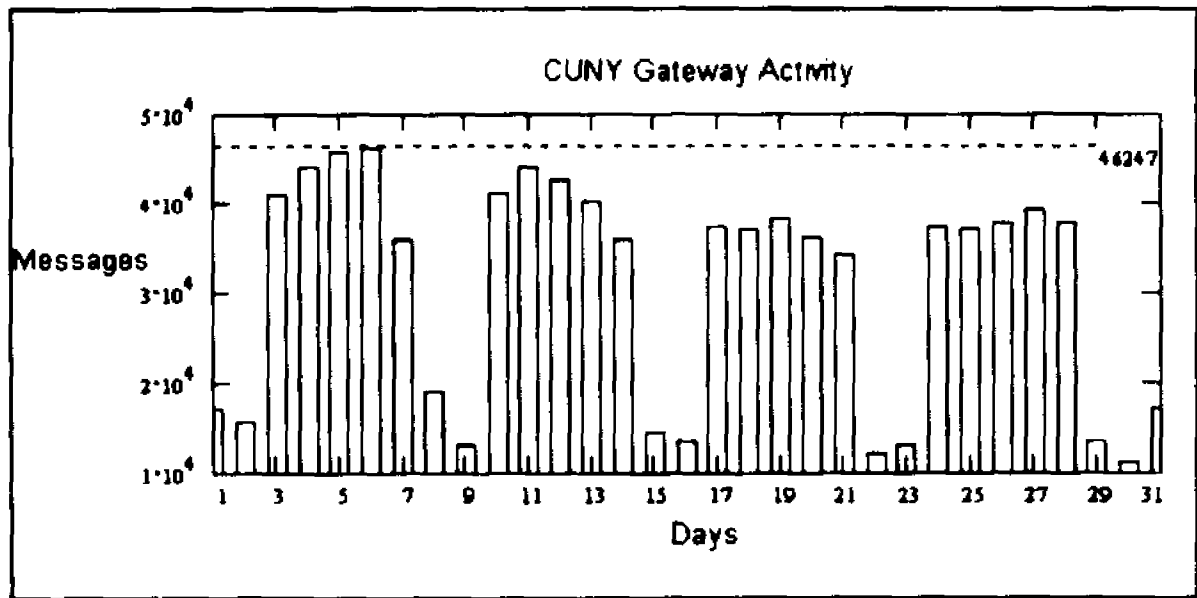


Figure 4. 5. Traffic activity for May 1993 at CUNY gateway

node outside the region are considered. This traffic information is used as a guide in the simulation studies.

Analysis of the traffic showed that most of the messages were transferred during weekdays (Fig. 4. 5), while message activity was lower during weekends and holidays. The peak message activity for the month was on May 6th 1993, a Thursday. That day 46,247 messages originating within the gateway region, passed through C.U.N.Y. with destinations all around the globe.

During the month of May 1993, the heaviest activity within a 24 hour weekday, for the most part, took place between the hours of 8:00 and 19:00. The message activity decreases during the night, with the least amount of activity occurring between 24:00 and

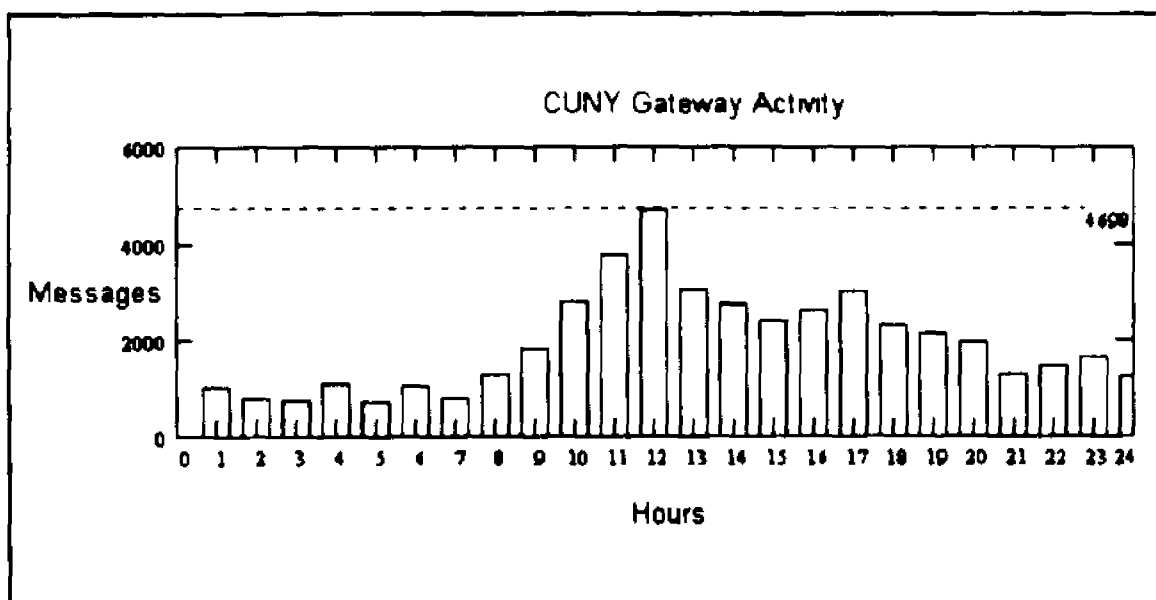


Figure 4.6 CUNY Gateway Traffic Activity for May 6, 1993 (number of messages)

6:00. Fig. 4.7 plots the message activity occurring during the 24 hour period on May 6th 1993, with a maximum of 4698 per hour between 11:00 and 12:00 and a minimum of 689 messages per hour between 03:00 and 04:00. The mean for the 24 hour period is 1850 messages per hour, with standard deviation of 1071. Since the number of messages vary from hour to hour, we can consider message arrivals a stochastic process thereby justifying the use of Poisson process in modelling arrival rates for the simulation models. Further analysis of the traffic shows that the average size of the messages also varies from hour to hour. The average message size has a maximum of 25,385.6 bytes per message from 02:00 to 03:00, and a minimum of 7,611.2 bytes per message between 17:00 and 18:00. The mean is 13680 bytes per message with standard deviation of 5778. Figure 4.8 plots the average message size for the 24 hour period mentioned above.

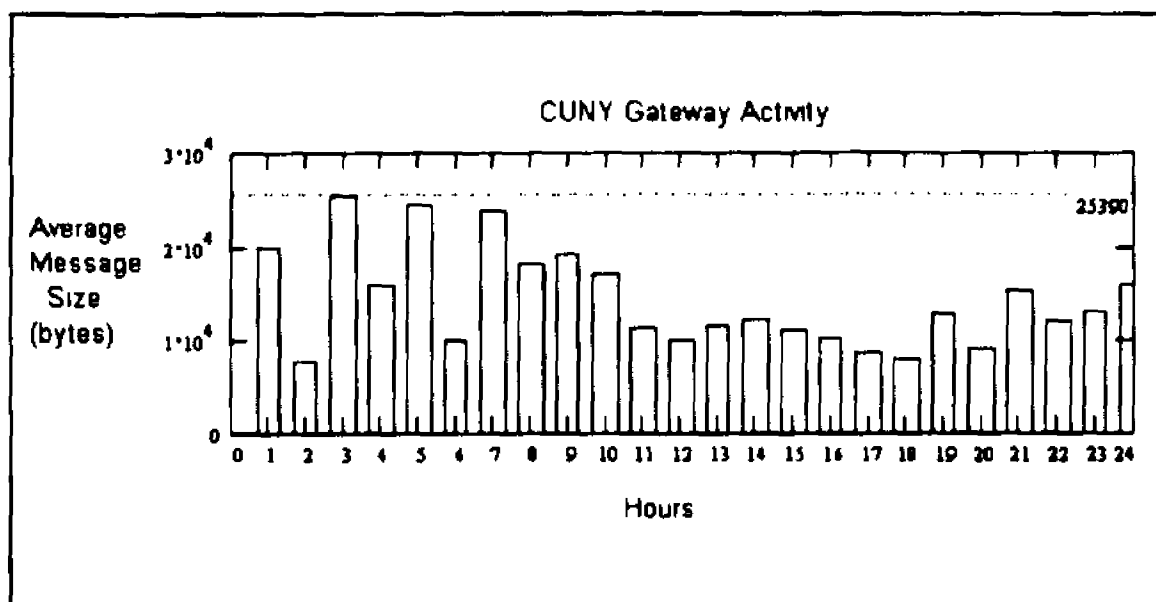


Figure 4.7. CUNY Gateway Traffic Activity for May 6, 1993 (average message size)

This analysis affirms our assumption utilizing the Poisson process to model for message service time which depends on the length of the message being processed.

From this analysis we can approximate the demand that may be placed on the network and use these figures as the basis of the traffic parameter for the simulation models. At this node, the busiest hour of traffic took place on May 6th, between 11:00 AM and 12:00 PM. During this hour there were 4698 messages with an average size of 9885.6 bytes. At the busy hour, the average number of messages per second is 1.305. When this number is multiplied with the average message size, we get an average of 103,205.7 bits transferred per second during the busy hour. This figure is the total number of messages arriving at C.U.N.Y. from the regional nodes destined to nodes in all other regions. Further analysis of the traffic yields busy hours for different links, as the destinations of the messages are global.

When specifying the traffic for the simulations, it was decided to normalize the actual traffic so that the traffic intensity could be varied without compromising the integrity of the data.

To achieve this goal we define *network topology constant* as equivalent to the mean arrival rate of 1024 bytes per minute per virtual circuit in the network. Since the number of virtual circuits in a network depends on the number nodes, the traffic intensity also depends on the number of nodes. For both of the simulation models, there is a virtual circuit connection from each node to all other nodes in the network, resulting in $n*(n-1)$ virtual circuits for n nodes. Therefore, the network topology constant for a model is equal to $1024 \text{ bytes/minute} * 8 \text{ bits/byte} * 1/60 \text{ min/sec} * n*(n-1)$. For the first model this constant is

$$\frac{1024 * 8 * 6 * 5}{60} = 4096 \text{ bits/sec}$$

It is 18022 bits/sec for the second model. Furthermore, the traffic intensity of n is defined to be equal to n times the network topology constant. The traffic intensity can be increased and decreased to achieve higher or lower traffic rates (e.g. 5 for five times the traffic, or 0.20 for one fifth the traffic). Under this schema, the busy hour traffic we have described above would be equivalent to a traffic intensity of $103,205.7 / 4096 = 25.19$ for the first and $103,205 / 18,022 = 5.73$ for the second simulation model.

4.2.4 Nodes

The nodes in a network model represent the sources and destinations of the traffic. In addition, nodes typically perform some communication processing function such as switching, routing, multiplexing or bridging. For each node in the simulation models the following node attributes are specified.

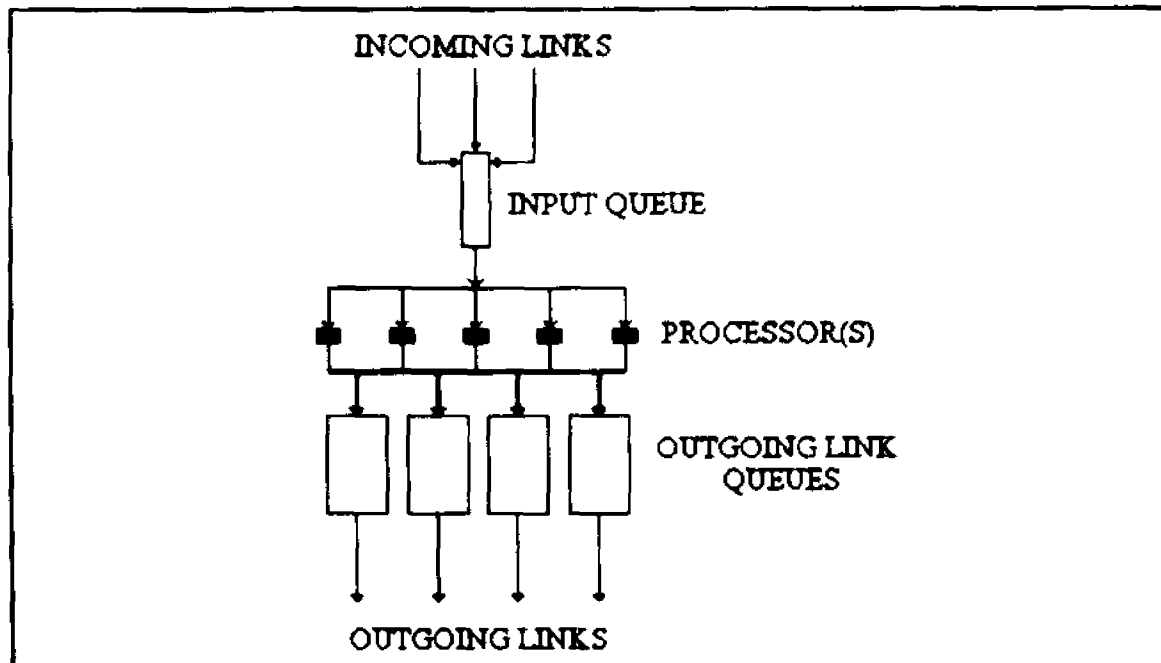


Figure 4. 8 Packet Switch Node Organization

ID Code and Name

Each node is identified by a unique identification code, which is a maximum of 6 characters and a name which fully identifies the node.

Number of Processors

A single packet switch node can have one or more processors to serve arriving packets. Packets arriving at a node enter a single queue ordered by packet priority. The queue is served by a collection of processors that operate in parallel. A packet waiting at the head of the queue is served by the first processor to become idle. The time required by a single processor to serve a packet is given by either the total packet processing time or the virtual-call setup packet processing time, depending on the type of packet.

Buffer Size (bytes)

The packet buffer capacity is the total amount of space available at a node for packets that are waiting for transmission over an outgoing link. If the packet buffer capacity is unspecified it is assumed there is no limit on packet buffer capacity. The buffer capacity and packet sizes are defined in units of bytes (8 bits) which are octets (CCITT) or characters.

Packet Processing Time per Packet (ms)

The packet processing time per packet is the time required by a single processor at a packet switching node to process a packet. For example, a packet switch that has a capacity of 1,000 packets per second requires about one millisecond to process a packet, assuming a single processor and ignoring queueing delays.

Packet Processing Time per Virtual Call Setup (ms)

The processing time per virtual call setup packet is the time required to process a virtual call request packet. For a packet-switched network using virtual circuit operation, a call request packet typically takes longer to process than an ordinary data packet, since a route must be selected for the call request packet. Ordinary data packets just follow the virtual circuit path established by the call-request and call-connected packets.

Packetizing Delay (ms)

The packetizing delay is the time between the formation of the packets in a message. It is dependent on the speed of the processor at the node.

4.2.5 Links

A link is a transmission facility connecting two or more nodes. A link group is composed of one or more identical links. Multiple link groups can connect the same nodes. For example, a pair of nodes may be connected by a group of microwave links and a group of satellite channels. The following sections give the characteristics for each type of link.

ID Code

Each link group is identified by a unique identification code.

Protocol Type

The access protocol type determines whether a link can connect exactly two nodes (point-to-point operation) or an arbitrary number of nodes. These simulation studies only consider point-to-point links. Aside from point-to-point link access protocol, there is ALOHA, which is used mainly for satellite links; Carrier Sense Multiple Access/Collision Detect (CSMA/CD), which is used for IEEE 802.3 (Ethernet) types of LANs; Token Passing, which is used by IEEE 802.5 (Token Ring) types of LANs; and Polling, which is typically used to model multidrop telephone lines that use a centralized computer to poll the other devices (such as dumb terminals) on the line.

Point-to-point links connect exactly two nodes. Circuit-switched traffic can only be routed over point-to-point link groups, whereas packet-switched traffic can be routed over any

of the link types. The capacity of a point-to-point link group is defined by the number of links (circuits) in the group for circuit switching or the total bandwidth (kb/s) of the group for packet switching

Transmission Rate (kb/s)

The transmission rate is the data transfer speed, in kilobits per second, of an individual link in the link group

Full Duplex

Full duplex link groups can transmit simultaneously in opposite directions using separate channels. When specified, the simulation package models the separate channels during the simulation.

Transmission Return Rate (kb/s)

The transmission return rate is the transmission speed in the reverse direction for a full duplex link group. This rate is the same for both directions in the simulation studies

Transmission Queues

Packets waiting at a node for transmission on a link group are assigned to a transmission queue based on the packet's priority. All packets within a single transmission queue are in priority order. The link group serves the transmission queues in order of decreasing minimum queue priority. A non-empty transmission queue, for which a non-zero

maximum queue service time is specified, is served until the queue service time threshold is exceeded. Service of the queue is interrupted nonpreemptively whenever packets are waiting in a higher priority queue that has a zero service time threshold. Multiple transmission queues can be used to model a mechanism used for a dynamically allocating bandwidth to integrated voice and data packets.

The simulation studies use only one transmission queue which, in effect, does not discriminate among the packets waiting to be transmitted. However, dynamic allocation of the bandwidth is an important aspect of network design and will be considered for future research.

Propagation Delay (ms)

The propagation delay is the time required for an electrical signal to be transmitted over a link group. There is an approximate delay of one millisecond for every 100 miles, which is negligible for most terrestrial circuits. However, for a satellite circuit, the propagation delay (up and down) is about 250 milliseconds.

Maximum Virtual Circuits

Each link group has a limit on the number of virtual circuits that can be established within the group. When the limit is exceeded virtual call setup is retried after the virtual call retry interval.

Time-Until-Failure Distribution (min)

The time-until-failure is the period during which the link group is operational. The link group is operational at the start of the simulation and remains so for an amount of time set by the time-until-failure distribution. The link group then remains in a failed state for an amount of time specified by the time-to-repair distribution. The link continues to alternate between operational and failed states for the duration of the simulation.

Time-To-Repair Distribution (min)

The time-to-repair is the period after a failure during which the link group is unavailable.

Frame Minimum (bytes)

When a transmitting node has a frame ready for transmission over a link with a minimum frame size, it adds enough "pad" characters so that the transmitted frame equals the minimum frame size. This feature is utilized in some access protocols, such as IEEE 802.3, to make it easier to distinguish valid frames from invalid ones.

Frame Maximum (bytes)

The frame maximum is the size limit of a link-level transmission unit. The maximum number of packet bytes that can be carried by a frame is given by the frame maximum minus the frame overhead. A packet too big to fit in a single frame is sent in as many frames as necessary. If no frame maximum is specified, each packet is transmitted as a

frame and the frame overhead is added to the packet size. When a frame maximum is specified, the transmitting node combines small packets into a single frame, and decomposes big packets into multiple frames.

Frame Overhead (bytes)

Frame overhead corresponds to the link-level protocol bytes included in the frame maximum. The frame overhead is added to the number of packet data bytes in a frame for purposes of computing frame transmission time. The frame overhead is never stored at a node; it is added to packet data bytes prior to transmission on a link and stripped off when the frame completes transmission.

Frame Error Probability

The frame error probability is the likelihood that a maximum-size frame requires retransmission. For frames smaller than the maximum, the probability of retransmission is reduced linearly by the ratio of the transmitted frame size to the maximum frame size. If no maximum frame size is specified, the frame error probability is the possibility that a packet (which would be sent as one frame) requires retransmission, regardless of its size. When a frame's transmission time has elapsed, and it is determined that the frame requires retransmission, the next transmission attempt is scheduled to occur in the simulation model without any delay.

Buffer Maximum (bytes)

The maximum buffer use is the maximum space available at a node for packets queued for the link group. The maximum buffer use allowed can be utilized to insure that packets waiting for a single link group do not monopolize the packet switch's buffer space

The links in the simulation models are point-to-point digital links with bandwidth ranging from 9.6 kb/s to 1544 kb/s. For the first iterative simulation model there are three types of links: half duplex, full duplex, and with failure. A half duplex link can transmit data in one direction at a time, whereas a full duplex link can transmit both ways simultaneously. A link with a failure is one that ceases to operate within a certain time into the simulation and remains non-operational. The purpose of link failure is to observe its effect on the rest of the links in the network.

The maximum buffer space for outgoing packets is unbounded since, in reality, it is main memory and can be allocated or deallocated dynamically, as demand dictates.

The data is transmitted over the links in frames. In the simulations there are no minimum or maximum frame sizes. Having minimum and maximum frame sizes tends to increase throughput at the expense of delay times.

In order to obtain simulation results that yield the upper bounds for the network performance criteria certain parameters that introduce delay or reduce throughput are omitted or are set to zero. These parameters are packet switching delay, call setup time, packetizing and propagation delays, frame minimum/maximum and overhead, frame error probability. It is fully acknowledged that in real systems these parameters are not omitted and contribute substantially, albeit negatively, to the overall performance of the network. Therefore, in order to investigate the behavior of a network under the "best" conditions and find its limitations, it was necessary to neutralize the parameters listed above.

4.2.6 Routing

The routing utilized in the simulations is adaptive routing with minimum end-to-end delay. Under this routing scheme, each node keeps a table of all paths to all other nodes, and updates the end-to-end delay time for each entry in the table periodically. Alternative routing methods for packet switched service include node by node routing, minimum cost routing and flooding. Each of these routing methods may have different results and is a possible future research topic.

4.3 Network Performance Measures

The simulation package produces several reports summarizing network utilization and network service. The reports that are relevant to packet switching operations are The Circuit Group Report which shows link utilization, frames sent, frames resent, and packet queue statistics by link group. The Buffer Use Report shows buffer use statistics for each

outgoing port at each node. The Transmission Queue Report shows packet queue times for each transmission queue at each port at each node. The Packet Switch Report shows buffer utilization, packet switching delays, and packet throughput statistics for each node in the network. Virtual Call Reports show the network service received by each category of call. Examples of the measures of network service for a message category are the average and maximum response time and the percentage of responses exceeding a predefined threshold. The Message Report also gives total network throughput aggregated over all message categories.

In this research, an overall performance rather than individual node and link performance evaluation of the network is considered. Therefore the overall throughput of the network as well as average and maximum delay times, are the only considerations. In addition a measure of network Quality of Service (QoS) called "the percent of messages over three second delay" is also considered. The individual node and link statistics are not considered since the overall network performance is under research.

V. SIMULATION RESULTS.

The Simulation study results are in two sets corresponding to the two different models used. The first set of results consists of the preliminary research conducted with a simple core network model. The second set of results is from a more complex version (level) of the core network model. Results from these two models expound the differences and similarities, and assemble a set of guidelines that enhance network performance.

5.1 Categories of results

The simulation results consist of three general groups throughput, delay times, and Quality of Service indicator. This indicator is defined as the percent of messages with delay of over three seconds. Throughput measures the amount of data that is delivered by the network within a certain unit amount of time. It is measured in kilobits per second and can be used as an indication of the cost effectiveness of the network.

The delay times consist of maximum and average message delay times. Average delay time indicates the average length of time it takes a message to be delivered from source to destination node. Maximum message delay time indicates the amount of time it takes for the slowest message to be delivered from its source to its destination. The delay times are measured in seconds.

Percent of messages with delay over three seconds measures the percentage of all messages that have arrived to their destination after three seconds of delay. This is a

measure of Quality of Service as described in Section 2.1.4. Three seconds was selected as the threshold after which human beings would start noticing the message delays. This result is expressed as a percentage.

In the second model, which has a larger number of nodes and links and consequently a different topology, same methodology is used to produce the simulation results.

5.2 Simulation Results

Each of the two models generates three groups (throughput, delay times and QoS) of results for four different data rates of 9.6 kb/s, 56 kb/s, 380 kb/s and 1544 kb/s and different link properties (full duplex, half duplex and link failure) within each data rate. Hence the number of simulation results is very large. To provide easy access to the findings we have provided graphs for each set of simulation results. The results from the simulation studies are plotted in graphs for easy readability. Each set of result produces three graphs. The first one is the throughput of the network, the second one is the average and maximum message delays combined, and the third one is the percentage of messages with delay over three seconds.

5.2.1 First Simulation Model Results.

In the first simulation model, the traffic intensity is gradually increased to observe its effects on the network performance. This method is repeated for three different versions of the model: full duplex links, half duplex links, full duplex links with a link failure.

5.2.1.1 Full Duplex Links

5.2.1.1.1 9.6 kb/s Links

The throughput simulation results for 9.6 kb/s full duplex links are displayed in Figure 5.1. In this figure, the throughput of the network rises with the increase of traffic intensity until the traffic intensity reaches 10. At that point, the slope of the throughput curve becomes gentler and at a traffic intensity of 20 levels out at about 58 kb/s and remains there for higher intensities of traffic, possibly indicating that the links have reached their saturation point.

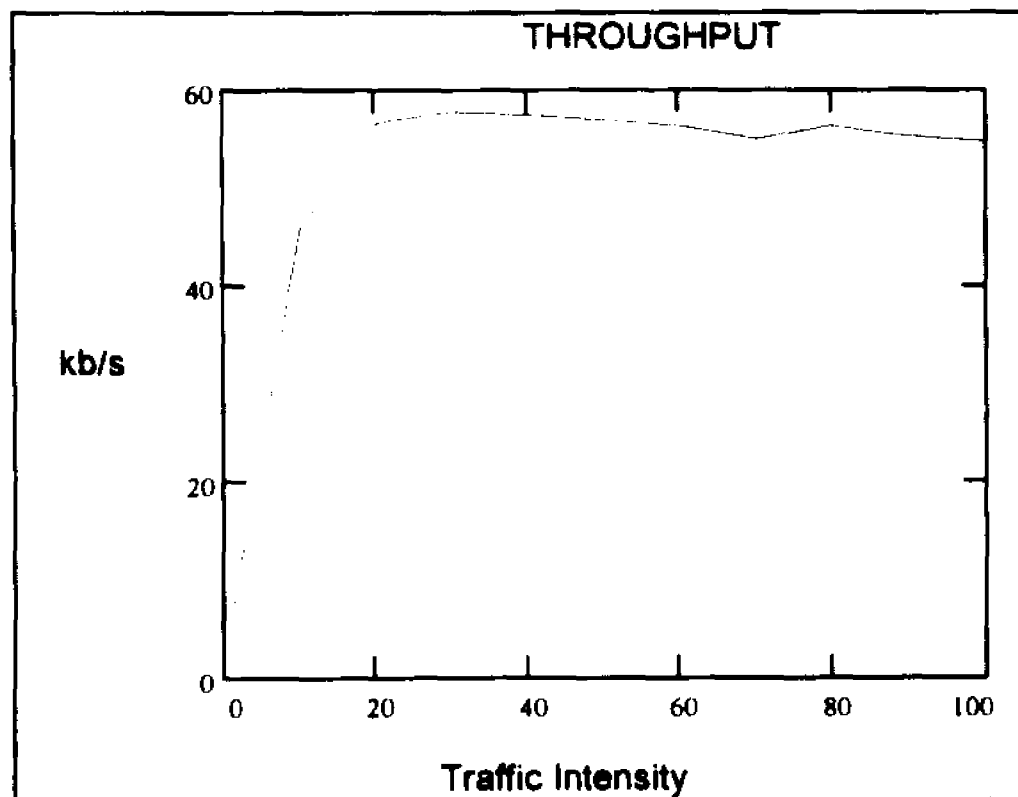


Figure 5.1. Full Duplex with 9.6 kb/s links

The delay times results for 9.6 kb/s full duplex links are displayed in Figure 5.2 with two curves. The average message delay, climbs to 15 sec at traffic intensity of 30 and then levels out rising only very slightly for higher traffic intensities. The maximum message delay curve also increases more steeply than average delay, to about 50 seconds at about traffic intensity of 30 and then levels out around 55 seconds and stays at that level for higher traffic intensities.

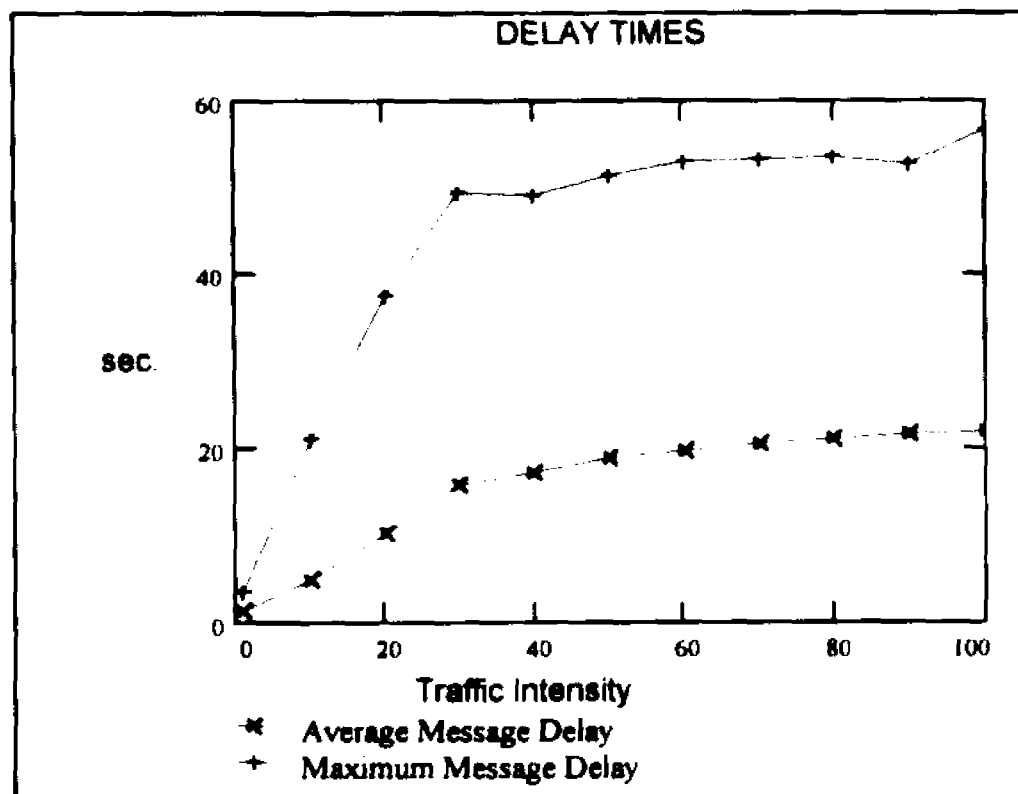


Figure 5. 2. Full Duplex with 9.6 kb/s links

The percentage of messages delayed over three seconds (percent of messages over three seconds) result for 9.6 kb/s full duplex links is displayed in Figure 5.3. This curve is similar to the throughput curve in that it starts with a steep gradient, lessening at traffic intensity of 10. At traffic intensity 30, it flattens out near 95% and remains there for higher traffic intensities.

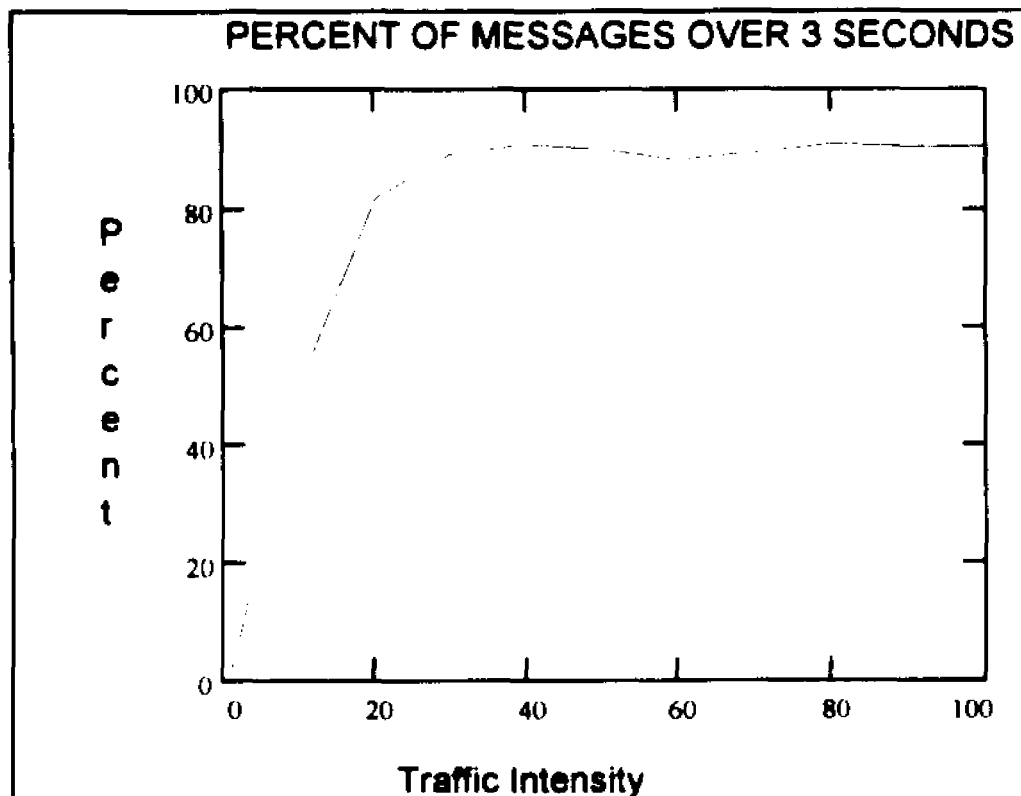


Figure 5.3 Full Duplex with 9.6 kb/s links

5.2.1.1.2 56 kb/s Links

The throughput results for 56kb/s full duplex links are displayed in Figure 5.4. The throughput curve goes linearly up to around 230 kb/s at traffic intensity of 40, then veers off slightly continuing its increase until about 300 kb/s at traffic intensity of 70. It then flattens out, increasing slightly overall, for the higher traffic intensities.

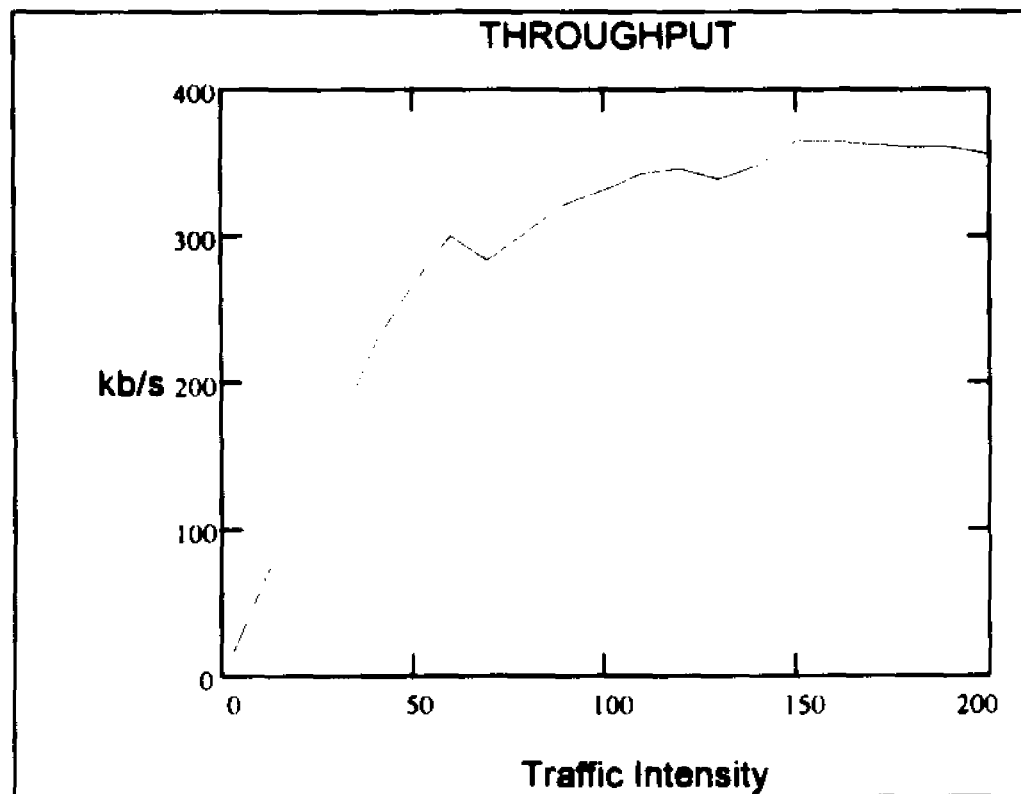


Figure 5.4. Full Duplex with 56 kb/s links

The average message delay and maximum message delay curves for 56kb/s Full Duplex links are displayed in Figures 5.5. The average message delay increases to 0.52 sec at traffic intensity of 40. At that point the slope of the curve increases and continues to increase for higher traffic intensities. The maximum message delay also increases slightly to 2.8 sec. at traffic intensity of 40, then it starts climbing up steeply to 4.83 sec at traffic intensity of 70 and then levels off, increasing slightly for higher traffic intensities.

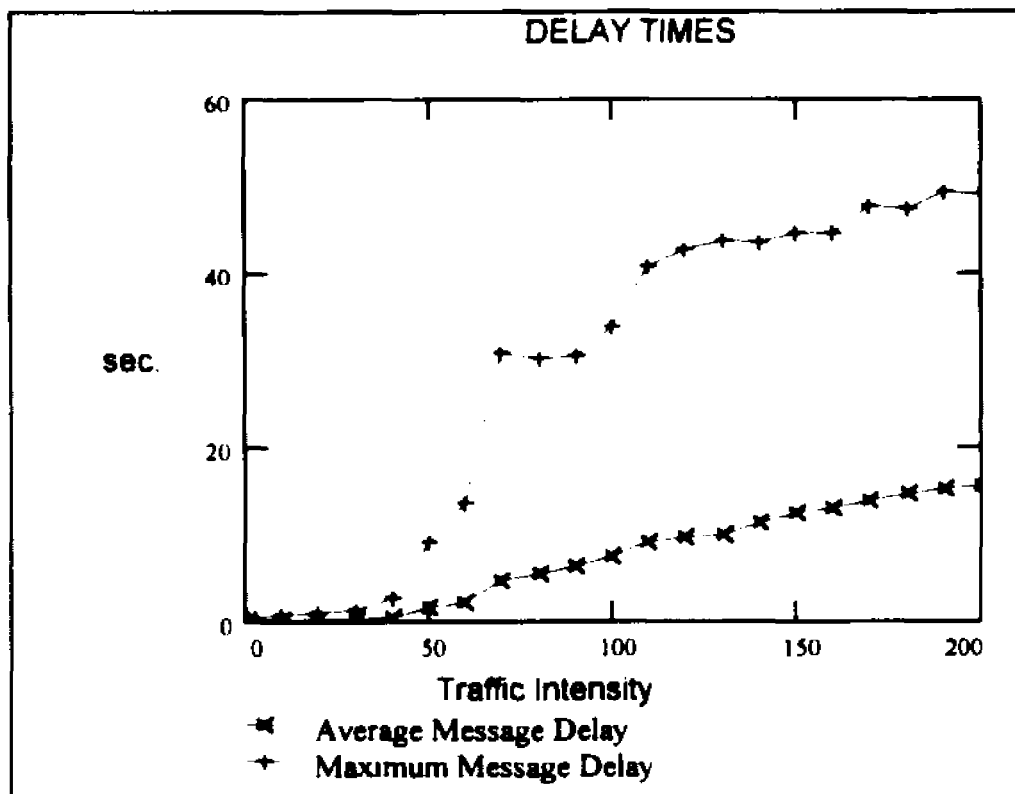


Figure 5.5 Full Duplex with 56 kb/s links

The percent of messages over three seconds curve for 56 kb/s full duplex links, seen in Figure 5.6, starts increasing at traffic intensity of 40, up to 65% at traffic intensity of 90, smoothing off at that point and continuing until traffic intensity of 140 at about 83% and remains in the 80% range for the higher traffic intensities

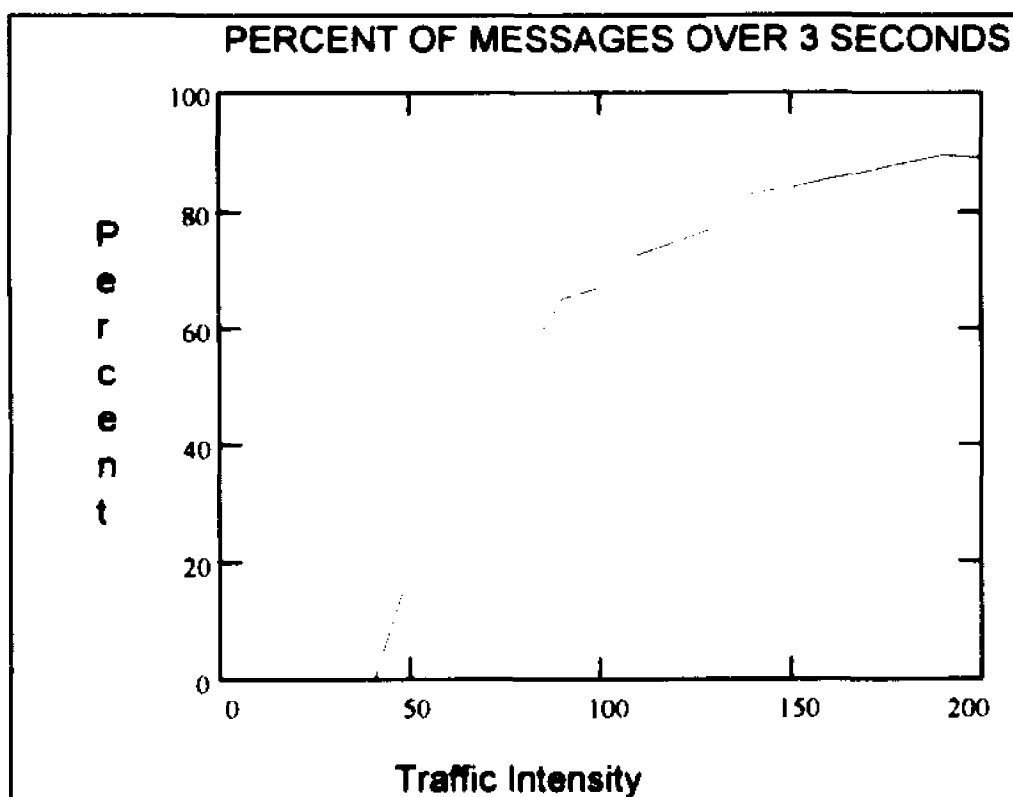


Figure 5. 6: Full Duplex with 56 kb/s links

5.2.1.1.3 380 kb/s Links

The throughput for 380 kb/s full duplex links (Figure 5.7), increases linearly to 1500 kb/s at traffic intensity of 250, then the slope becomes more gradual for higher traffic intensities

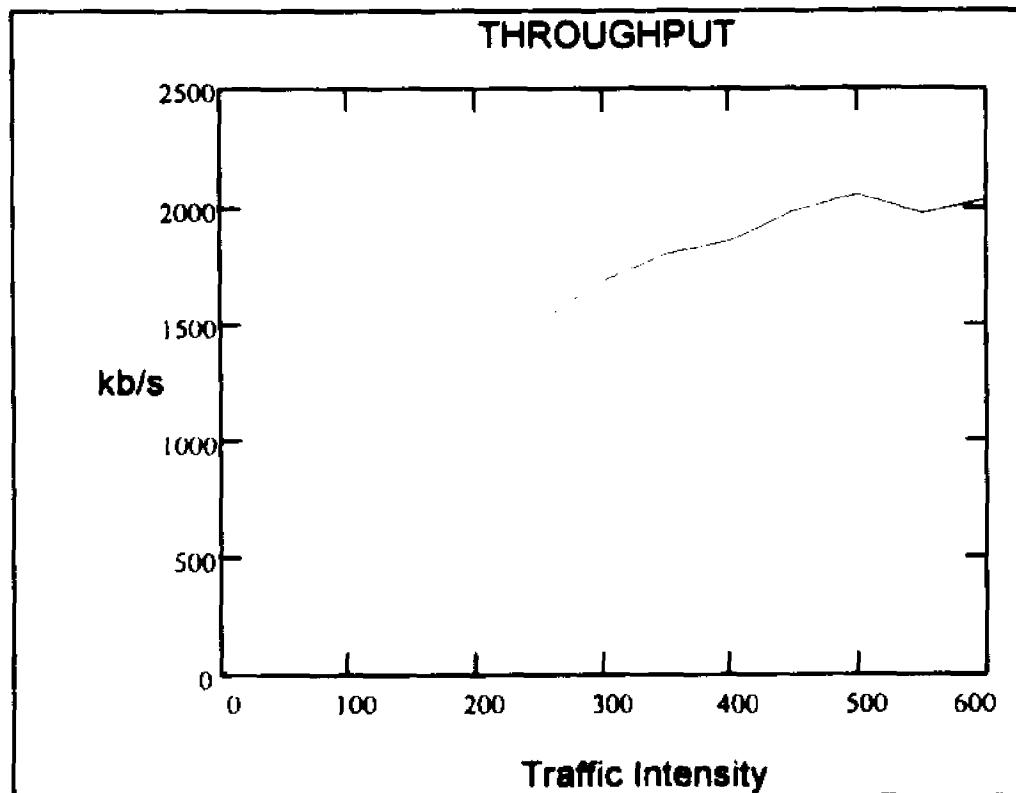


Figure 5.7. Full Duplex with 380 kb/s links

The average message delay and maximum message delay curves for 380 kb/s full duplex links are plotted in Figure 5.8. The average message delay curve starts increasing at traffic intensity of 250 and continues its increase at a linear rate for higher traffic intensities. The maximum message delay curve however, shows a slight increase between traffic intensities 1 and 250, then increases at a higher rate up to about 25% at traffic intensity of 400, and then gradually increasing at a gentler slope for higher traffic intensities.

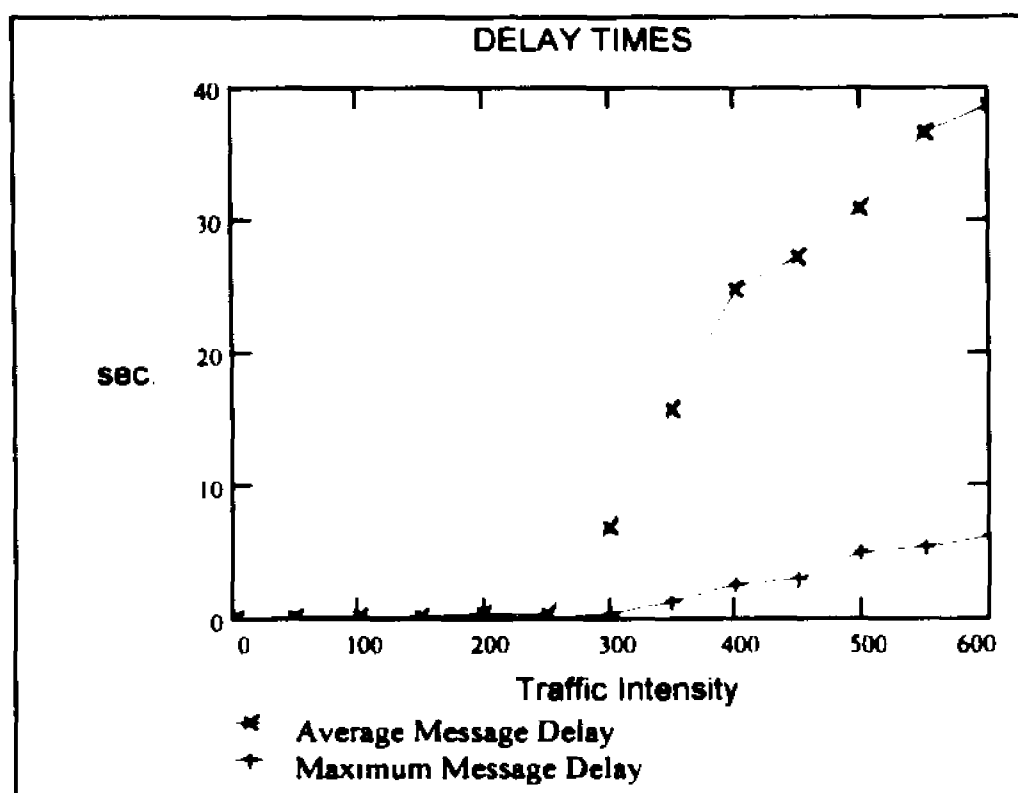


Figure 5.8 Full Duplex with 380 kb/s links

The percent of messages over three seconds curve for 380 kb/s full duplex links (Figure 5.9), shows a slight increase between traffic intensity of 250 and 300, and then increases at a steep slope between traffic intensities 300 to 400 and gradually leveling off at higher traffic intensities.

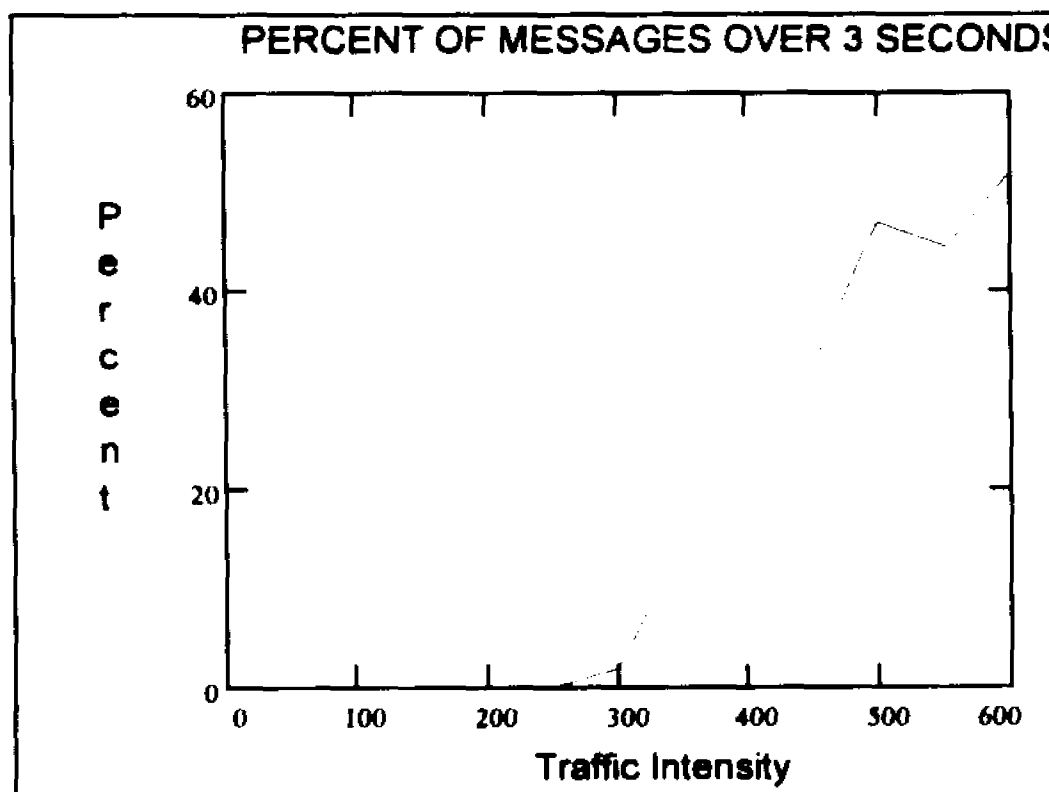


Figure 5.9 Full Duplex with 380 kb/s links

5.2.1.1.4 1544 kb/s Links

The throughput curve for 1544 kb/s full duplex links is plotted in Figure 5.10. It rises linearly from traffic intensity 1 to traffic intensity 1150, then dropping off slightly at traffic intensity 1200.

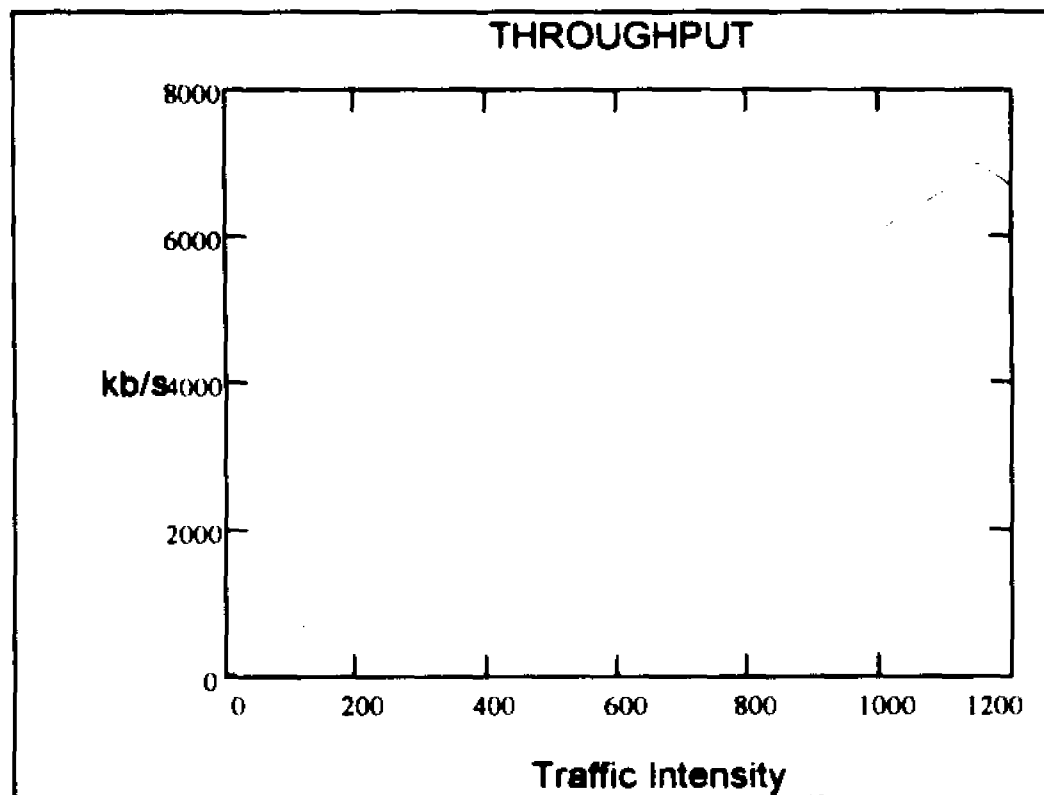


Figure 5.10 Full Duplex with 1544 kb/s links

The results for average message delay and maximum message delay for 1544 kb/s full duplex links is shown in Figure 5.11. The average message delay, rises ever so slightly from 0.09 sec at traffic intensity of 1 to 0.06 sec at traffic intensity of 1100 with a big increase to 0.27 sec at traffic intensity of 1200. The maximum message delay, also increases slightly up to 0.07 sec at traffic intensity of 650, then its slope increasing for the higher traffic intensities with a big jump to 7.56 sec at traffic intensity of 1200.

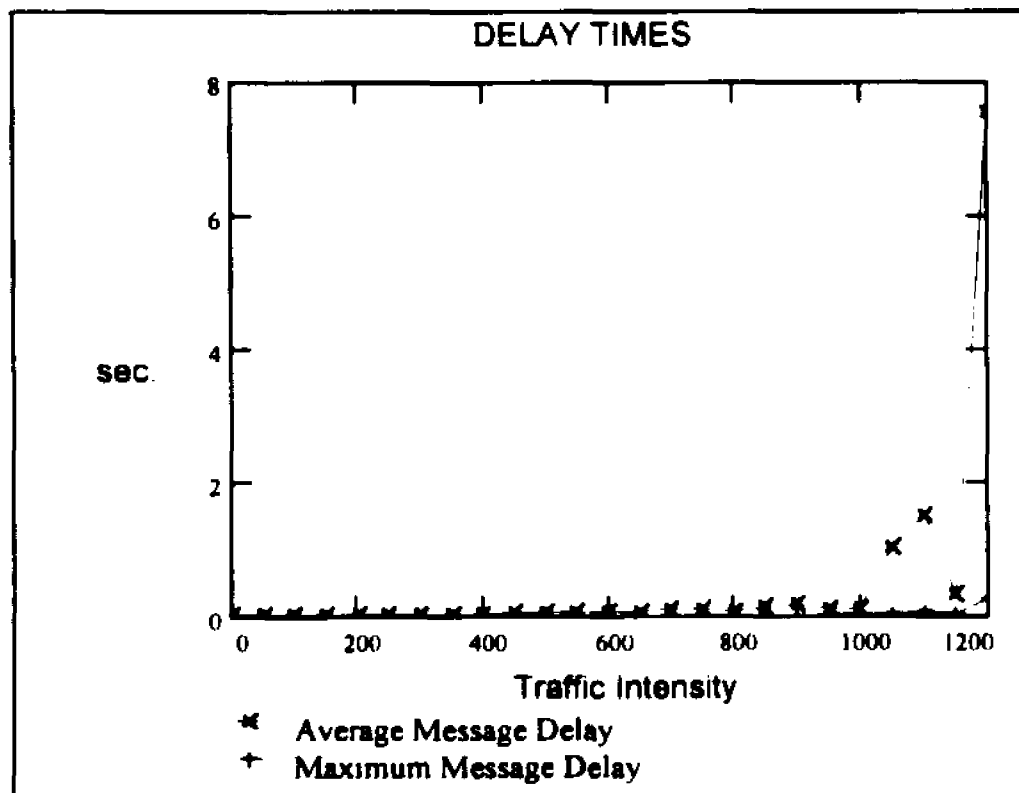


Figure 5.11. Full Duplex with 1544 kb/s links

The percent of messages over three seconds curve remains at 0 for traffic intensities 1 through 1150 with a sharp increase to 2.06 % at traffic intensity of 1200

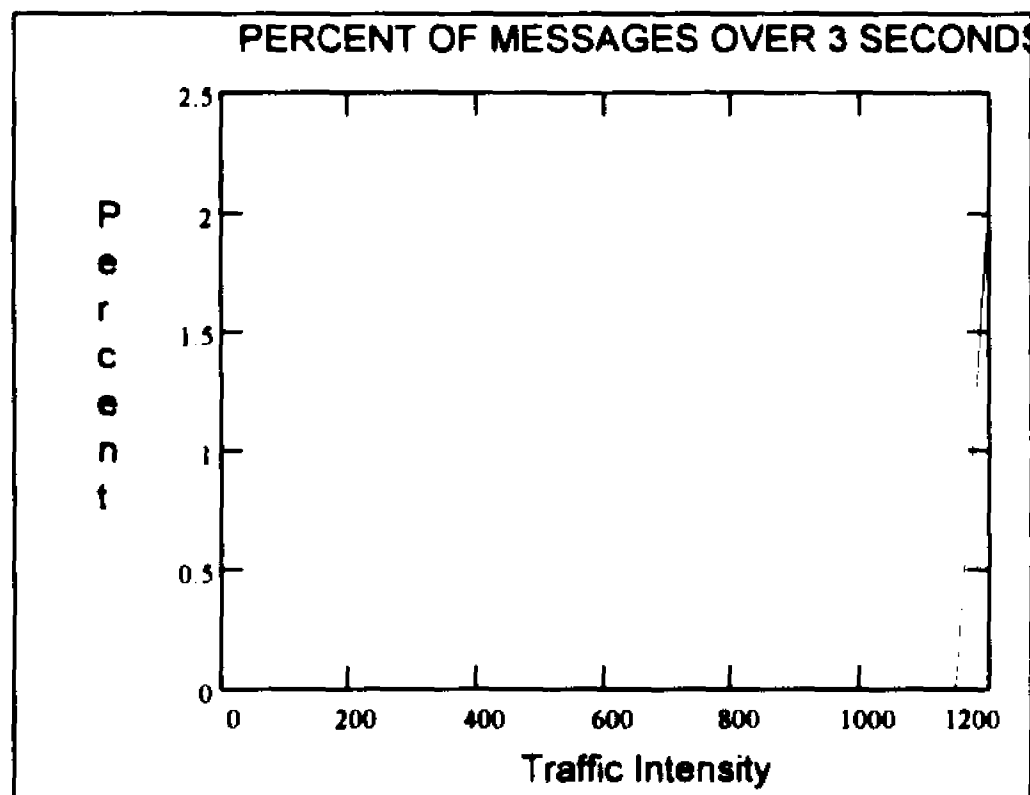


Figure 5.12 Full Duplex with 1544 kb/s links

5.2.1.2 Half Duplex Links

5.2.1.2.1 9.6 kb/s Links

The throughput curve for half duplex 9.6 kb/s links, displayed in Fig 5.13, show that the throughput of the network increases sharply from traffic intensity 1 to 10 at 28.3 kb/s then levels out staying around 30 kb/s for higher traffic intensities.

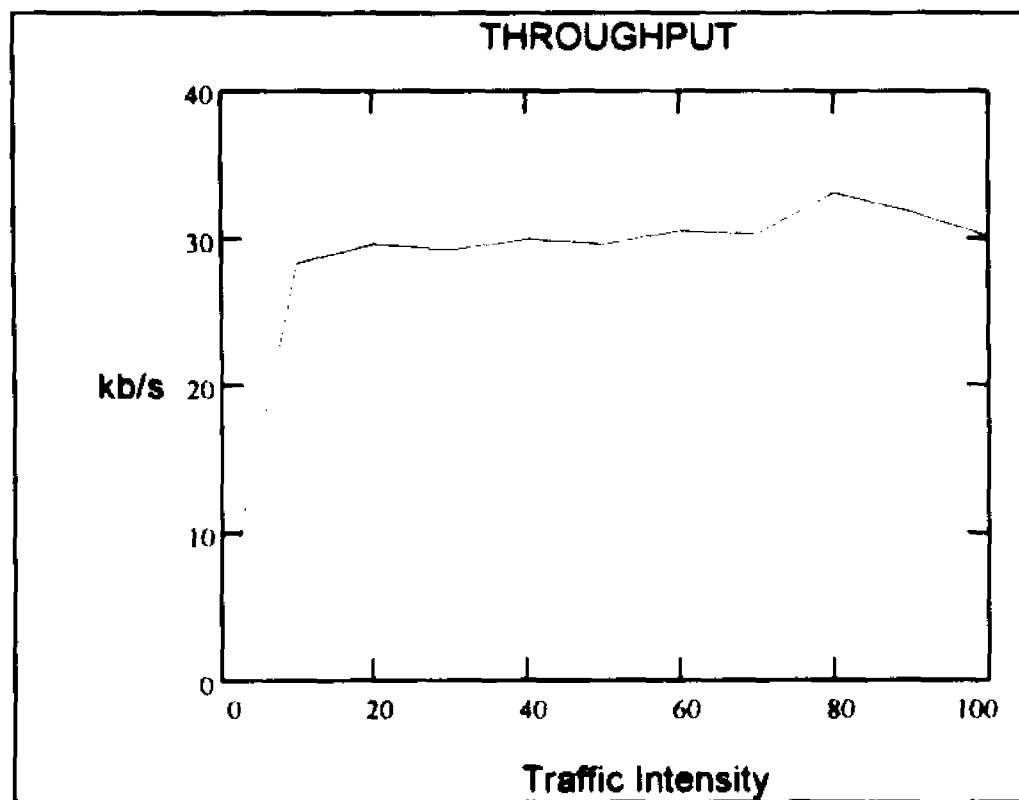


Figure 5.13 Half Duplex with 9.6 kb/s links

The average message delay for half duplex 9.6 kb/s (Fig 5.14) increases from 1.59 sec to 15.82 sec for traffic intensity 1 through 20. At traffic intensity 20, the curve starts increasing very slightly to a maximum of 22.33 sec. at traffic intensity 100

The maximum message delay also increases sharply, to 35.37 sec. at traffic intensity 10. It then increases less steeply between traffic intensity 10 and 30, to 52.78 sec. For traffic intensities above 30, it remains constant fluctuating around 55 sec, reaching a maximum of 56.61 at traffic intensity 90.

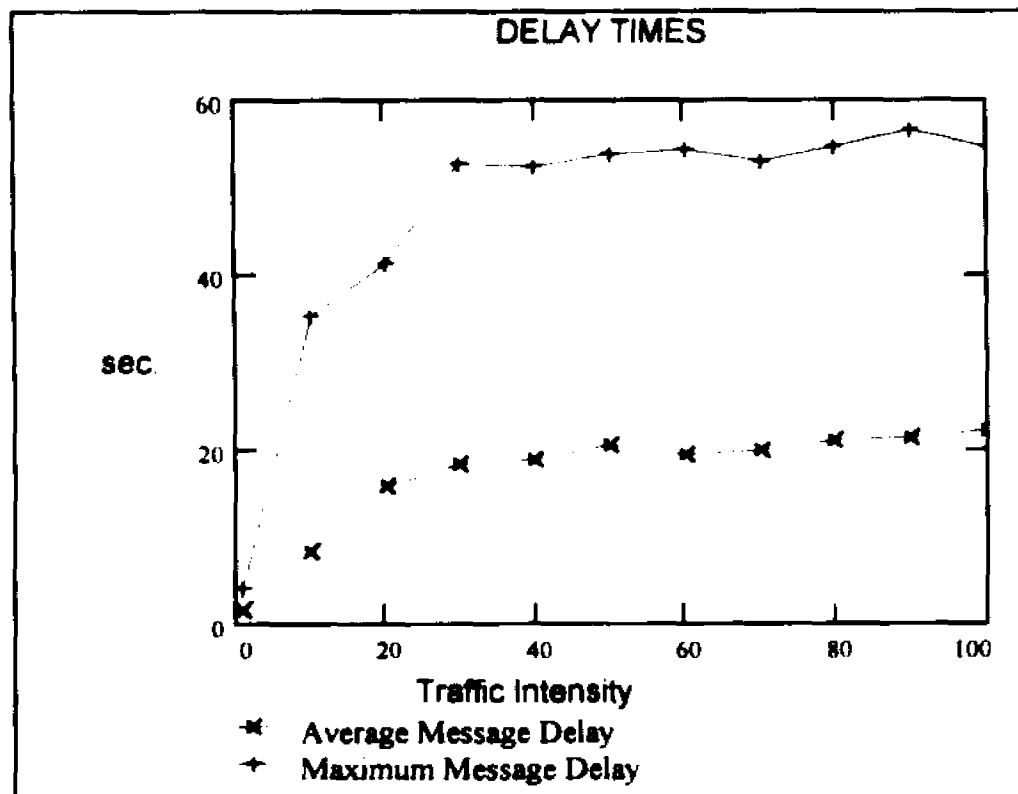


Figure 5.14. Half Duplex with 9.6 kb/s links

The percent of messages over three seconds curve depicted in Figure 5.15, shows a steep incline between traffic intensity 1 and 10 to 67.63 %. Between traffic intensity 10 and 20, it continues to increase but at a slower rate to 86.7 %. It remains at high 80 percentile for higher traffic intensities reaching its maximum of 90.13 % at traffic intensity 90

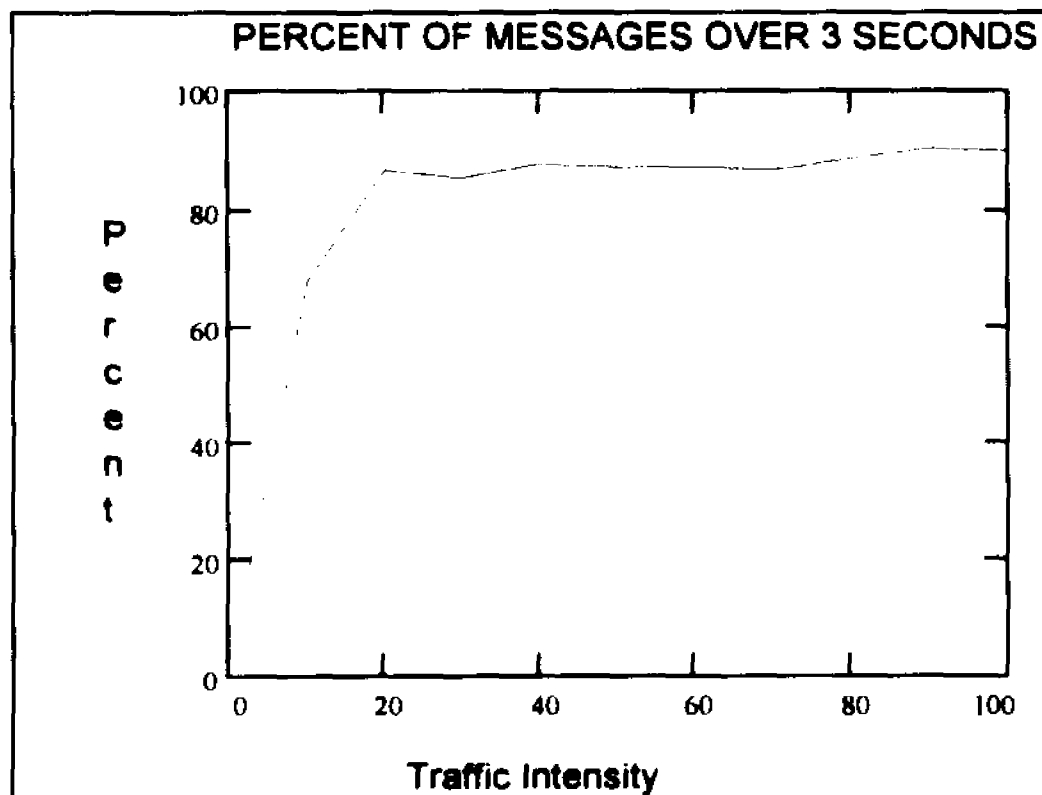


Figure 5.15. Half Duplex with 9.6 kb/s links

5.2.1.2.2 56 kb/s Links

The throughput curve for half duplex 56 kb/s links, displayed in Fig 5.16, show that the throughput of the network increases sharply between traffic intensity 1 and 30 to 152.6 kb/s. It continues to increase, but at a less steep gradient between traffic intensity 30 and 70 to 181.7 kb/s. For traffic intensities greater than 70, it remains relatively constant around 170 kb/s, decreasing slightly for higher traffic intensities over 150.

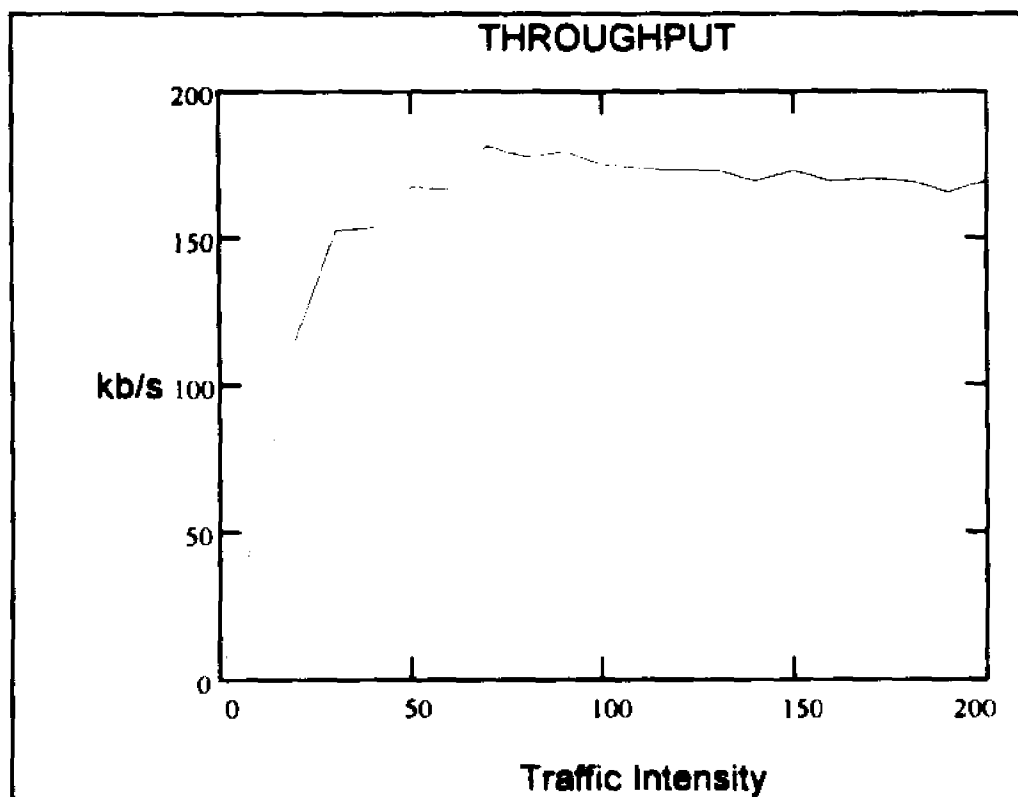


Figure 5.16 Half Duplex with 56 kb/s links

The average message delay for half duplex 56 kb/s links (Fig. 5.17) remains low (less than 1 sec) for traffic intensities 1 through 30. Between traffic intensity 30 and 50 there is marked increase to 8.09 sec. At traffic intensities above 50, this increase becomes more gradual reaching a maximum of 20.39 sec at traffic intensity 200.

The maximum message delay for half duplex 56 kb/s links remains low (less than 1 sec) between traffic intensities 1 and 10. At traffic intensity 10, it increases rapidly reaching 39.01 sec. at traffic intensity 50. This rapid increase becomes more gradual between traffic intensities 50 and 130 where it reaches 53.31 sec. At traffic intensities above 130 it flattens out around 53 sec. attaining a maximum of 54.57 sec at traffic intensity 190.

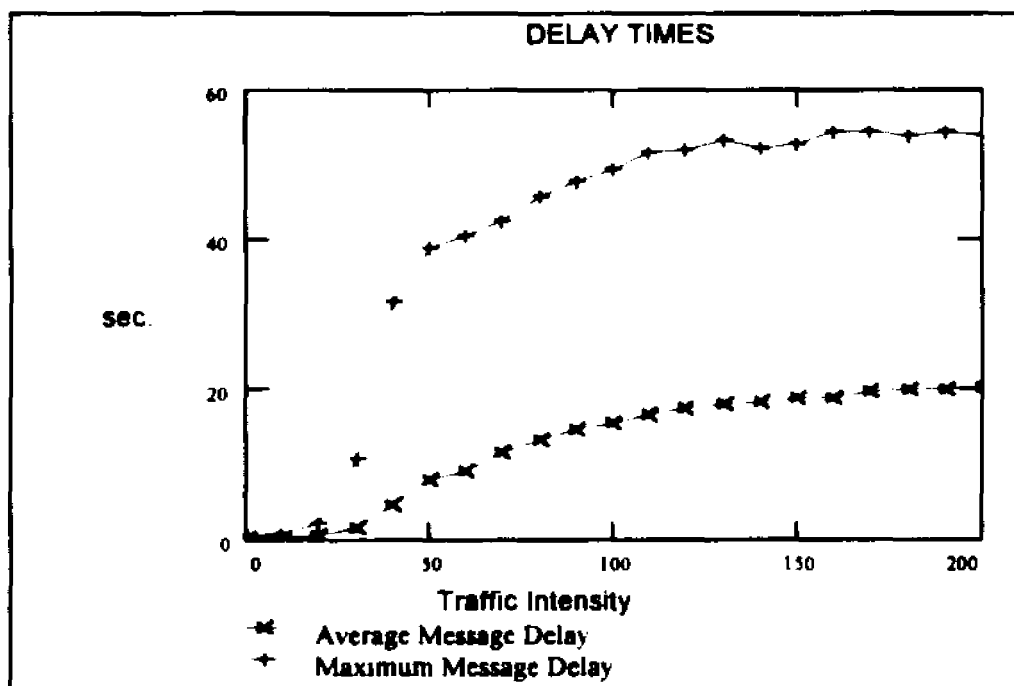


Figure 5.17 Half Duplex with 56 kb/s links

The percent of messages over three seconds curve for 56kb/s half duplex links(Fig. 5.18) remains at 0 % between traffic intensities 1 and 10. After that it starts a steep increase to 82.39 % at traffic intensity of 70. This increase becomes very gradual for higher traffic intensities and flattening out and fluctuating around 90 % between traffic intensities of 120 and 200 reading a maximum of 91.34 % at traffic intensity 170

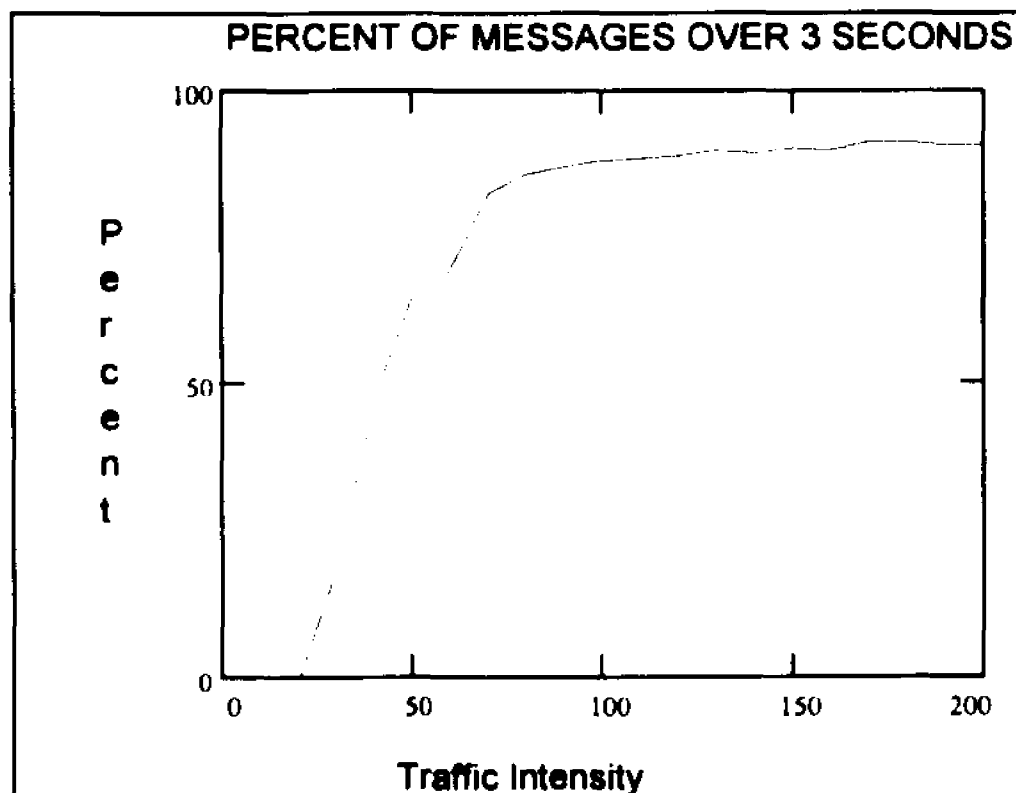


Figure 5.18: Half Duplex with 56 kb/s links

5.2.1.2.3 380 kb/s Links

The throughput for 380 kb/s half duplex links increases linearly between traffic intensity of 1 and 200 to 1108 kb/s. It then levels out for higher traffic intensities remaining around 1180 kb/s reaching a maximum of 1230 kb/s at traffic intensity of 700

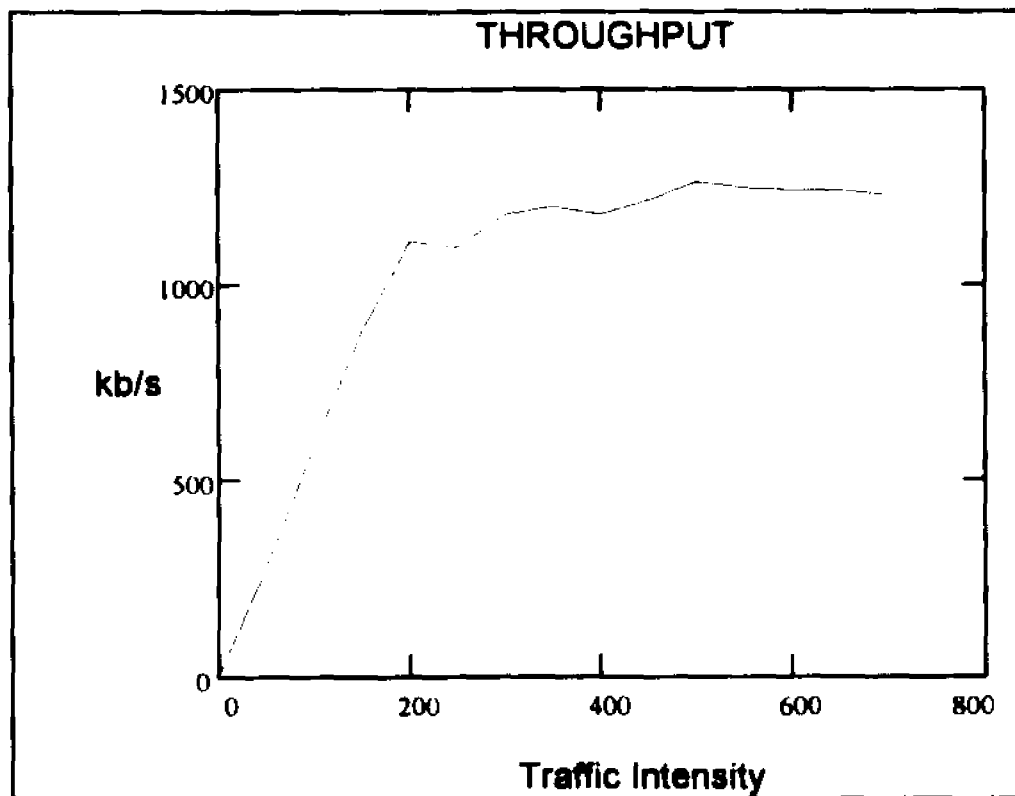


Figure 5 19: Half Duplex with 380 kb/s links

The average message delay for 380 kb/s half duplex links (Fig 5.20) remains at less than 1 sec for traffic intensities 1 to 200. Between traffic intensity 200 and 250 it increases sharply to 4.01 sec. from 0.76 sec. This increase more becomes gradual at traffic intensities above 250 reaching a maximum of 15.67 at traffic intensity 700.

The maximum message delay for 380 kb/s half duplex links is low (less than 1 sec) for traffic intensities 1 to 100. It increases slowly at first to 7.6 sec at traffic intensity of 200, then sharply increasing to 45.02 sec at traffic intensity of 400 before showing a small dip and gradually reaching a maximum of 51.29 sec at traffic intensity of 700.

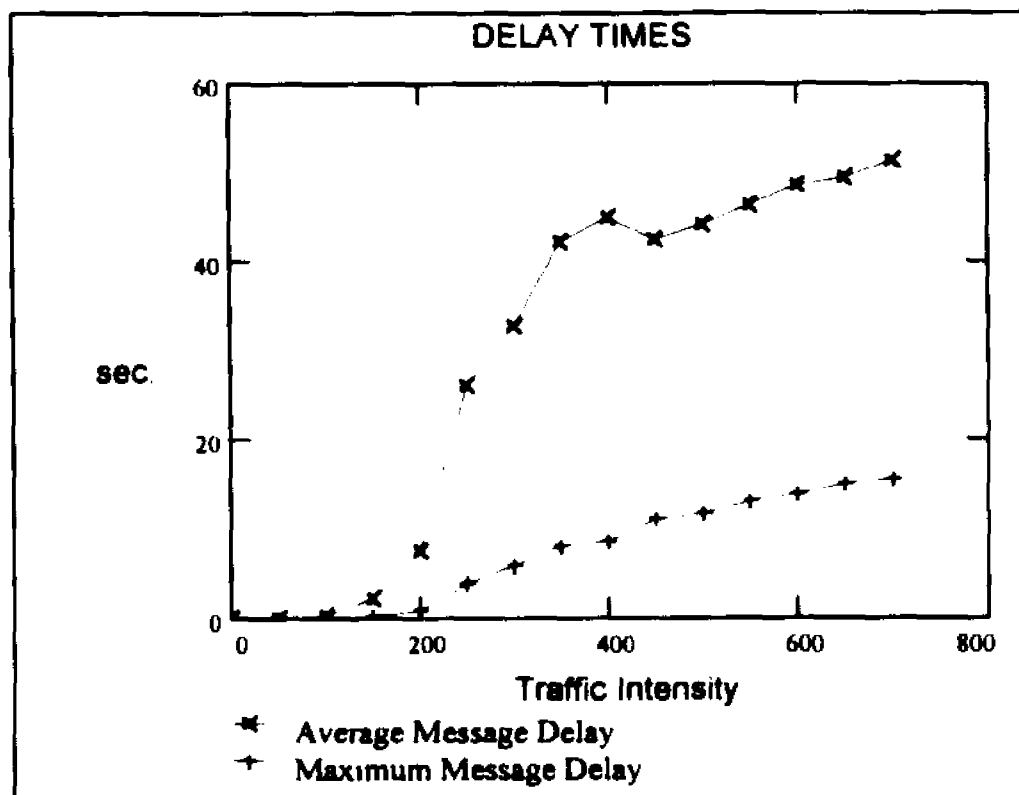


Figure 5.20 Half Duplex with 380 kb/s links

The percent of messages over three seconds for 380 kb/s half duplex links (Fig. 5.21) remains at 0% between traffic intensity 1 and 150. From traffic intensity 150 to 200, it rises gradually to 6.47%. Between traffic intensities of 200 and 350, it increases rapidly reaching 67.32%. After traffic intensity of 350, the increase becomes gradual again hitting 88.24% at traffic intensity of 700

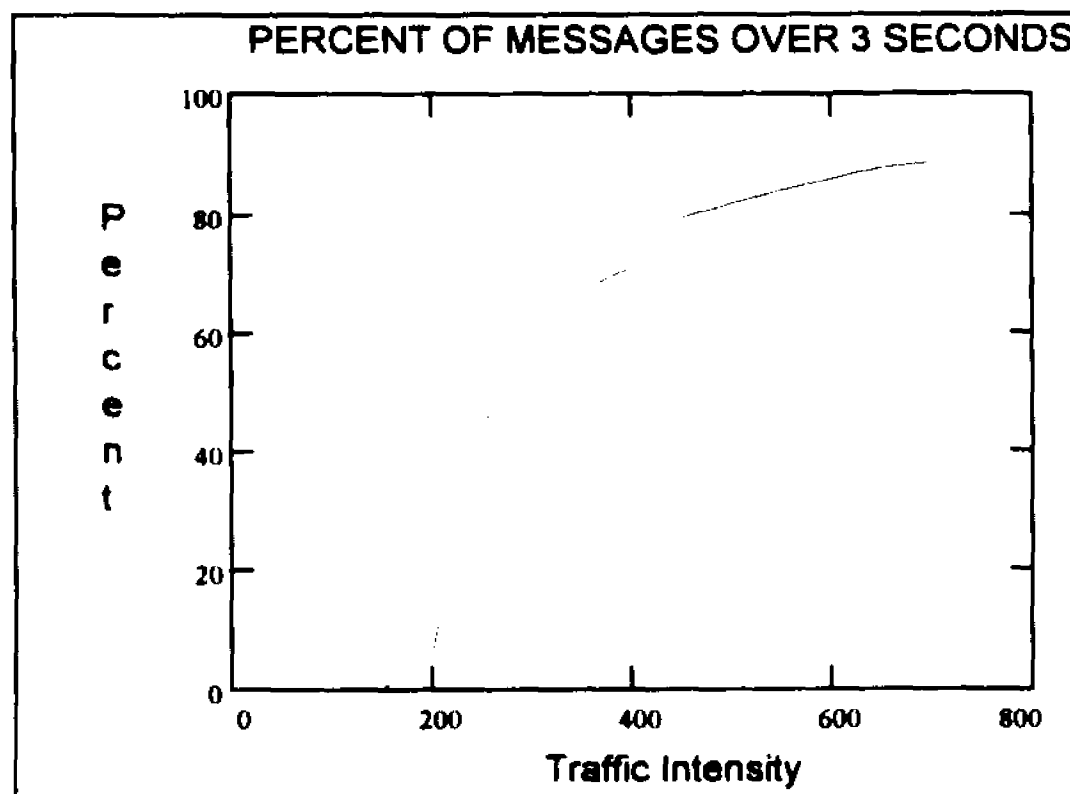


Figure 5.21 Half Duplex with 380 kb/s links

5.2.1.2.4 1544 kb/s Links

The throughput for 1544 kb/s half duplex links, which is represented in Fig 5.22, increases linearly between traffic intensities of 1 and 600 to 3605 kb/s. This rate of change, decreases between traffic intensities 600 and 750. After traffic intensity of 750, it further decreases, reaching 4194 kb/s from 4265 kb/s at traffic intensity of 1000

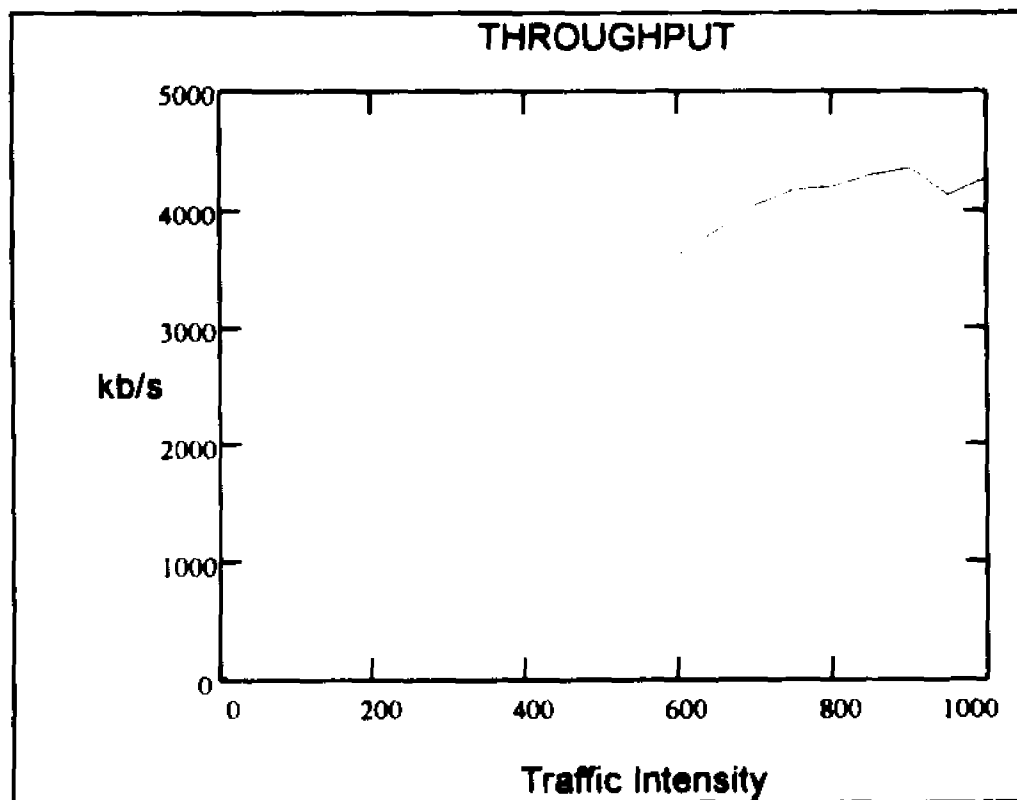


Figure 5.22: Half Duplex with 1544 kb/s links

The average message delay for 1544 kb/s half duplex links remains very low (less than 0.1 sec) for traffic intensities between 1 through 500. Between traffic intensity of 500 and 600, it increases slightly to 0.03 sec. After traffic intensity of 600 it increases sharply, reaching 4.21 sec. at traffic intensity of 1000.

The maximum message delay for 1544 kb/s half duplex links also remains low (less than or equal to 0.1 sec) for traffic intensities between 1 and 450. Between traffic intensity of 450 and 600 there is a gradual increase to 0.53 sec. After that there is a steep increase to 27.53 sec at traffic intensity of 1000.

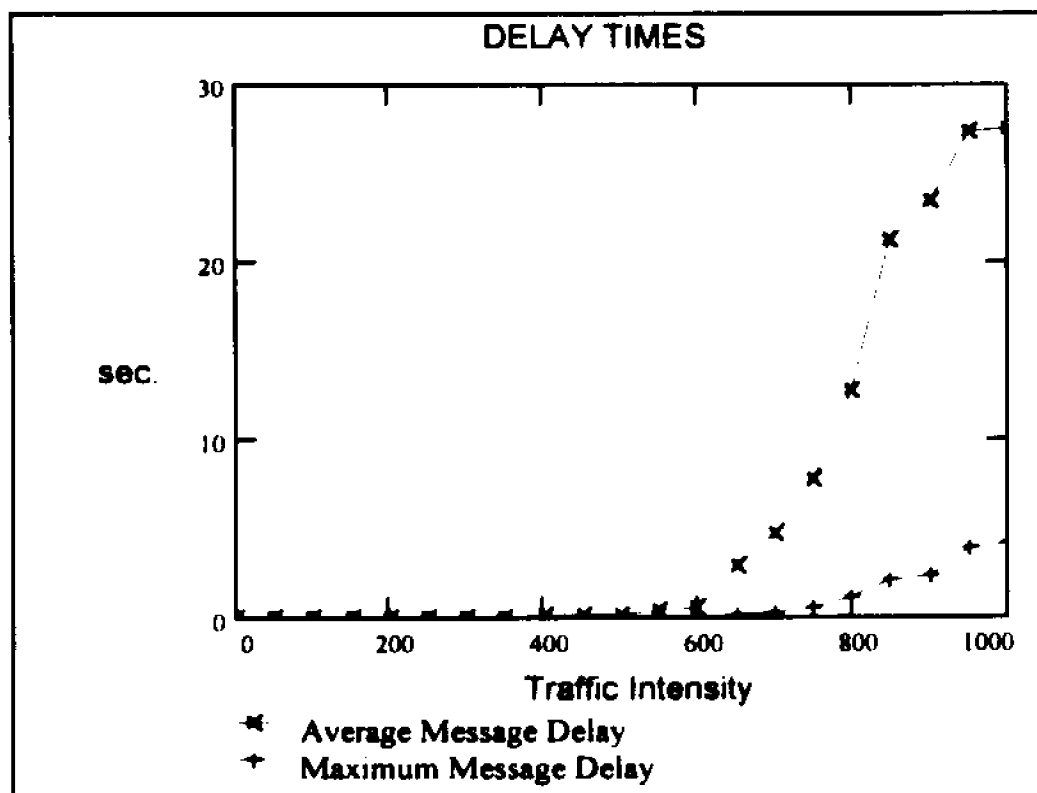


Figure 5.23: Half Duplex with 1544 kb/s links

The percent of messages over three seconds for 1544 kb/s half duplex links shown in Fig 5.24 remains at 0 % between traffic intensities of 1 and 650. After that it first increases gradually to 1.13 % at traffic intensity of 700, then increases sharply to 43.83 % at traffic intensity of 1000

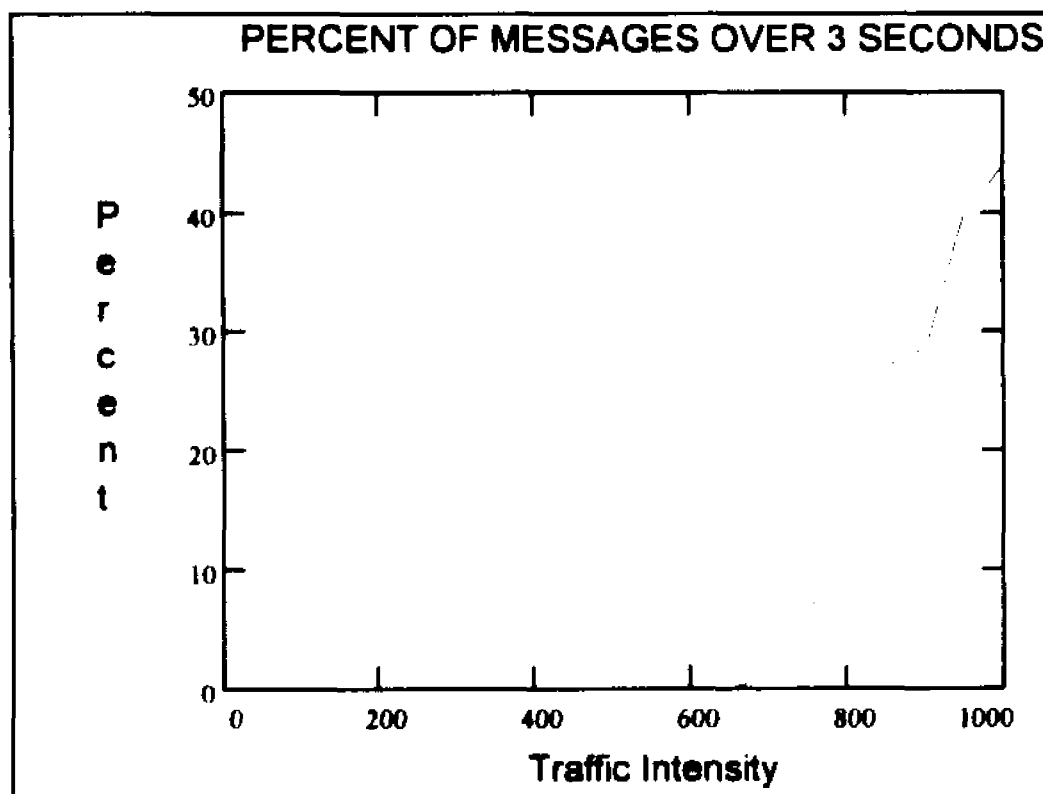


Figure 5.24 Half Duplex with 1544 kb/s links

5.2.1.3 Link Failure

5.2.1.3.1 9.6 kb/s Links

The throughput for 9.6 kb/s links with link failure, shown in Fig 5.25 increases sharply between traffic intensities 1 and 10 to 33.9 kb/s. It then gradually increases to 40.4 kb/s at 40. For traffic intensities higher than 40, it remains constant around 42 kb/s with attaining a maximum of 42.7 kb/s at traffic intensity of 100.

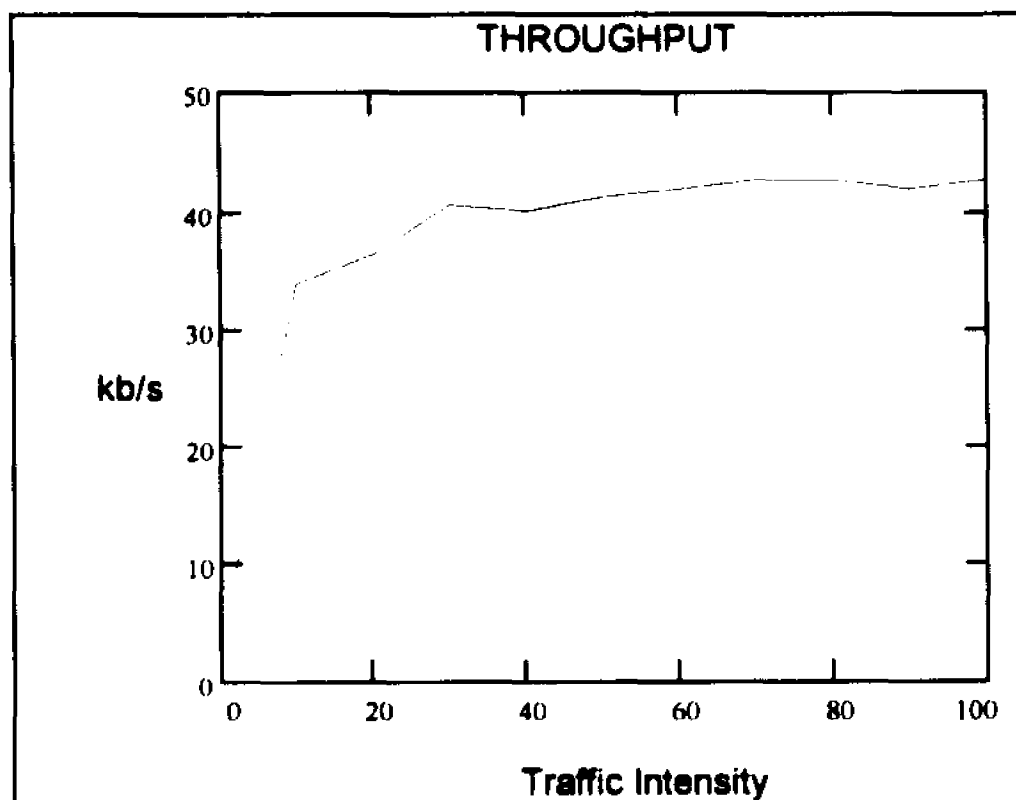


Figure 5.25 Link Failure with 9.6 kb/s links

The average message delay for 9.6 kb/s links with link failure (Fig 5.26) increases from 1.94 sec. at traffic intensity of 1 to 13.03 at traffic intensity of 20. After that the increase becomes moderate until traffic intensity of 70 with 17.98 sec. Then it remains flat around 17.5 sec for traffic intensities up to 100.

The maximum message delay for 9.6 kb/s links with link failure increases between traffic intensities 1 and 20 from 4.55 sec. to 47.34 sec. After that it flattens out, remaining around 17 sec., with maximum of 18.56 sec at traffic intensity of 100.

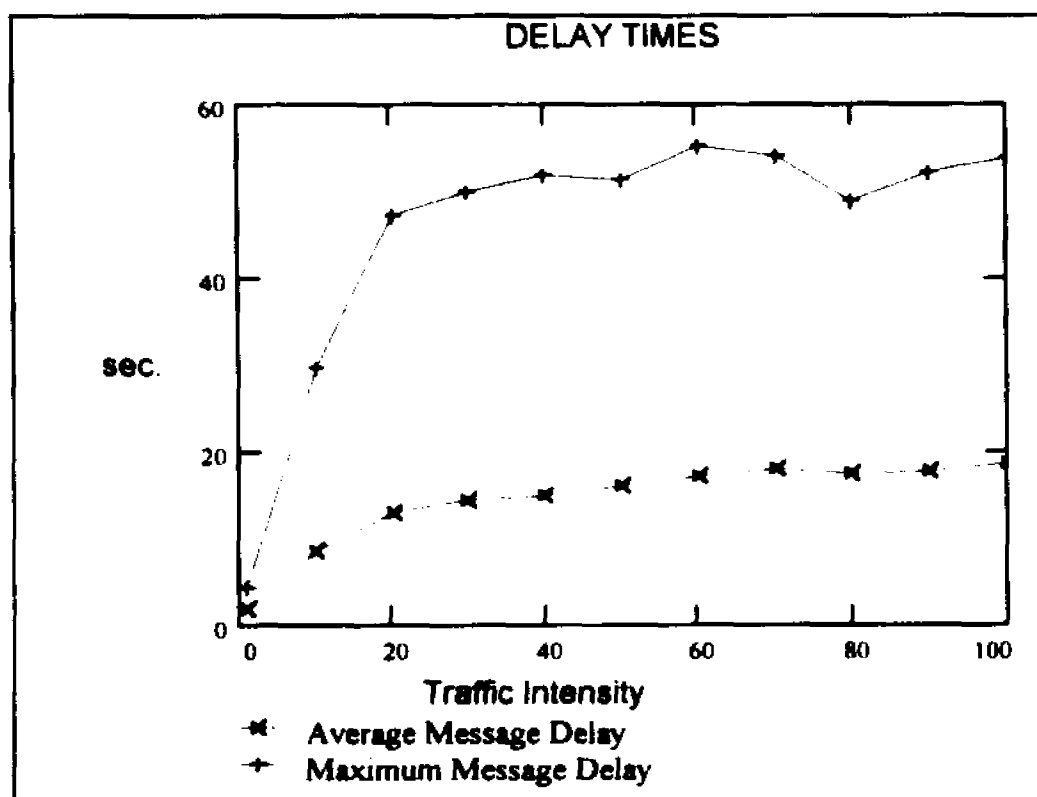


Figure 5.26. Link Failure with 9.6 kb/s links

The percent of messages over three seconds for 9.6 kb/s links with link failure increases sharply between traffic intensity of 1 and 10 to 71.26%. After that the increase becomes very gradual for the higher traffic intensities, reaching a maximum of 88.24% at traffic intensity of 90.

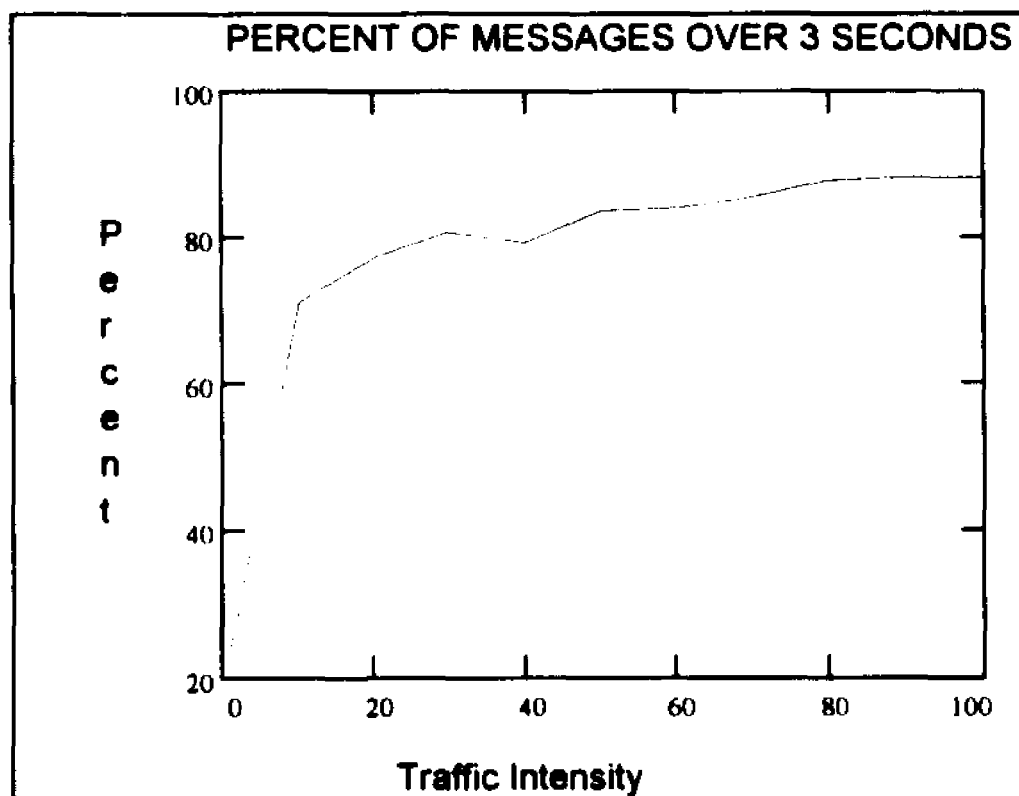


Figure 5.27 Link Failure with 9.6 kb/s links

5.2.1.3.2 56 kb/s Links

The throughput for 56 kb/s links with link failure shows a steep increase up to 115 kb/s at traffic intensity of 20. At that point the increase becomes progressively gradual reaching 186 kb/s at traffic intensity of 40. After this point, the throughput increases residually to a maximum 249 kb/s at traffic intensity of 200

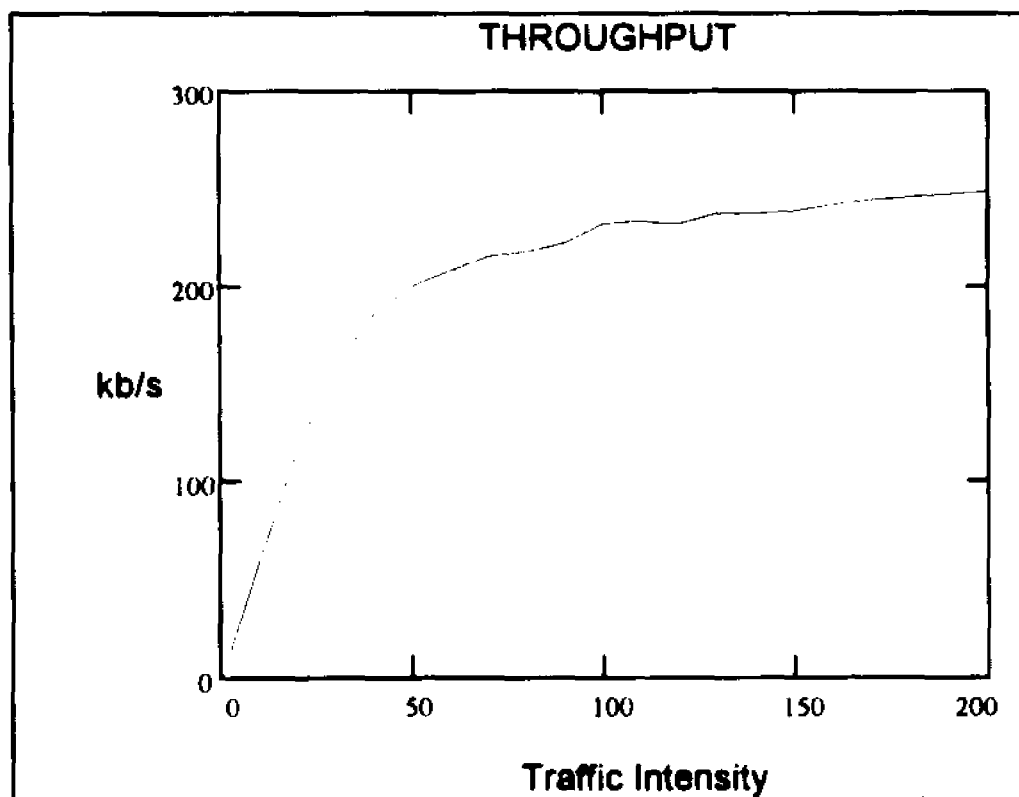


Figure 5.28. Link Failure with 56 kb/s links

The average message delay for 56 kb/s links with link failure increases slightly between traffic intensity of 1 and 30, from 0.32 sec to 1.03 sec. Between traffic intensity of 30 and 70, the increase becomes more apparent, reaching 9.07 sec. At traffic intensities higher than 70, the average message delay, increases gradually again, reading a maximum of 14.59 sec at traffic intensity of 190.

The maximum message delay for 56 kb/s links with link failure increases slightly from 0.75 sec. to 2.07 between traffic intensities 1 and 20. After that it increases sharply up to 34.14 sec. at traffic intensity of 60. For traffic intensities greater than 60 the increase becomes more gradual, leveling off at about 56 sec., with the maximum occurring at 56.84 sec at traffic intensity of 200.

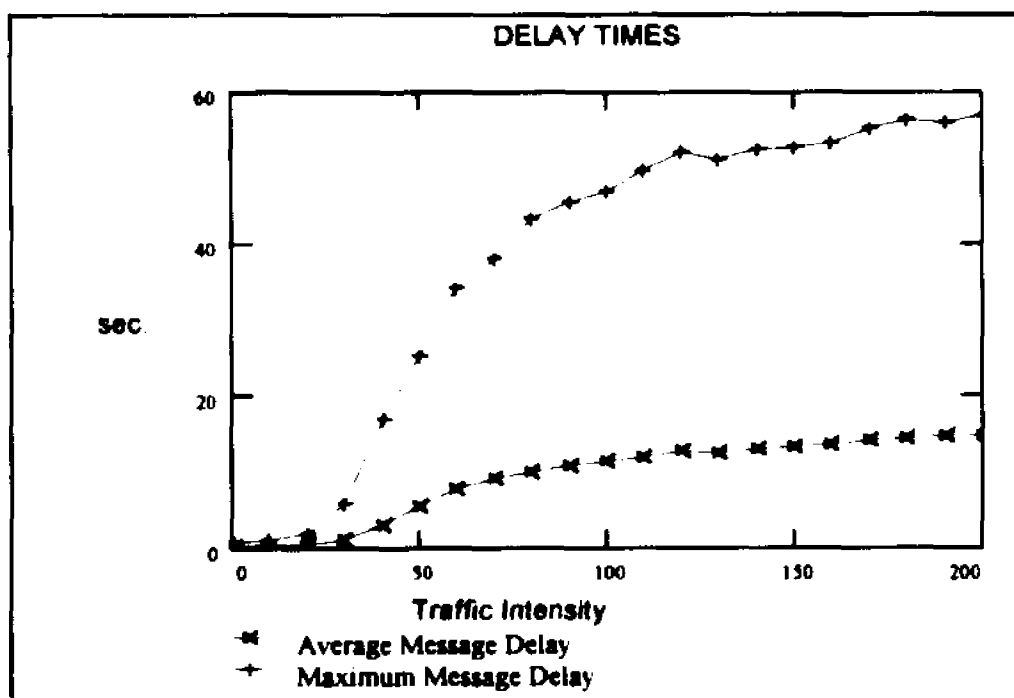


Figure 5.29 Link Failure with 56 kb/s links

The percent of messages over three seconds for 56 kb/s links with link failure remains at 0 % between traffic intensities of 1 and 20. It increases sharply from 3.58 % at traffic intensity of 30 to 69% at traffic intensity of 70. It increases slightly for the remaining higher traffic intensities, reaching a maximum of 79.55% at traffic intensity of 190

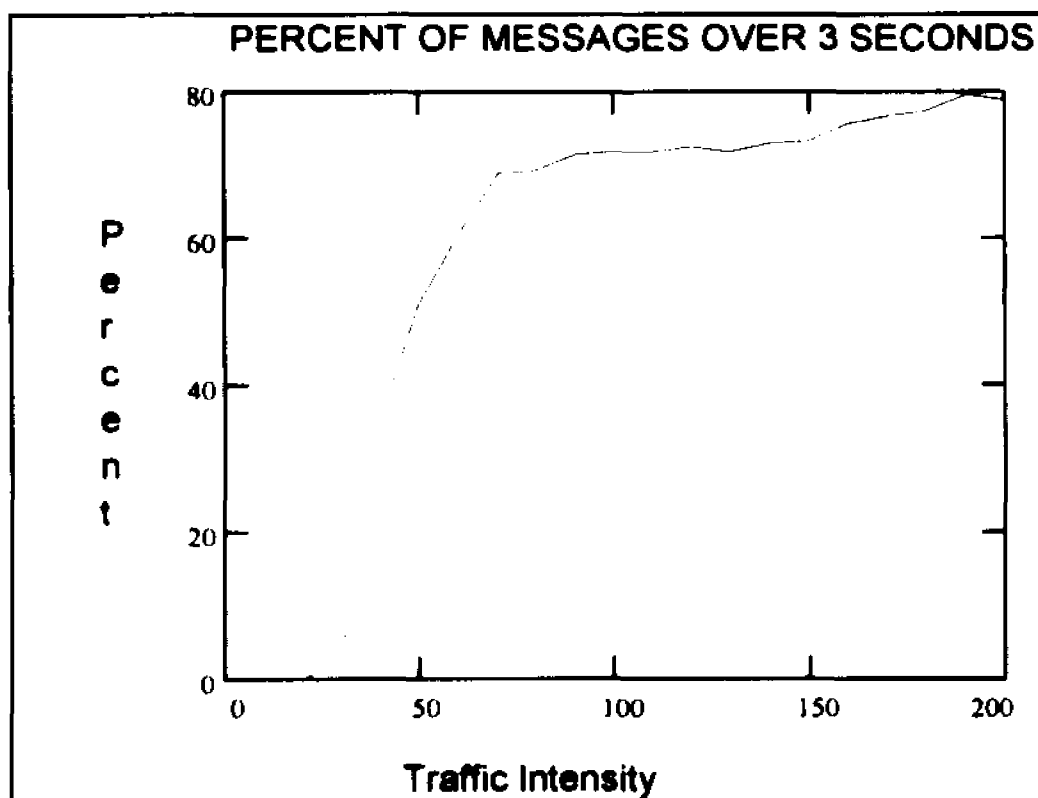


Figure 5.30. Link Failure with 56 kb/s links

5.2.1.3.3 380 kb/s Links

The throughput for 380 kb/s links with link failure increases linearly between traffic intensity of 1 and 150 from 6.9 kb/s to 895.2 kb/s. After traffic intensity of 150, it starts leveling off gradually reaching 1619.1 kb/s at traffic intensity of 700

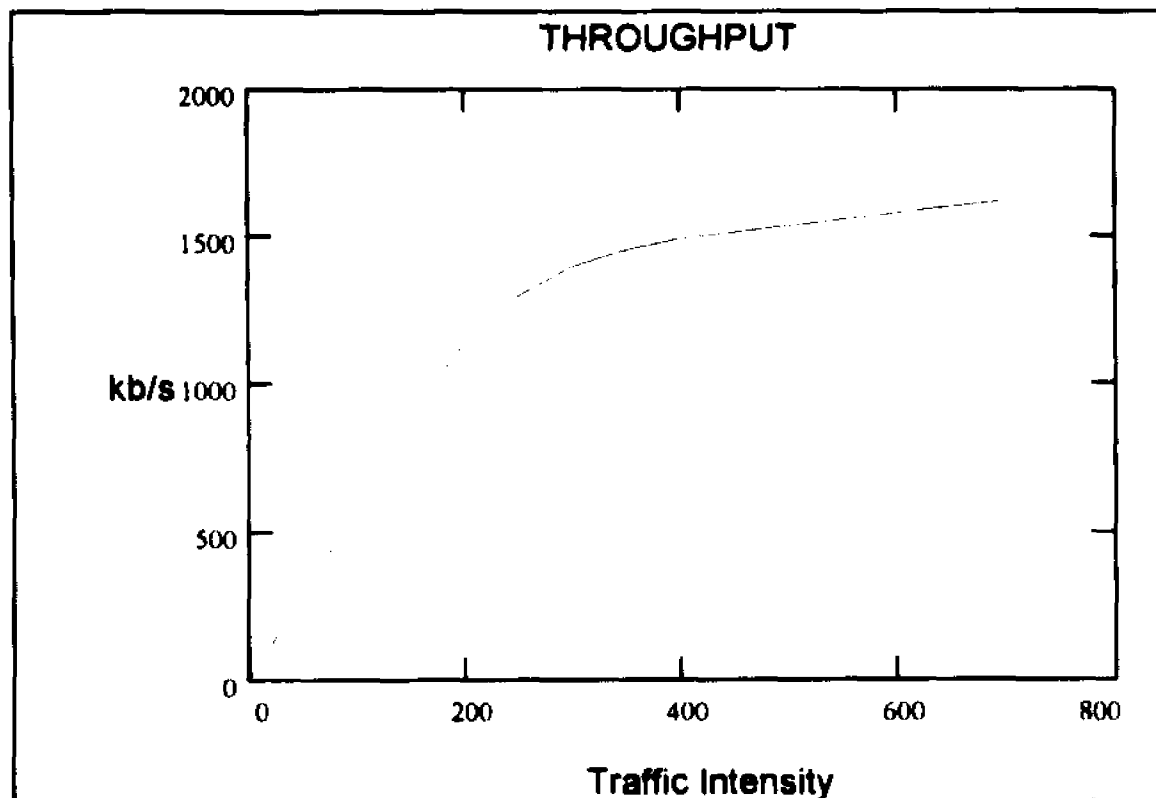


Figure 5.31. Link Failure with 380 kb/s links

The average message delay for 380 kb/s links with link failure remains below 1 sec. for traffic intensities between 1 and 150. After that it increases gently to 3.72 sec at traffic intensity of 200 starts to increase reaching 11.08 sec at traffic intensity of 700.

The maximum message delay for 380 kb/s links with a link failure remains below 1 sec to traffic intensity of 200. At that point it increases slightly to 3.72 sec at traffic intensity of 250, then sharply to 26.02 sec. at 350, ending up at 48.94 sec. at traffic intensity of 700.

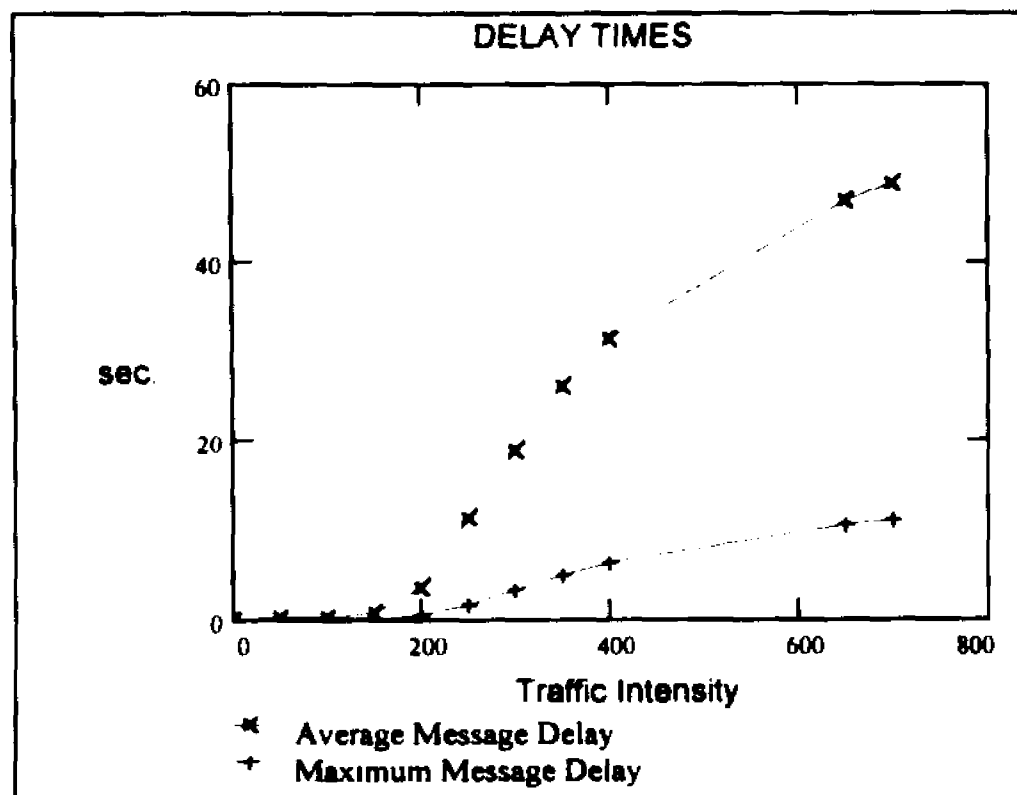


Figure 5.32 Link Failure with 380 kb/s links

The percent of messages over three seconds for 380 kb/s links with link failure remains below 1 % for the first 200 traffic intensities. At 250, there is a small increase to 1.83 % followed by a larger increase for higher traffic intensities ending at 70.53 % at traffic intensity of 700.

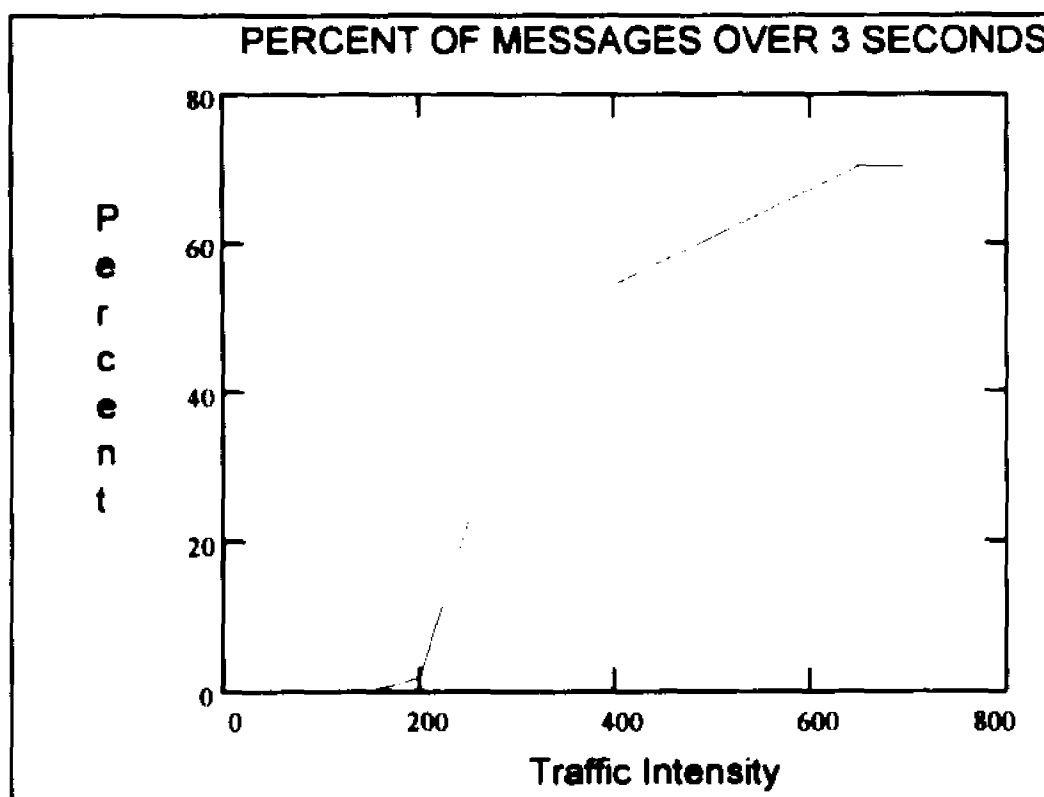


Figure 5.33 Link Failure with 380 kb/s links

5.2.1.3.4 1544 kb/s Links

The throughput for 1544 kb/s links with link failure shows a steady increase from 6.9 kb/s at traffic intensity of 1 to 4413.8 kb/s at traffic intensity of 750. It tapers off slightly and ends at 5425.7 kb/s at traffic intensity of 1050.

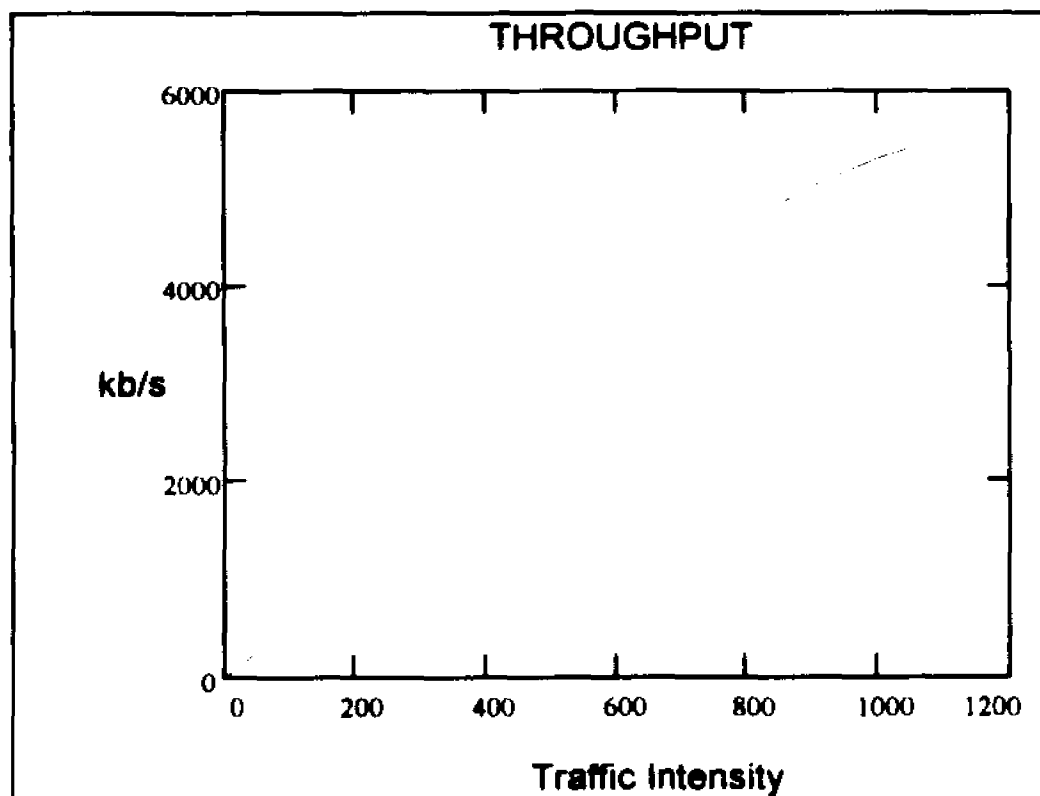


Figure 5.34 Link Failure with 1544 kb/s links

The average message delay time for 1544 kb/s links with link failure stays under .05 sec for traffic intensities 1 to 650. At 700, there is a small increase to 0.06 sec., which is followed by a larger increase which reaches 1.89 sec. at traffic intensity of 1050

The maximum message delay time for 1544 kb/s links with link failure remains below 1 sec. for traffic intensities 1 to 650. At 700, it climbs to 1.05 sec and continues to increase to 12.09 sec at traffic intensity of 1050

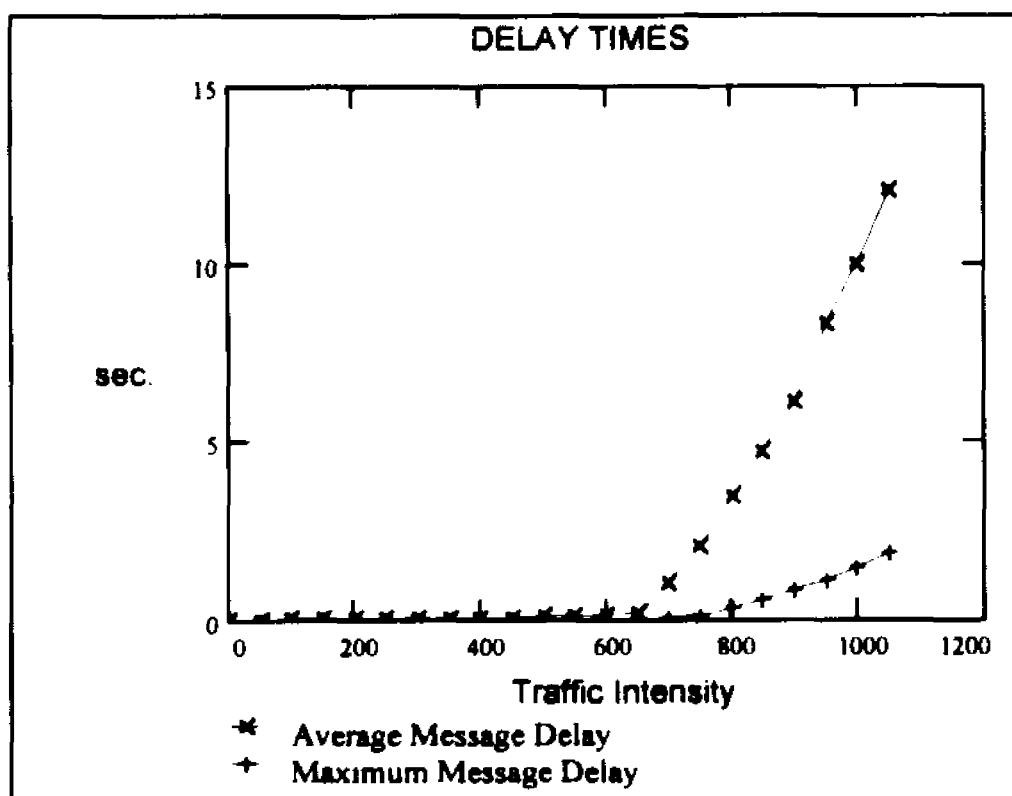


Figure 5.35 Link Failure with 1544 kb/s links

The percent of messages over three seconds for 1544 kb/s links with link failure remains at 0 % up to traffic intensity of 750. After that it increases sharply to 25.41 % at traffic intensity of 1050.

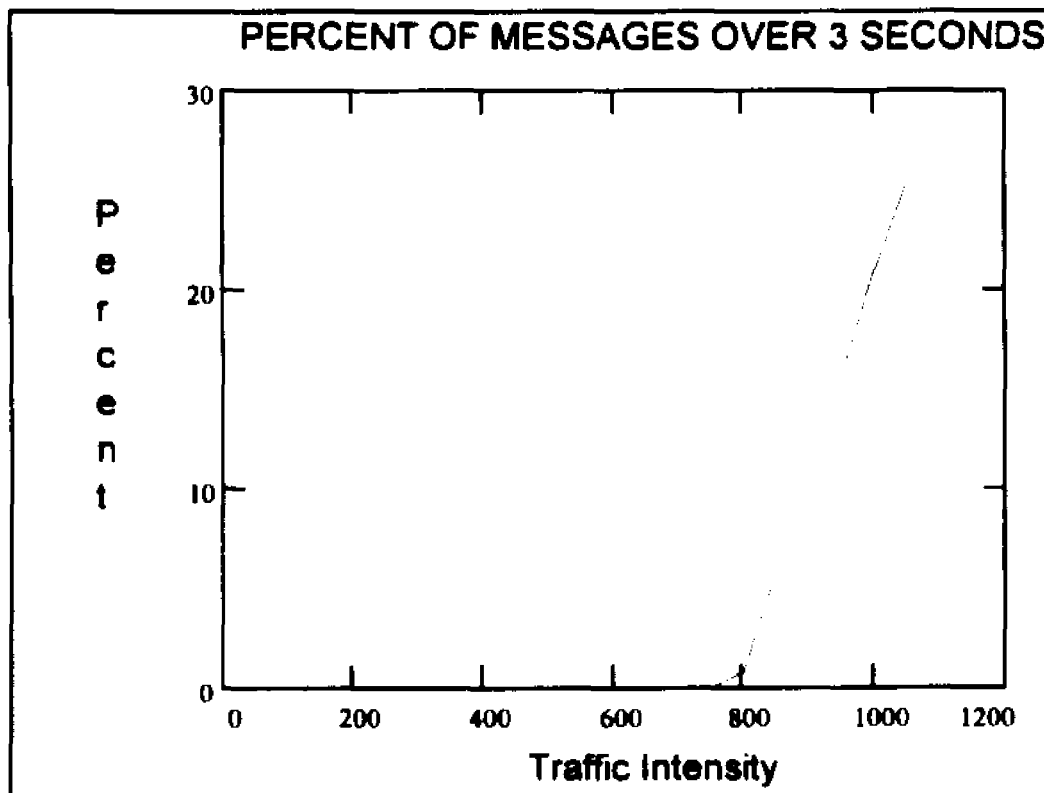


Figure 5.36. Link Failure with 1544 kb/s links

5.2.2 Second Simulation Model Results

5.2.2.1 Full Duplex Links

5.2.2.1.1 9.6 kb/s Links

The throughput for 9.6 kb/s full duplex links shows a rapid increase from traffic intensities 1 to 10 reaching 111.4 kb/s. The increase becomes gentler reaching 151.7 kb/s at traffic intensity of 30. It then levels off at around 160 kb/s reaching a maximum of 161.1 kb/s at traffic intensity of 60.

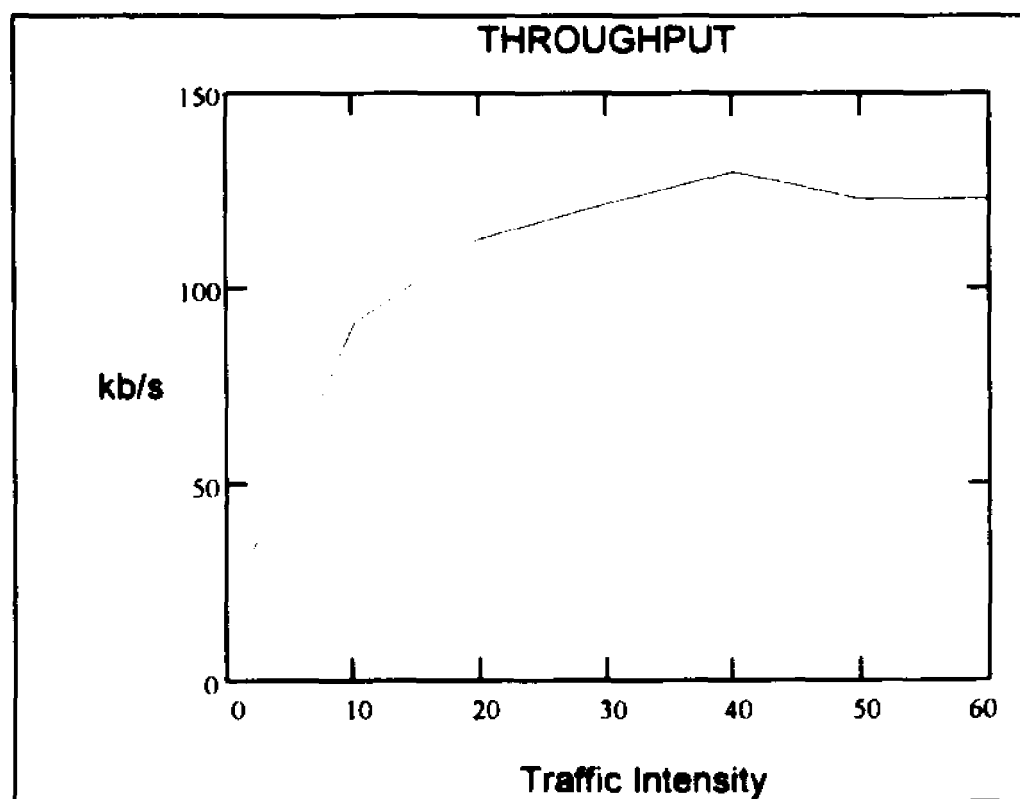


Figure 5.37. Full Duplex with 9.6 kb/s links

The average message delay for 9.6 kb/s full duplex links increases from 2.59 sec. at traffic intensity of 1 to 9.6 sec at traffic intensity of 10. It then increases to 16.36 sec at traffic intensity of 50 and remains at that range reaching a maximum of 16.86 sec at traffic intensity of 60.

The maximum message delay for 9.6 kb/s full duplex links increases sharply from 8.67 sec at traffic intensity of 1 to 44.14 sec at traffic intensity of 10. It then gradually increases to 56.28 sec at traffic intensity of 30 and remains at around 55 sec for higher traffic intensities, reaching 56.11 sec at traffic intensity of 50.

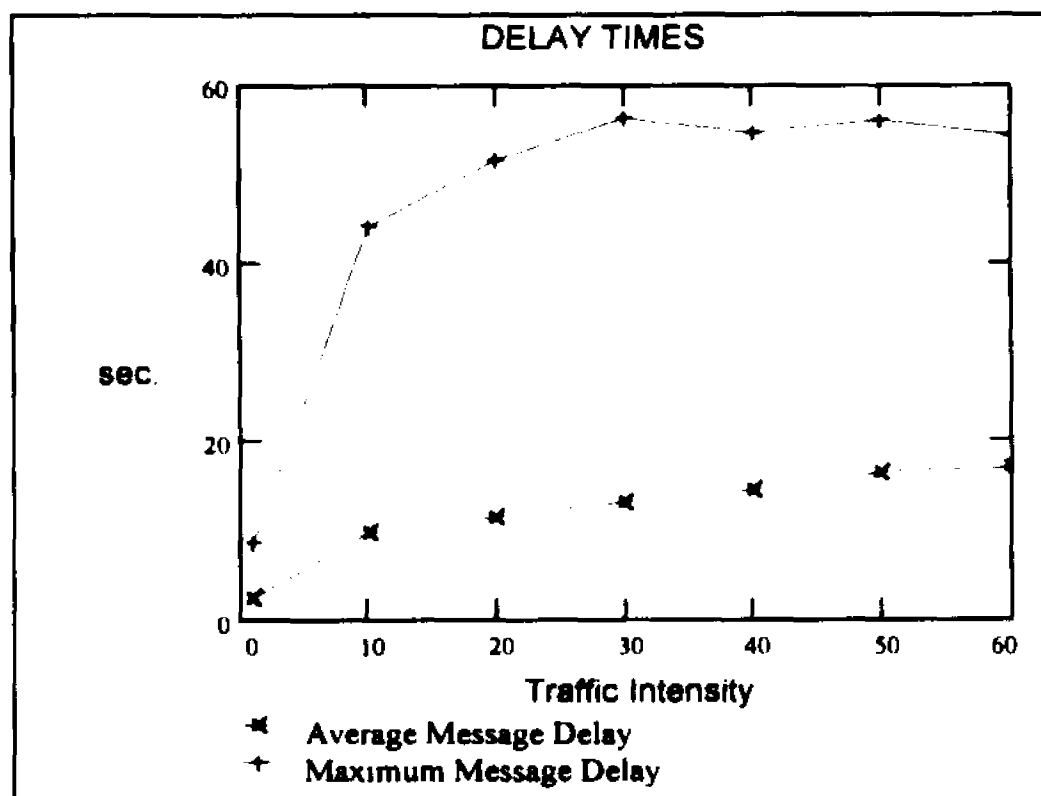


Figure 5.38 Full Duplex with 9.6 kb/s links

The percent of messages over three seconds for 9.6 kb/s full duplex links increases from 35.03 % to 63.36% between traffic intensities 1 and 10. After the increase becomes less steep reaching 82.13% at traffic intensity of 40, it ends at 87.02% at traffic intensity of 60.

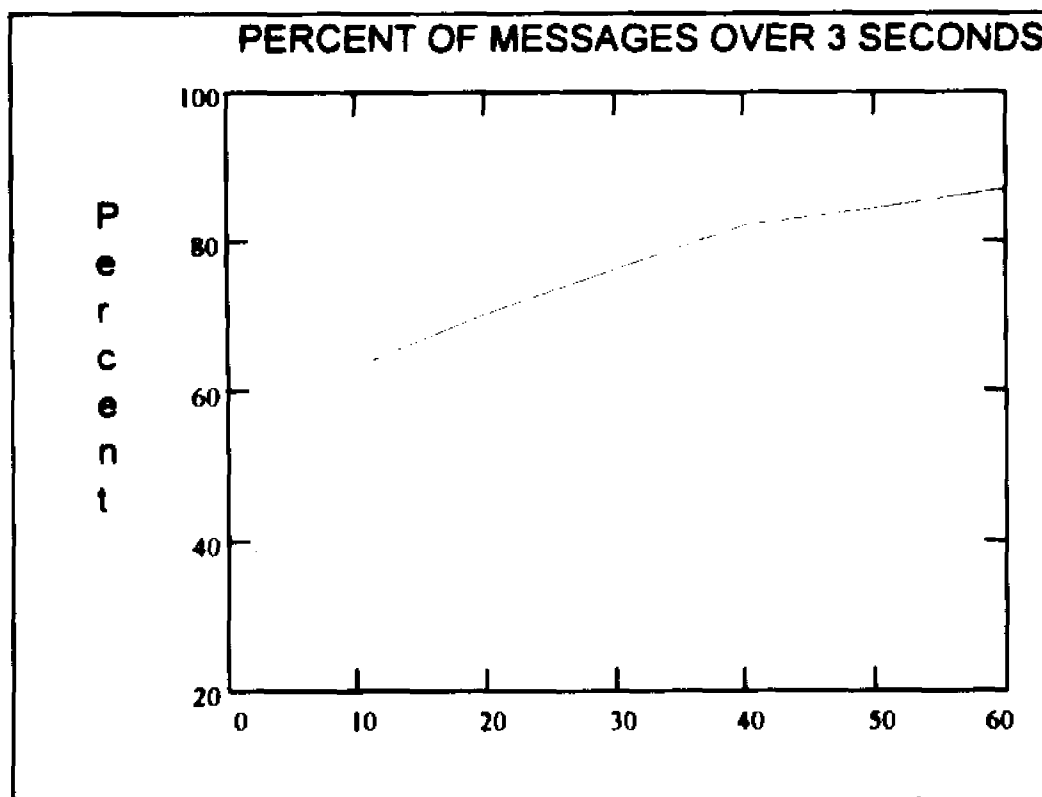


Figure 5.39 Full Duplex with 9.6 kb/s links

5.2.2.1.2 56 kb/s Links

The throughput for 56 kb/s full duplex links increases from 29 kb/s at traffic intensity of 1 to 254.1 kb/s at traffic intensity of 10. After this point, it starts leveling off gradually, eventually reaching a maximum of 690 kb/s at traffic intensity of 80.

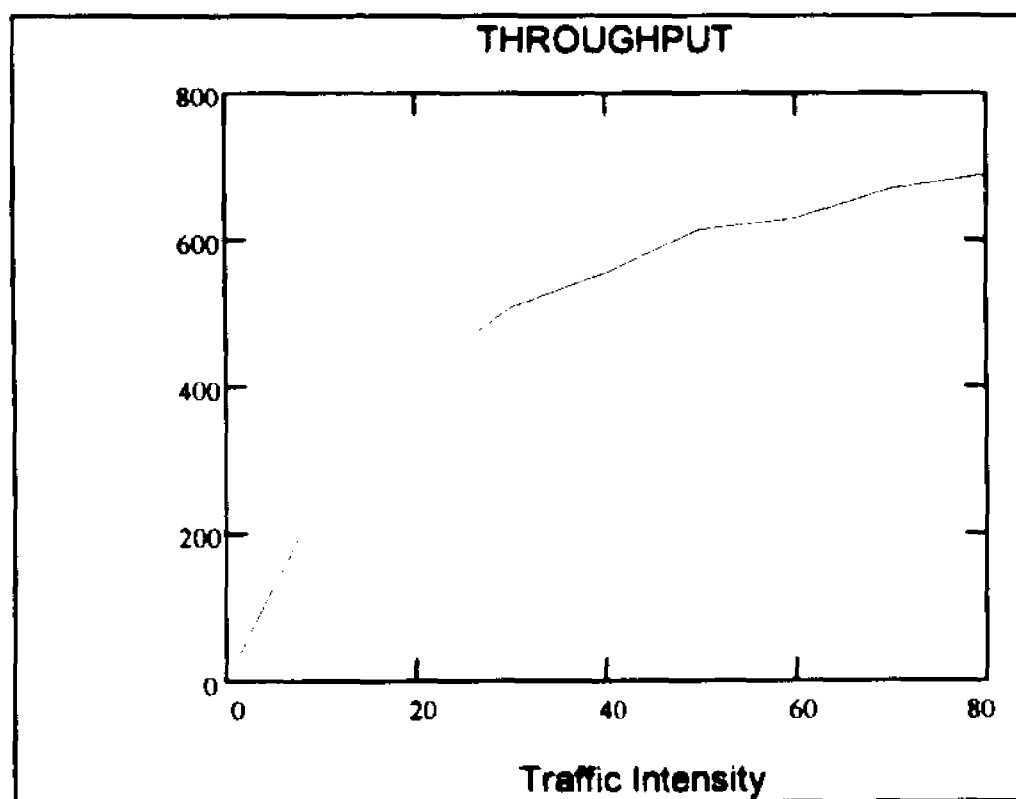


Figure 5 40 Full Duplex with 56 kb/s links

The average message delay for 56 kb/s full duplex links starts off at 0.42 sec. at traffic intensity of 1. It increases slightly to 0.56 sec. at traffic intensity of 10 before increasing drastically to 3.97 sec. at traffic intensity 20. It continues increasing for higher traffic intensities reaching a maximum of 10.56 sec at traffic intensity of 80.

The maximum message delay for 56 kb/s full duplex links increases from 1.28 sec at traffic intensity of 1 to 4.77 sec. at traffic intensity of 10. It then increases drastically, reaching 50.15 sec at traffic intensity of 50, giving way to less drastic increase reaching 54.57 sec at traffic intensity of 80.

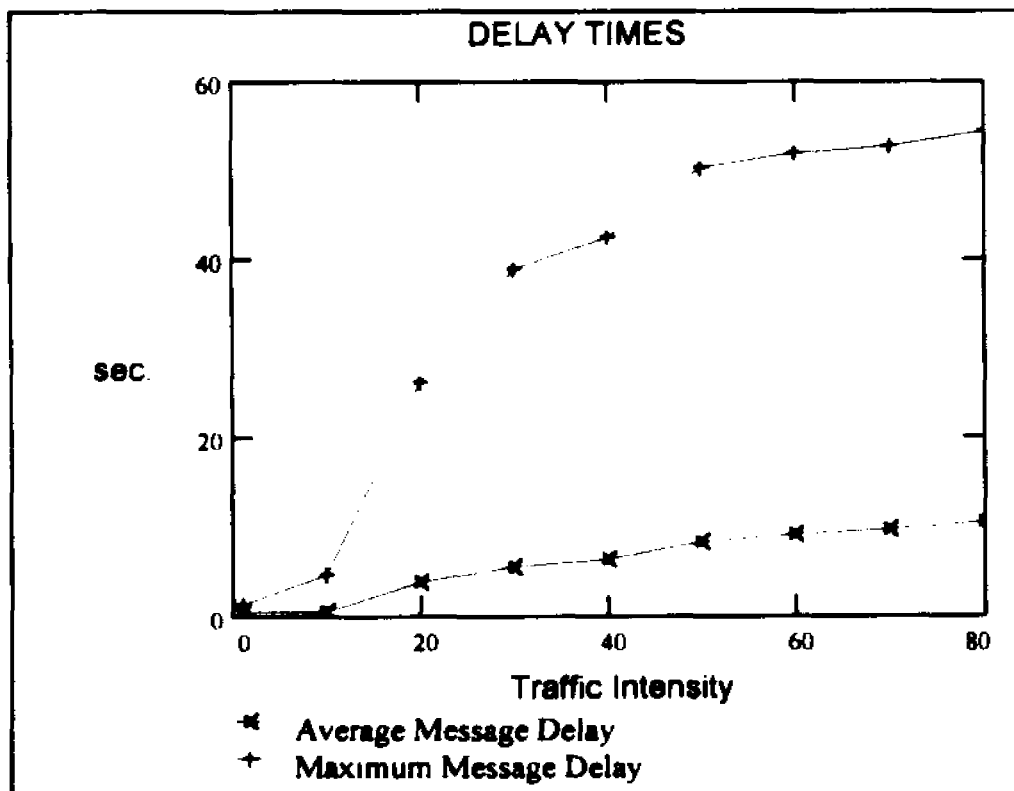


Figure 5.41. Full Duplex with 56 kb/s links

The percent of messages over three seconds for 56 kb/s full duplex links increases from 0% at traffic intensity of 1 to 1.61% at traffic intensity of 10. After that it starts increasing sharply reaching a maximum of 64.45% at traffic intensity of 80.

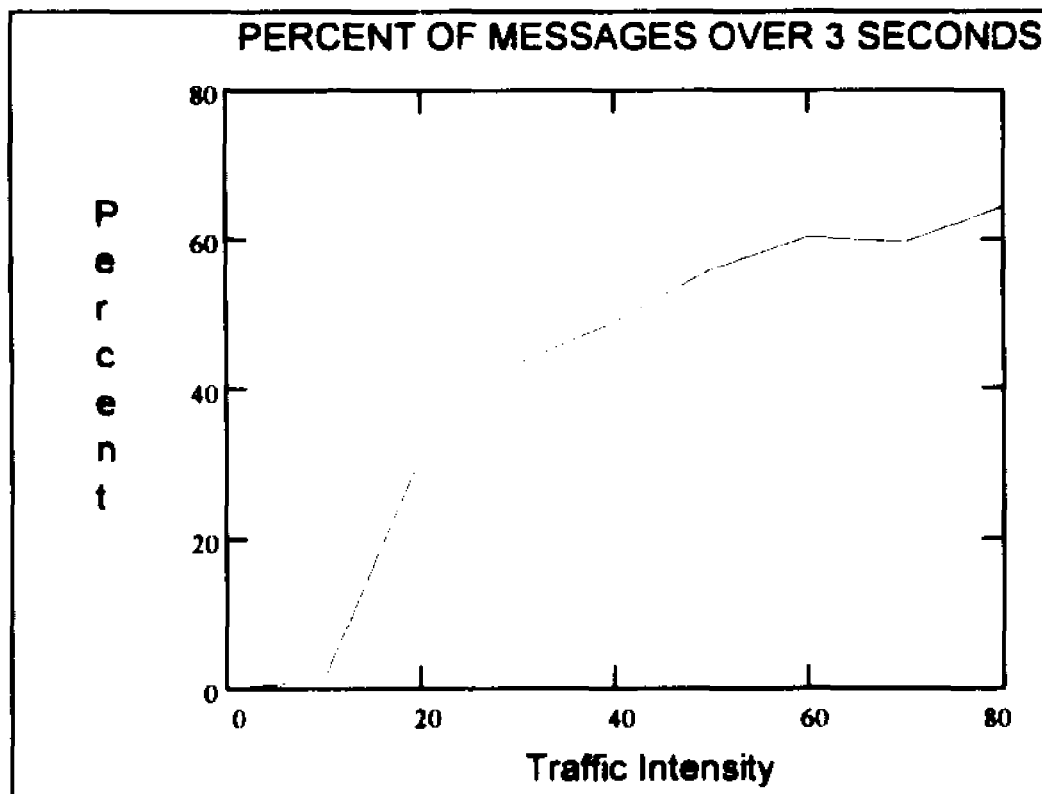


Figure 5.42: Full Duplex with 56 kb/s links

5.2.2.1.3 380 kb/s Links

The throughput for 380 kb/s full duplex links increases from 29.3 kb/s at traffic intensity of 1 to 2063.3 kb/s at traffic intensity of 80. Then it levels off slightly reaching 2586.4 kb/s at traffic intensity of 110.

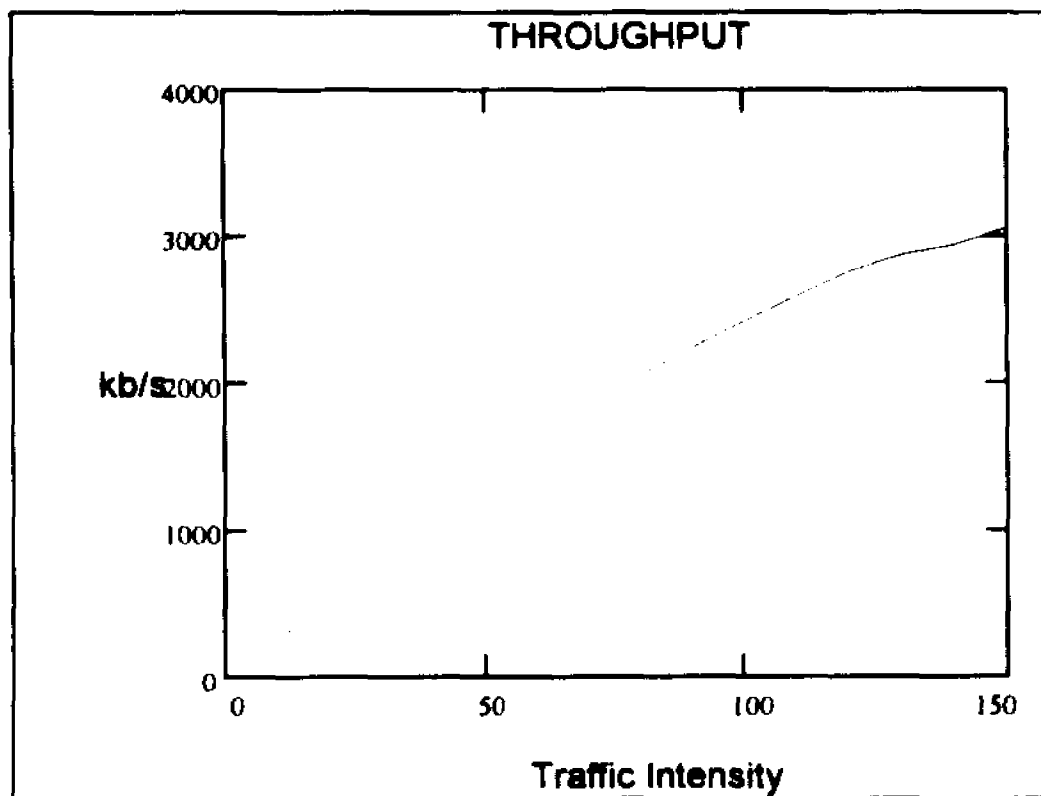


Figure 5.43: Full Duplex with 380 kb/s links

The average message delay for 380 kb/s full duplex links remains under 0.1 sec between traffic intensities 1 and 60. It then increases to 1.91 sec at traffic intensity of 100, finally reaching 2.21 sec. at traffic intensity 110.

The maximum message delay remains under 0.5 sec between traffic intensities 1 and 50. It then increases slightly to 2.37 sec at traffic intensity 70 before starting to increase steeply to 19.36 sec. at traffic intensity of 110.

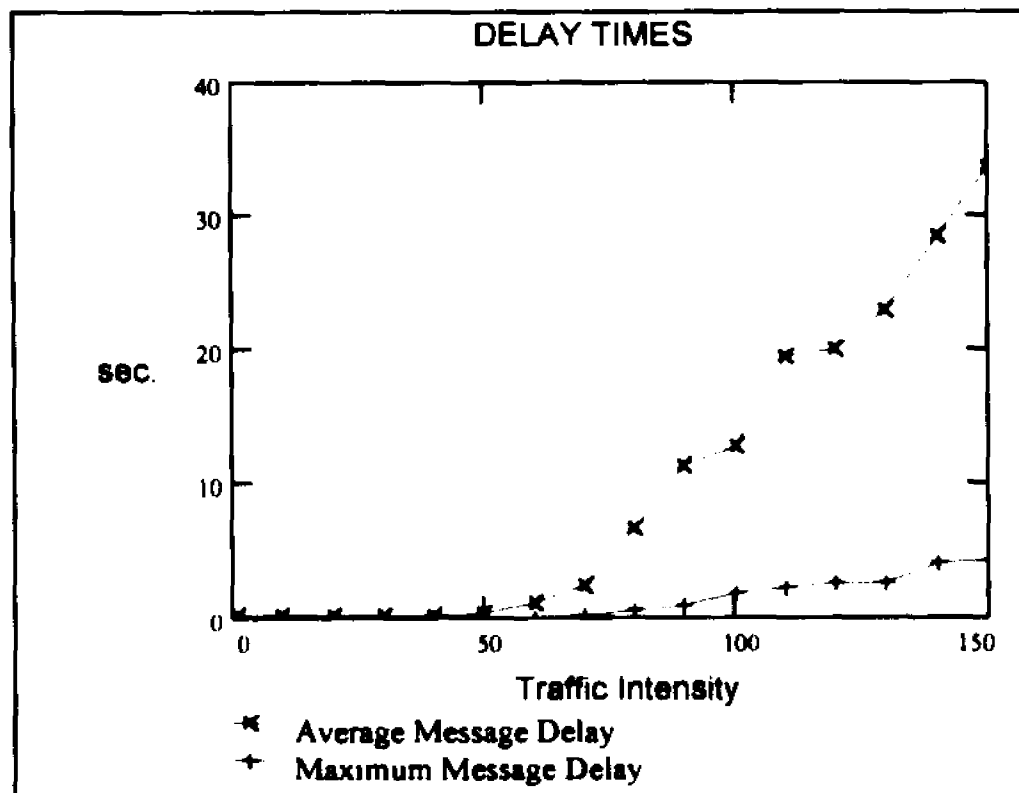


Figure 5.44. Full Duplex with 380 kb/s links

The percent of messages over three seconds remains at 0% between traffic intensities 1 and 70. After that it increases to 23.84% at traffic intensity of 100. It then levels off increasing slightly reaching 24.74% at traffic intensity of 110

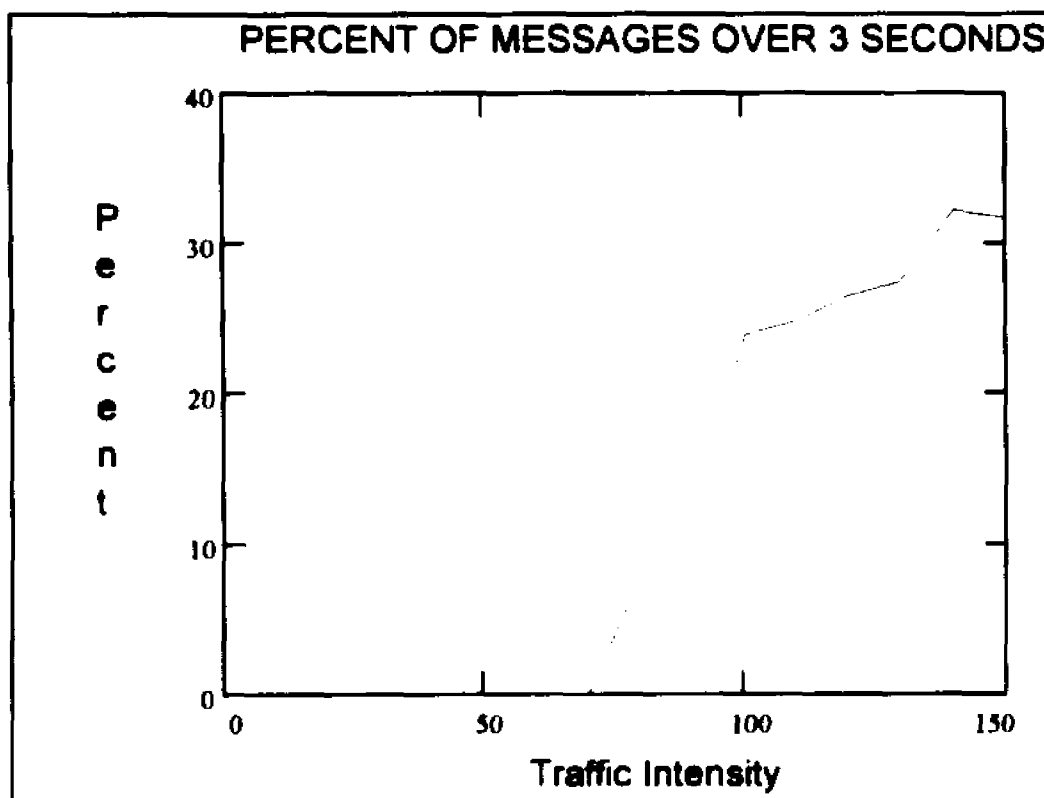


Figure 5.45: Full Duplex with 380 kb/s links

5.2.2.1.4 1544 kb/s Links

Throughput for 1544 kb/s full duplex links increases from 29.3 kb/s at traffic intensity of 1 to 7958.4 kb/s at traffic intensity of 300.

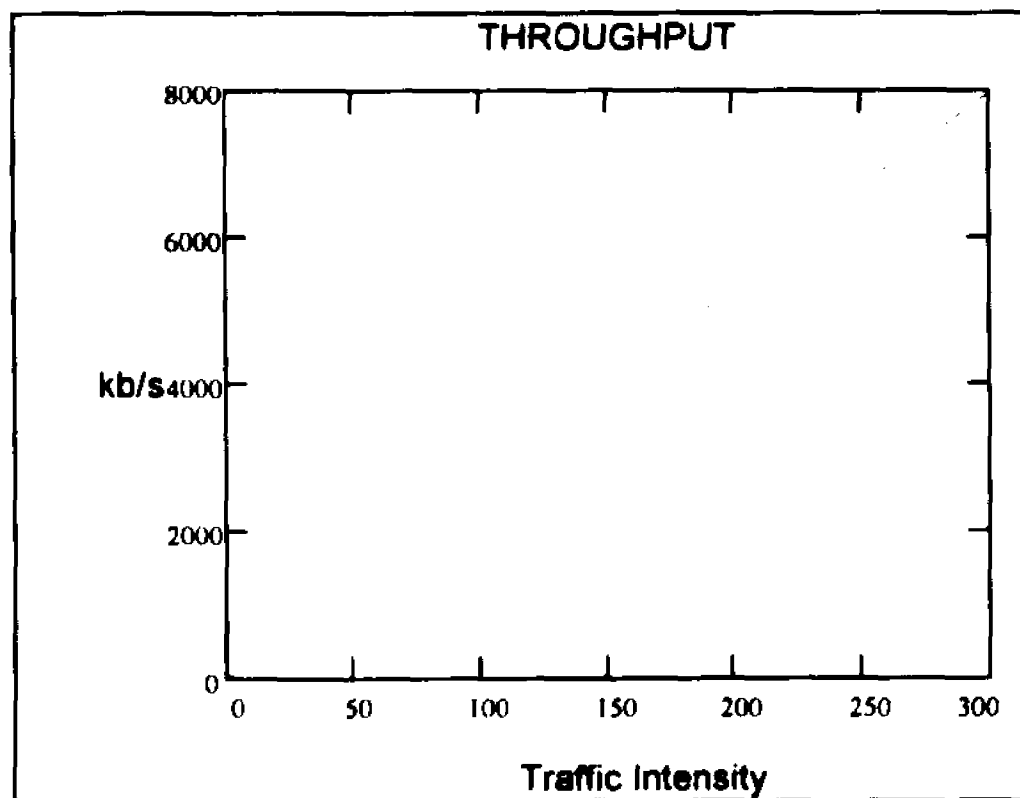


Figure 5 46 Full Duplex with 1544 kb/s links

The average message delay for 1544 kb/s full duplex links remains at 0.01 sec for traffic intensity 1 through 200. It then increases first to 0.02 sec at 250, then to 0.15 sec at traffic intensity of 300.

The maximum message delay for 1544 kb/s full duplex links remains below 0.05 sec for traffic intensities below 100. After that it first increases to 0.09 sec at traffic intensity of 250, then to 2.81 sec at traffic intensity of 300.

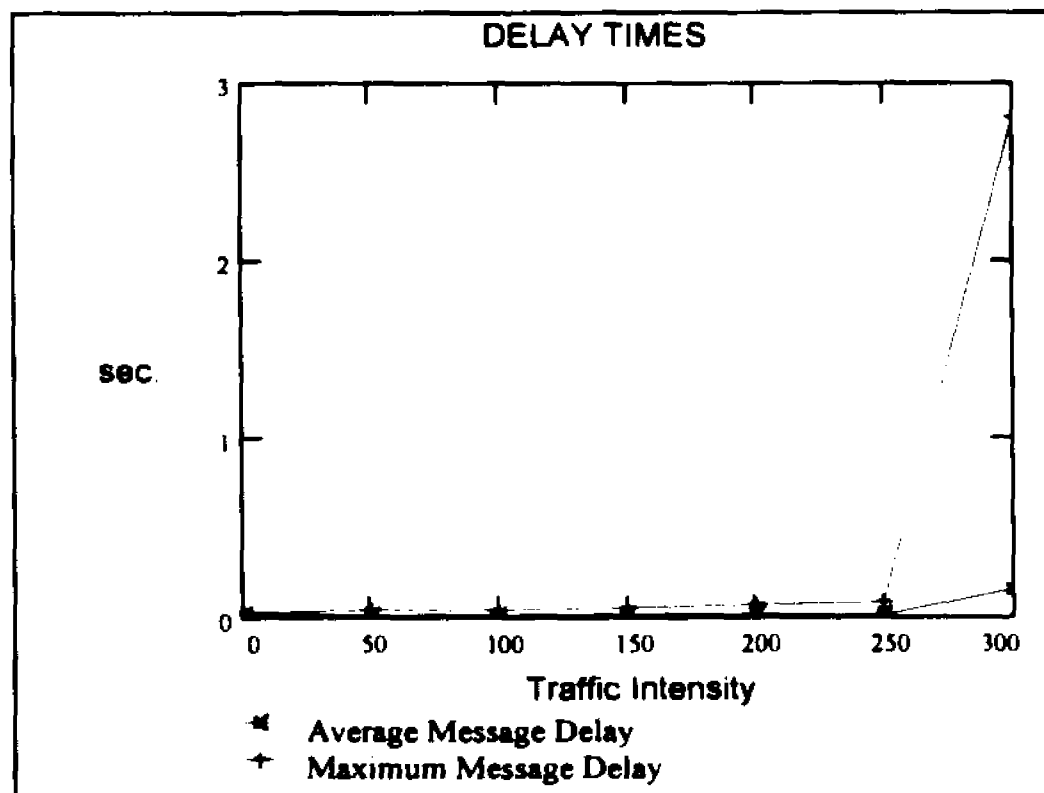


Figure 5.47. Full Duplex with 1544 kb/s links

The percent of messages over three seconds remains at 0% for traffic intensities 1 through 300.

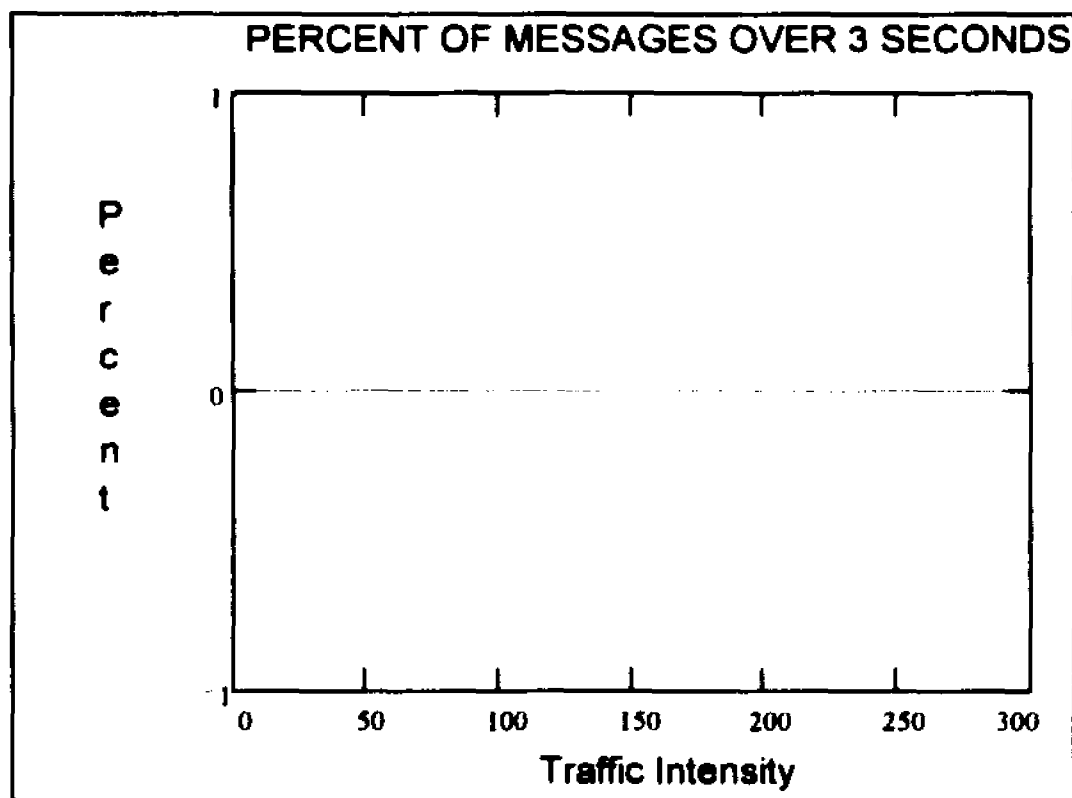


Figure 5 48 Full Duplex with 1544 kb/s links

5.2.2.2 Link Failure

5.2.2.2.1 9.6 kb/s Links

The throughput for 9.6 kb/s links with link failure increases from 25.9 kb/s at traffic intensity 1 to 91 kb/s at traffic intensity 10. It then starts leveling off for higher traffic intensities reaching a maximum of 129.7 kb/s at traffic intensity of 40.

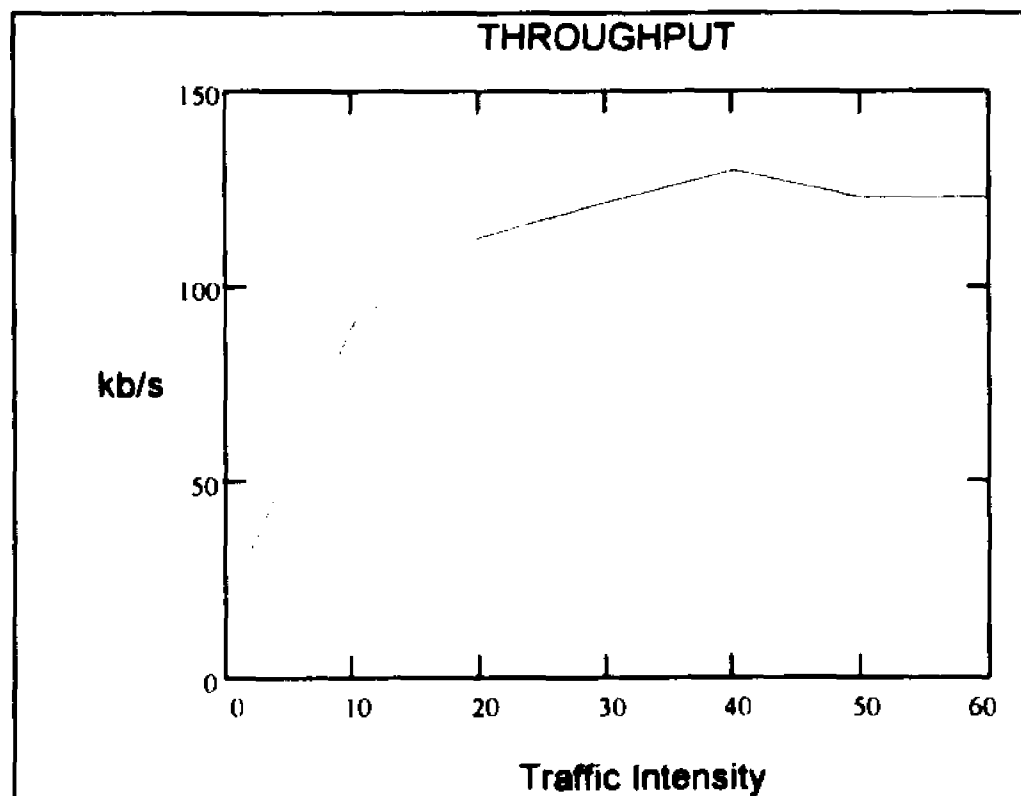


Figure 5.49. Link Failure with 9.6 kb/s links

The average message delay for 9.6 kb/s links with link failure increases from 3.37 sec at traffic intensity of 1 to 9.25 at traffic intensity of 10. It then steadily increases reaching 16.66 sec at traffic intensity of 250, before ending up at 15.05 sec at traffic intensity of 300.

The maximum message delay increases from 11.91 sec at traffic intensity of 1 to 52.03 sec at traffic intensity of 10. It remains at around 55 sec. range for higher traffic intensities reaching a maximum of 56.28 sec at traffic intensity of 300.

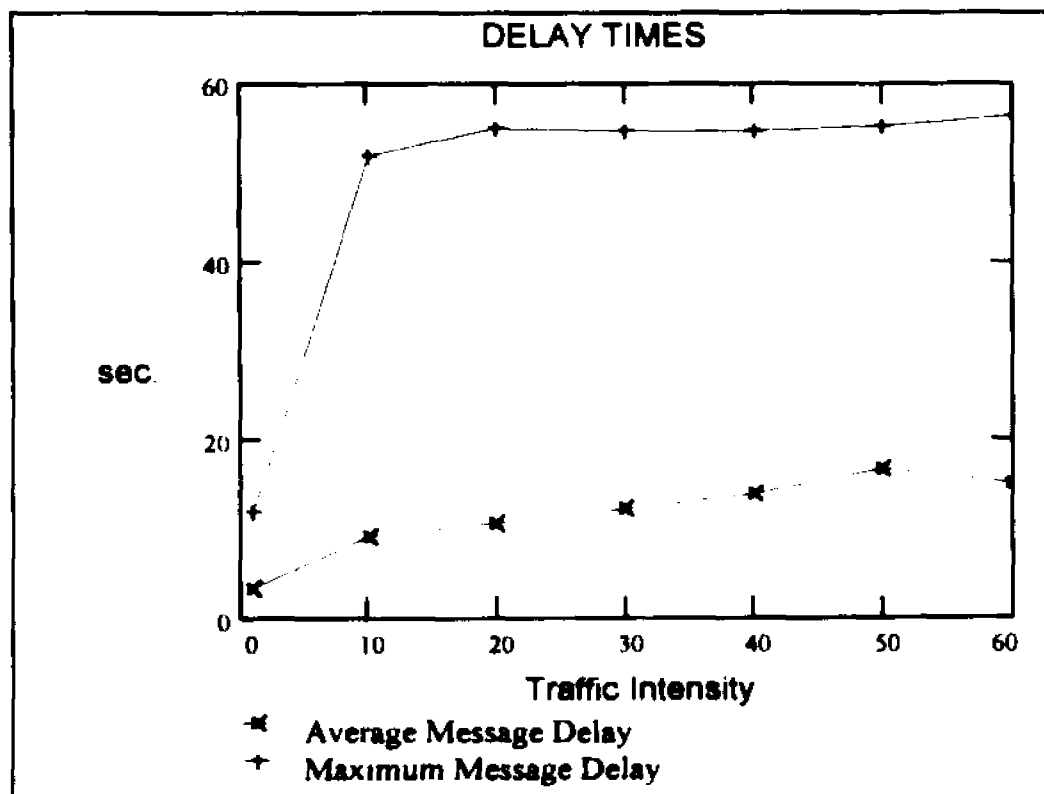


Figure 5.50 Link Failure with 9.6 kb/s links

The percent of messages over three seconds for 9.6 kb/s links with link failure increases steadily from 51.32% at traffic intensity of 1 to 87.44% at traffic intensity of 50. It ends up at 85.08% for traffic intensity of 60.

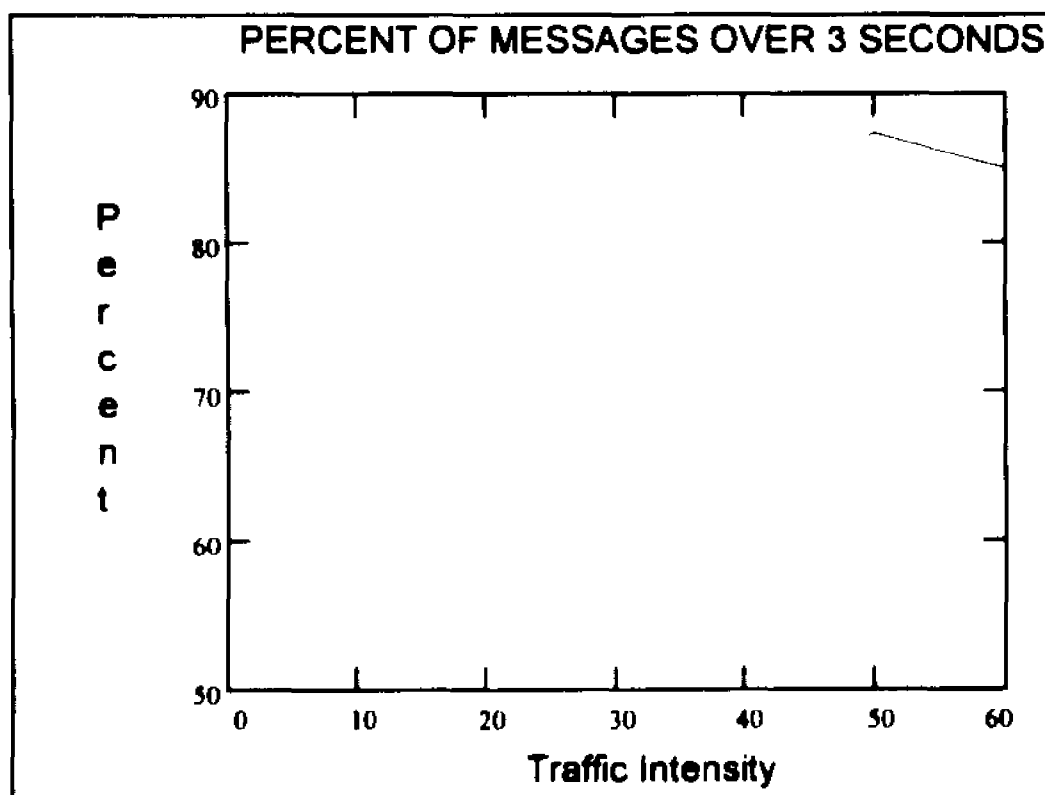


Figure 5.51 Link Failure with 9.6 kb/s links

5.2.2.2.2 56 kb/s Links

The throughput for 56 kb/s links with link failure increases from 28.9 kb/s at traffic intensity 1 to 224.3 kb/s at traffic intensity 10. After that it continues to increase less steeply reaching 627.8 kb/s at traffic intensity of 110.

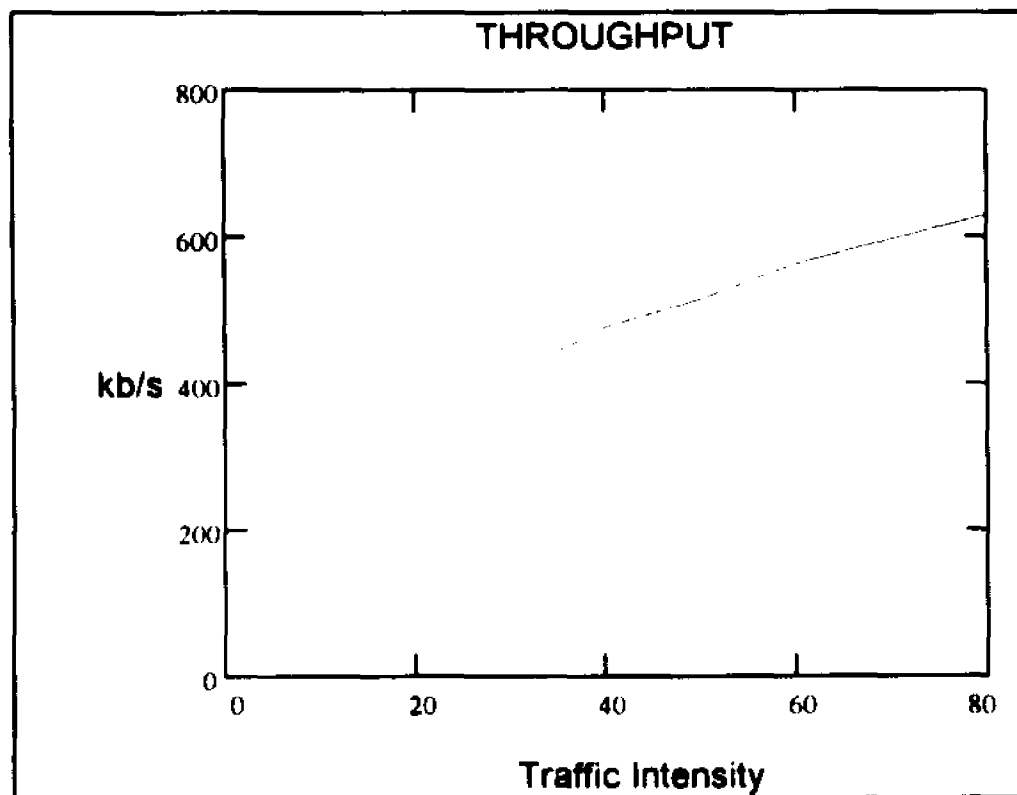


Figure 5.52 Link Failure with 56 kb/s links

The average message delay for 56 kb/s links with link failure increases slightly to 1.89 sec. at traffic intensity of 10 from 0.45 sec. at traffic intensity of 1. It then shows a rapid increase to 5.54 sec. at traffic intensity of 20, before leveling off again, reaching 9.14 sec at traffic intensity of 80.

The maximum message delay increases steeply between traffic intensity 1 and 30 from 1.23 sec. to 45.5 sec. It then continues to increase at a lower rate reaching 56.13 sec at traffic intensity of 80.

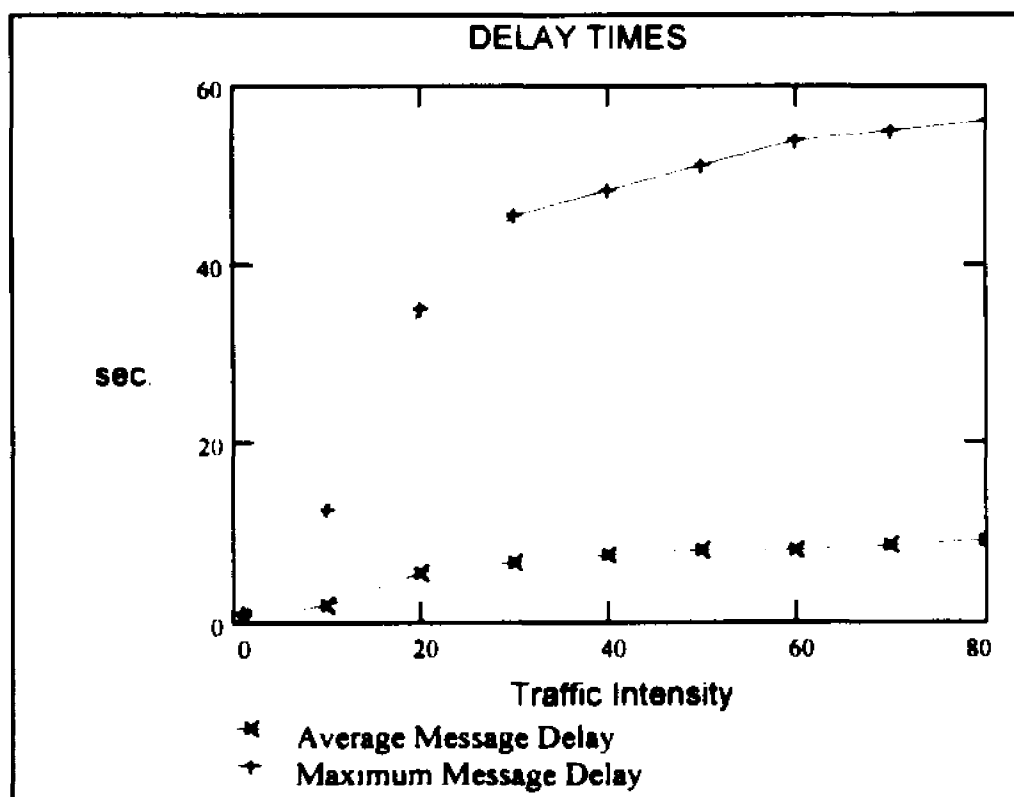


Figure 5.53 Link Failure with 56 kb/s links

The percent of messages over three seconds for 56 kb/s links with link failure increases from 0% at traffic intensity of 1 to 34.75% at traffic intensity of 20. After that the increase becomes less steep reaching 53.9% at traffic intensity of 80.

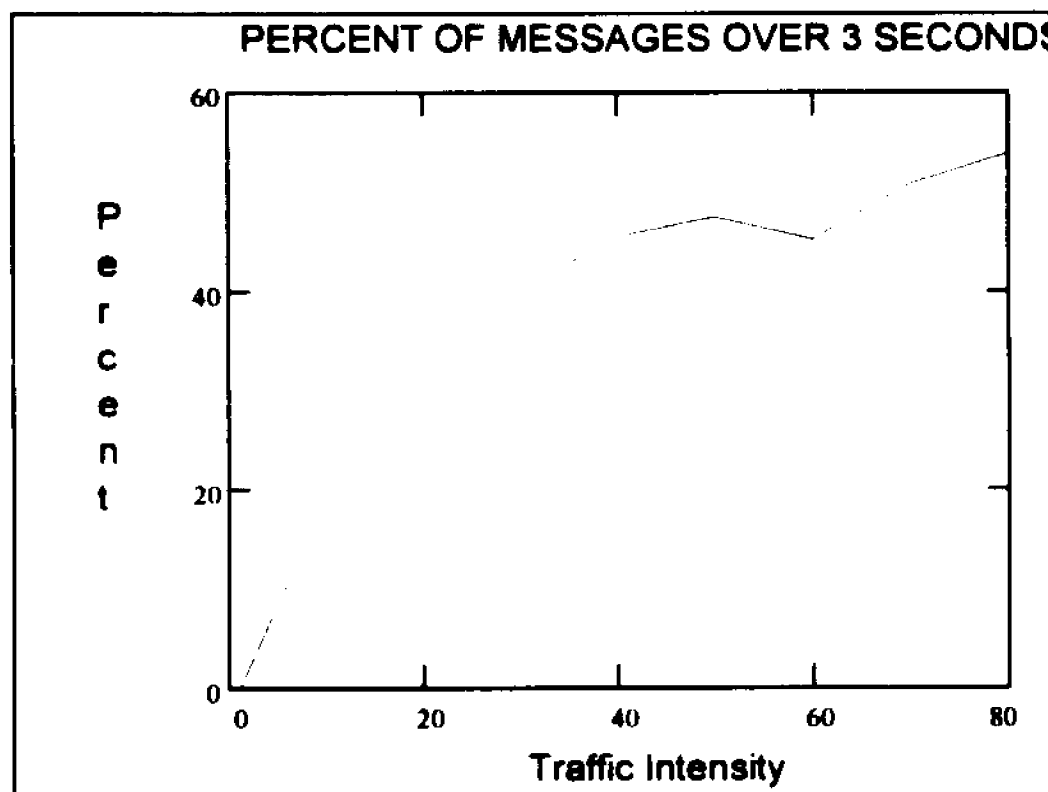


Figure 5.54 Link Failure with 56 kb/s links

5.2.2.2.3 380 kb/s Links

The throughput for 380 kb/s links with link failure increases from 29.3 kb/s at traffic intensity of 1 to 1055.8 kb/s at traffic intensity of 40. Then it levels off slightly, reaching 1968.1 kb/s at traffic intensity of 110.

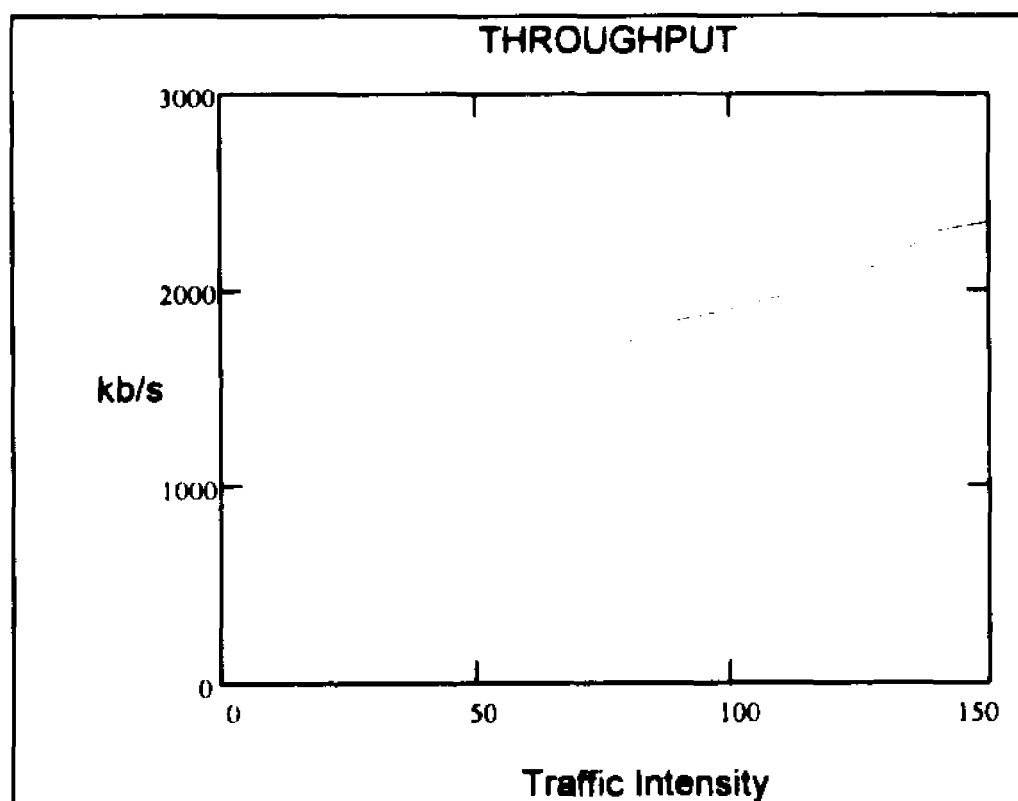


Figure 5.55 Link Failure with 380 kb/s links

The average message delay for 380 kb/s links with link failure remains below 0.1 sec. for traffic intensities below 40. It then has a small increase to 0.42 sec. at traffic intensity of 50, followed by a larger increase reaching 5.29 sec. at traffic intensity of 100. It eventually levels off at 5.8 sec. at traffic intensity of 110.

The maximum message delay for 380 kb/s links with link failure remains under 1 sec. between traffic intensities 1 and 40. It then starts to increase sharply reaching 35.05 sec. at traffic intensity of 110.

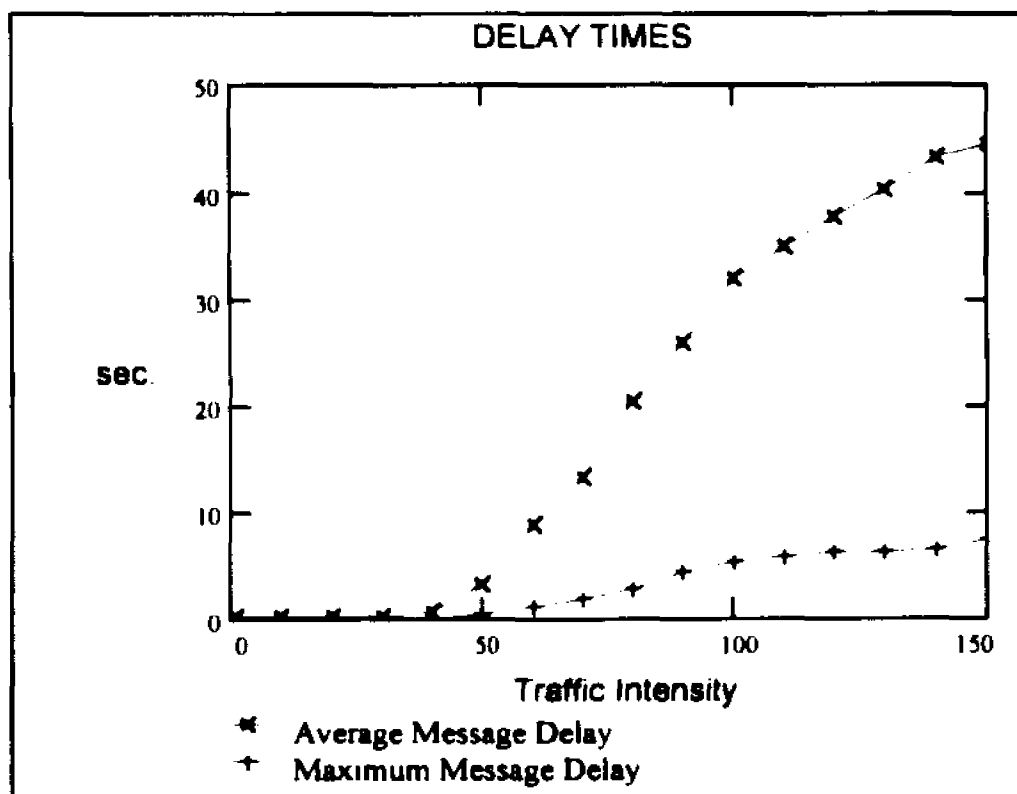


Figure 5.56. Link Failure with 380 kb/s links

The percent of messages over three seconds remains at 0% for traffic intensities 1 to 40. This is followed by a small increase to 0.07% at traffic intensity of 50. It then starts increasing sharply, reaching 46.93% at traffic intensity of 110.

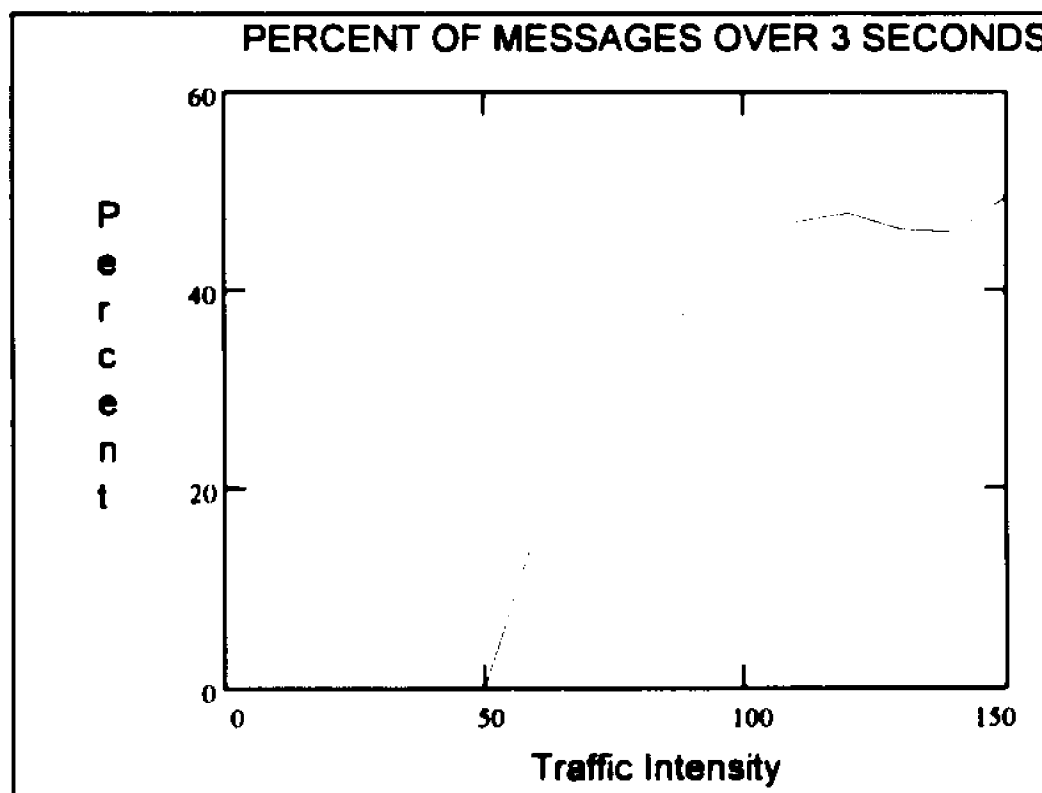


Figure 5.57 Link Failure with 380 kb/s links

5.2.2.2.1 1544 kb/s Links

The throughput for 1544 kb/s links with link failure steadily increases from 29.3 kb/s at traffic intensity of 1 to 5179.5 kb/s at traffic intensity of 200. Then it levels off slightly reaching 6107.5 kb/s at traffic intensity of 250.

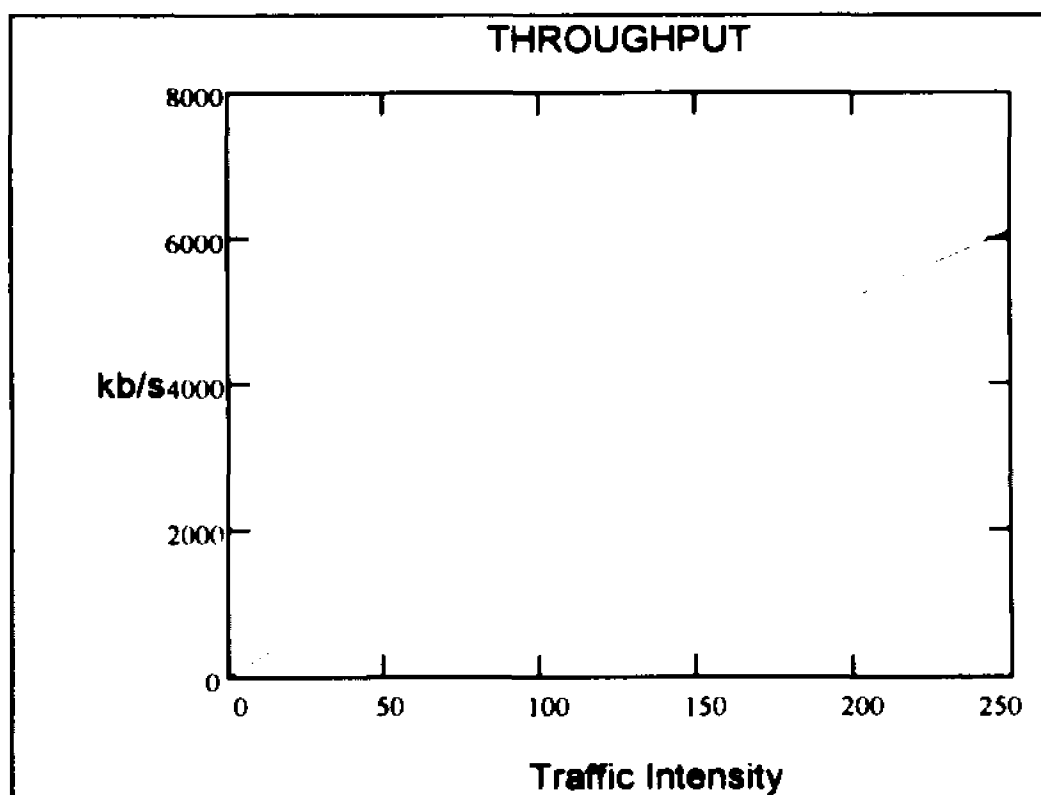


Figure 5.58 Link Failure with 1544 kb/s links

The average message delay for 1544 kb/s links with link failure remains at 0.02 sec for traffic intensities 1 through 150. It then increases to 0.25 sec at traffic intensity of 200 before reaching 1.02 sec at traffic intensity of 250.

The maximum message delay for 1544 kb/s links with link failure increases from 0.04 sec. at traffic intensity 1 to 0.06 sec at traffic intensity 100. It then increases slightly to 0.2 sec. at traffic intensity of 150 followed by steeper increases to 2.65 sec at traffic intensity of 200 and 7.9 sec at traffic intensity of 250.

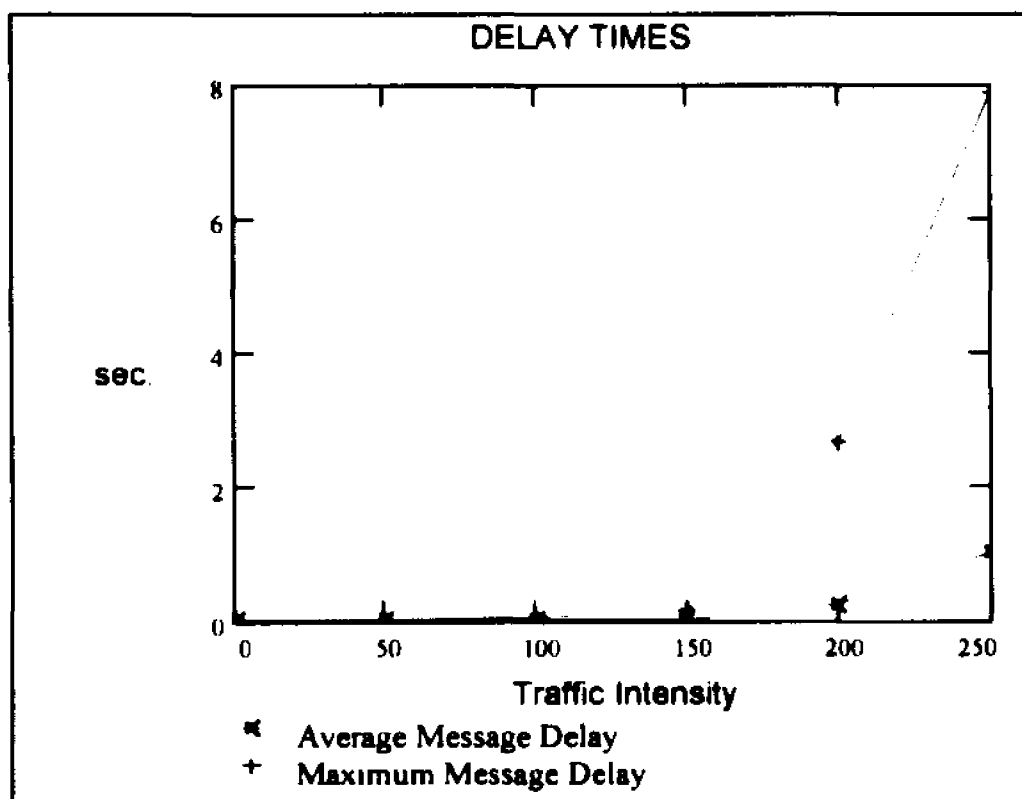


Figure 5.59. Link Failure with 1544 kb/s links

The percent of messages over three seconds for 1544 kb/s links with link failure remains at 0% for traffic intensities 1 through 200. It then increases steeply, reaching 14.85% at traffic intensity of 250.

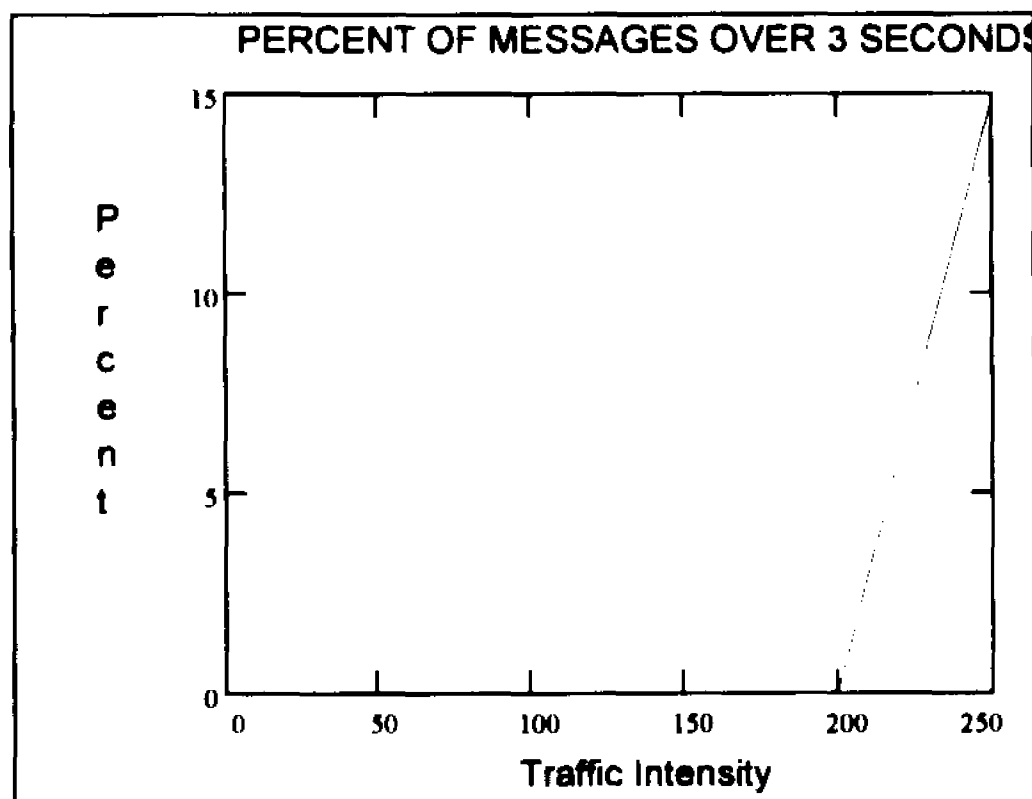


Figure 5.60. Link Failure with 1544 kb/s links

VI. DISCUSSION OF RESULTS

The network simulation results obtained in the simulation studies show that the link bandwidth and direction of transmission have an impact on throughput and delay times as well as the Quality of Service (QoS). The following discussion of results are categorized first according to each of the simulation models, then within each model divided into throughput, delay times and QoS.

6.1 Network Results for the First Simulation Model

6.1.1 Throughput

At the outset, the throughput of the network in this model increases with the traffic intensity. This increase appears to be linear, until it reaches a point where it curves away from its trajectory, and remains virtually constant for increased traffic levels. This particular pattern can be seen for each of the four link data rates for all three different link properties of the network model: with full duplex links (Figure 6.1); with half duplex links (Figure 6.2) and with link failure (Figure 6.3).

This point where the throughput increase deviates from the previous increases can also be observed for network models with different link rates. It is also noted that for each of the link property of the network model, the separation point of actual throughput from the previously traced linear increase occurs at lower traffic intensities for lower link rates. This result implies that there is a strong correlation between link rates and network throughput and we demonstrate numerically that the implication is indeed accurate and valid.

After the point of separation, the increase in network throughput gradually approaches zero. This implies that as the traffic intensity is increased, the throughput of a network reaches approaches its maximum.

It is also noted that, after the point of separation, the decrease in the slope of the graph, (eventually to zero) takes place much faster for lower link speeds than higher ones. This which may suggest that the throughput capacity of the network is reached more quickly for lower link speed network than higher link speed networks.

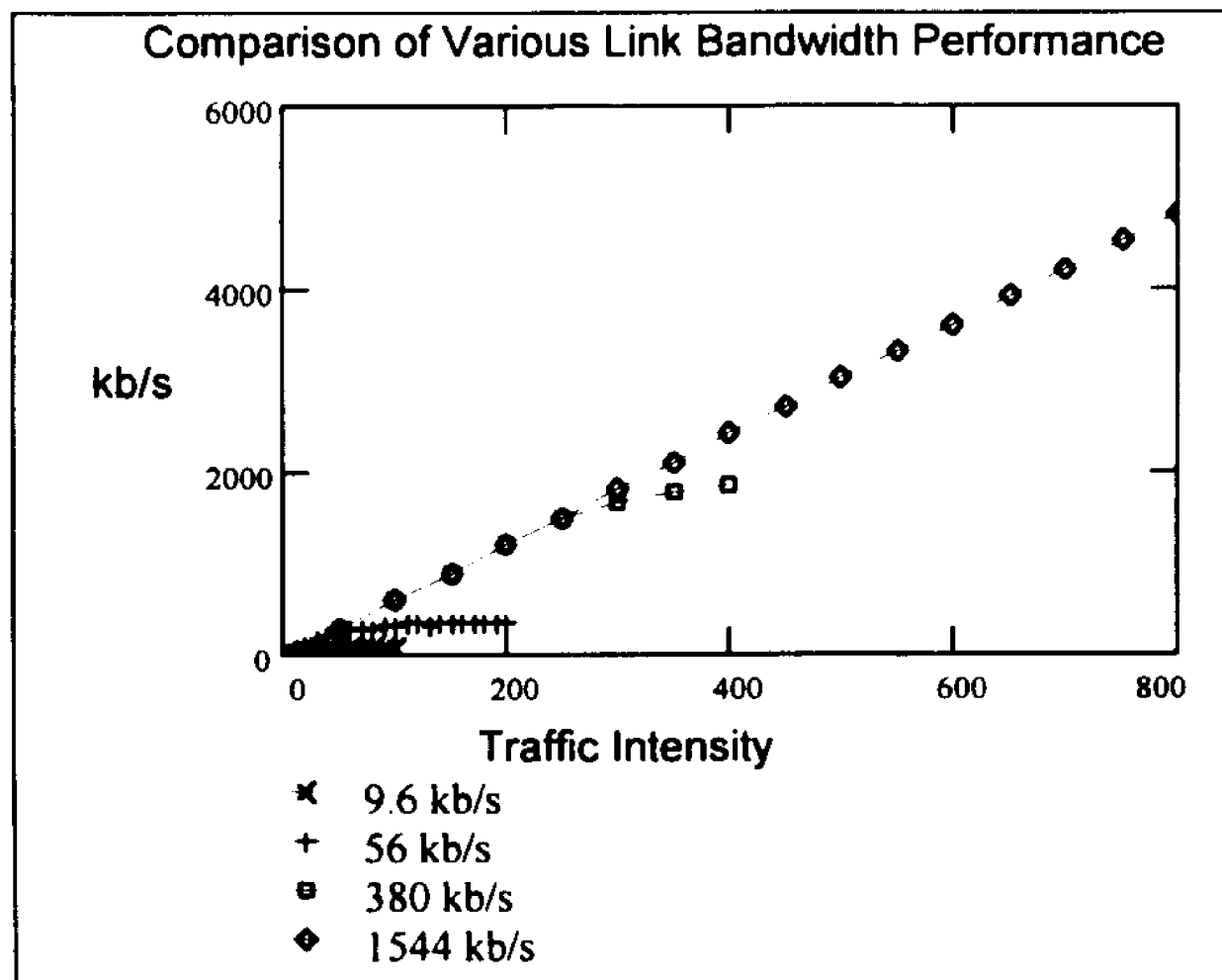


Figure 6-1 Throughput for Full Duplex links

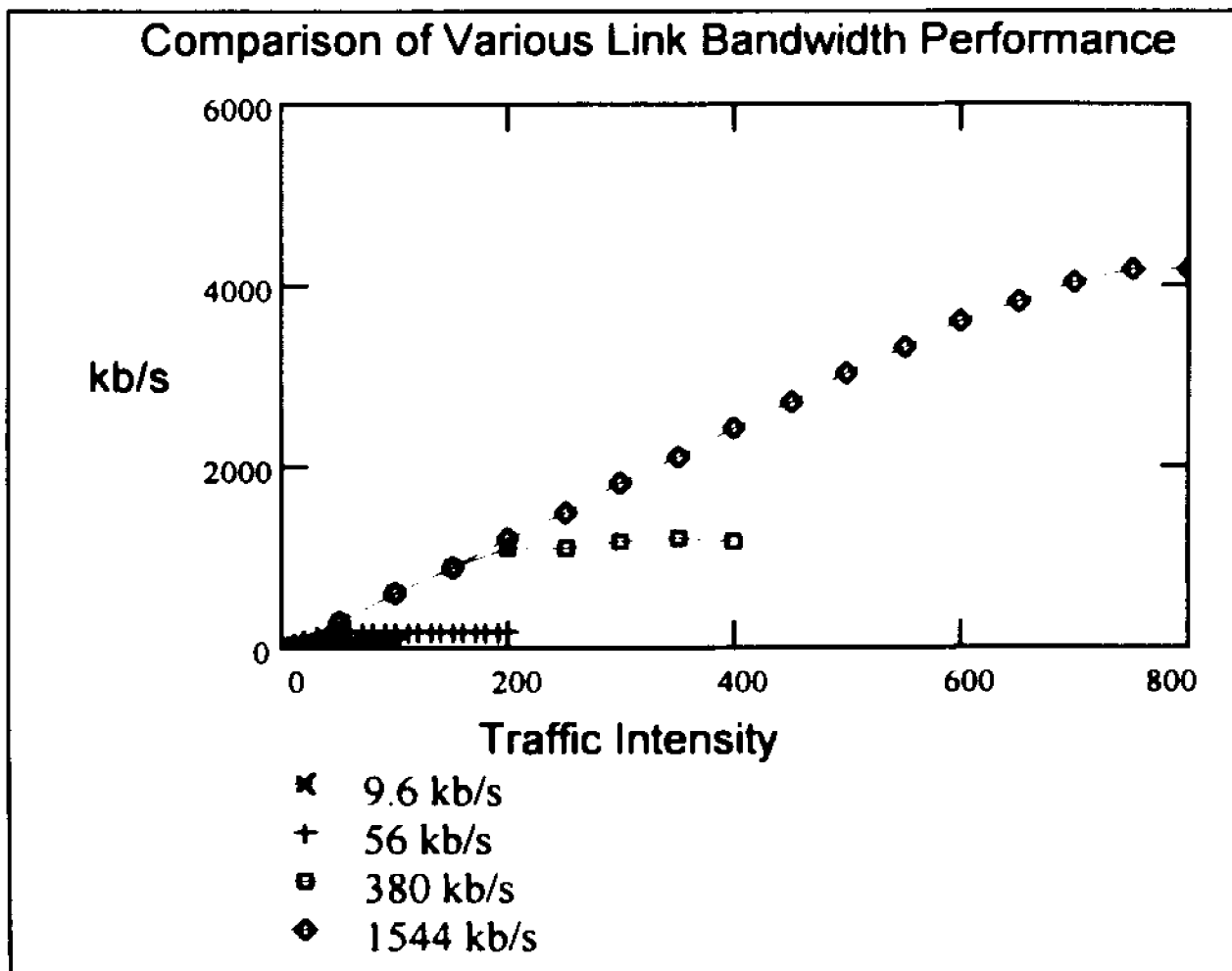


Figure 6. 2. Throughput for Half Duplex links

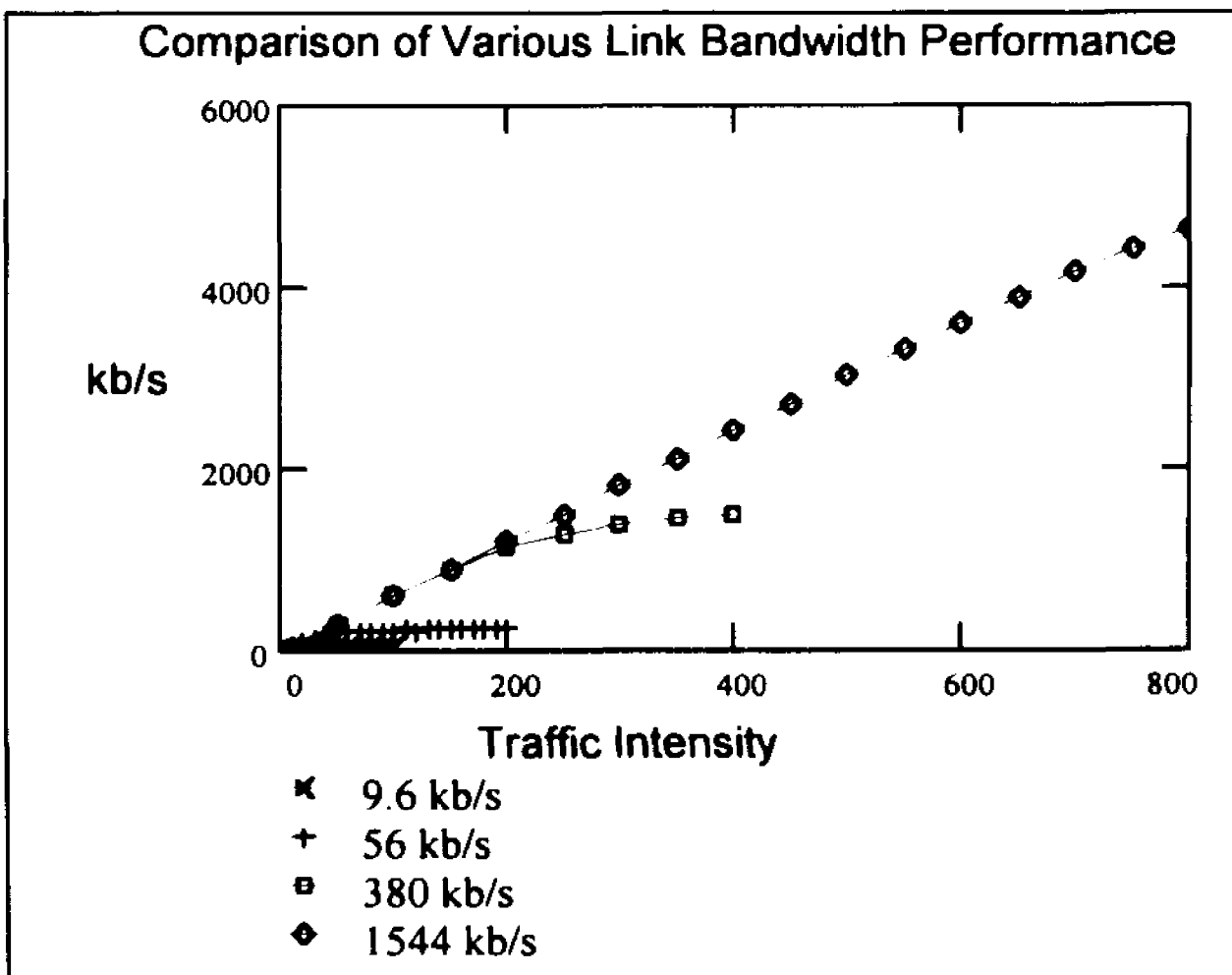


Figure 6 3 Throughput for Link Failure

The second parameter that affects throughput is the direction of transmission. The results show that, for a particular line speed, throughput levels off at lower traffic intensities for half duplex links than for full duplex links. This indicates that the maximum throughput level occurs at lower traffic intensities for half duplex links. The traffic intensity at the point where the throughput for the half duplex link splits from the linear increase is about half that of the point where the full duplex link throughput splits from the linear increase. This suggests that substitution of a full duplex line for a half duplex line increases the potential throughput by a factor of two.

Another aspect of a network that affects throughput is link failure. Though temporary, link failures cause traffic to be rerouted to other links, thus increasing the load on other links. It is important to investigate how this phenomenon affects the operation of a network. Link failure, can be considered the absence of a transmission link between two nodes. From the results of the simulations, it is evident that a link failure has a negative impact on throughput.

The throughput for a network with a link failure demonstrates the same pattern as for a network with full or half duplex links. After the point of separation is reached, the maximum throughput for the network is between the maximum throughput levels of half duplex and full duplex network models. This observation leads to the following inequality:

$$THR_{FD} \geq THR_{LFI} \geq THR_{HD}$$

where THR_{LFI} is failure of n links, THR_{FD} is full duplex, and THR_{HD} is half duplex linked network throughput respectively. We can see that this inequality holds true for the lower bandwidths of 9.6 kb/s (Fig. 6.4) and 56 kb/s (Fig. 6.5), as well as the high bandwidths of 380 kb/s (Fig. 6.6) and 1544 kb/s (Fig. 6.7).

Since a link failure implies the absence of a link, and this may suggest that for a well connected network, as far as network throughput is concerned, it is better to have fewer full duplex links, than a full complement of half duplex links. However it is important to point out that the above inequality may not hold if the number of failed links were greater than one, or the failed link is a link cut-set of the topology (thus resulting in two separated sub-topologies).

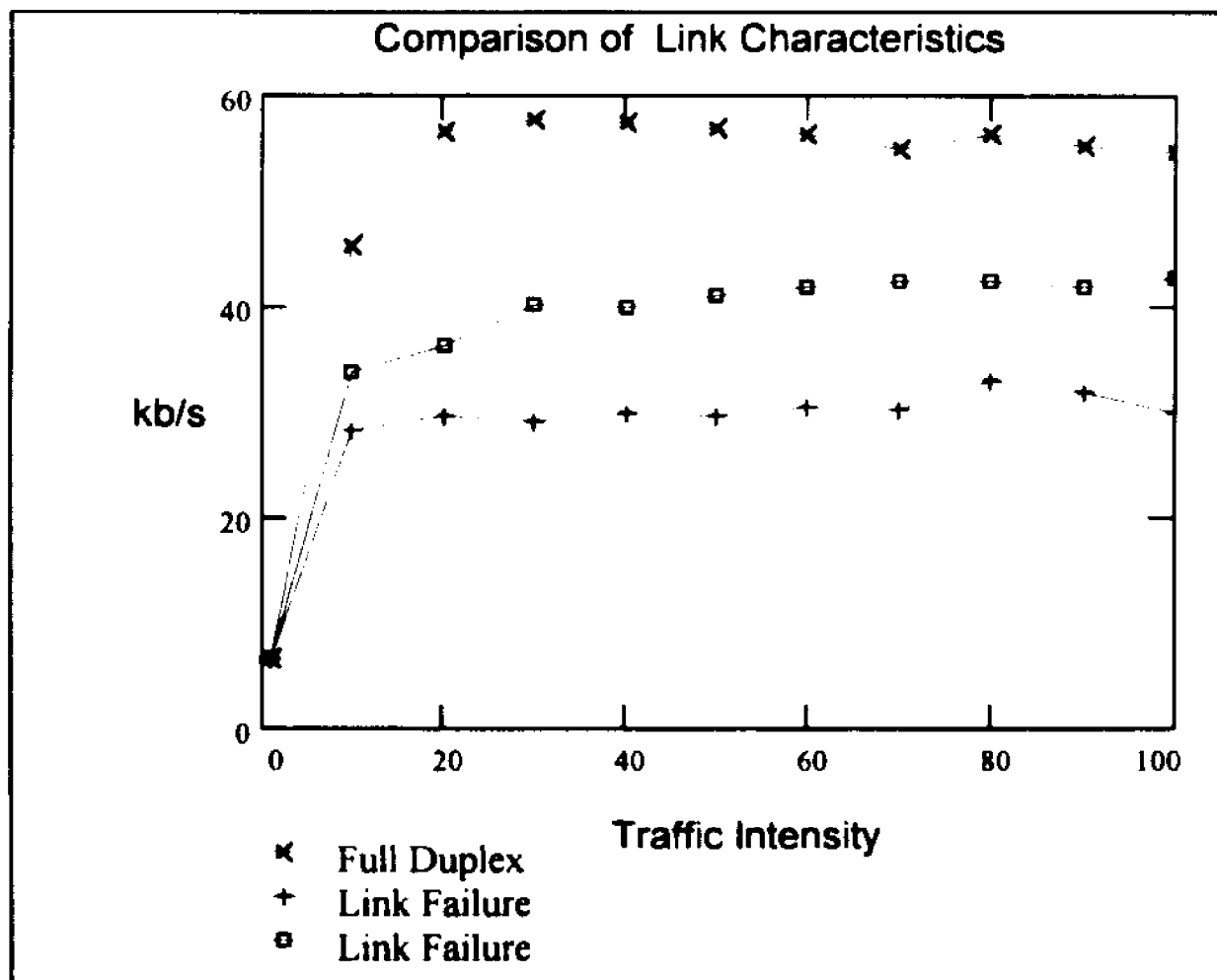


Figure 6. 4 Throughput for 9.6 kb/s links

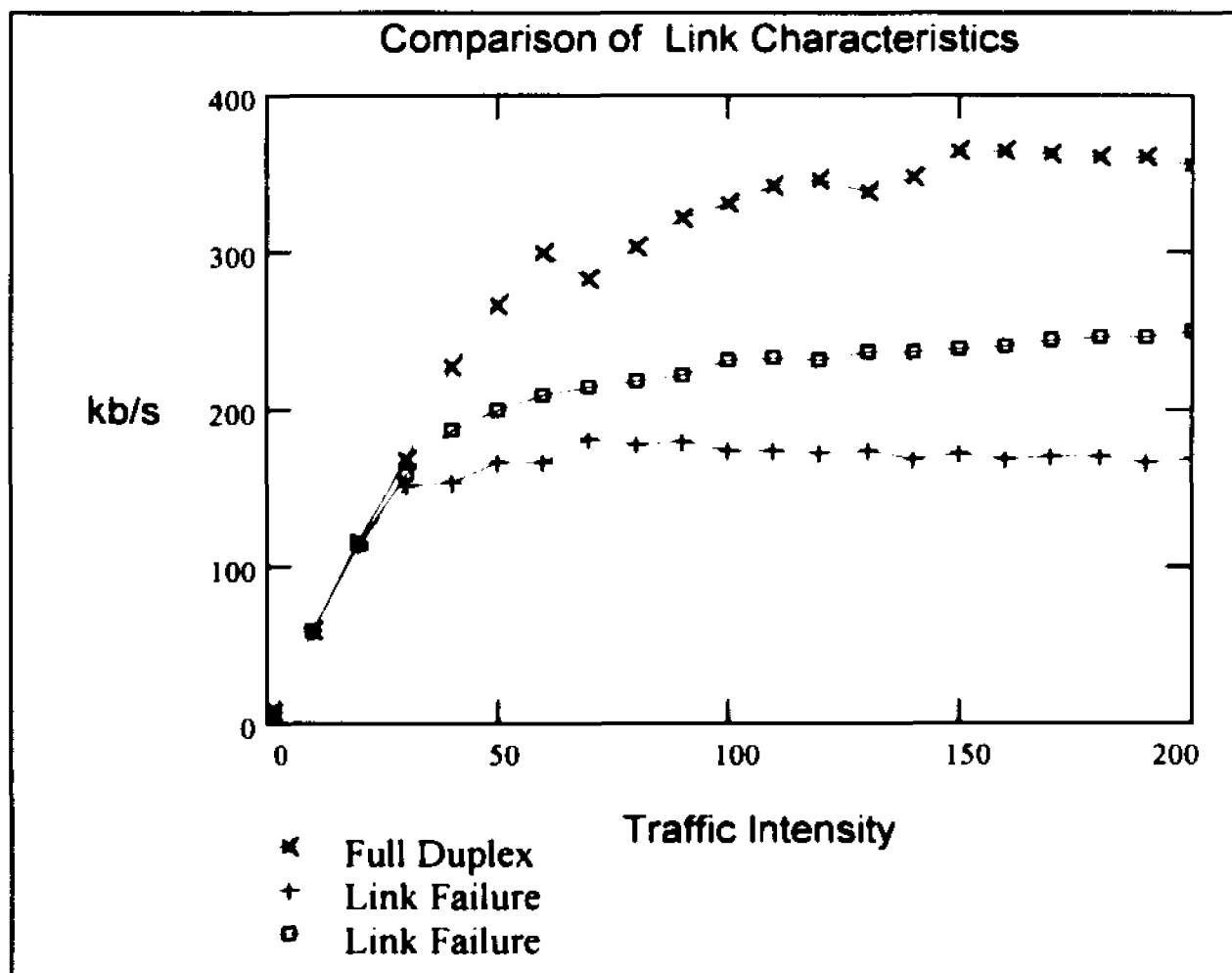


Figure 6 5 Throughput for 56 kb/s links

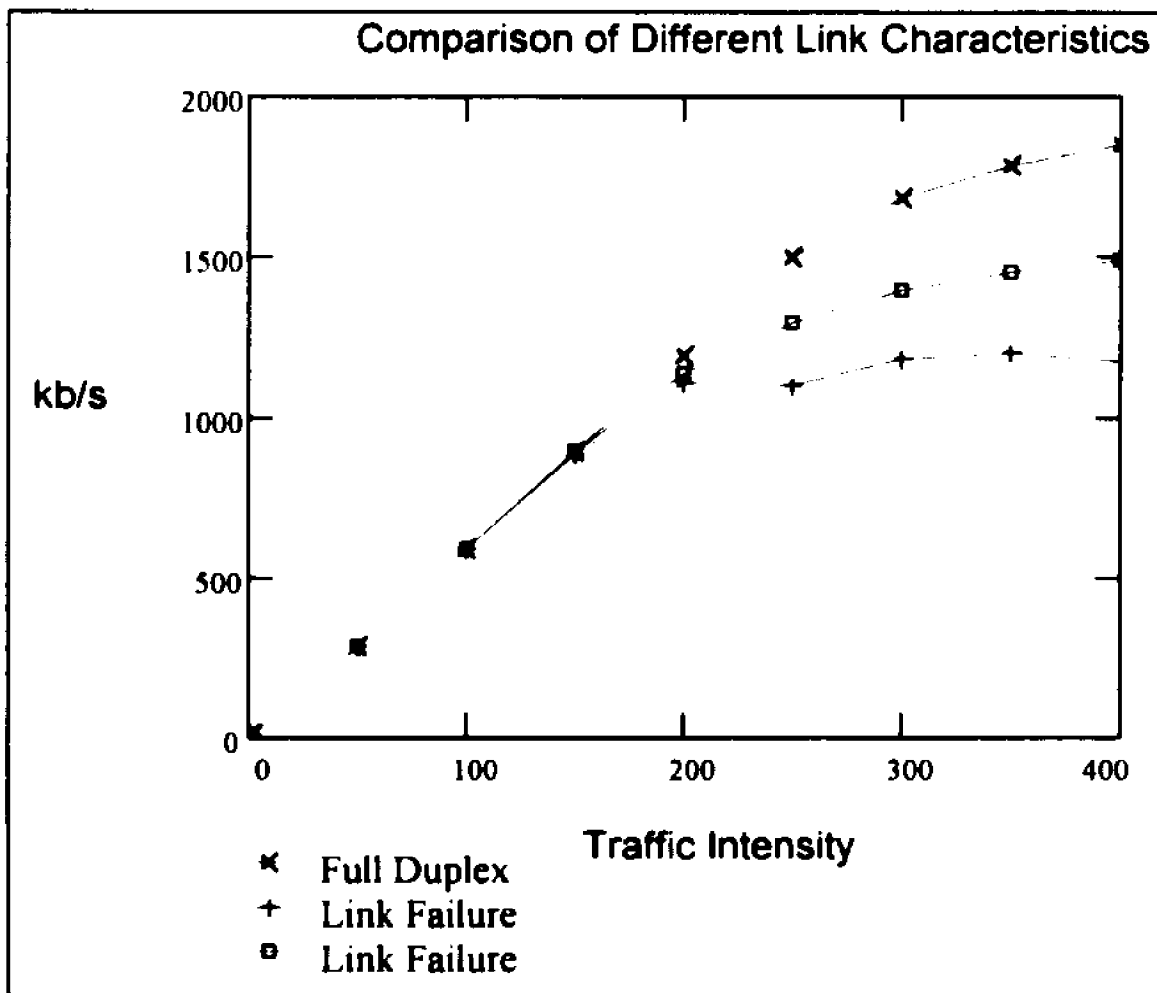


Figure 6 6 Throughput for 380 kb/s links

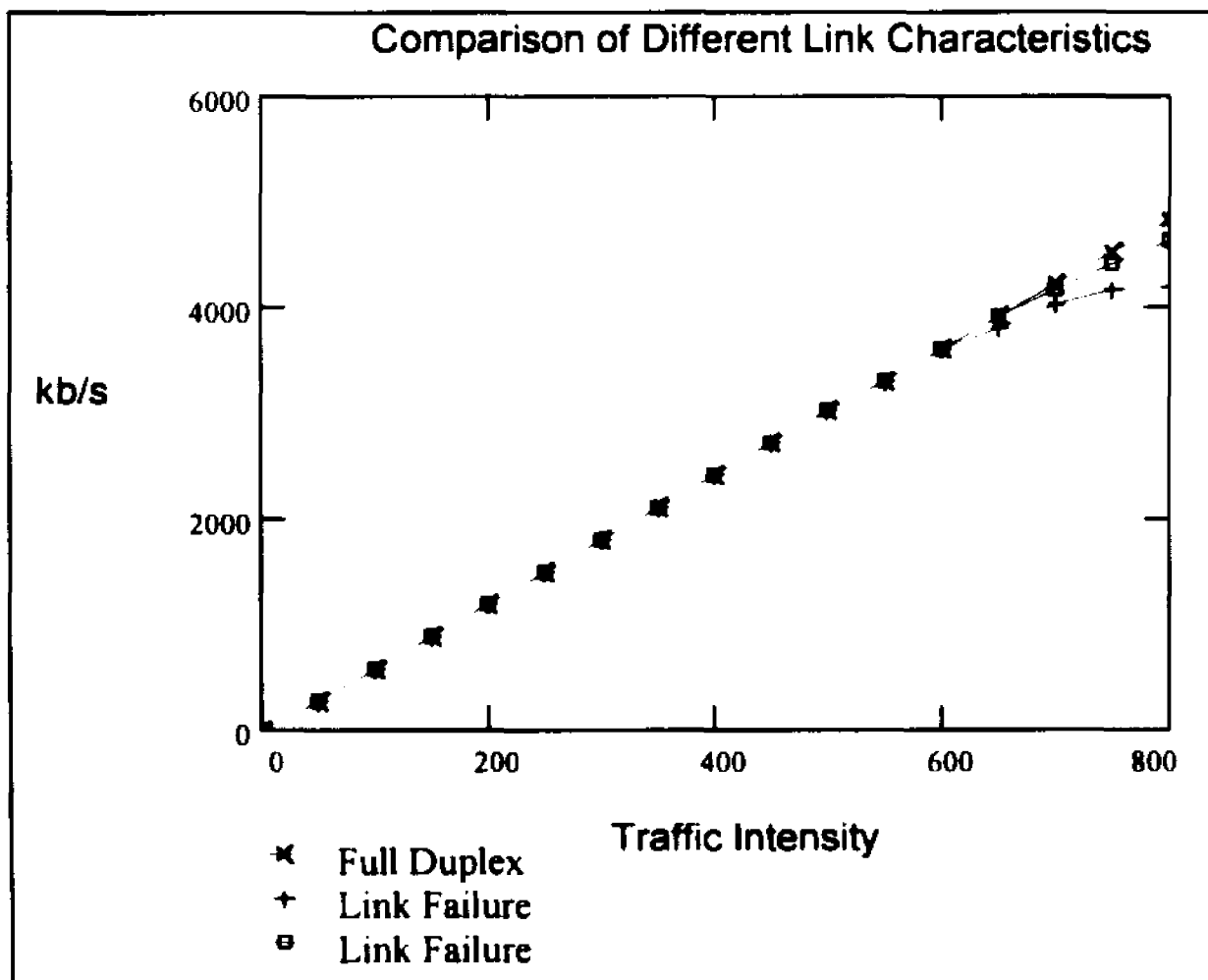


Figure 6. 7 Throughput for 1544 kb/s links

6.1.2 Message Delay Times

6.1.2.1 Average Message Delay

Average Message Delay indicates the average time it takes for a message to reach its destination. It is a measure of network performance as far as message delivery is concerned. The results indicate that for low traffic intensities the average message delay is either zero or low. This is followed by a rapid increase as the traffic intensifies, eventually giving way to a less rapid increase for high traffic intensities. This pattern remains the same for different configurations of models with full duplex links (Fig 6.8), half duplex links (Fig 6.9) and link failures (Fig 6.10). The pattern that emerges from these graphs resembles an "S", and is called a saturation curve. In a typical saturation curve, the first part of the curve, where there is little increase, corresponds to the priming of the system. Once there are enough entities in the system to affect the operation of the system, the steep part of the curve occurs. This continues until the system reaches its capacity. The curve then starts to level out, indicating that the system has reached its limit and cannot handle any more entities. The saturation curve is a natural phenomenon, seen not only in traffic theory, but electromagnetic theory as well¹.

¹ Cullwick E.G., *The Fundamentals of Electro-magnetism* (Cambridge: The University Press, 1966) p.198

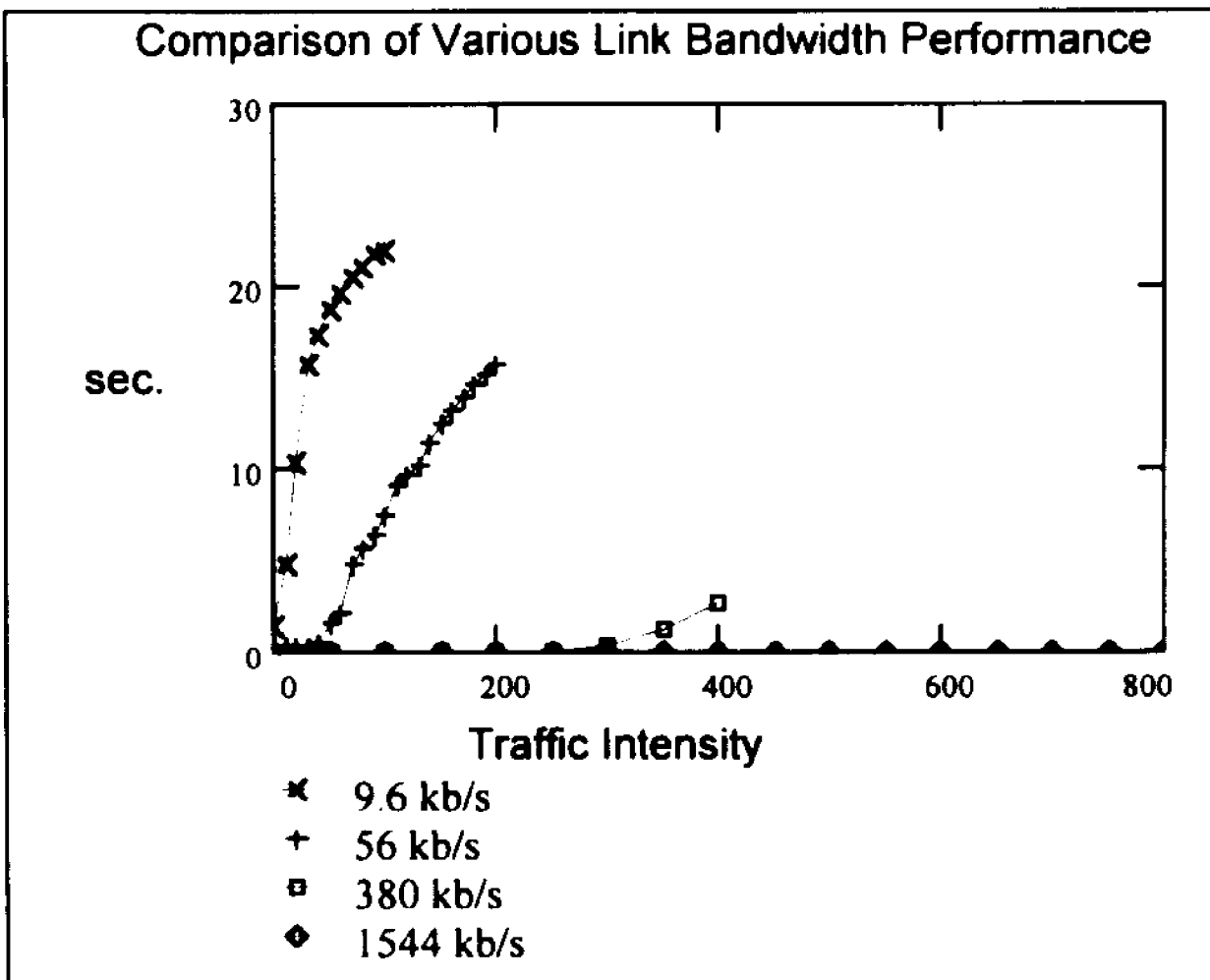


Figure 6-8 Average Message Delay for Full Duplex links

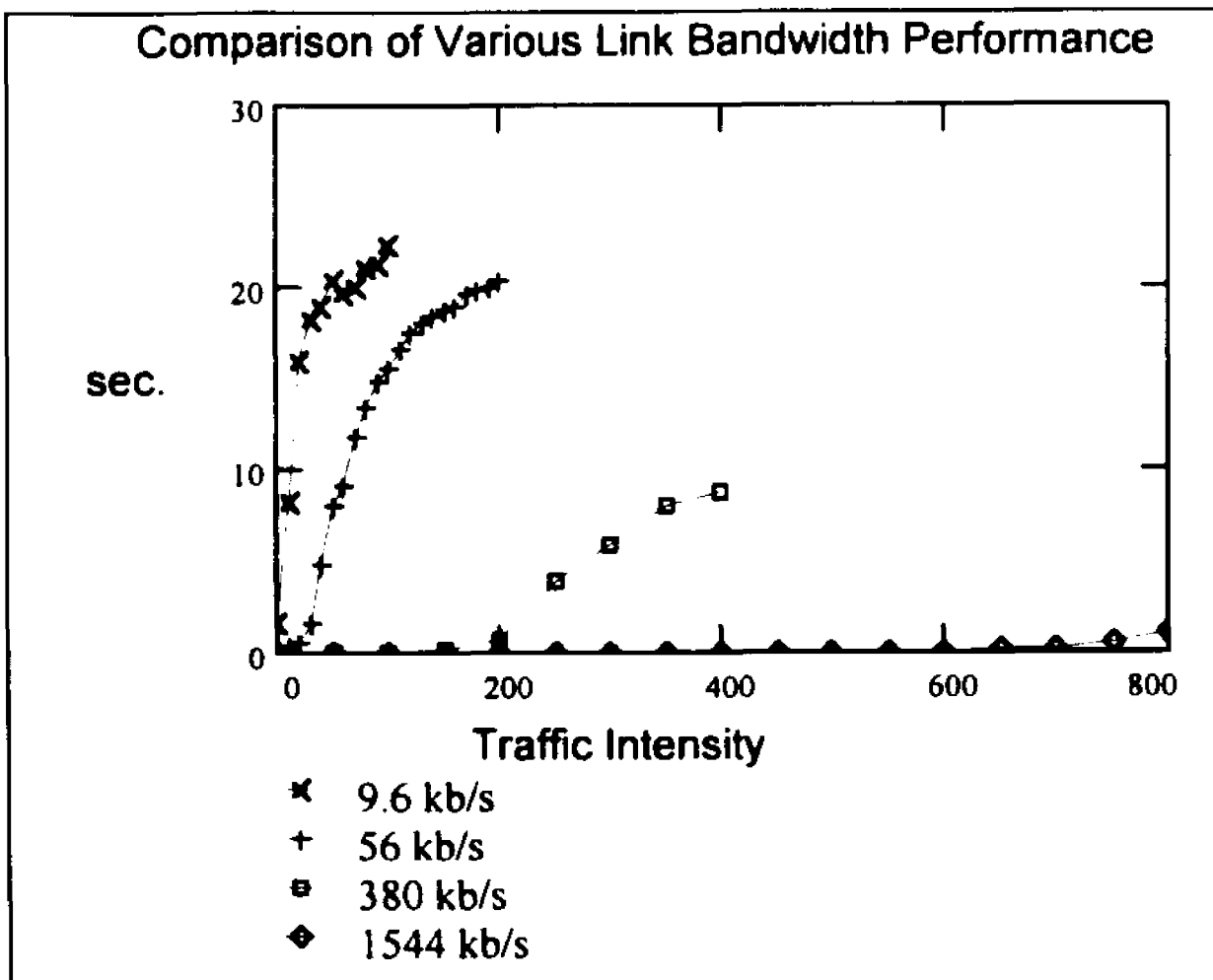


Figure 6 9 Average Message Delay for Half Duplex links

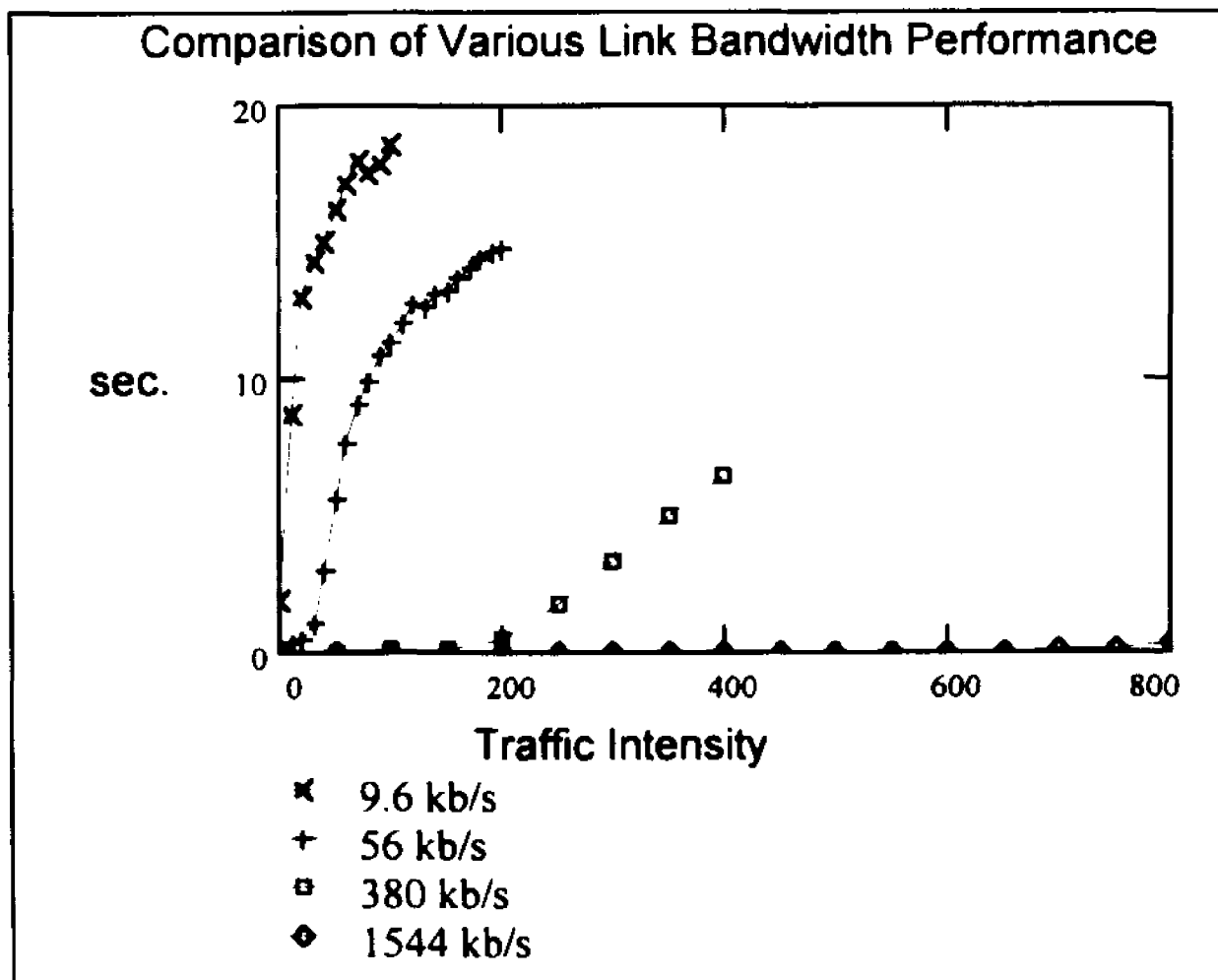


Figure 6 10 Average Message Delay for Link Failure

The delay saturation curves resulting from the simulation results indicate that for higher levels of traffic intensities the average message delay continues to increase approaching some asymptote at a very slow rate. It can be observed that for low bandwidths of 9.6 kb/s (Fig. 6.11) and 56 kb/s (Fig. 6.12) a steep increase takes place over a small range of traffic intensity, resulting in tall, thin curve. This indicates that high average delay times are reached at low traffic intensities. Conversely, for higher link bandwidths of 380 kb/s (Fig. 6.13) and 1544 kb/s (Fig. 6.14) high average message delays do not occur until higher traffic intensities. It then can be concluded that networks with higher data rate links have better average message delay times than networks with lower speed links.

For a network it is desirable to have an elongated delay curve, which means that the average message delay increases gradually as traffic intensities increase. The areas of interest in this curve are the two points of inflection, at which the curve bends away from its previous course. The first is at a point where the delay is no longer gradual (and linear), and starts to increase at a more rapid rate than before. The second point is where this steep increase gives way to a more gradual increase.

It is also desirable to compare the average message delay curves for different link bandwidths, in order to study the network behavior under varying conditions. For all bandwidths of 9.6 kb/s (Fig. 6.11), 56 kb/s (Fig. 6.12), 380 kb/s (Fig. 6.13) and 1544 kb/s (Fig. 6.14), it can be seen that the network with the full duplex links has the best average message delay times. The first point of inflection comes at a later traffic intensity than the half duplex and failed link version of the network (approximately at twice the traffic intensity of half duplex); and the increase after the first point is gradual; the second point of inflection occurs at a higher traffic intensity than both the half duplex and link failure networks.

When comparing a half duplex linked network to a network with a link failure it is clear that the half duplex network is a poor performer. Though the first delay point occurs at about the same traffic intensity, the increase in average message delay between the two points is steeper in the half duplex network. This, of course, results in a taller and thinner "S", making that type of network the least desirable of the three.

Average message delay is lower for a full duplex linked network than a network with a link failure. However, as the traffic intensity increases the average message delay for the full duplex linked network increases and surpasses its counterpart for a link failed network. This happens because there are fewer messages being transferred within the link failed network than the full duplex one. The rerouted messages cause congestion at the nodes along the reroute. This causes other messages passing through these nodes to be delayed more than usual. As a result, that reduces the load on the links and nodes not along the reroute, thus lowering the overall average message delay.

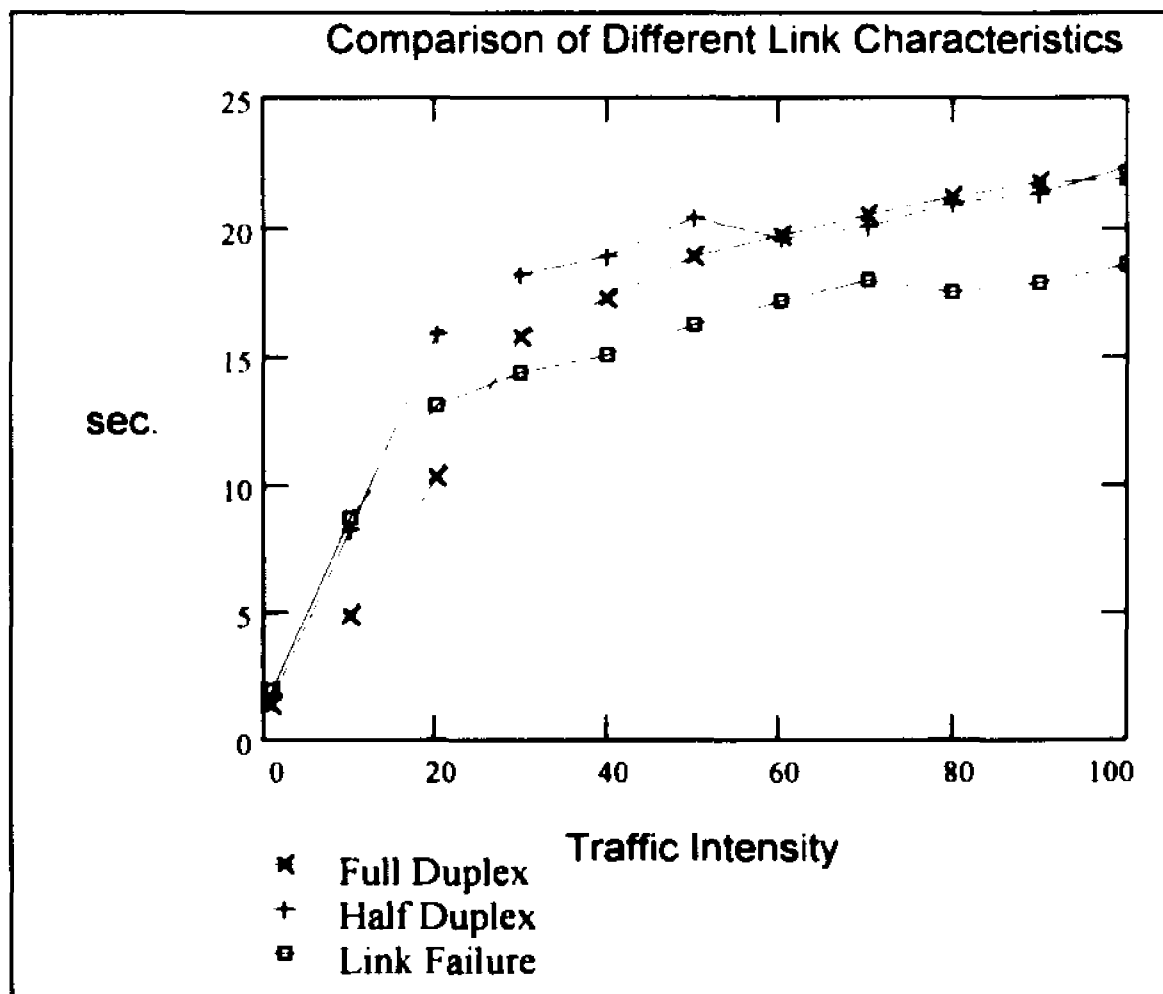


Figure 6.11 Average Message Delay for 9.6 kb/s links

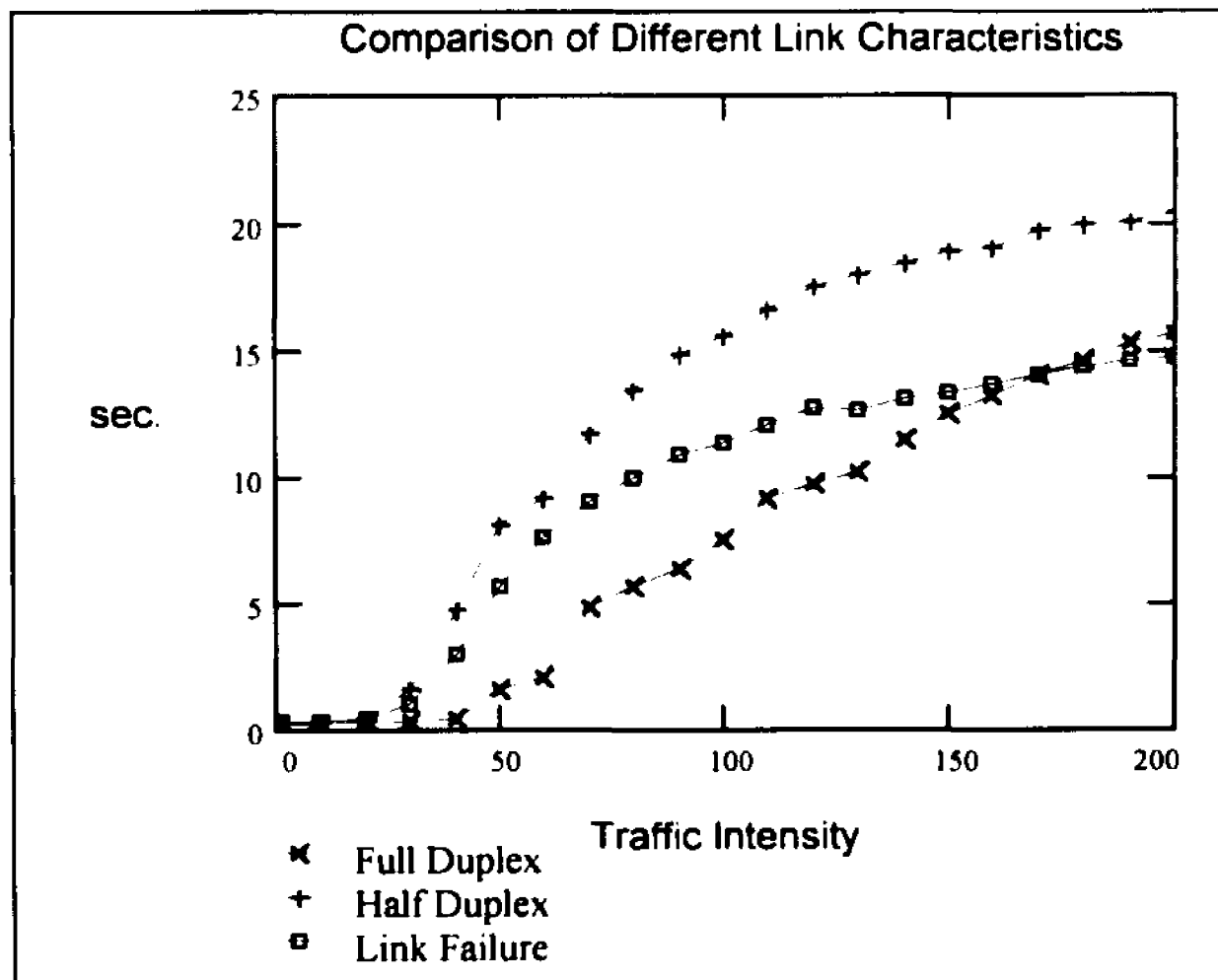


Figure 6.12 Average Message Delay for 56 kb/s links

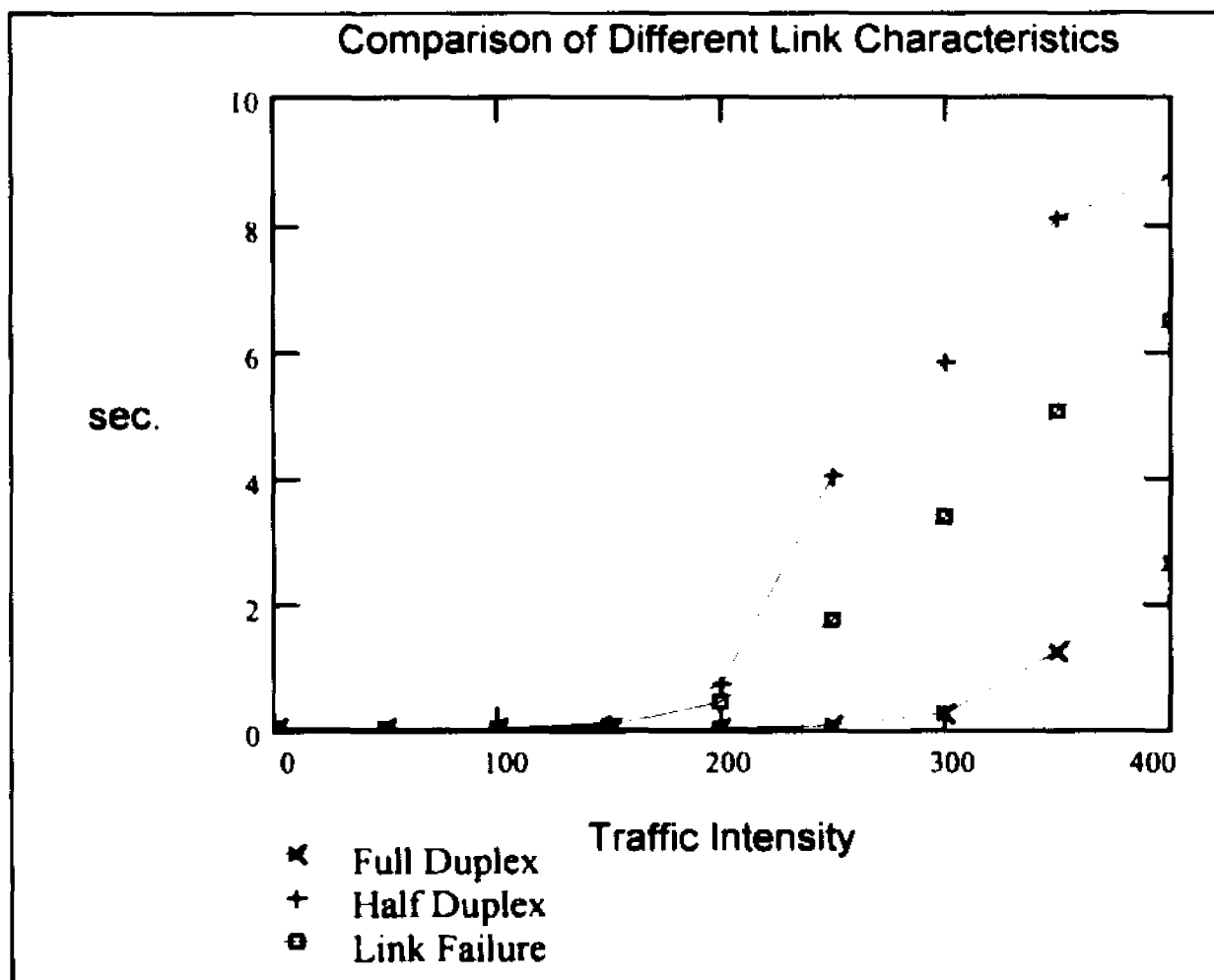


Figure 6 13 Average Message Delay for 380 kb/s links

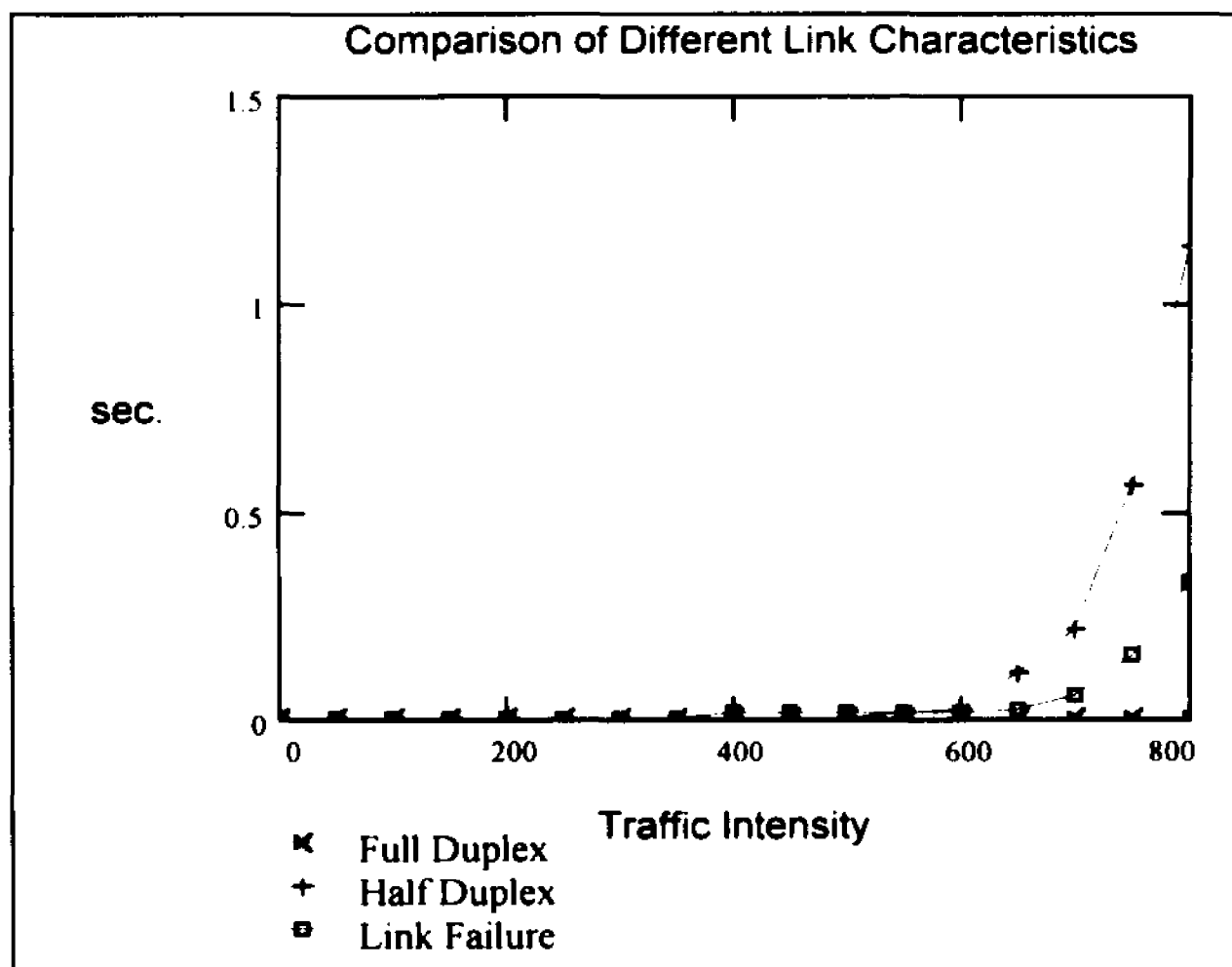


Figure 6.14 Average Message Delay for 1544 kb/s links

6.1.2.2 Maximum Message Delay

Maximum message delay indicates the maximum amount of time for a message to be delivered from its source node to its destination node, and is considered a worst-case indicator. The curves associated with this delay are similar to those found in the results of average message delay, where the curve is very gradual at first then, at some point, begins to increase rapidly up to a second point where it starts to level off.

When comparing the curves of different link speeds for the full duplex linked network (Fig. 6.15), half duplex linked network (Fig. 6.16) and a network with a link failure (Fig. 6.17), one observes the familiar "S" curve, with link rates reaching peak maximum message delay times at lower traffic intensities than the link rates.

Comparison of half duplex, full duplex and link failure networks at different link bandwidths of 9.6 kb/s (Fig. 6.18), 56 kb/s (Fig. 6.19), 380 kb/s (Fig. 6.20) and 1544 kb/s (Fig. 6.21), demonstrates that the full duplex linked network has the smallest maximum message delay of the three at all traffic intensities. While the half duplex delay time is greater than the link failed network at lower traffic intensities, this situation then reverses itself, with the link failed network exhibiting larger maximum message delay times. This situation occurs due to the congestion caused by the failed link. As mentioned in the previous section, the nodes and links along the reroute can not handle the overflow of messages, causing high maximum message delays.

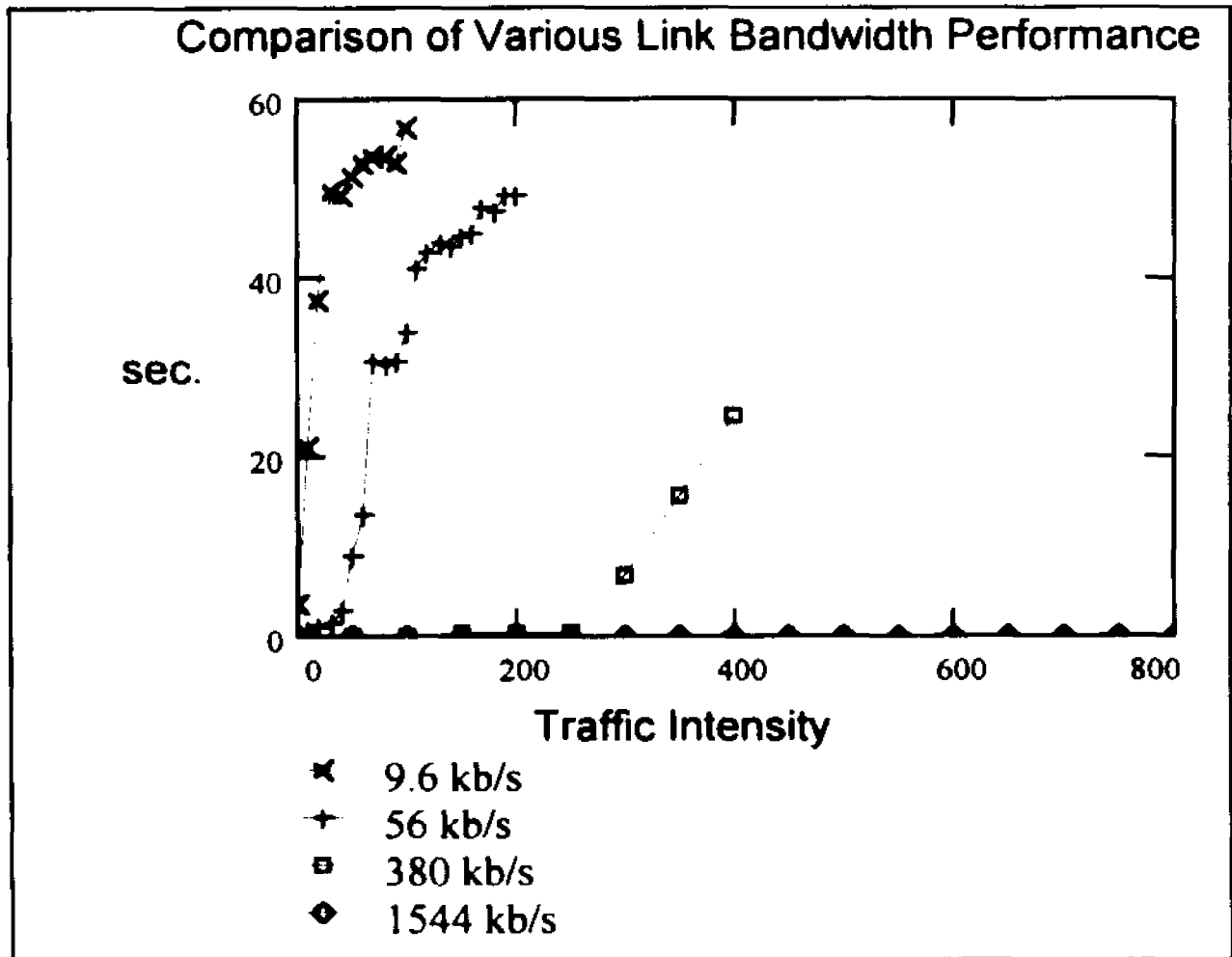


Figure 6.15: Maximum Message Delay for Full Duplex links

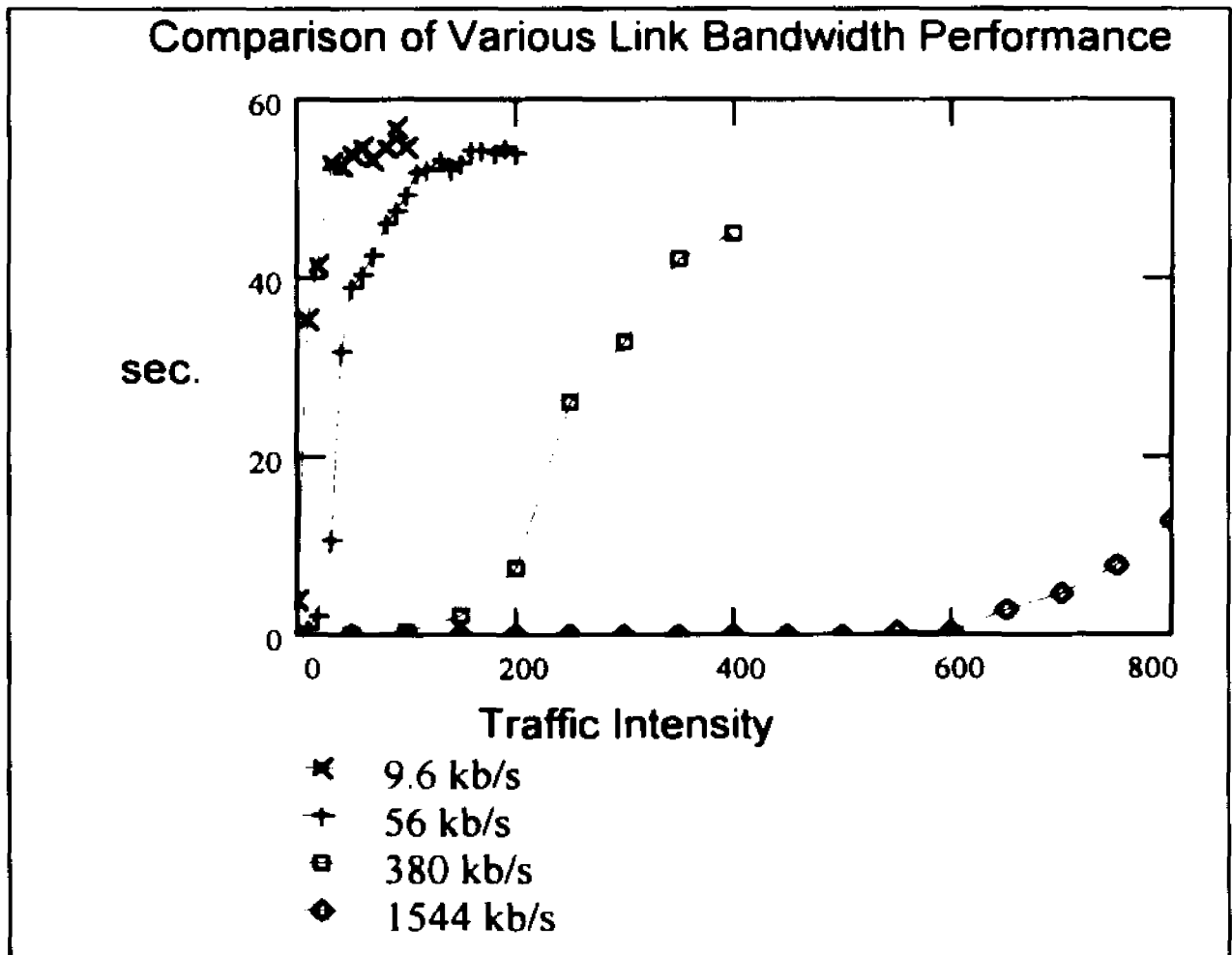


Figure 6.16: Maximum Message Delay for Half Duplex links

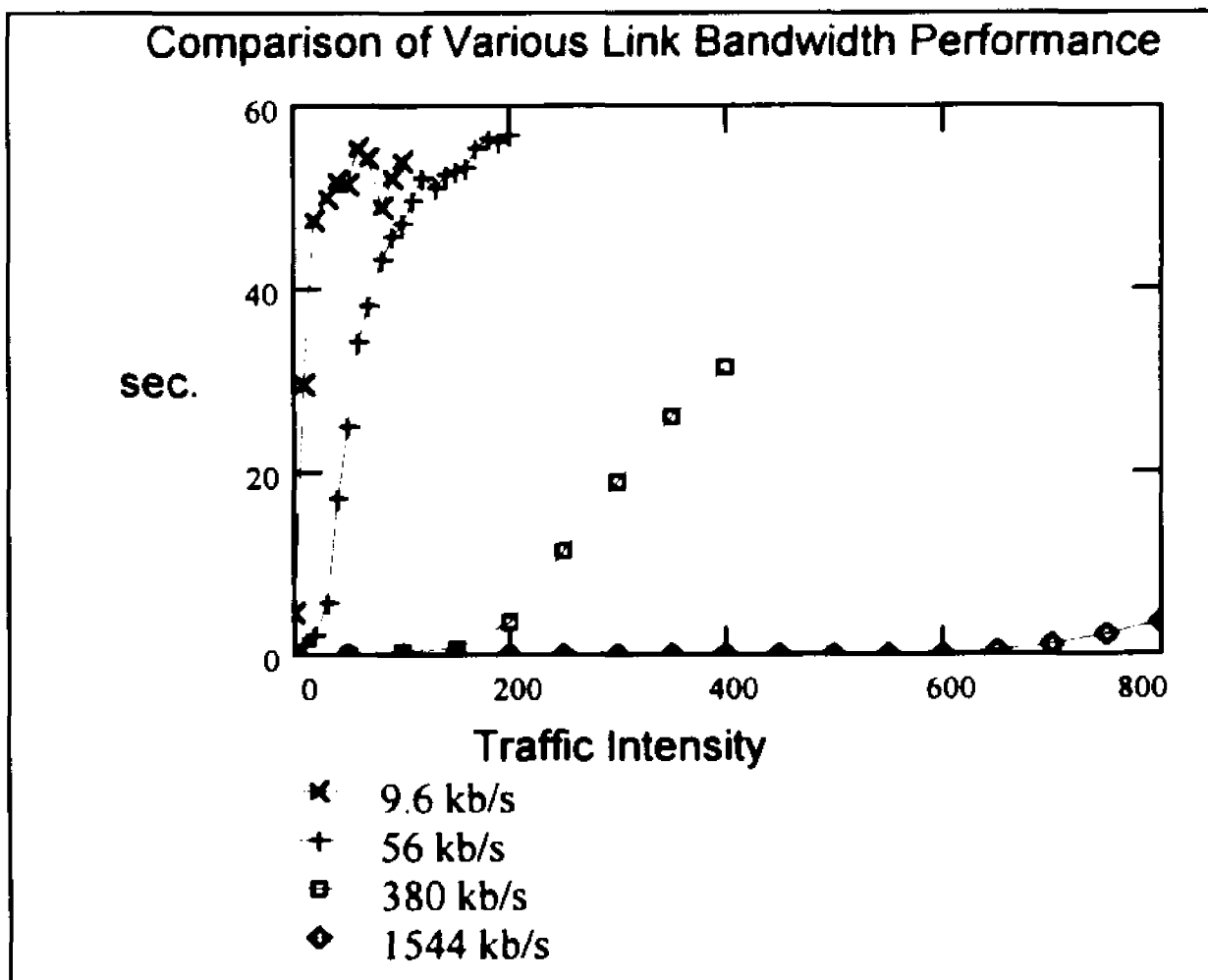


Figure 6.17: Maximum Message Delay for Link Failure

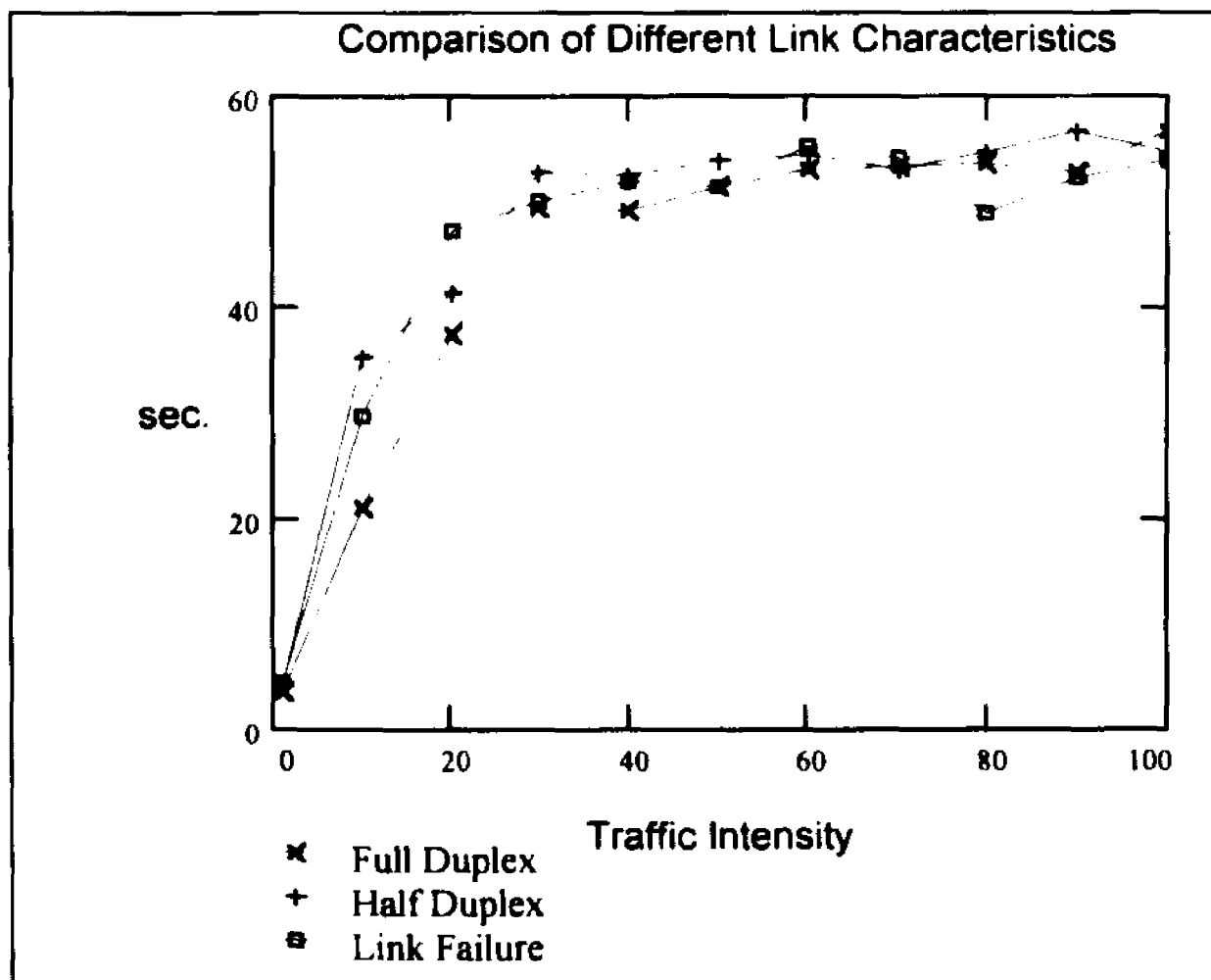


Figure 6.18: Maximum Message Delay for 9.6 kb/s links

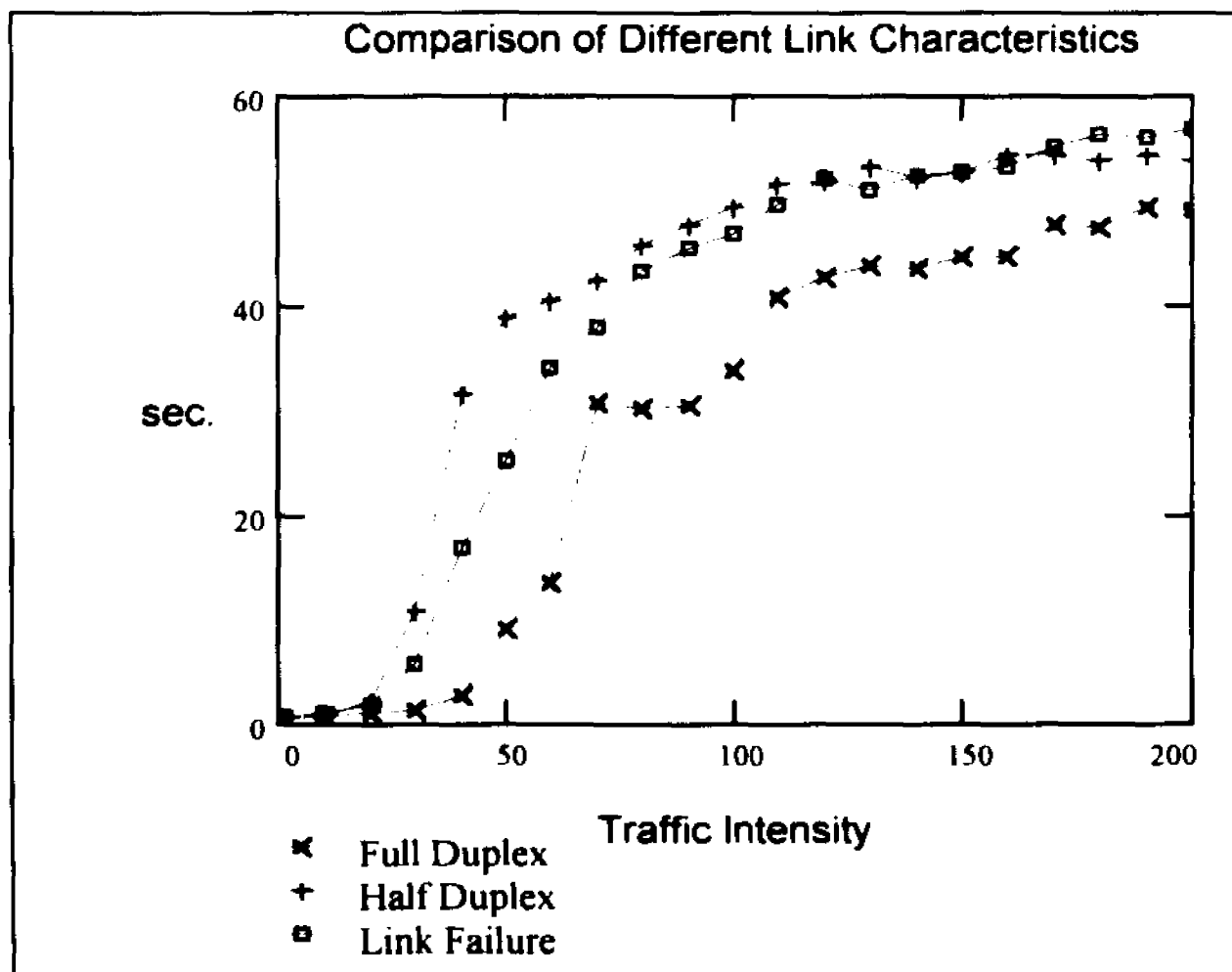


Figure 6.19. Maximum Message Delay for 56 kb/s links

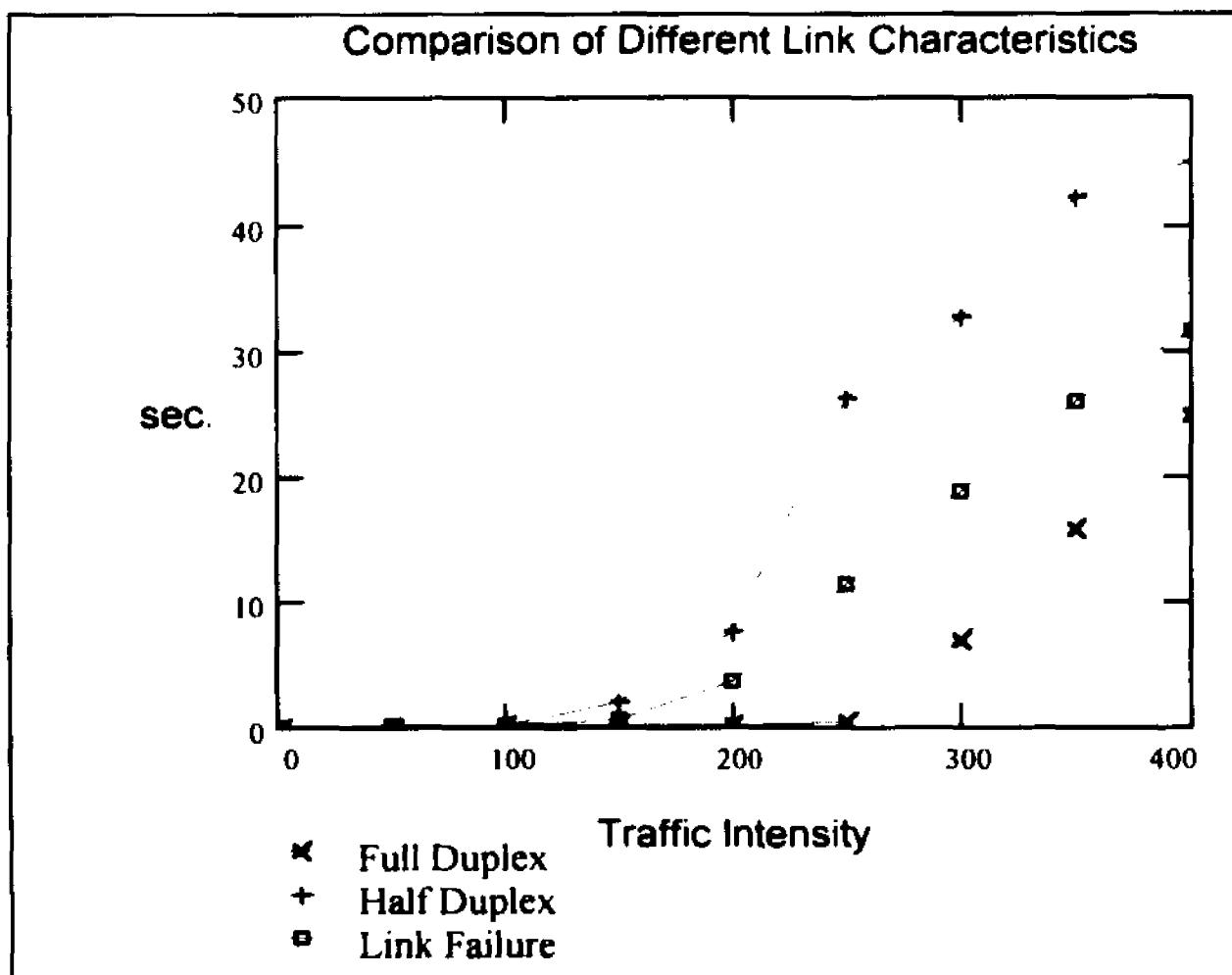


Figure 6.20 Maximum Message Delay for 380 kb/s links

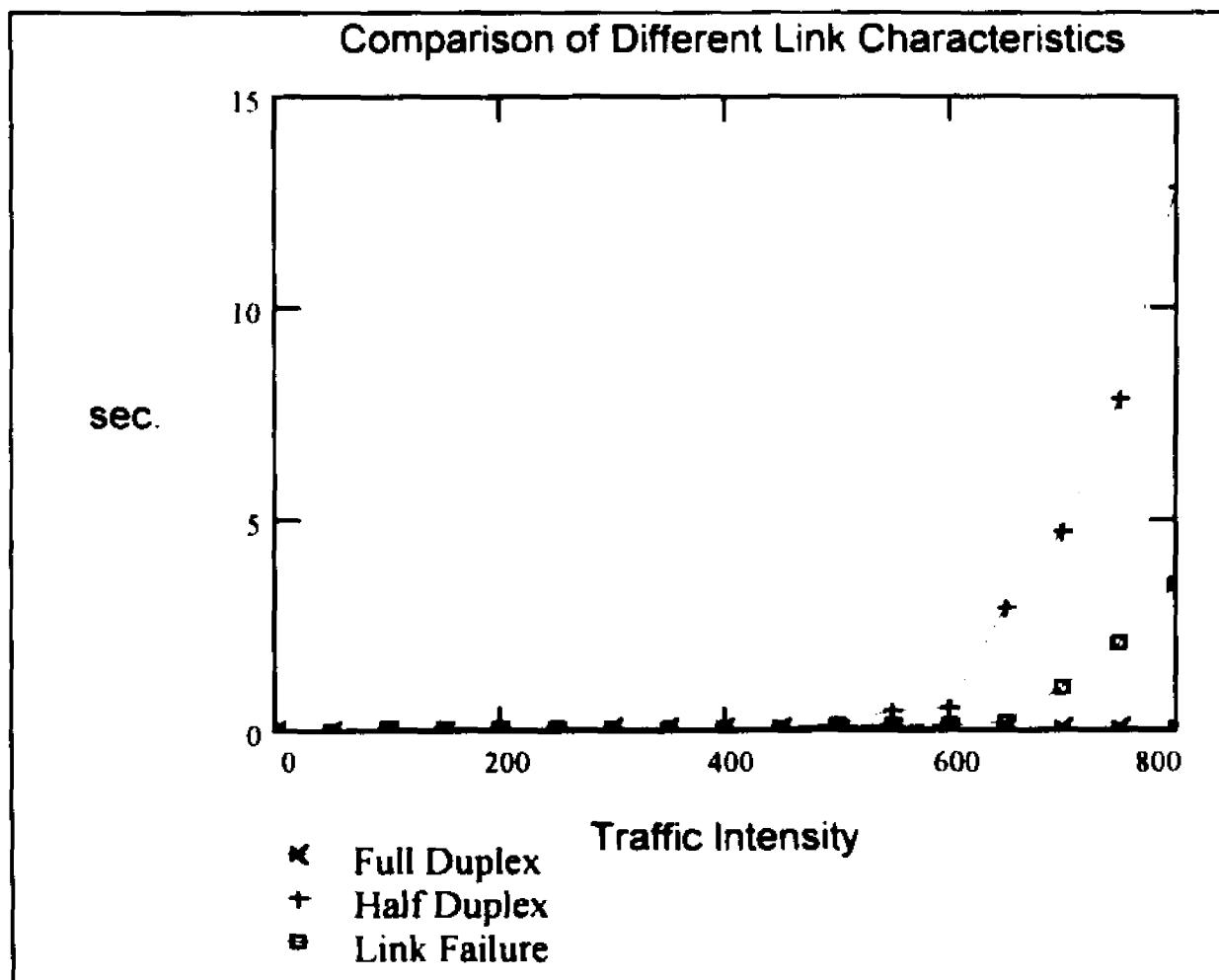


Figure 6.21. Maximum Message Delay for 1544 kb/s links

6.1.3 Percent of Messages over Three Seconds

The results for the percent of messages over three seconds statistic, which is a measure of a network's QoS, also produced saturation curves. These curves show that, initially, the full duplex linked (Fig. 6.22) and half duplex linked (Fig. 6.23) networks reach higher percentages at larger traffic intensities than link failed (Fig. 6.24) networks. This is due to fewer messages that can be processed by the network due to the congestion that the failed link causes, thereby lightening the load of the other links. The curves also show that for each of the three different types of networks, the saturation level remains the same regardless of the link bandwidths.

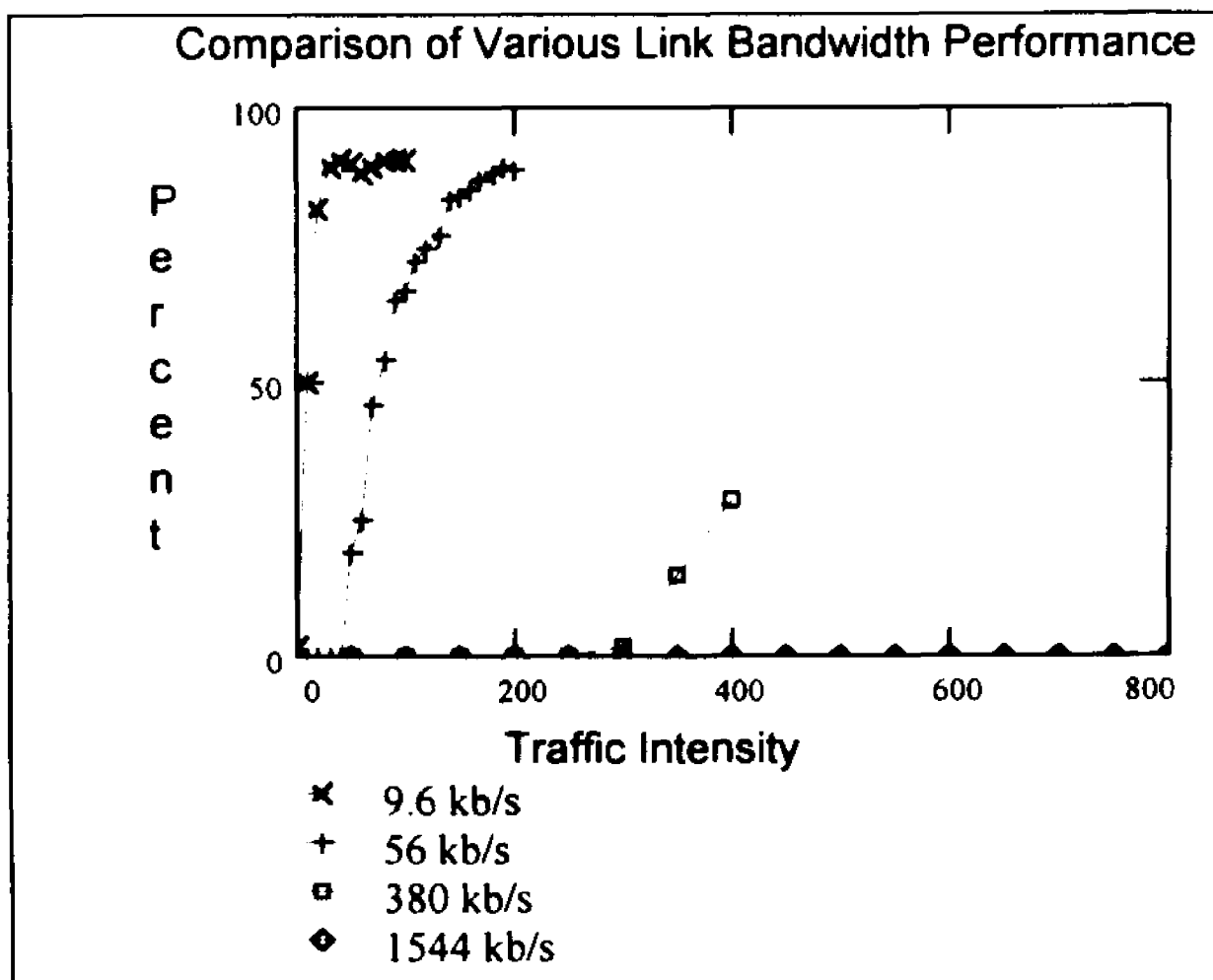


Figure 6.22: Percent of Messages over 3 sec. for Full Duplex links

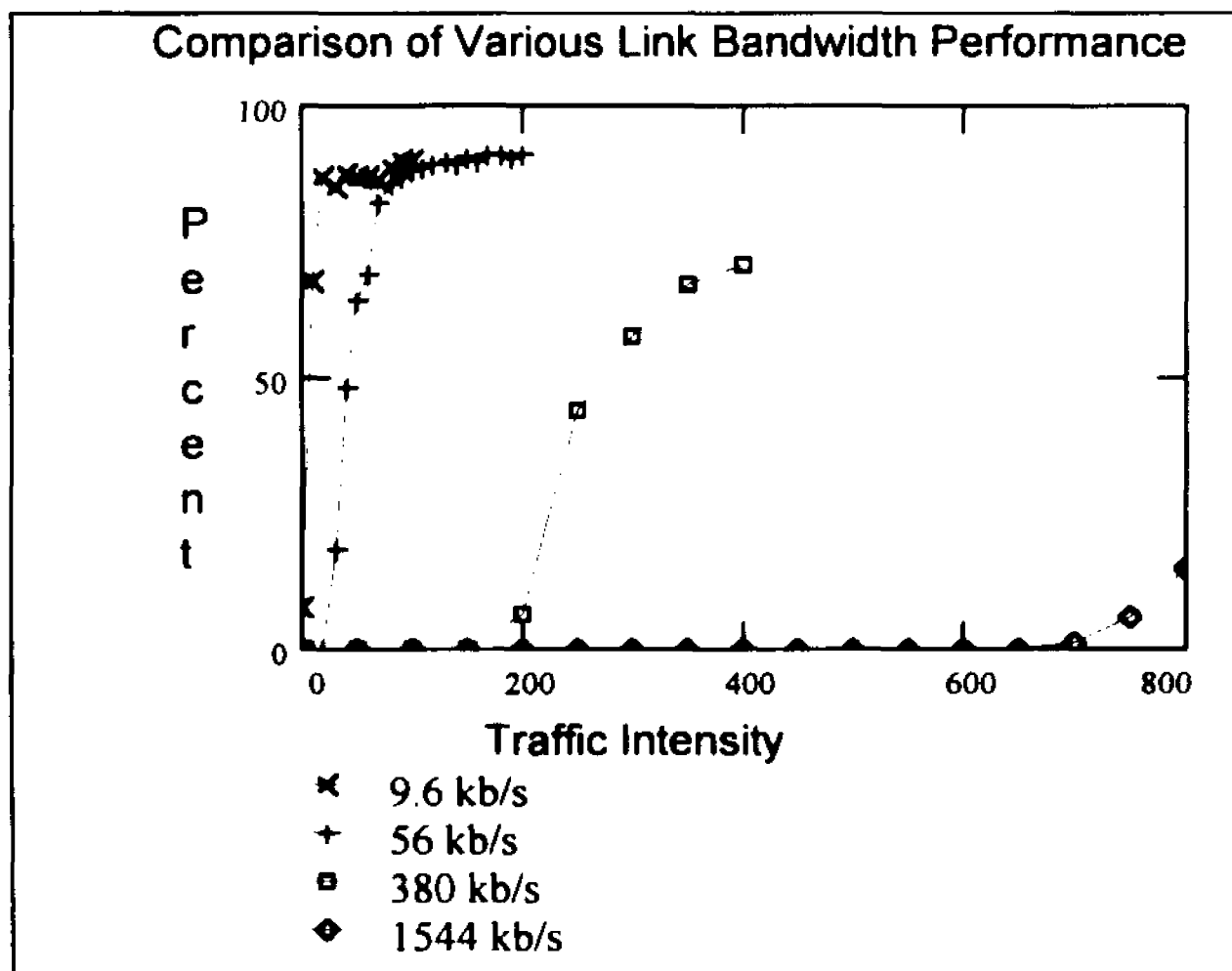


Figure 6.23. Percent of Messages over 3 sec. for Half Duplex links

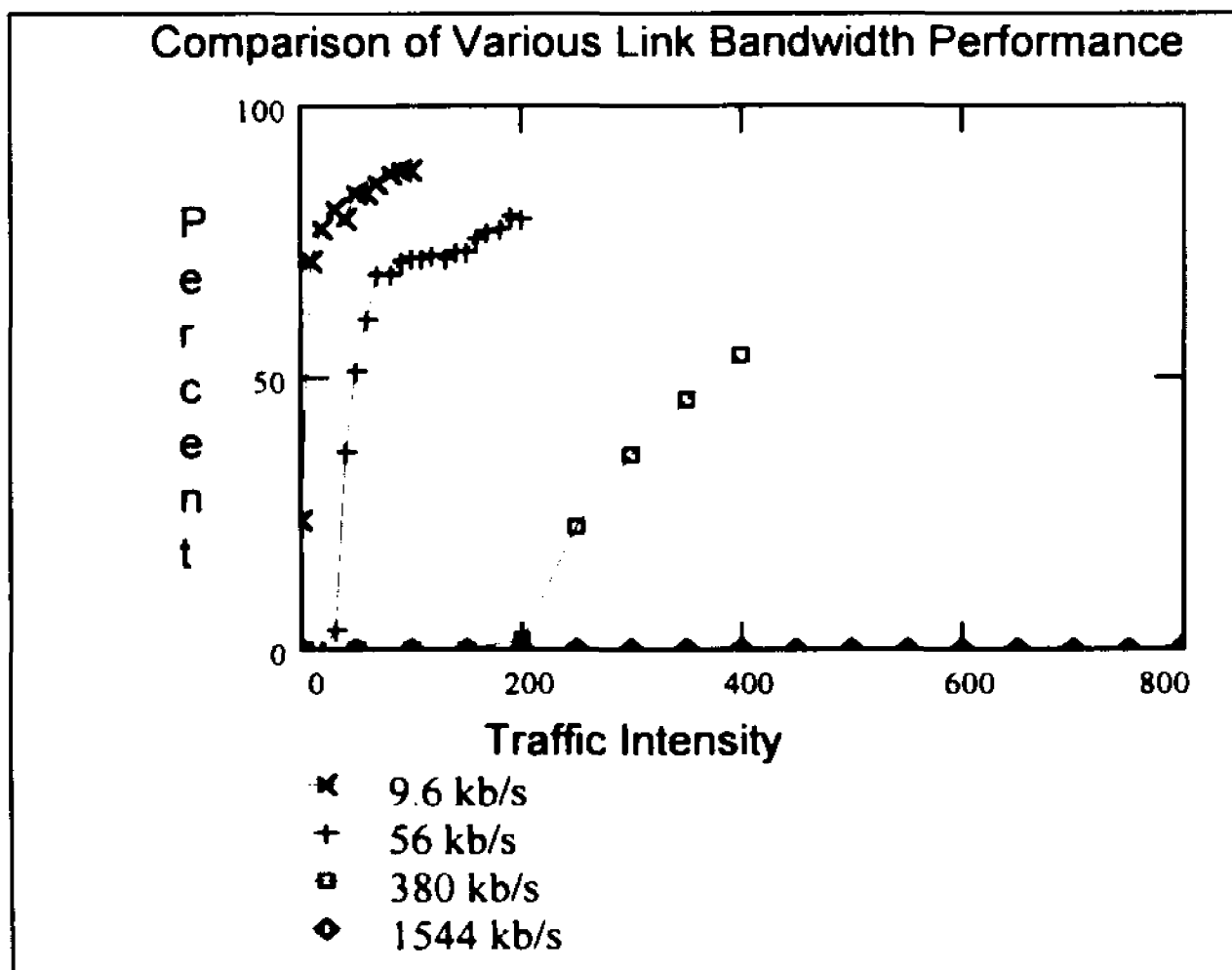


Figure 6.24 Percent of Messages over 3 sec for Link Failure

Further analysis of the data indicate that the percentages all types of network converge or start to converge for 9.6 kb/s (Fig. 6.25) and 56 kb/s (Fig. 6.26) links. This indicates that there is a common saturation point, but that different types of network arrive to that point differently. For the higher bandwidth links of 380 kb/s (Fig. 6.27) and 1544 kb/s (Fig. 6.28), the large increase in the percentage is preceded by a small increase. This small increase may be used to predict the imminent large increase.

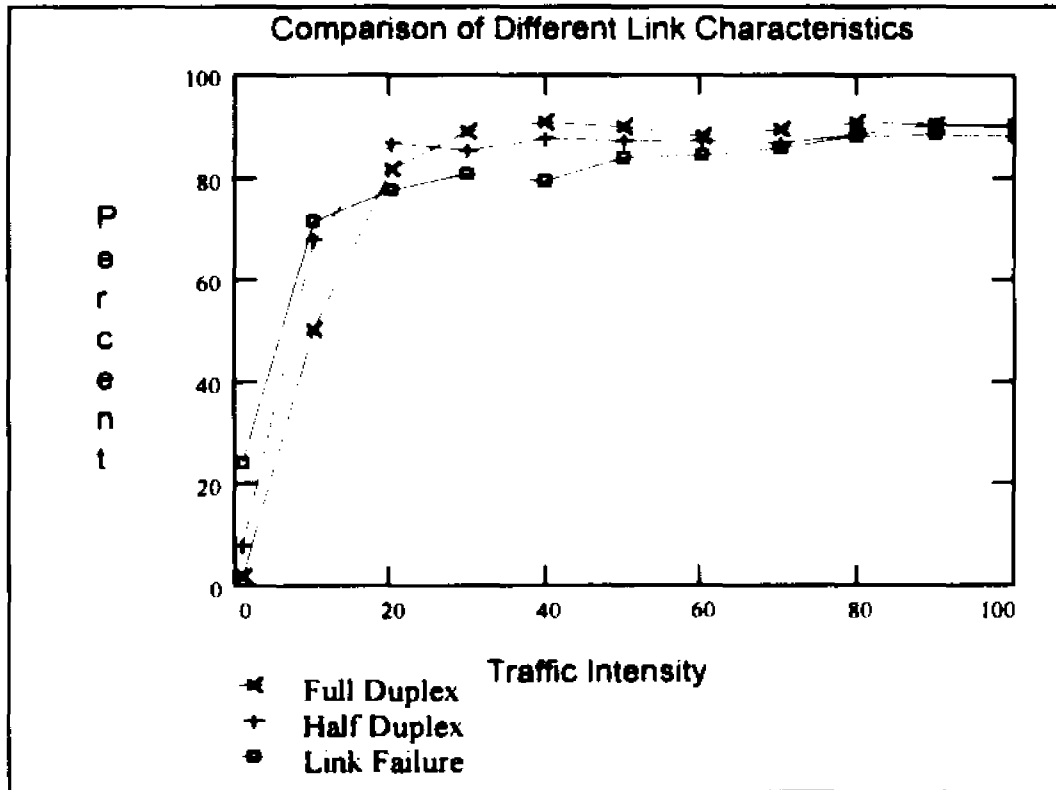


Figure 6.25. Percent of Messages over 3 sec for 9.6 kb/s links

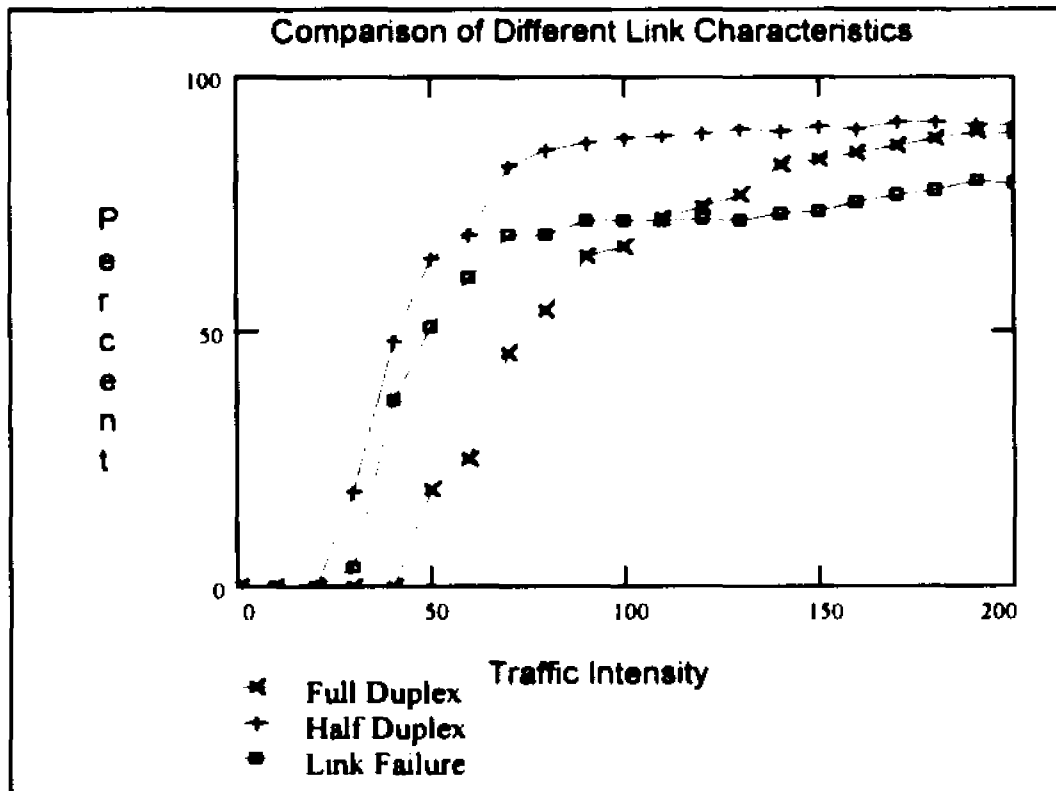


Figure 6.26. Percent of Messages over 3 sec for 56 kb/s links

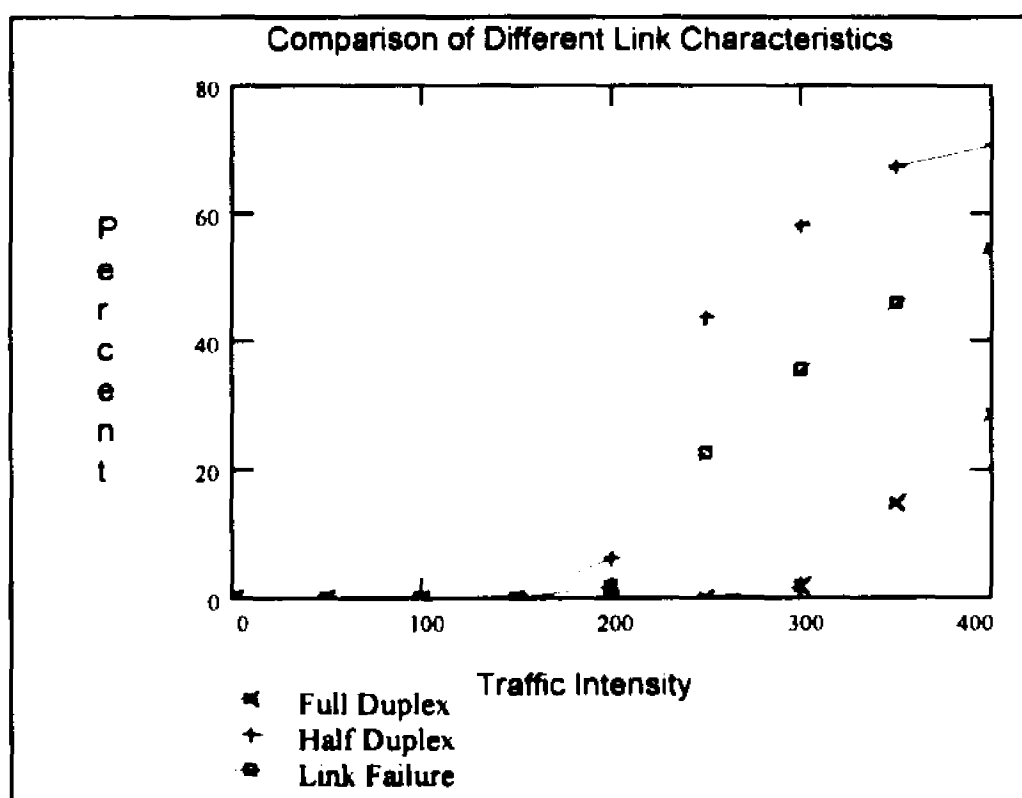


Figure 6.27. Percent of Messages over 3 sec for 380 kb/s links

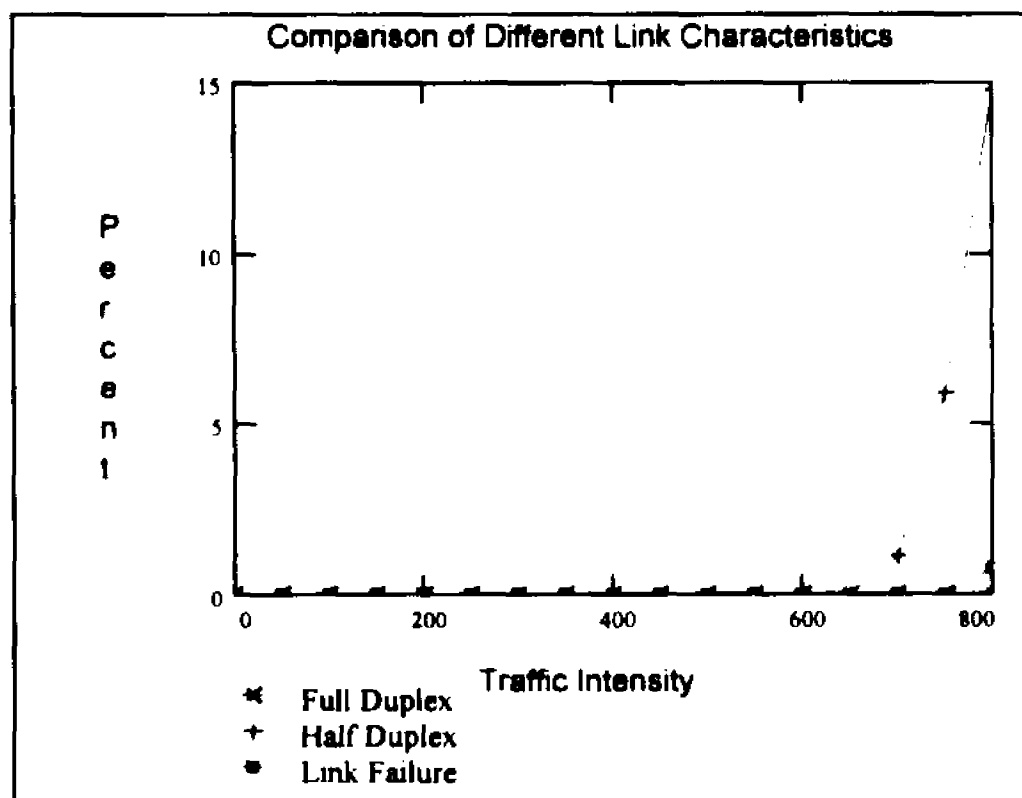


Figure 6.28. Percent of Messages over 3 sec for 1544 kb/s links

6.2 Network Results for the Second Simulation Model

The second iterative simulation model demonstrates that, not only the link bandwidth, but also the number of nodes affect throughput, delay times, and Quality of Service.

The model has six more nodes and six more links. The 100% increase in nodes over the previous model also increases the number of virtual circuits that can be setup between those nodes by 102 to a total of 132. Using the same traffic parameters for a virtual circuit as the previous model, the second simulation model has over four times the amount of traffic as the first simulation model at the same traffic intensity.

6.2.1 Throughput

This results from the model indicate that but the number of nodes in a network affect the throughput of the network. The increased number of nodes in the second model have resulted in larger throughput figures than the first model. This is due to the larger number of virtual circuits set up between the nodes, thereby increasing the traffic in the network.

The linear increase for throughput is still present confirming the previous results obtained from the first model. The full duplex linked (Fig. 6.29) network model produces higher throughput for all link bandwidths than link failure (Fig. 6.30) network model. The point of separation, however, occurs at lower traffic intensities and higher throughput rates than the first model. This result indicates that larger networks reach their throughput limits at lower levels of traffic intensities than smaller networks.

If we considered link failure, the absence of a link, as another topology, we can see how another link improves the throughput of a network. It can be observed that the point of separation for link failed network takes place at smaller traffic intensities than the full duplex network. This implies that the intensity of traffic where the point of separation takes place is related to the topology of the network.

Consequently, since a link failure is similar to removal of a link in a network, the drop in the throughput is proportional to the amount of data (which corresponds to the number of virtual circuits going through that link) that has to be diverted to other links. If the network has many nodes but few links, then it is likely that it will have many more virtual circuits, thus making a link failure more detrimental than for a smaller network. If, however, the network has many links between nodes that can handle the extra traffic, then the effect of a link failure may be minimal.

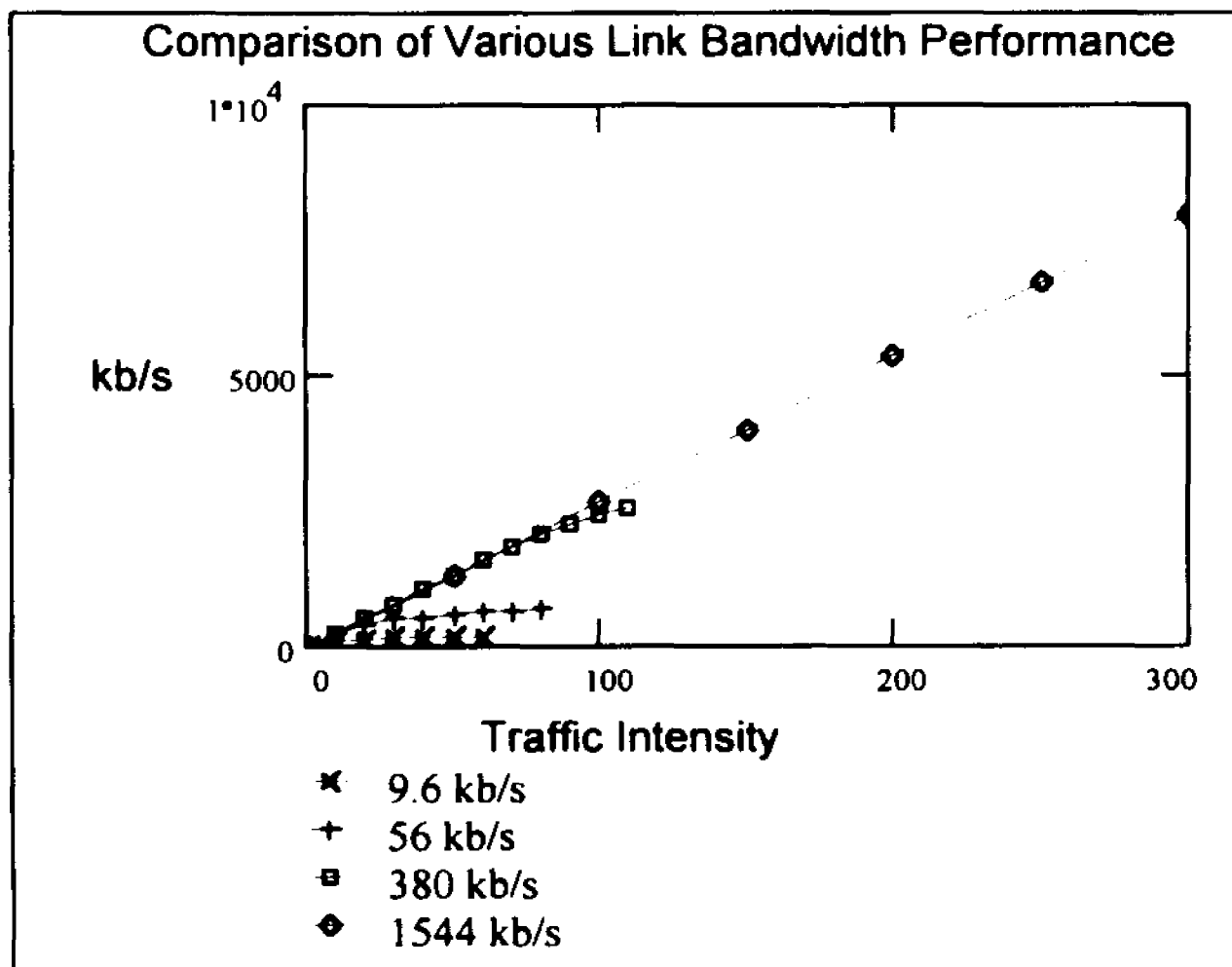


Figure 6.29: Throughput for Full Duplex links

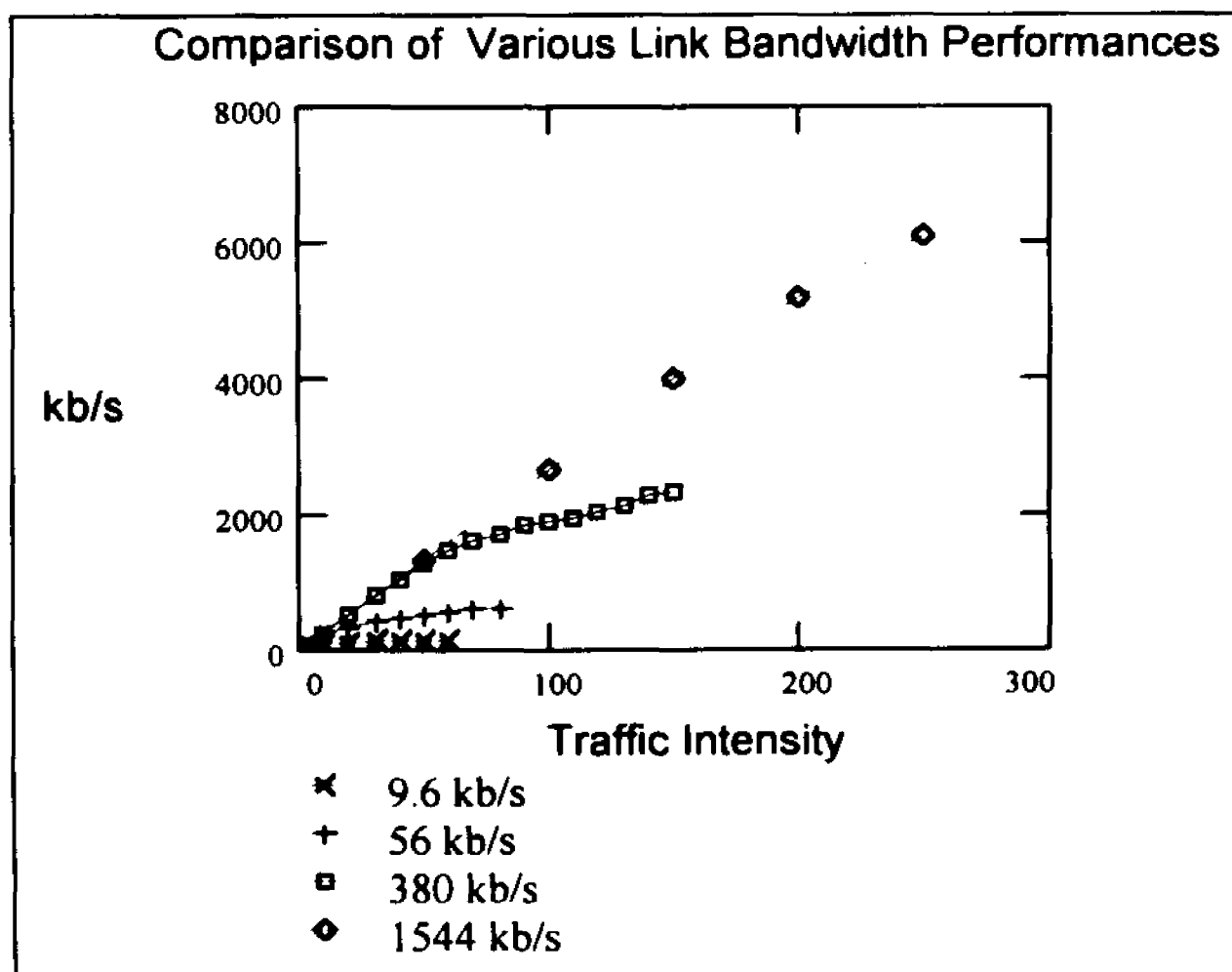


Figure 6.30. Throughput for Link Failure

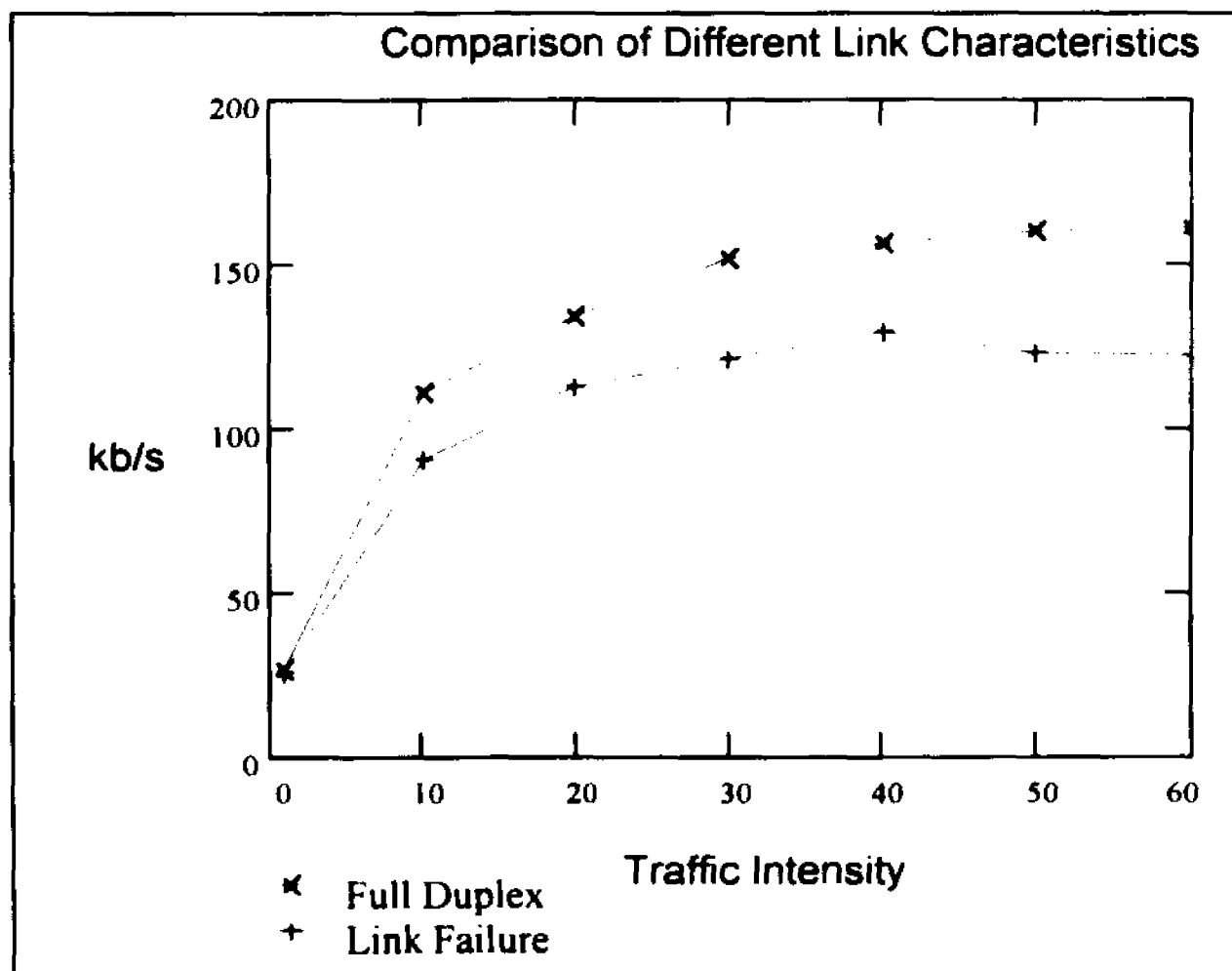


Figure 6.31. Throughput for 9.6 kb/s links

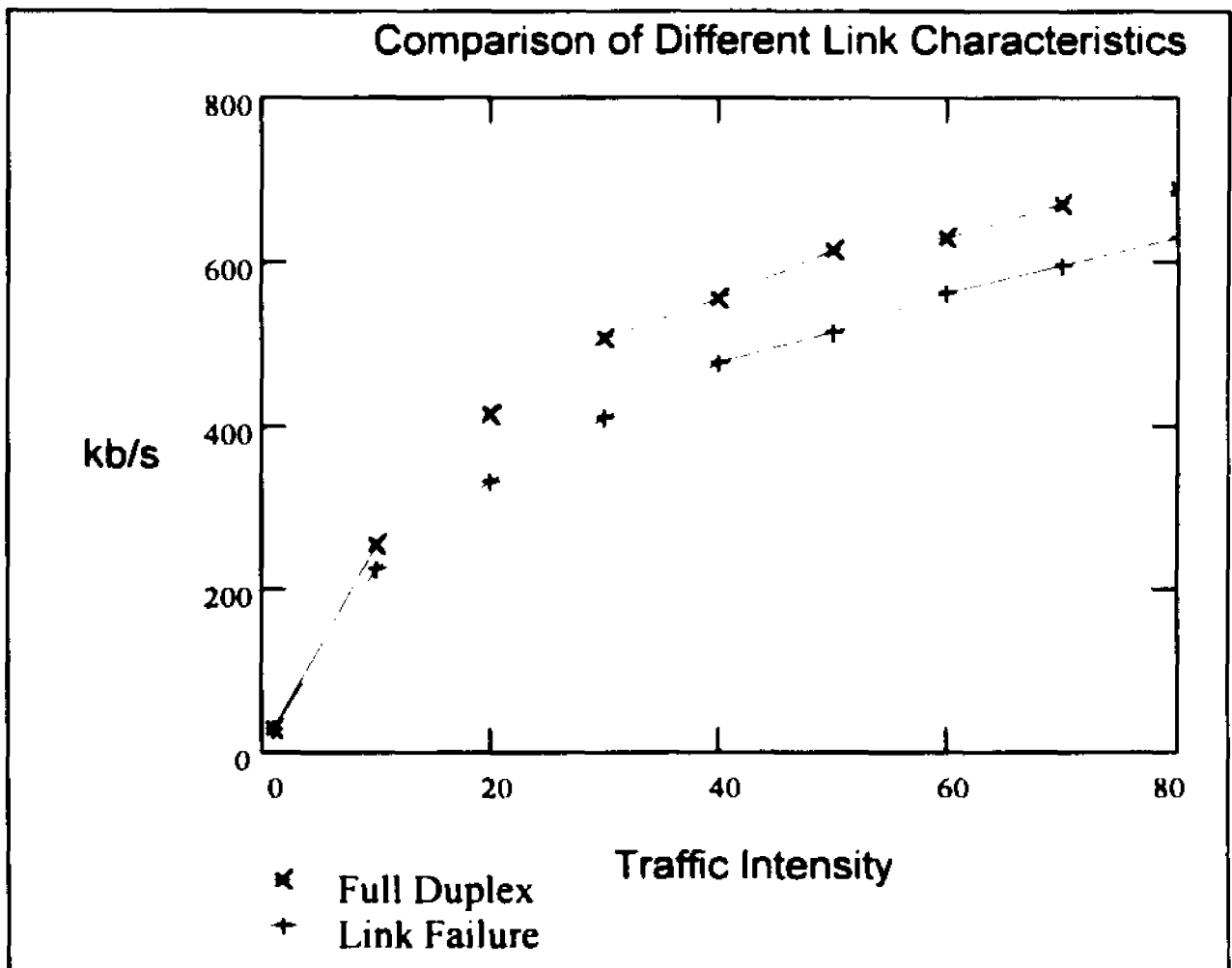


Figure 6.32: Throughput for 56 kb/s links

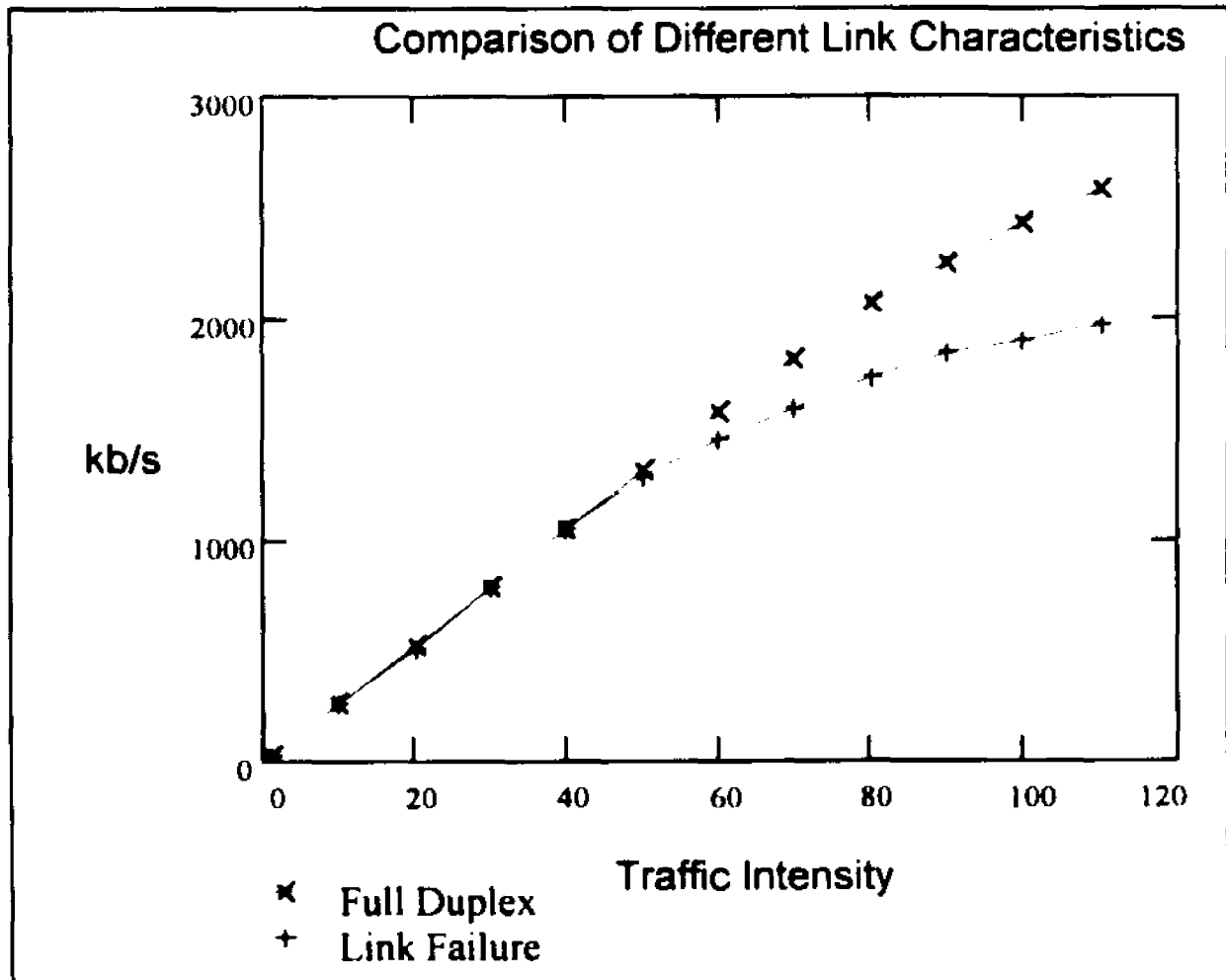


Figure 6 33 Throughput for 380 kb/s links

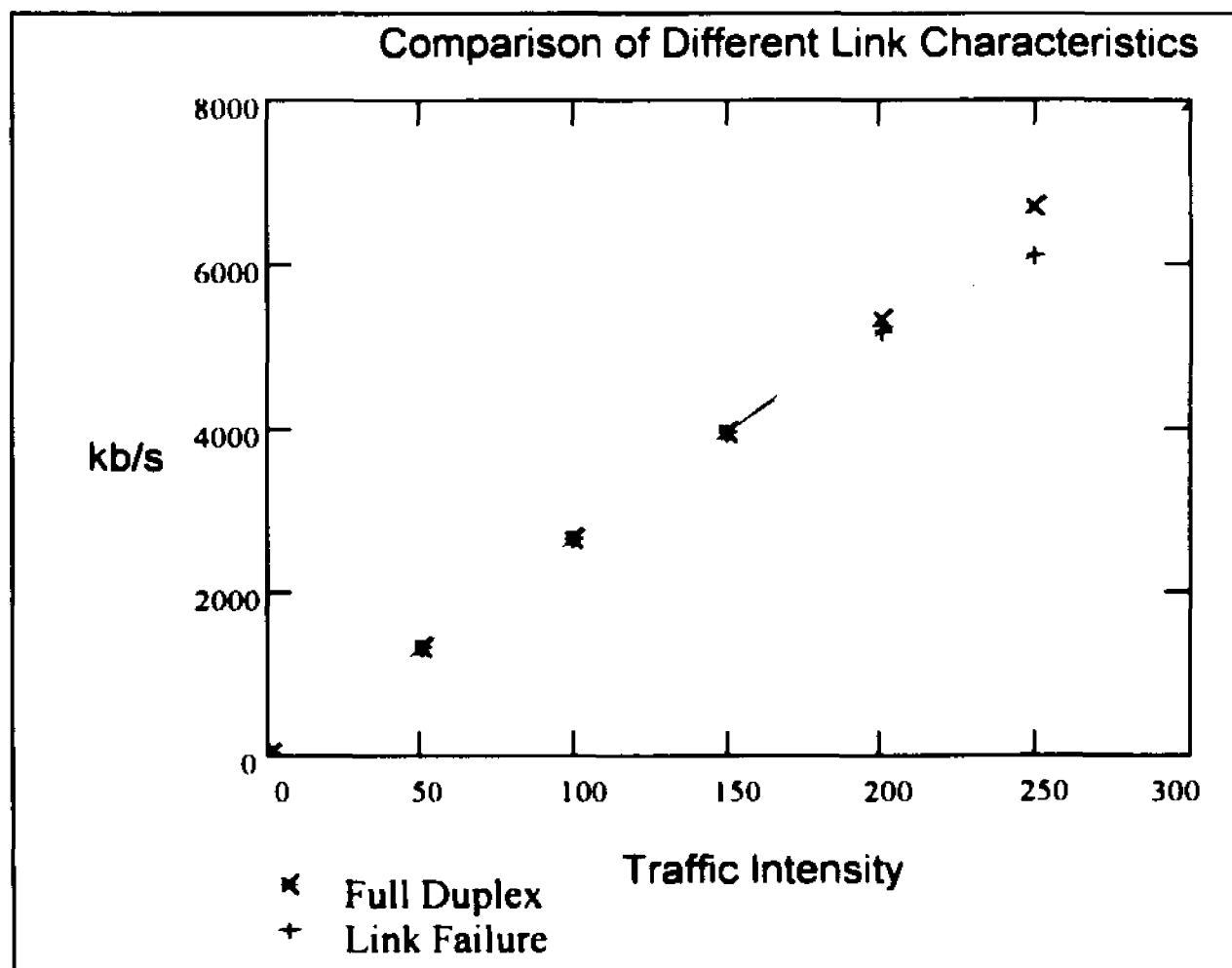


Figure 6.34 Throughput for 1544 kb/s links

6.2.2 Message Delay

6.2.2.1 Average Message Delay

The Average Message Delay results produced by second simulation model are consistent with the first model, with full duplex (Fig. 6.35) linked network registering a higher delay times than the failed link (Fig. 6.36) network. These delay times are not significantly different than the ones obtained from the first model. The reason for this is the increase in the number of links. Even though there are more nodes - therefore more virtual circuits - to produce more traffic, this extra traffic can be handled by the increased number of links serving the nodes. The results from the simulations show that if the number of nodes are doubled, the number of links must be squared to attain similar Average Message Delay times.

Comparing Average Message Delay times of a full duplex network and a link failed network, it can be observed that the difference in delay times for low bandwidth links of 9.6 kb/s (Fig. 6.37) and 56 kb/s (Fig. 6.38) is very small. The crossover point, where the delay for link failed network becomes less than the delay for the full duplex network, occurs at lower traffic intensities than its counterpart in the first model.

For the higher bandwidth links of 380 kb/s (Fig. 6.39) and 1544 kb/s (Fig. 6.40), the results show that the initial increase in the delay time occurs at a smaller traffic intensity than the previous model. This is due to the increased traffic precipitated by increased nodes

Finally, combining the results of the two simulation models, it can be said that, though it is possible to reduce Average Message Delay of a small network by removing a link, for larger networks with more links, this reduction can be small. Since the larger network has more links to ease the congestion caused by the link failure there are fewer bottlenecks, which in turn allow the network to operate at comparable delay rates with a network without any link failures

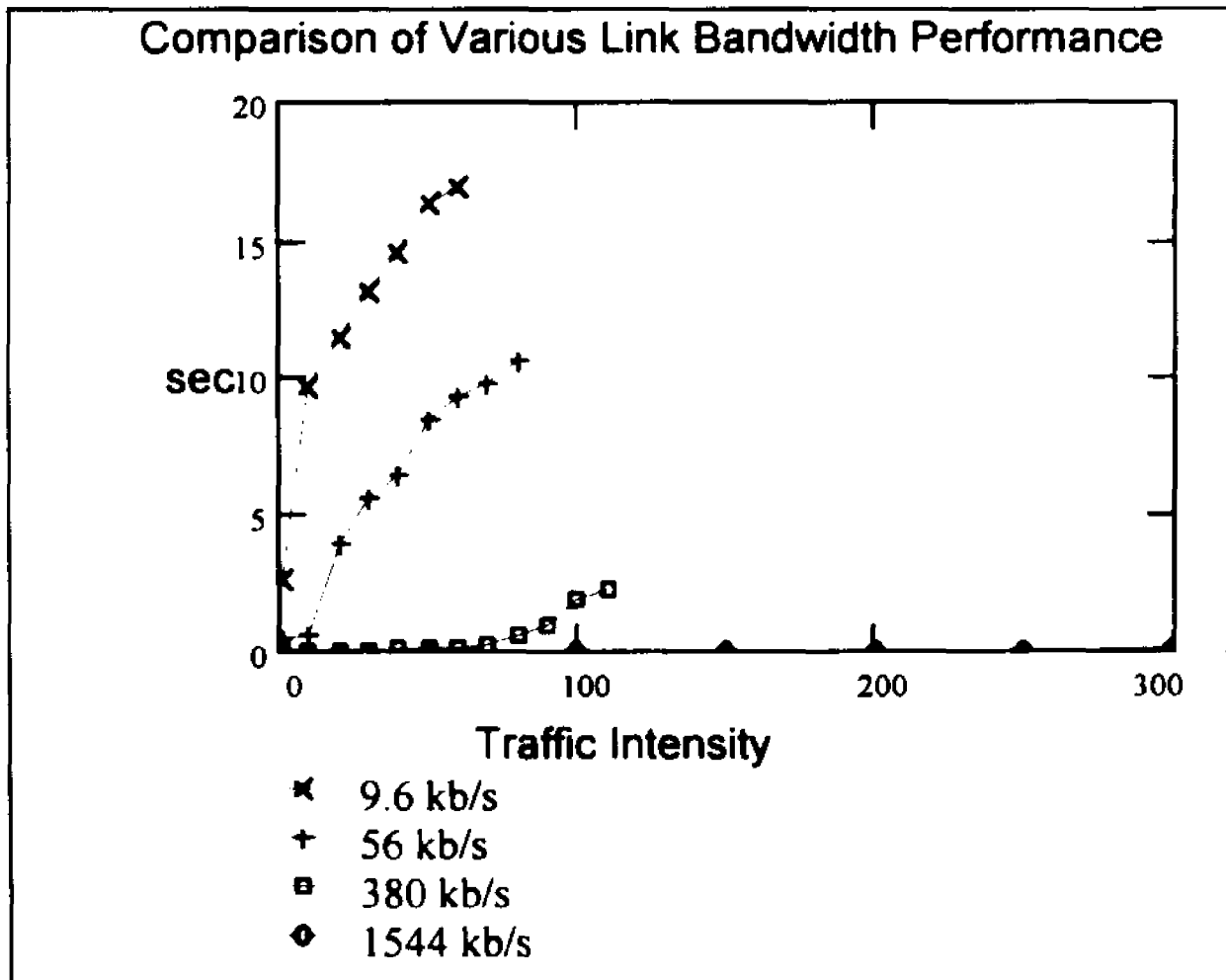


Figure 6.35. Average Message Delay for Full Duplex links

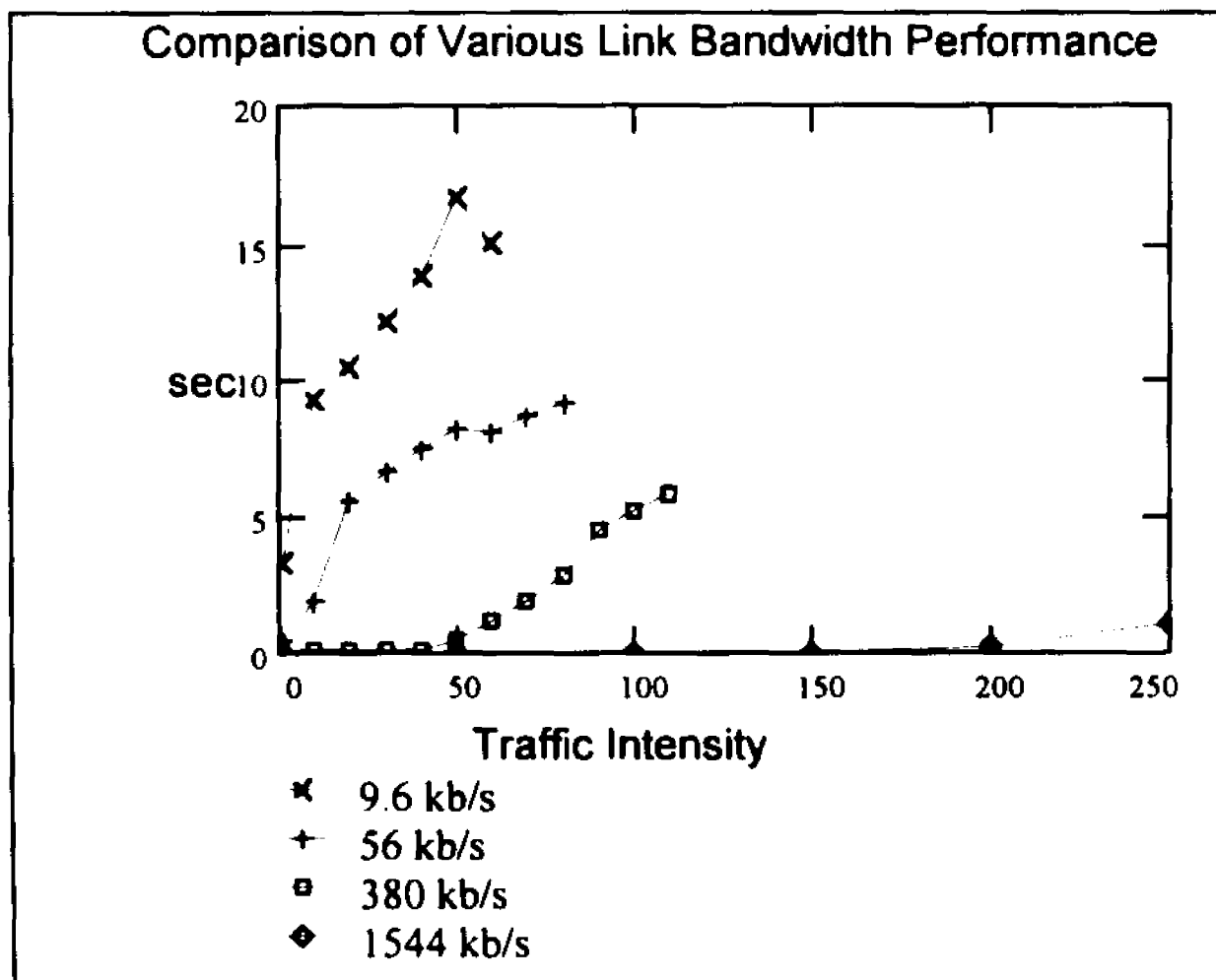


Figure 6.36: Average Message Delay for Link Failure

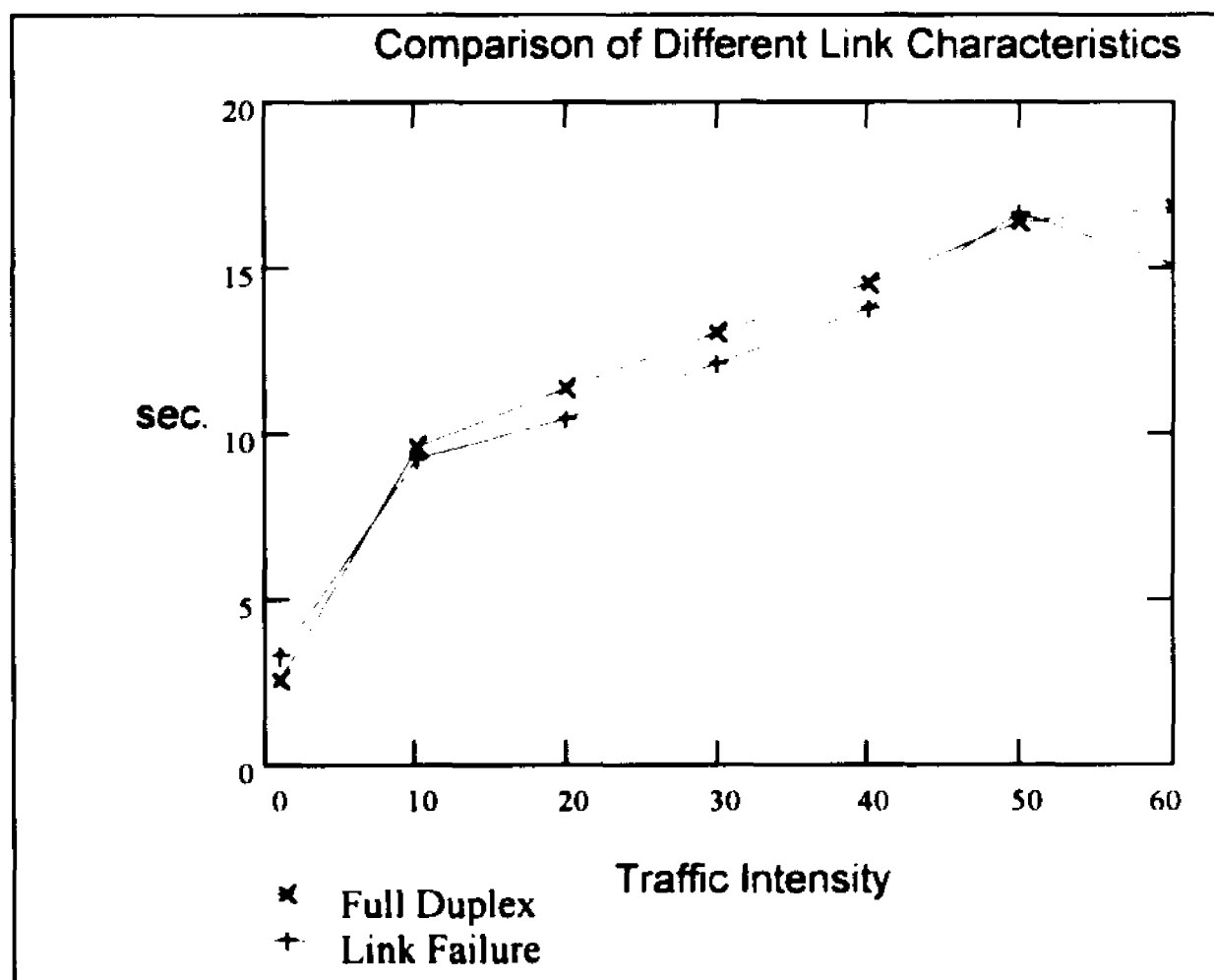


Figure 6.37. Average Message Delay for 9.6 kb/s links

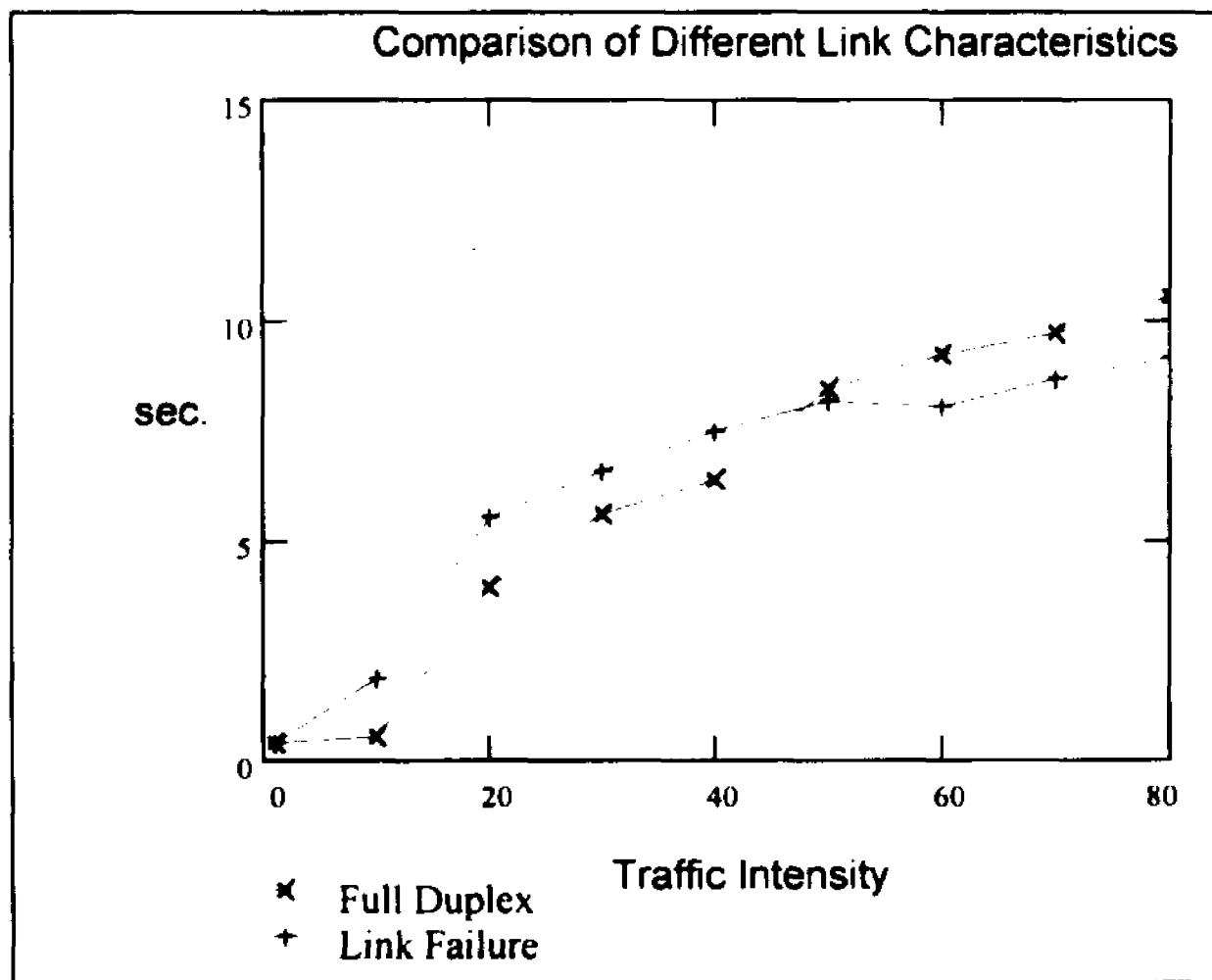


Figure 6.38: Average Message Delay for 56 kb/s links

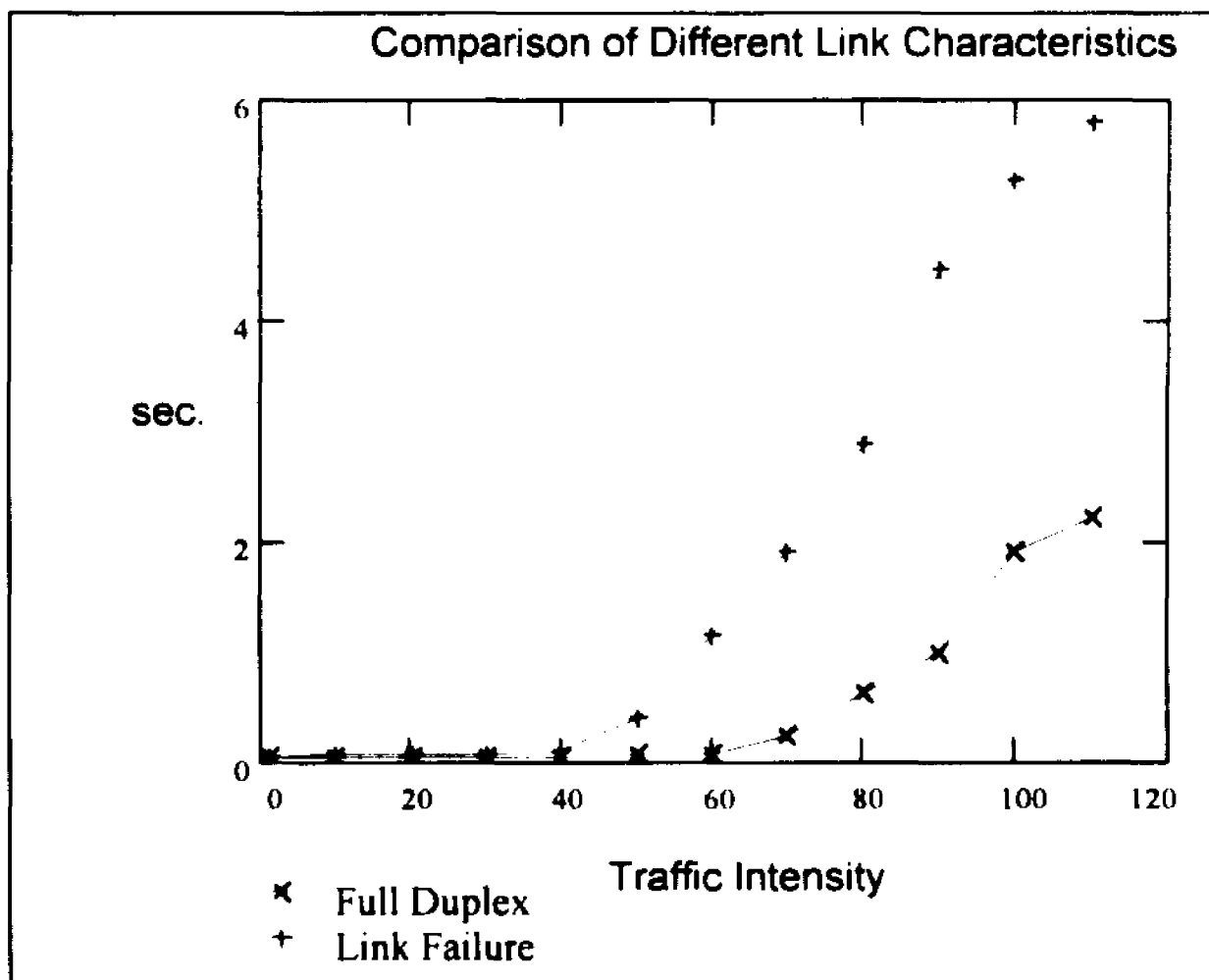


Figure 6.39 Average Message Delay for 380 kb/s links

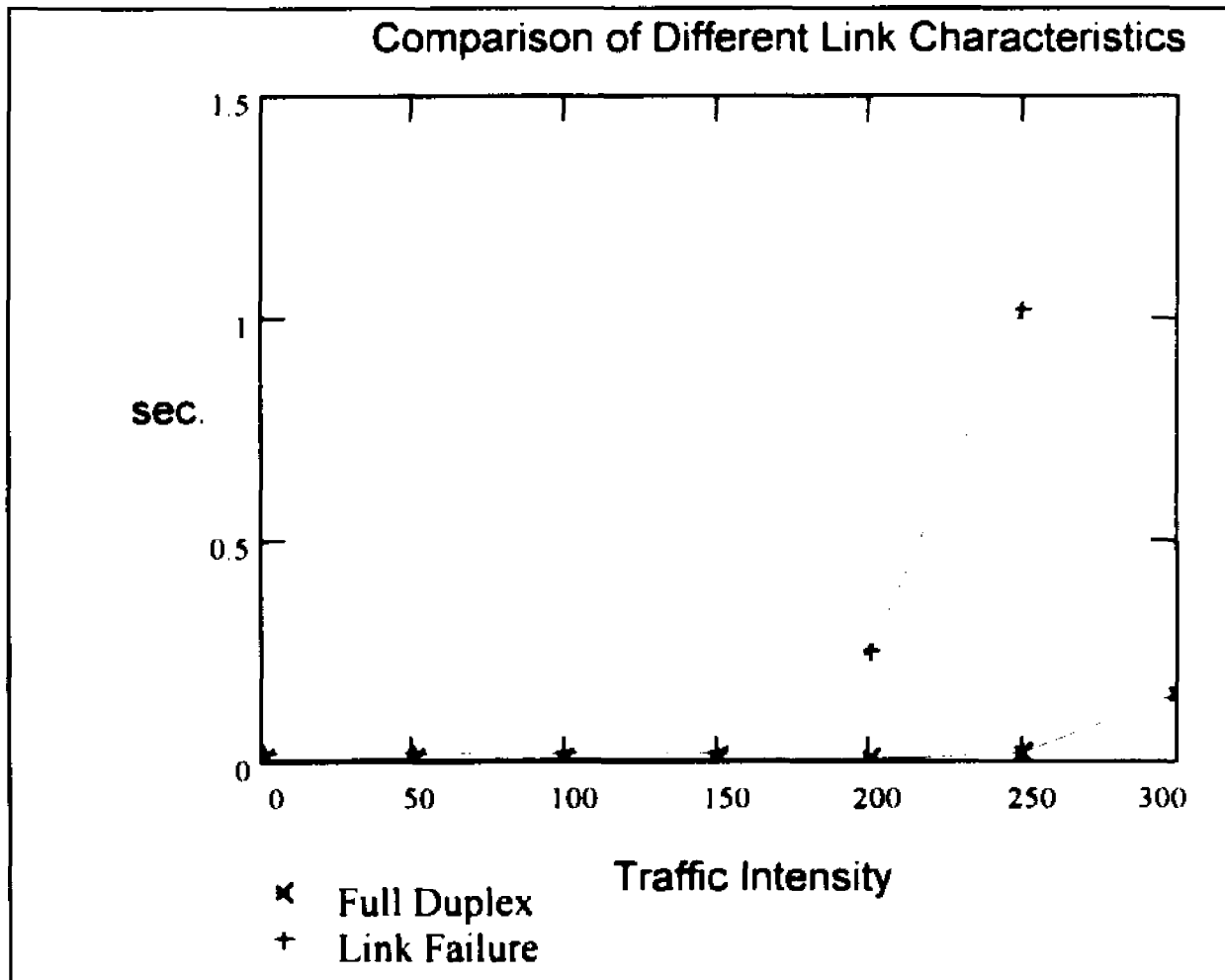


Figure 6 40 Average Message Delay for 1544 kb/s links

6.2.2.2 Maximum Message Delay

The Maximum Delay times obtained from the second model showed similarities to the results obtained from the first model. The delay time for the full duplex (Fig. 6.41) linked networks and link failure (Fig. 6.42) networks level off before sixty seconds, which is also the leveling off point for the networks in the first model. The larger network can achieve the same delay times because it has enough bandwidth to handle the increased traffic produced by the nodes

Results show that for both the low bandwidth links of 9.6 kb/s (Fig. 6.43) and 56 kb/s (Fig. 6.44) and high bandwidth links of 380 kb/s (Fig. 6.45) and 1544 kb/s (Fig. 6.46) a full duplex network has lower delay time than a link failed network. The difference between these times however, is not as large as the ones from the first model. This implies that a removal of a link has less of an effect on Maximum Message Delay time for large network than a smaller network. The main reason for this is the availability of more links to transport data, thereby not allowing any messages to remain in the outgoing link queue too long

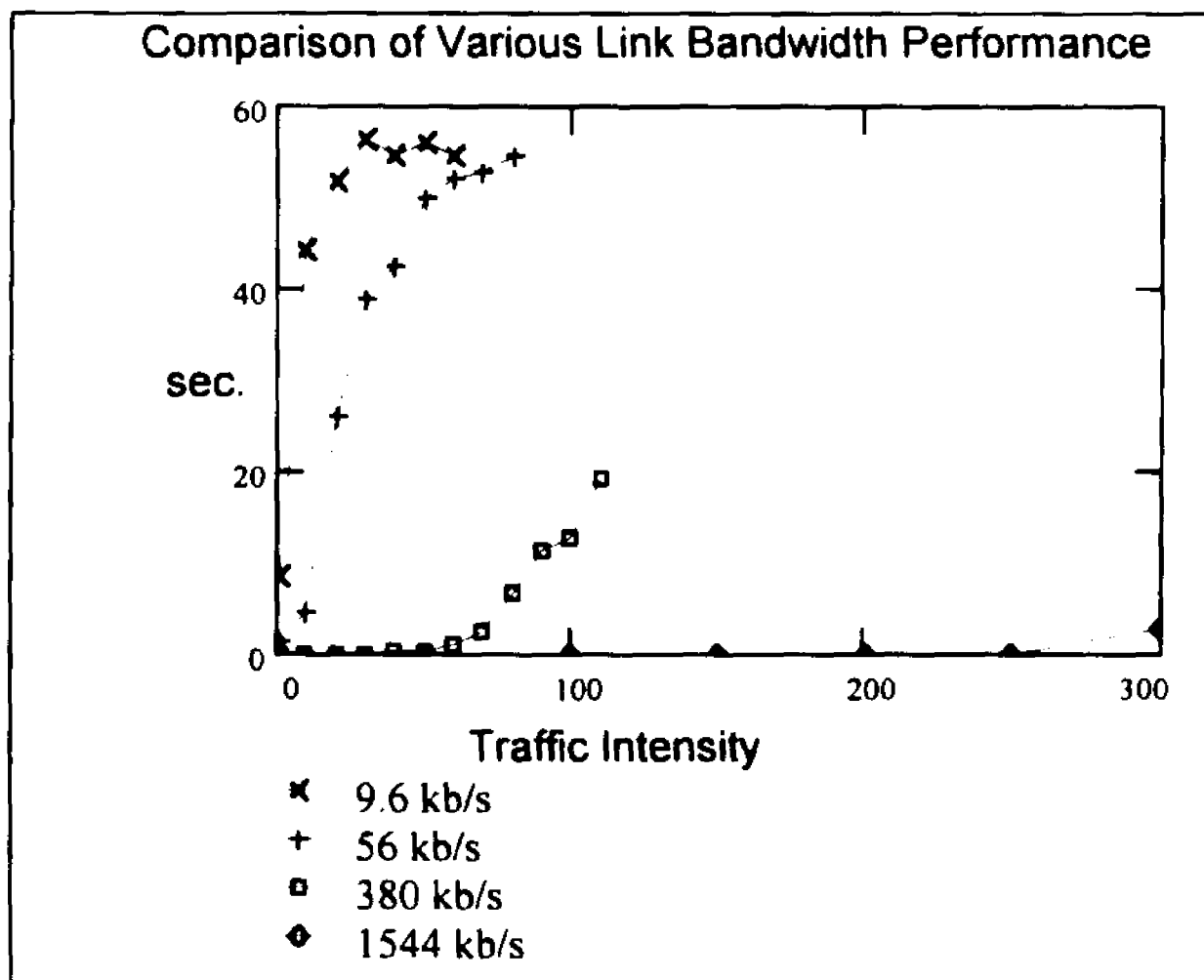


Figure 6.41. Maximum Message Delay for Full Duplex links

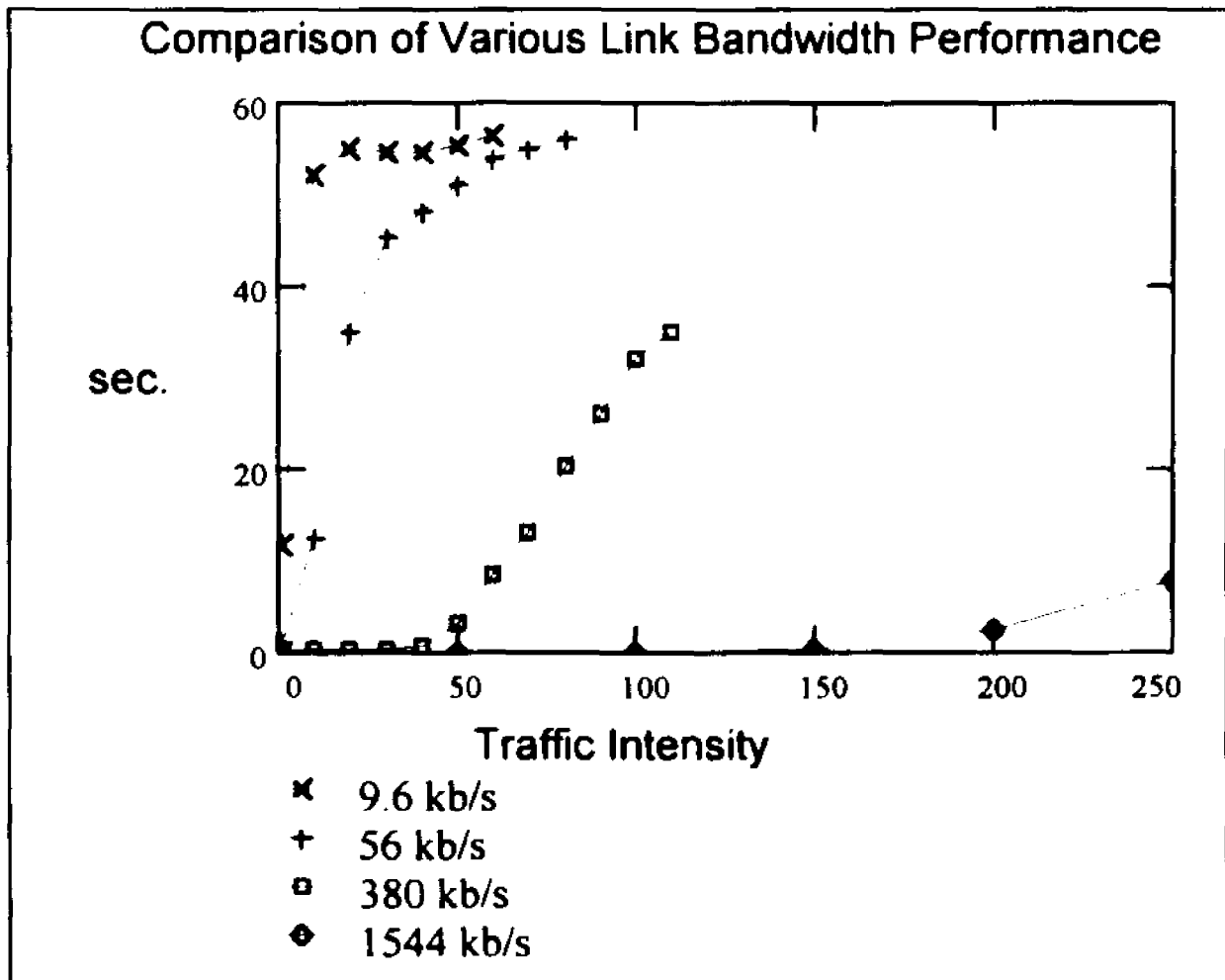


Figure 6.42: Maximum Message Delay for Link Failure

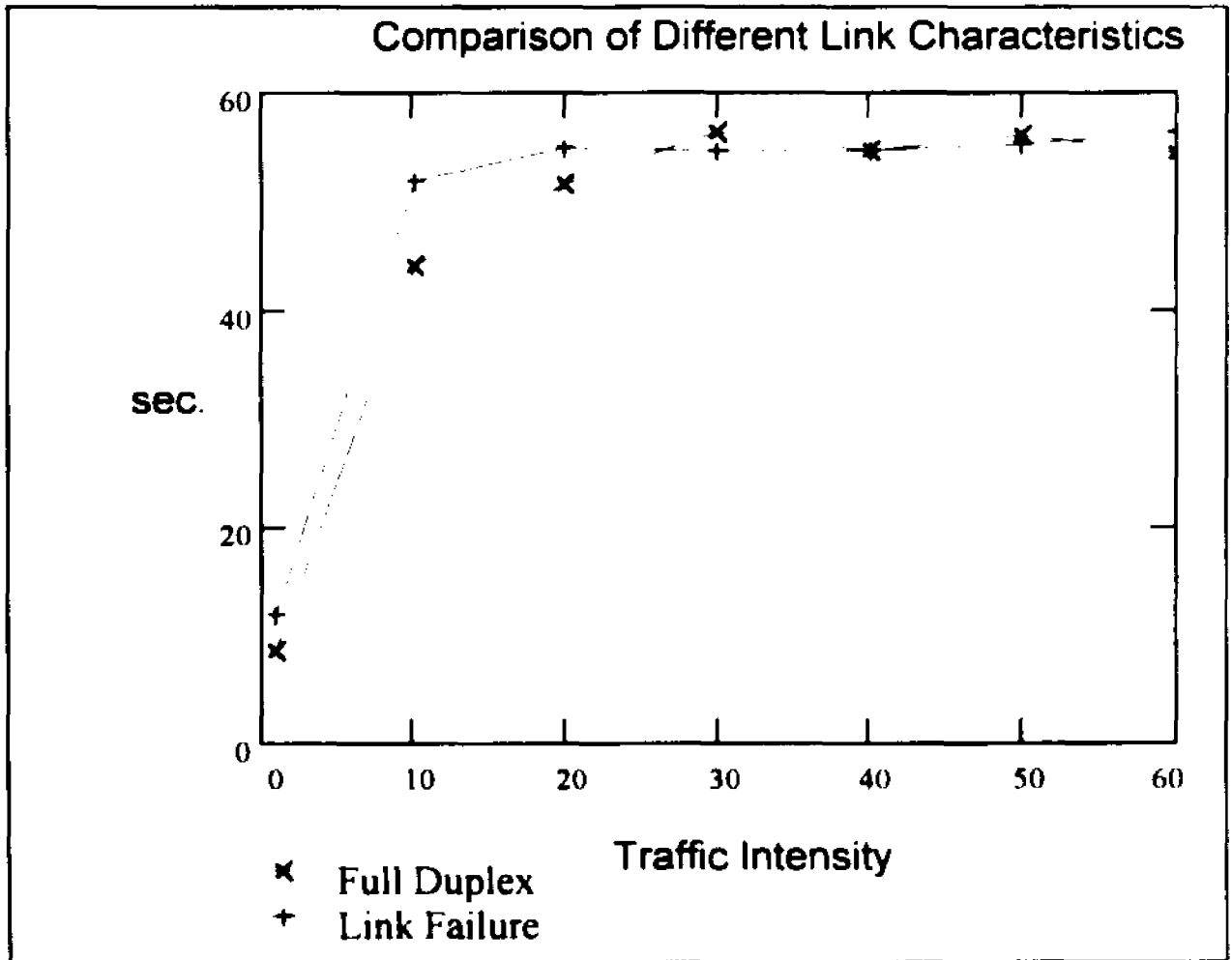


Figure 6 43. Maximum Message Delay for 9 6 kb/s links

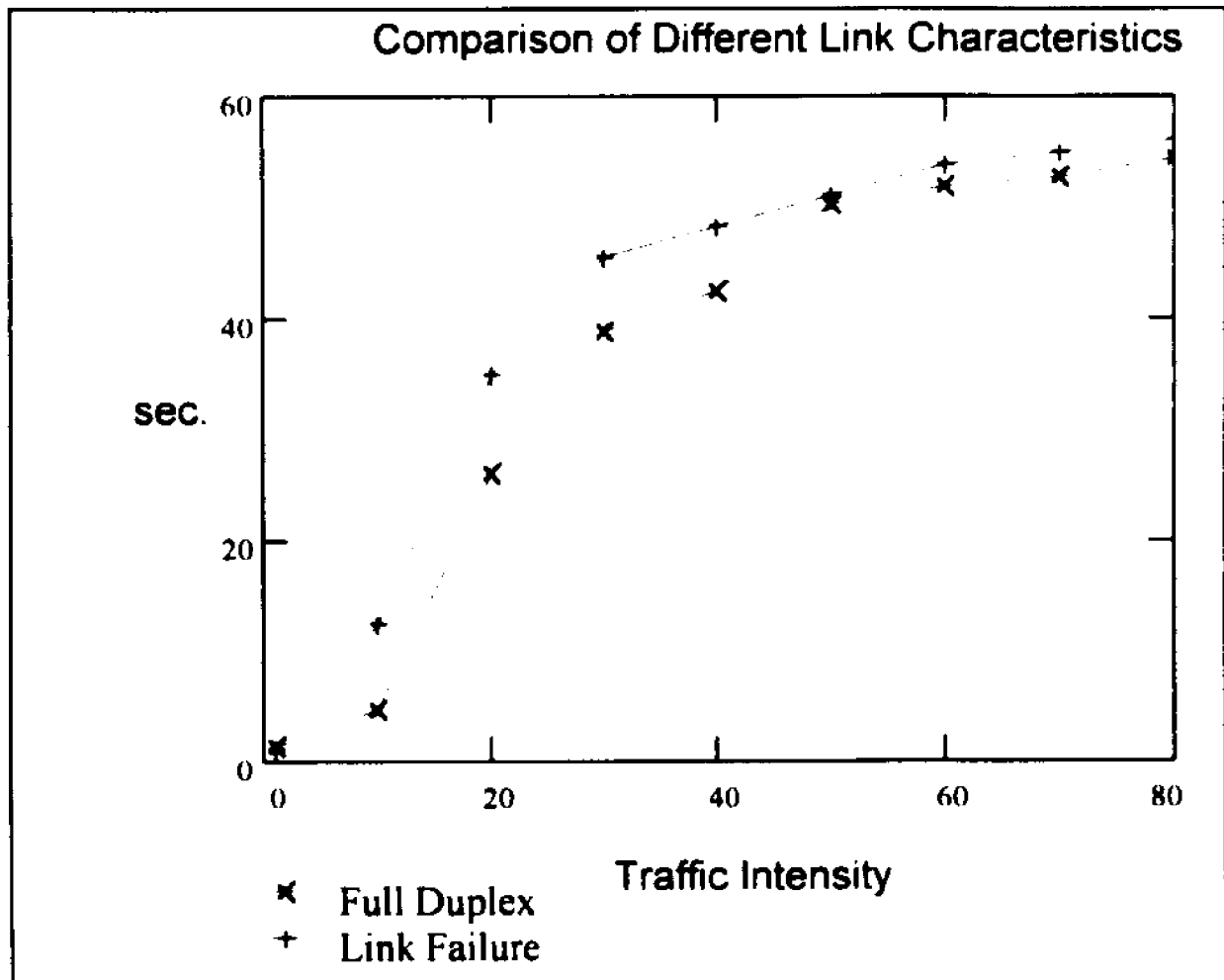


Figure 6.44. Maximum Message Delay for 56 kb/s links

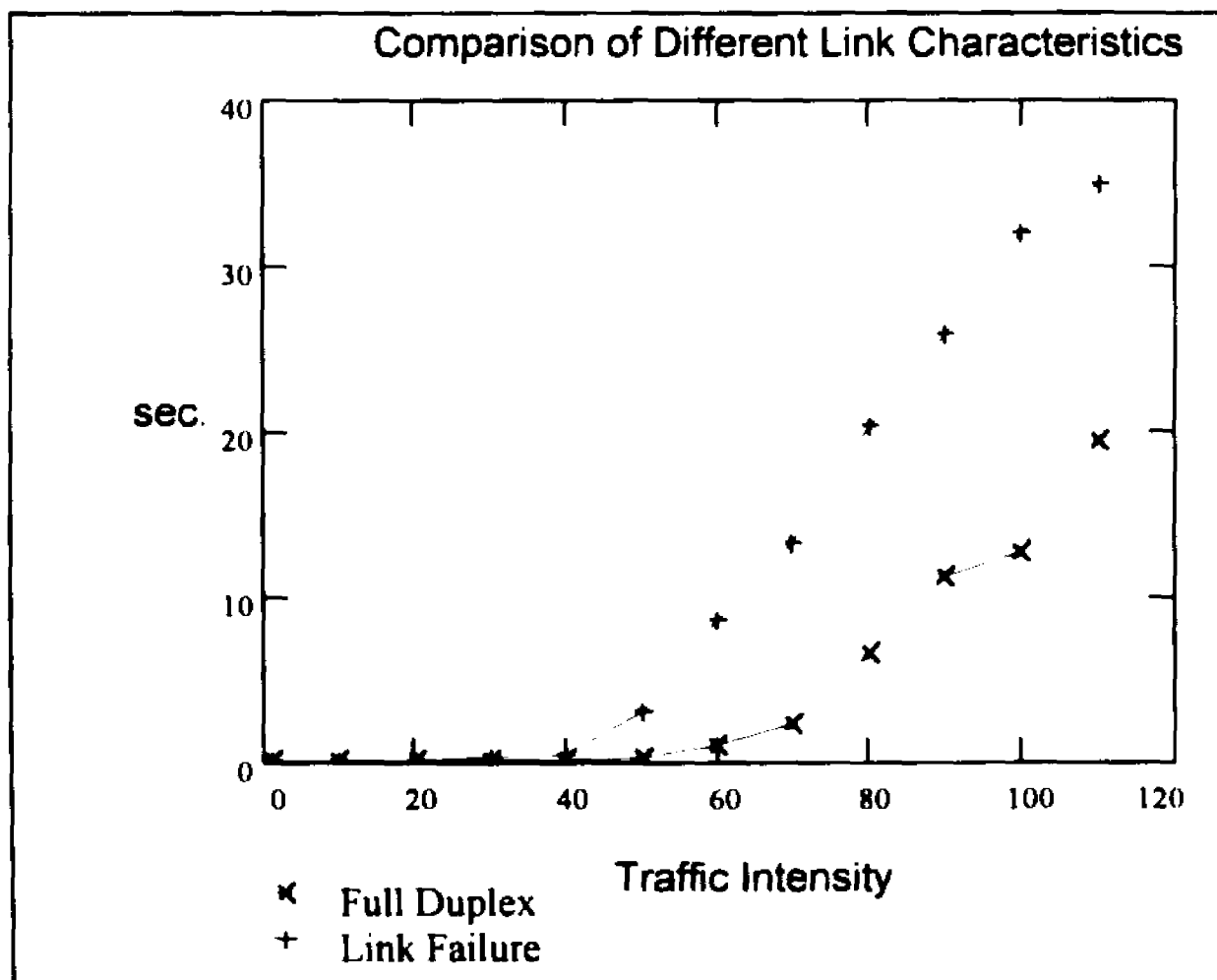


Figure 6.45. Maximum Message Delay for 380 kb/s links

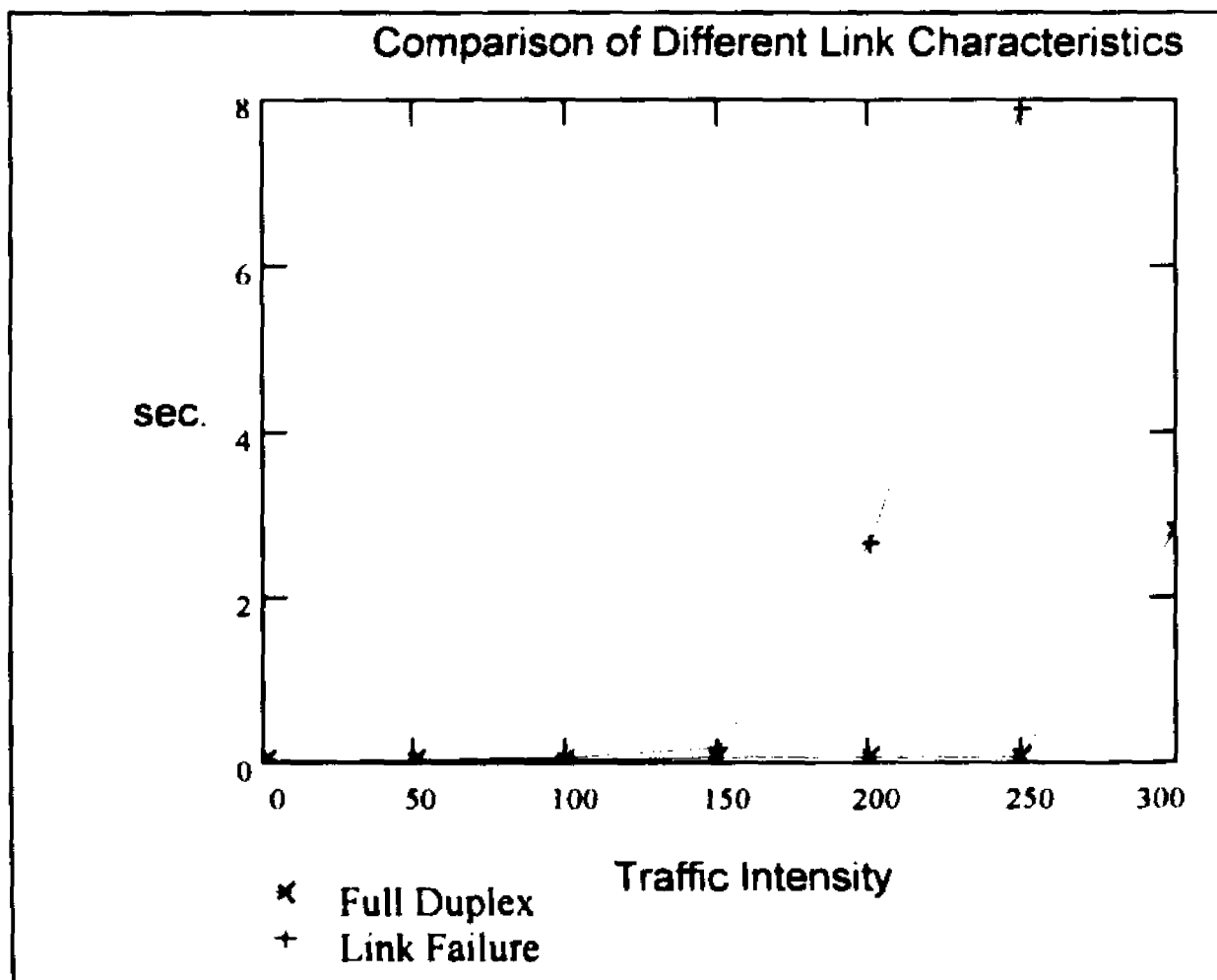


Figure 6.46 Maximum Message Delay for 1544 kb/s links

6.2.3 Percent of Messages over Three Seconds

The results for percent of messages over three seconds show that the larger network model has similar percentages for its messages as the smaller model. The second model reaches the high percentages at lower traffic intensities than the first one. This is due to the higher traffic activity produced by the second model. The full duplex (Fig. 6.47) linked and link failure (Fig. 6.48) networks reach 90% level at low bandwidths, with the link failure model's percentage starting to climb at a lower traffic intensity than the full duplex one.

Closer inspection of the networks at 9.6 kb/s (Fig. 6.49) and 56 kb/s (Fig. 6.50) show that, though the percentage for link failure network is high at low traffic intensities, as the traffic increases it remains level and eventually resulting in lower percentages than the full duplex network. For the higher bandwidth links of 380 kb/s (Fig. 6.51) and 1544 kb/s (Fig. 6.52), it can be seen that link failed networks not only start increasing at smaller traffic intensities, but also have consistently higher percentages than the full duplex linked networks.

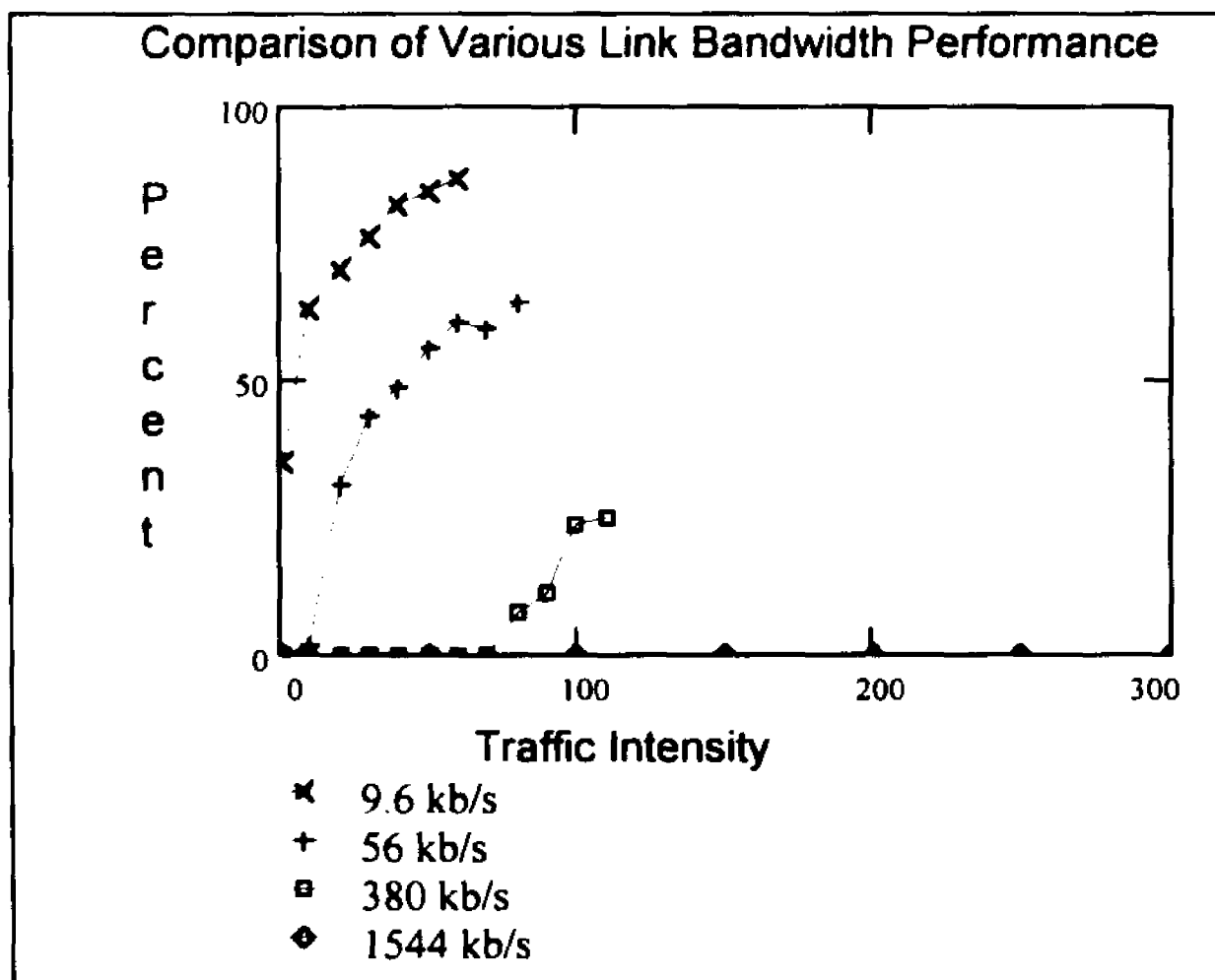


Figure 6.47: Percent of Messages over 3 sec for Full Duplex links

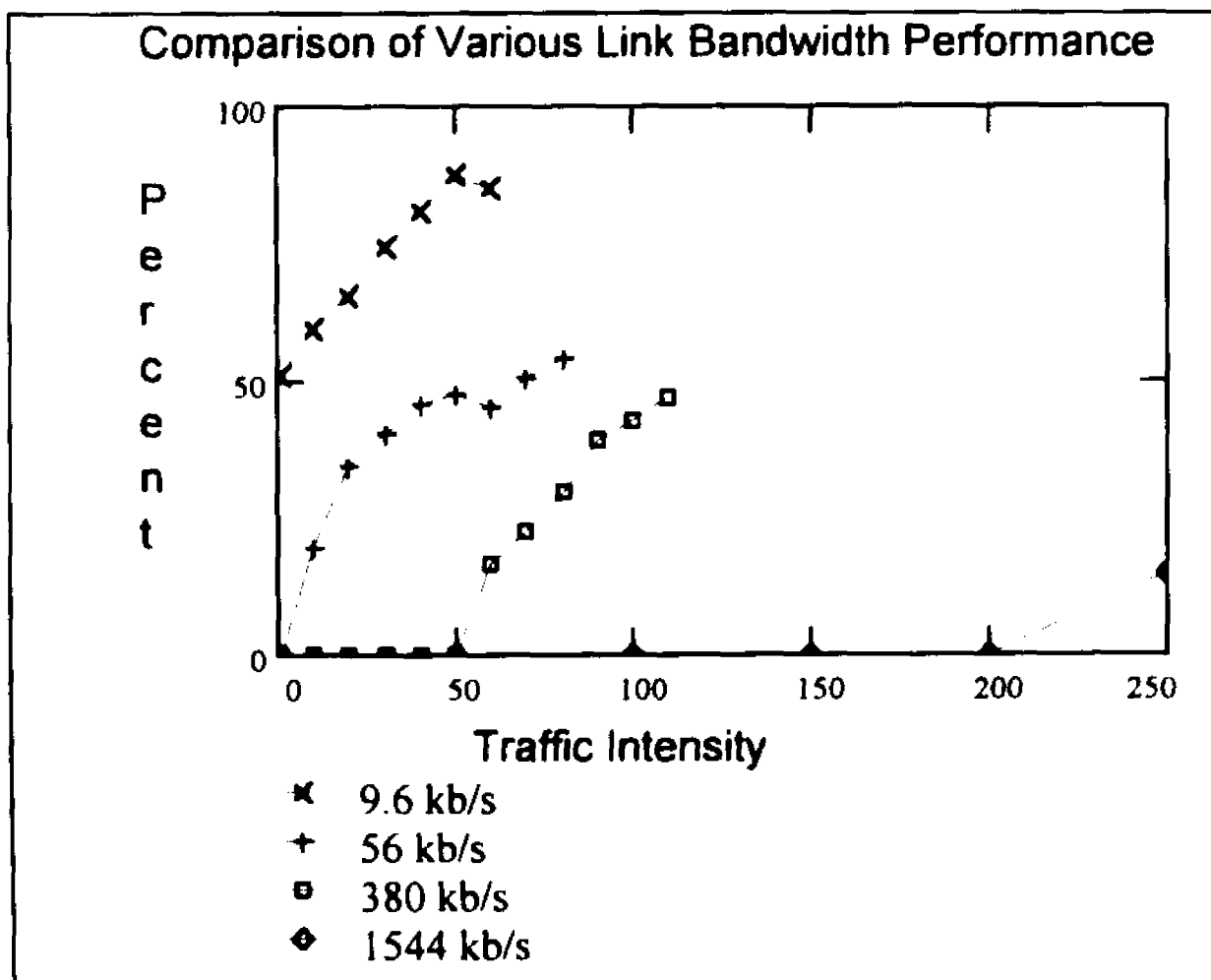


Figure 6 48 Percent of Messages over 3 sec for Link Failure

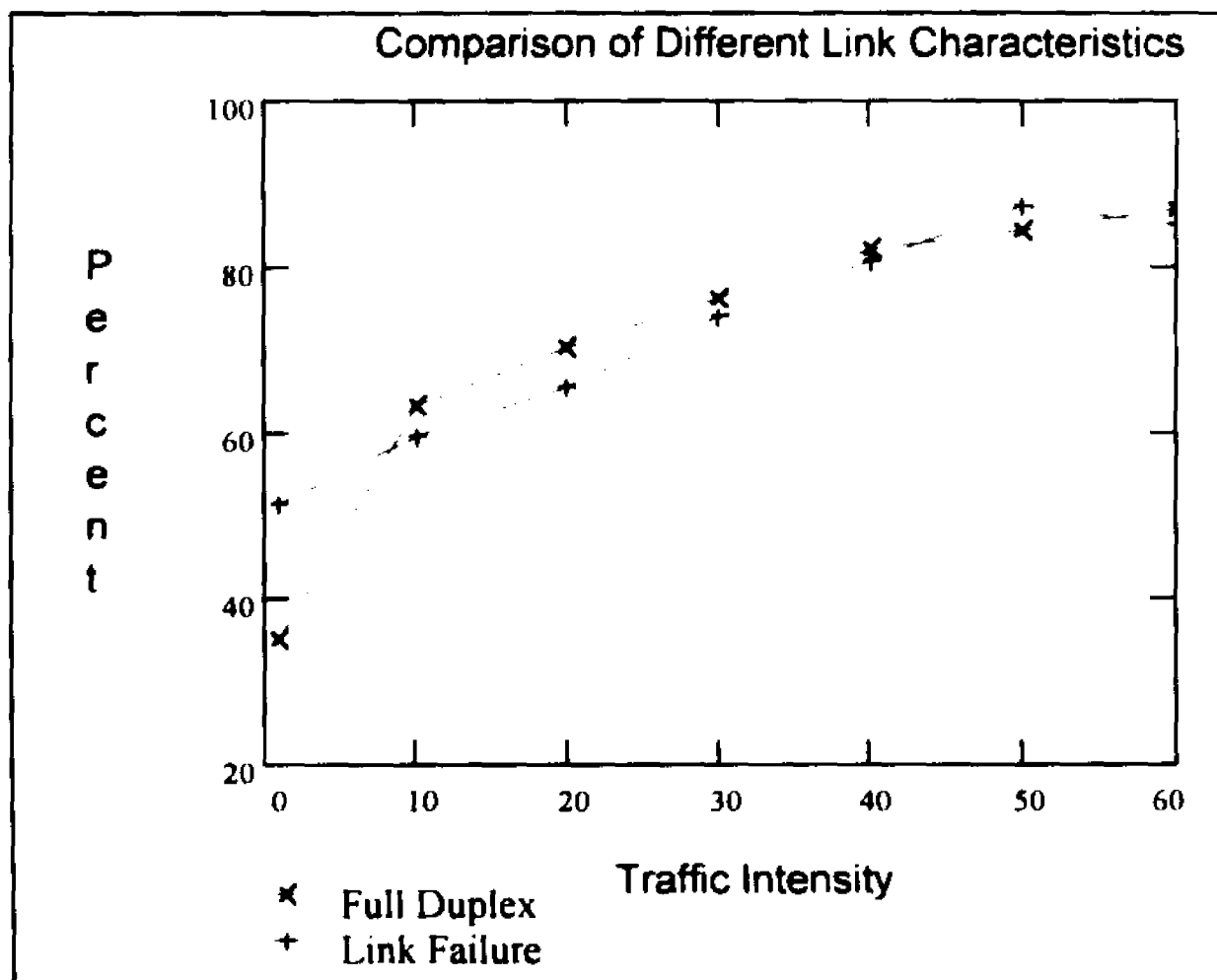


Figure 6.49 Percent of Messages over 3 sec for 9.6 kb/s links

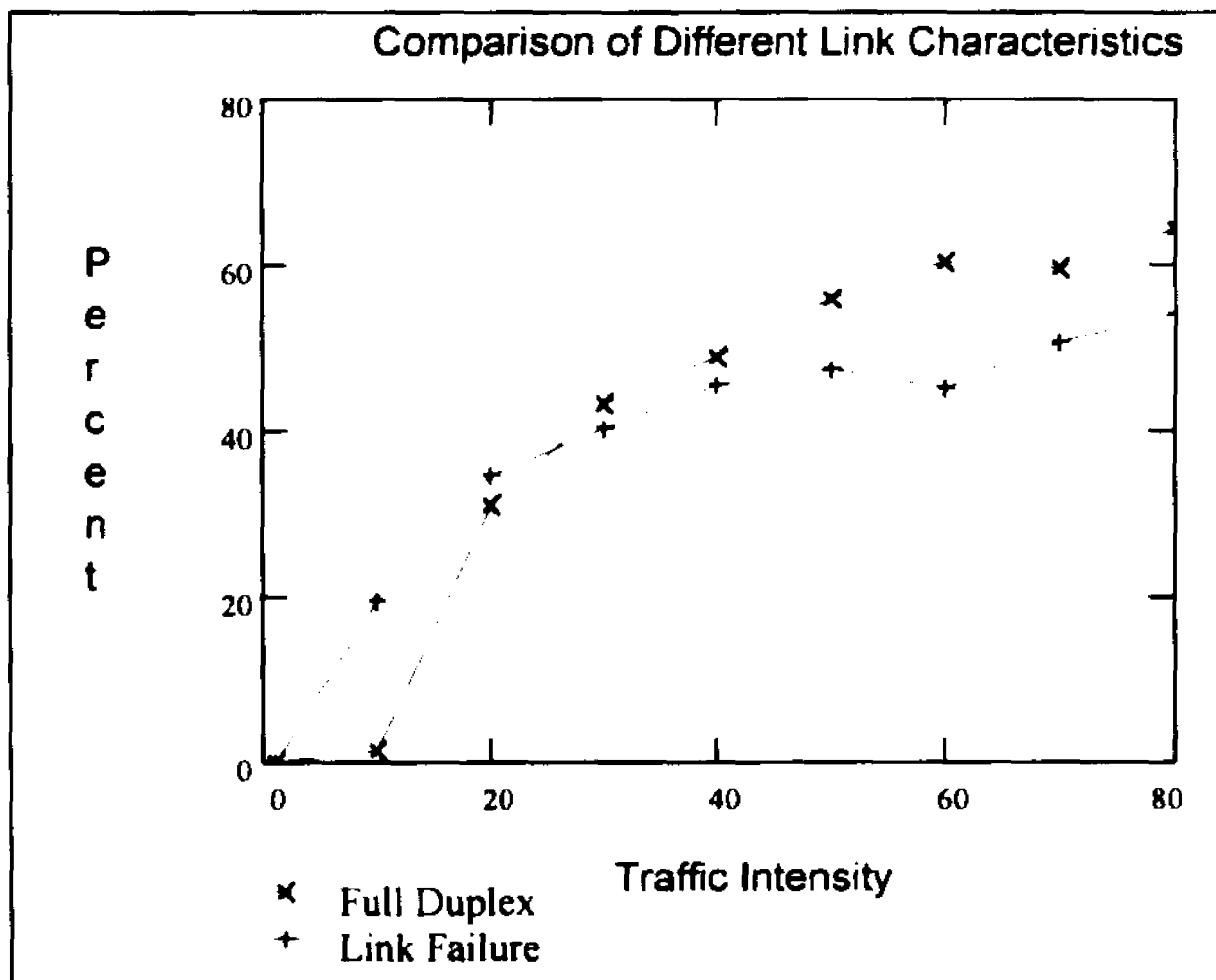


Figure 6.50 Percent of Messages over 3 sec for 56 kb/s links

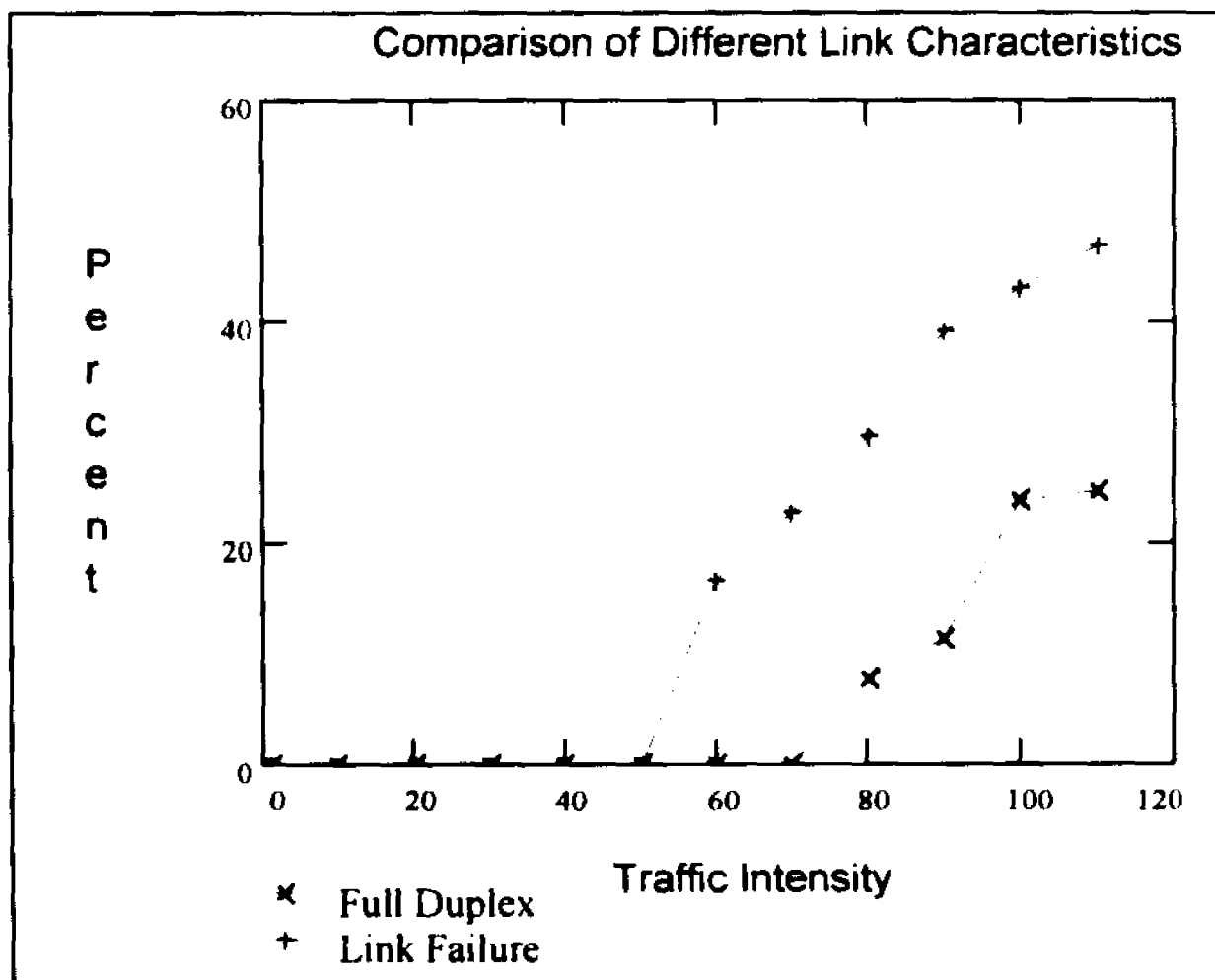


Figure 6.51: Percent of Messages over 3 sec for 380 kb/s links

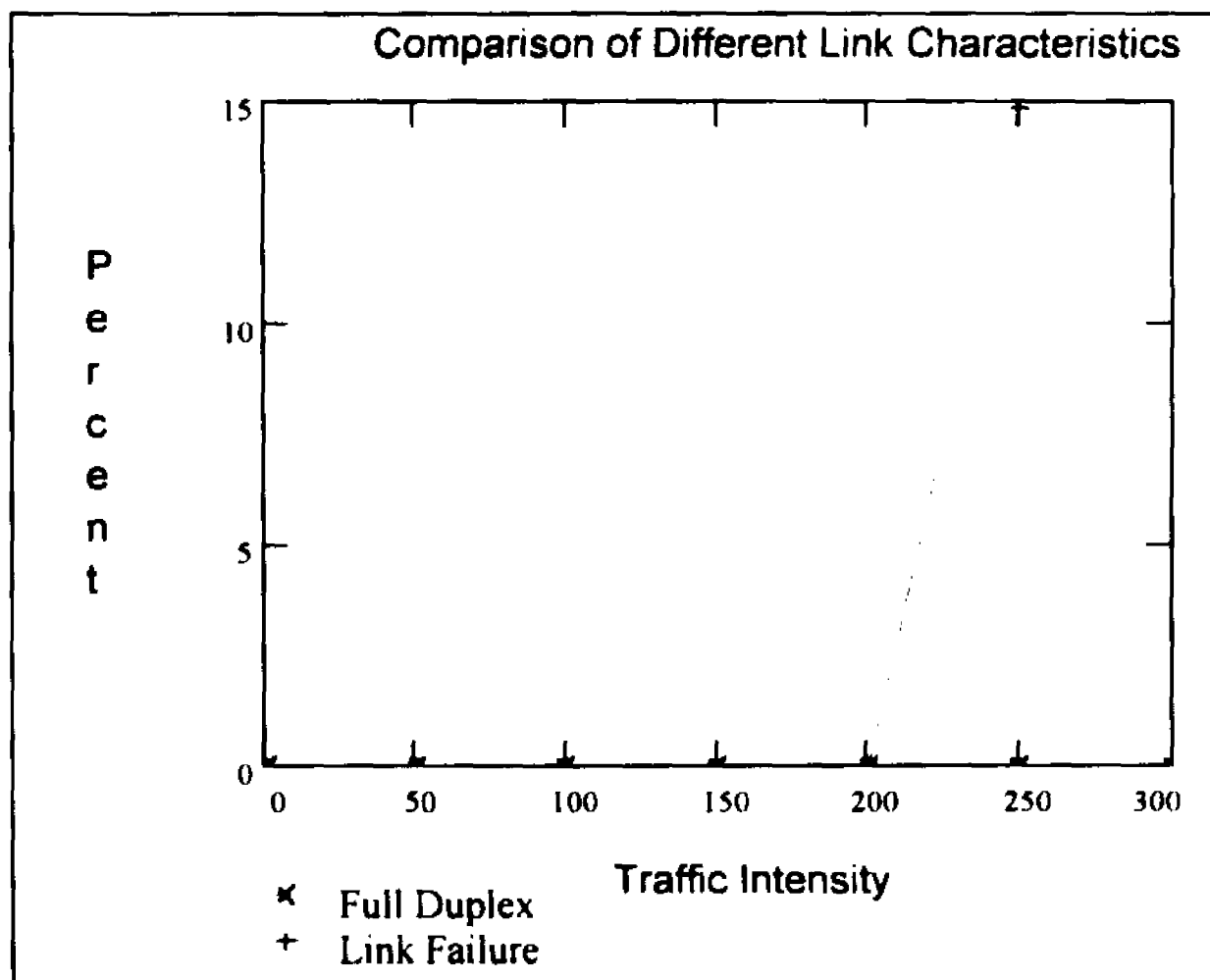


Figure 6.52: Percent of Messages over 3 sec for 1544 kb/s links

6.3 Further Analysis and Findings

Further analysis of results show that given a particular topology, it possible conclude that there is a strong relationship between throughput and message delay times. In addition both follow similar patterns of behavior for different link properties, which allow us to draw some general conclusions when approximating throughput or delay times. The analysis mainly focuses on the throughput since it is an important performance factor.

6.3.1 Throughput Separation

A quick examination of the results in chapter five reveals that a typical throughput curve has two important points to consider. The first one is where the steady increase of throughput starts lagging, which we call point of separation. The second one is the point where throughput ceases to increase even though the traffic intensity does so, which we call point of saturation. We show not only that the throughput level (kb/s) up to the point of separation increases linearly with traffic intensity for networks for different link properties and link speeds for a given topology, but also these relationships converge for different link speeds.

It is also shown that the delay times show a large increase at the same traffic intensity where the point of separation occurs, thereby enabling us to predict the increases in delay times by predicting point of separation. To find the point of separation for each network, two curves are plotted: rate of change of throughput versus traffic intensity and rate of change of delay versus traffic intensity.

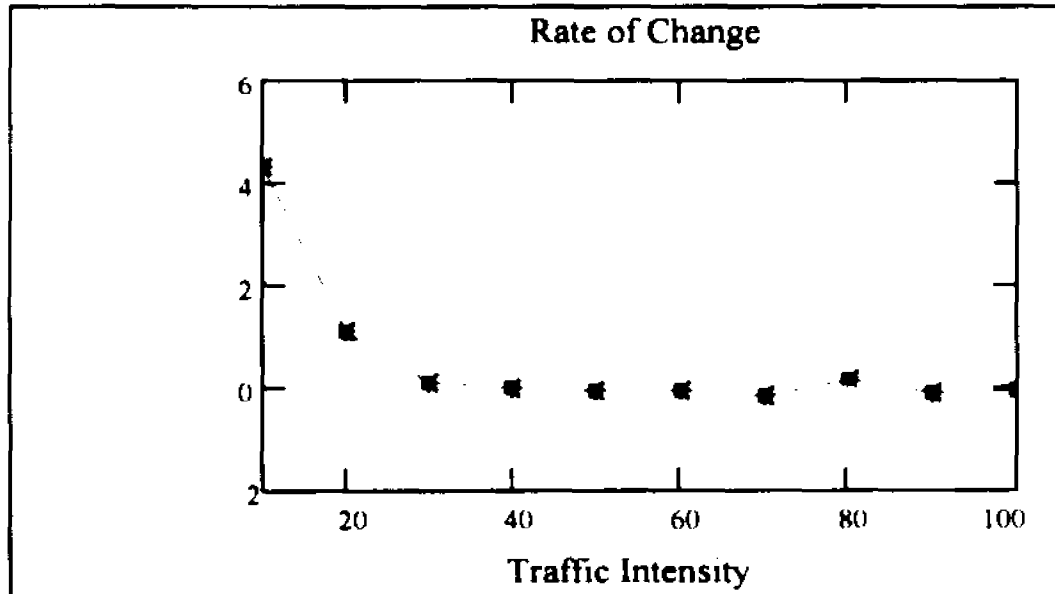


Figure 6.53. Rate of change of throughput (Model 1 - 9.6 kb/s, full duplex)

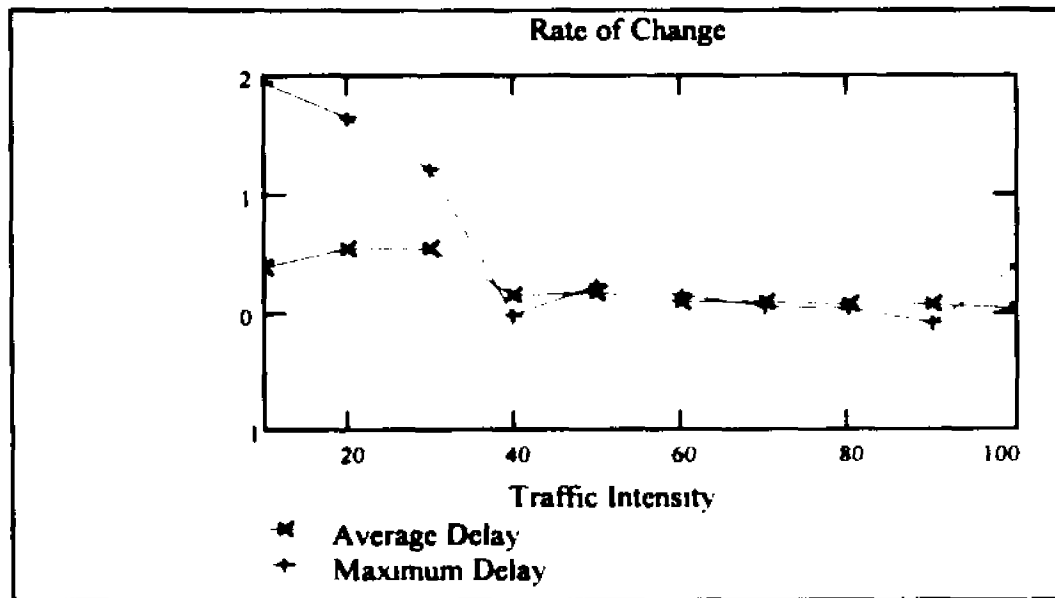


Figure 6.54. Rate of change of delay times (Model 1 - 9.6 kb/s, full duplex)

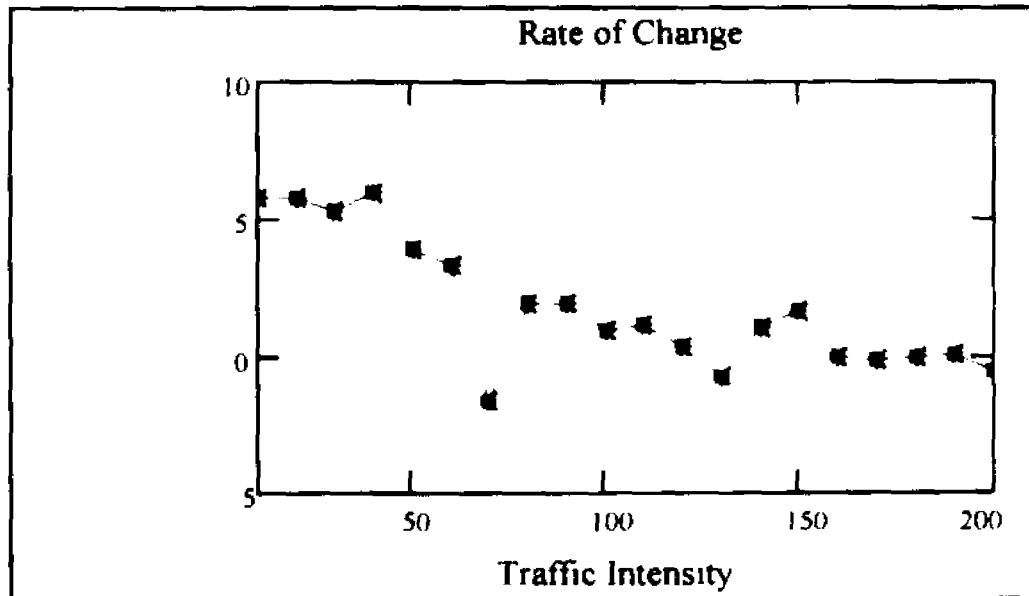


Figure 6.55 Rate of change of throughput (Model 1 - 56 kb/s, full duplex)

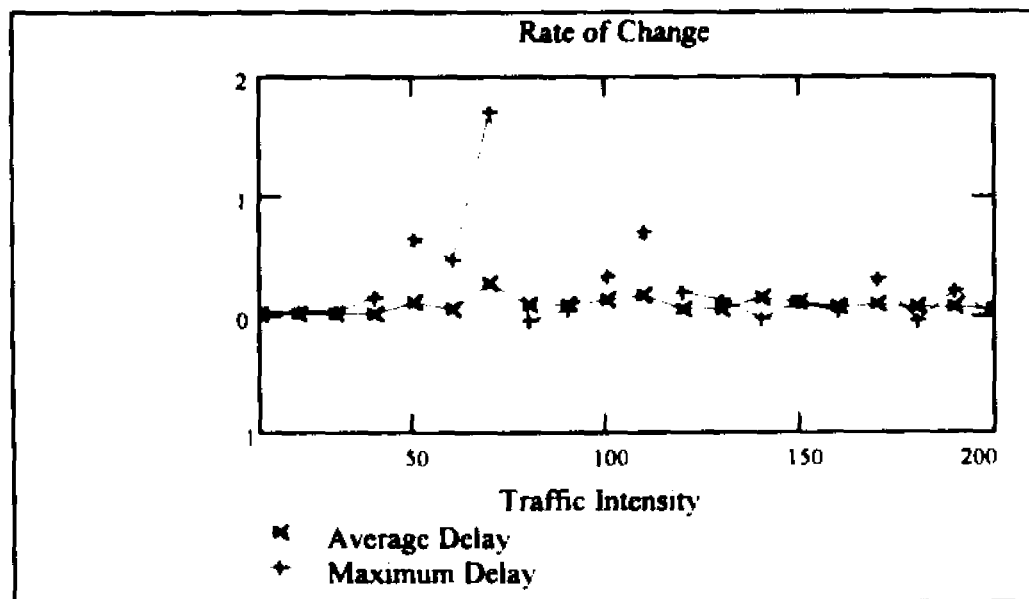


Figure 6.56 Rate of change of delay times (Model 1 - 56 kb/s, full duplex)

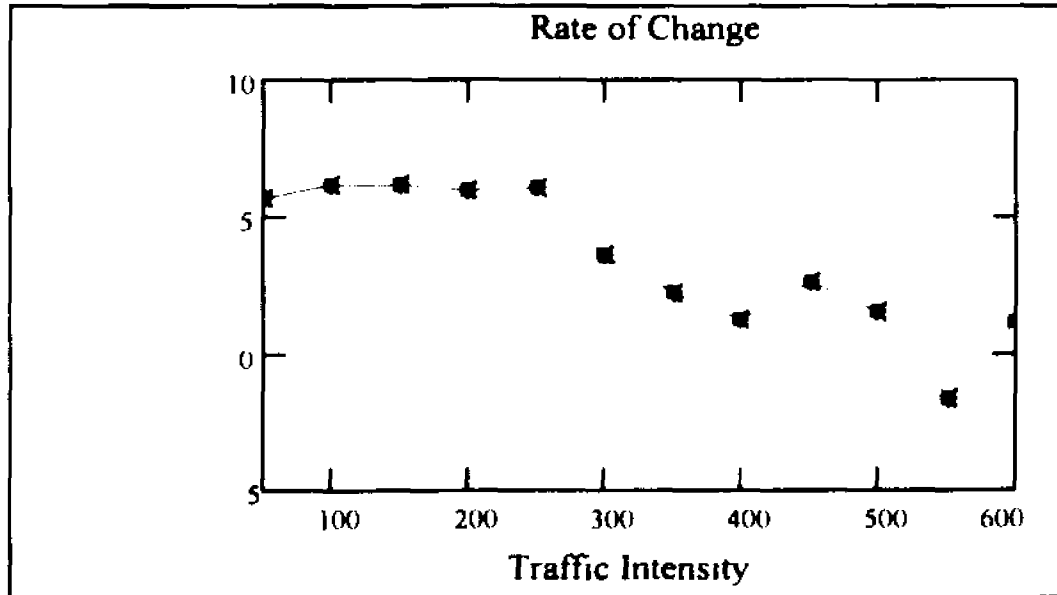


Figure 6.57 Rate of change of throughput (Model 1 - 380 kb/s, full duplex)

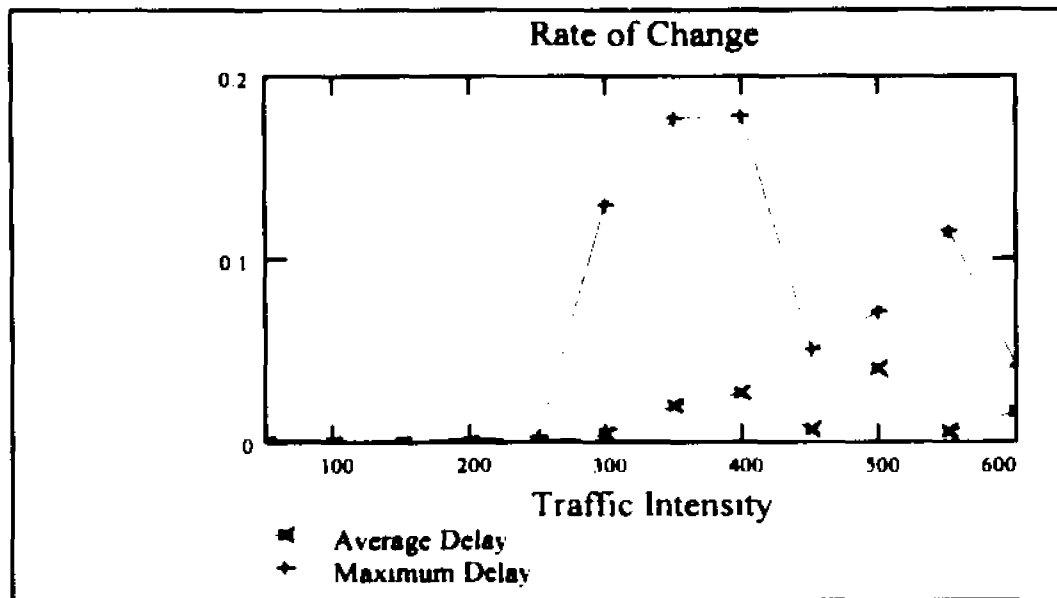


Figure 6.58: Rate of change of delay times (Model 1 - 380 kb/s, full duplex)

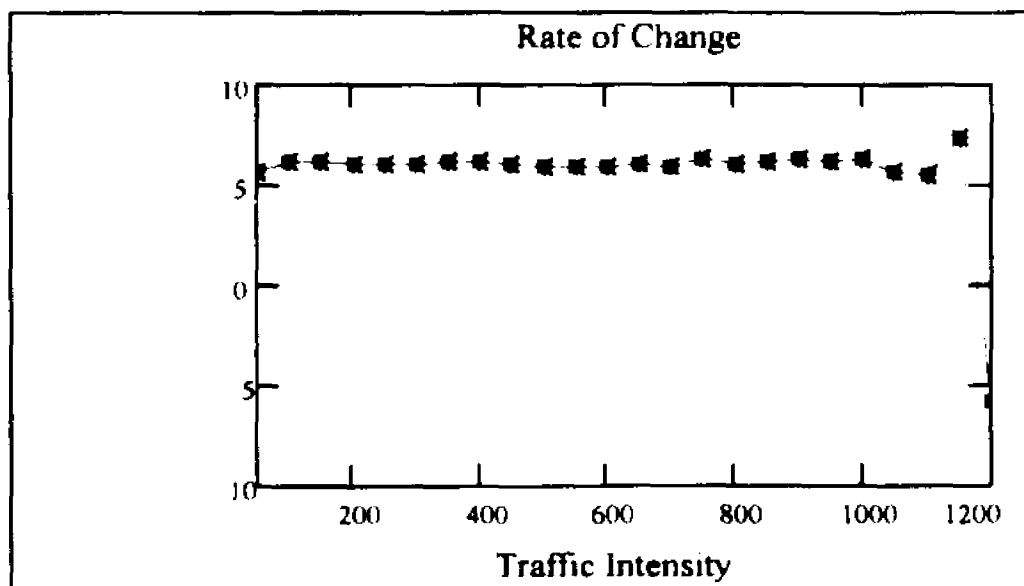


Figure 6.59 Rate of change of throughput (Model 1 - 1544 kb/s, full duplex)

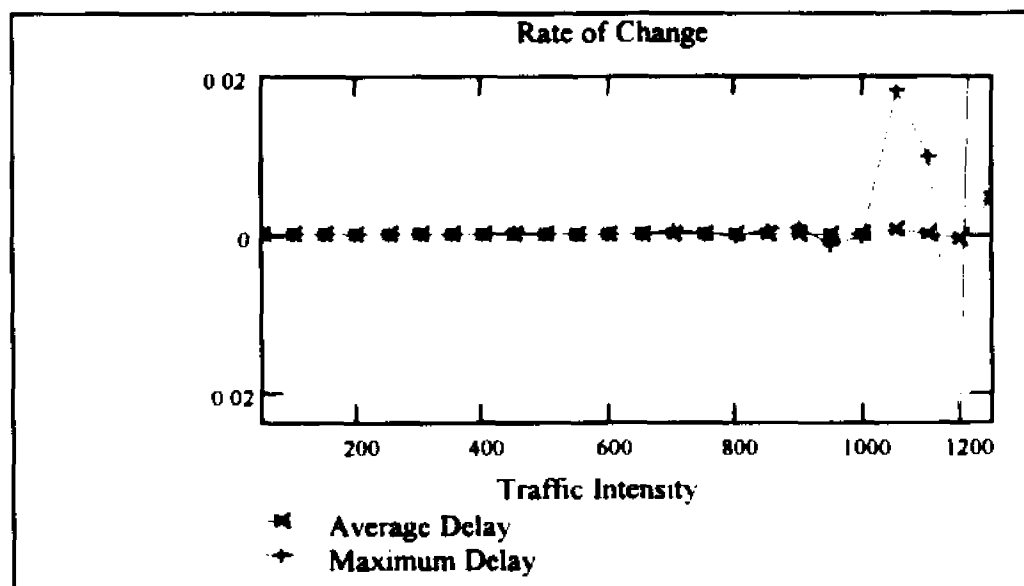


Figure 6.60 Rate of change of delay times (Model 1 - 1544 kb/s, full duplex)

Examining Figure 6.53, we observe that the rate of change of throughput decreases as traffic intensity is increased. This implies that the rate is not constant, and that the point of separation has already been reached at a lower traffic intensity. This conclusion is further supported by Figure 6.54 where the rates of message delay times have started to increase confirming that the point of separation has indeed already occurred.

Figure 6.55 shows that the rate of change of throughput remains constant for low traffic intensities. It is, however, difficult to determine at which traffic intensity (20 or 40) the point of separation occurs from that figure alone. Using Figure 6.56, we see that there is an increase in the rate of change of delay times at traffic intensity of 30, which implies that the point of separation occurring at traffic intensity of 20, before any of the delay times increase.

Figure 6.57 clearly shows that the point of separation occurs at traffic intensity of 250, a fact that is corroborated by Figure 6.58, where the increase in rate of change of delay times also occurs at the same traffic intensity.

Finally, Figure 6.59 and Figure 6.60 are used to deduce that the point of separation occurs at traffic intensity of 1000. Summarizing the findings for simulation model one with full duplex links we get table 6.1.

Link Speed (kb/s)	Traffic Intensity	Throughput (kb/s)
9.6	--	--
56	20	116
380	250	1501
1544	1000	6081

Table 6. 1: Point of Separation occurrences for full duplex links. (Model 1)

Combining Figures 6.53, 6.55, 6.57 and 6.59 in Figure 6.61, we see that not only that the rate of change of throughput before the point of separation is constant, but it is almost the same for different link speeds. Plotting the corresponding figures of different link properties for the first and the second simulation models, we see that this result recurs in each figure regularly.

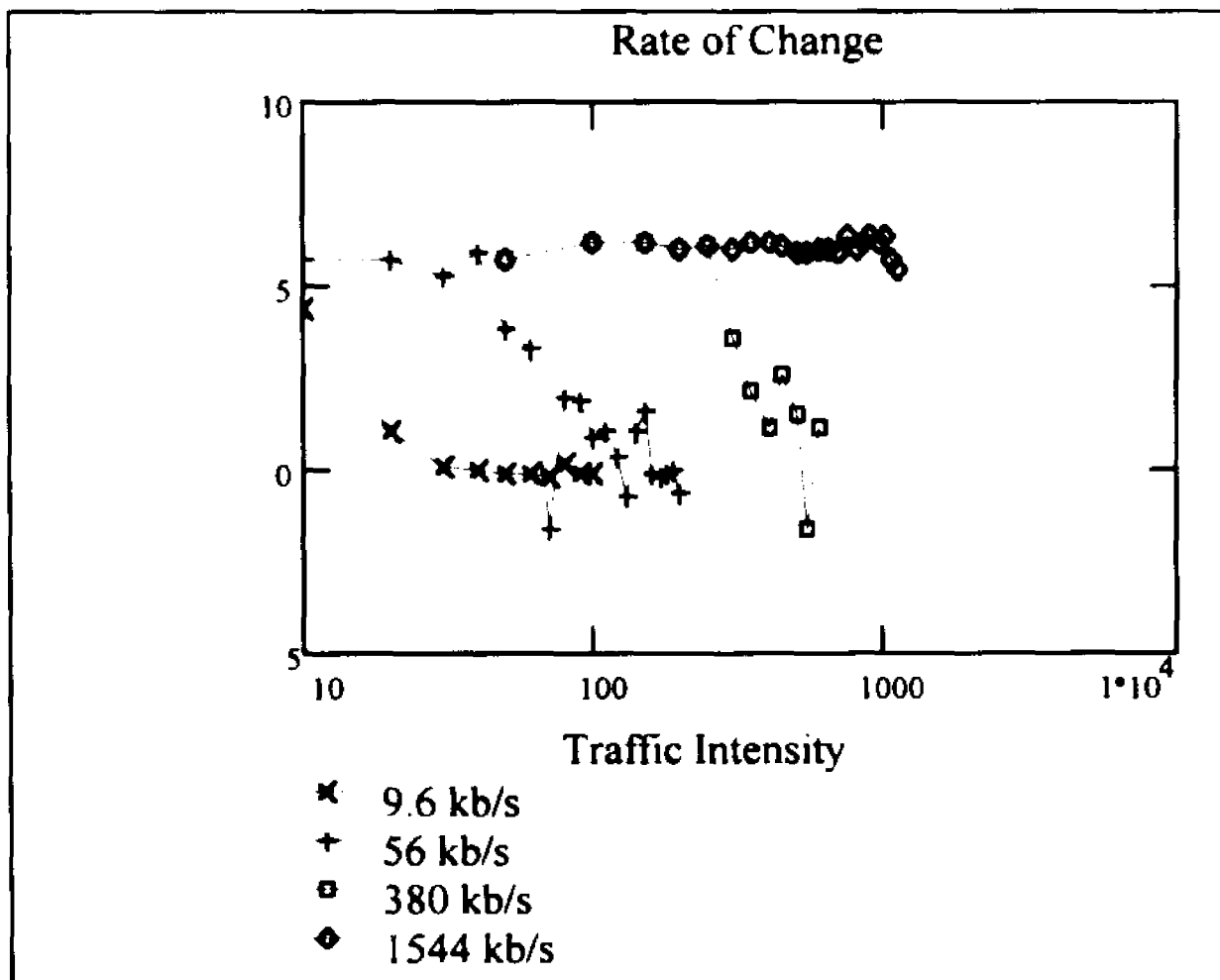


Figure 6.61: Rate of change of throughput (Model 1 - full duplex)

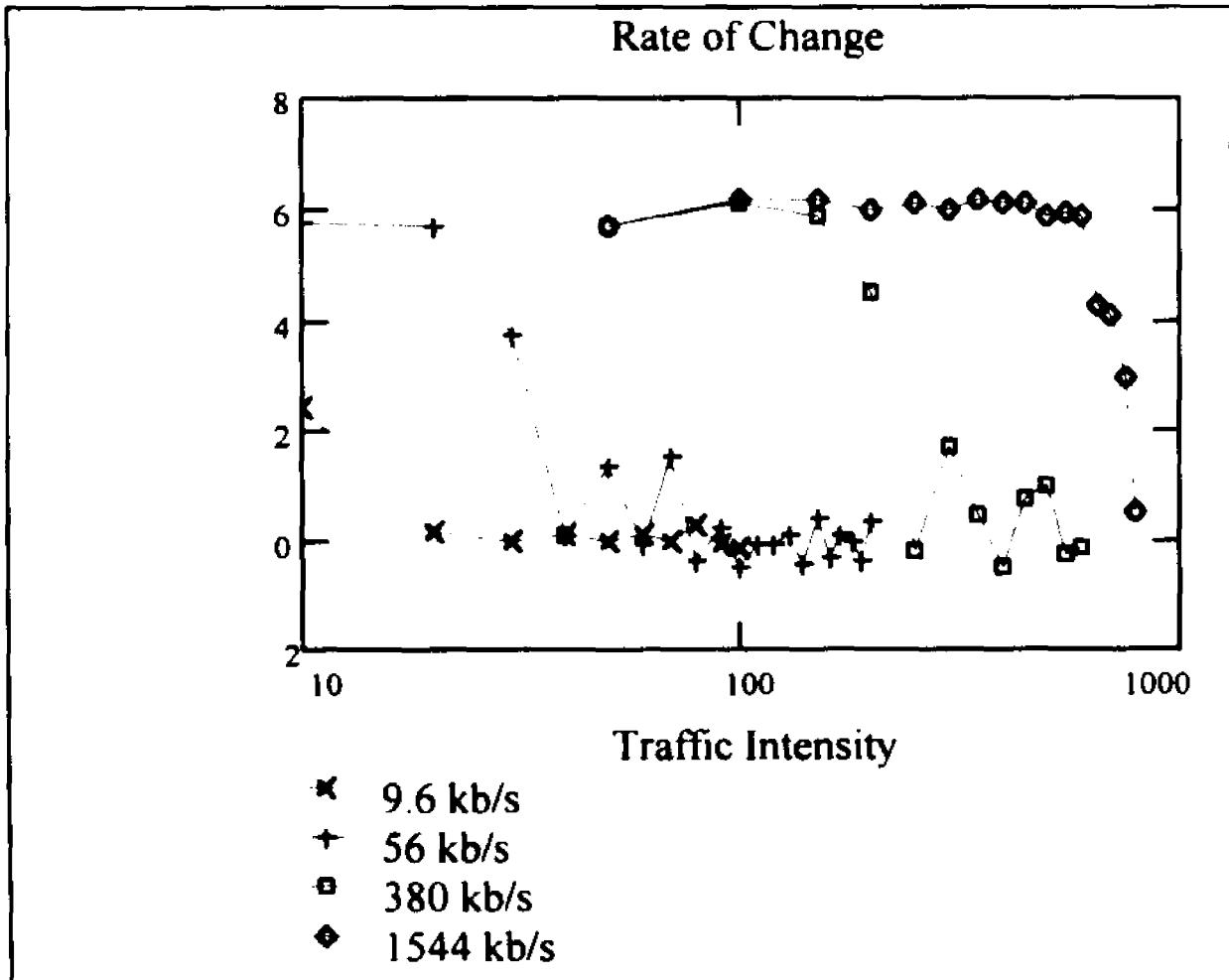


Figure 6.62 Rate of change of throughput (Model 1 - half duplex)

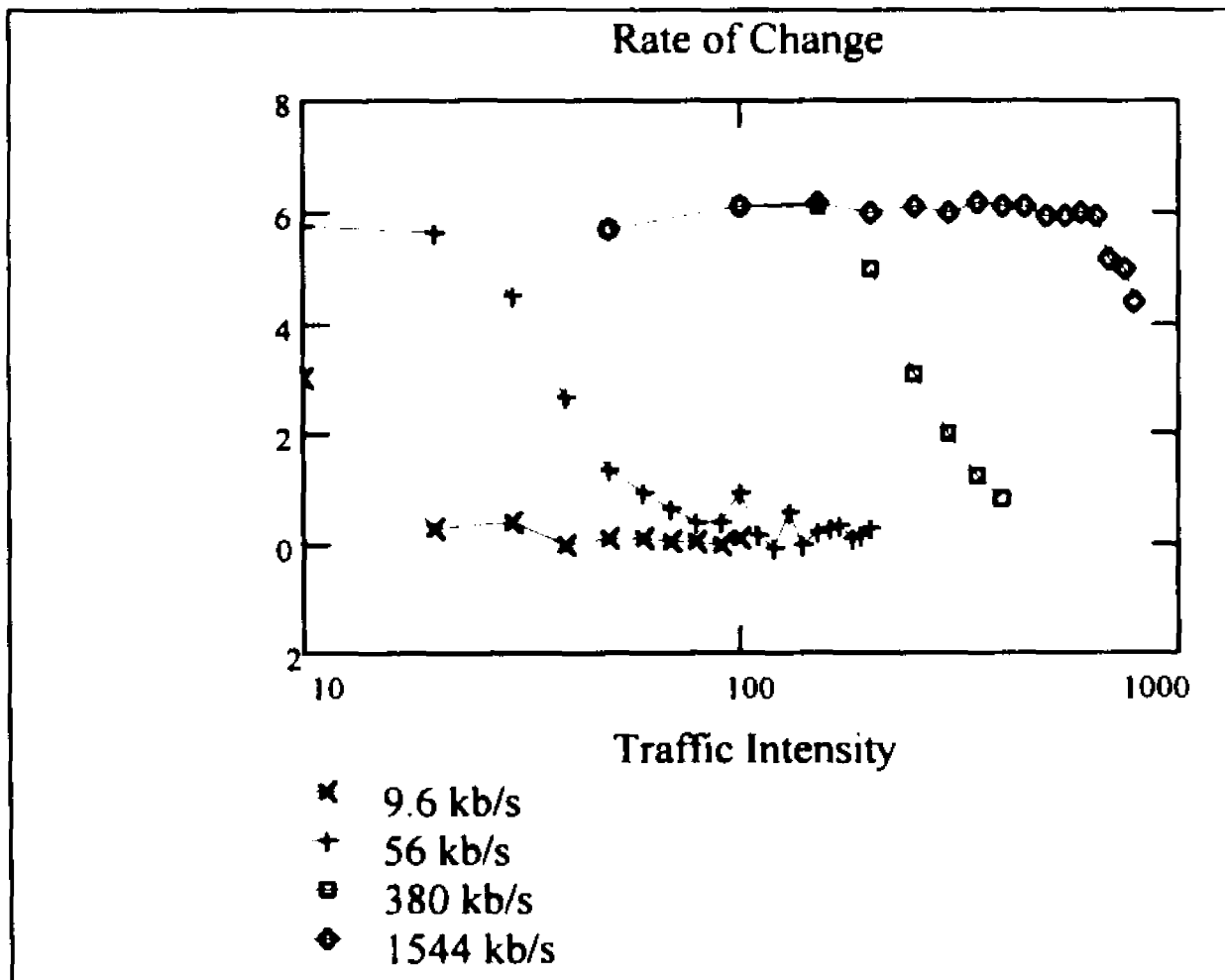


Figure 6.63. Rate of change of throughput (Model 1 - link failure)

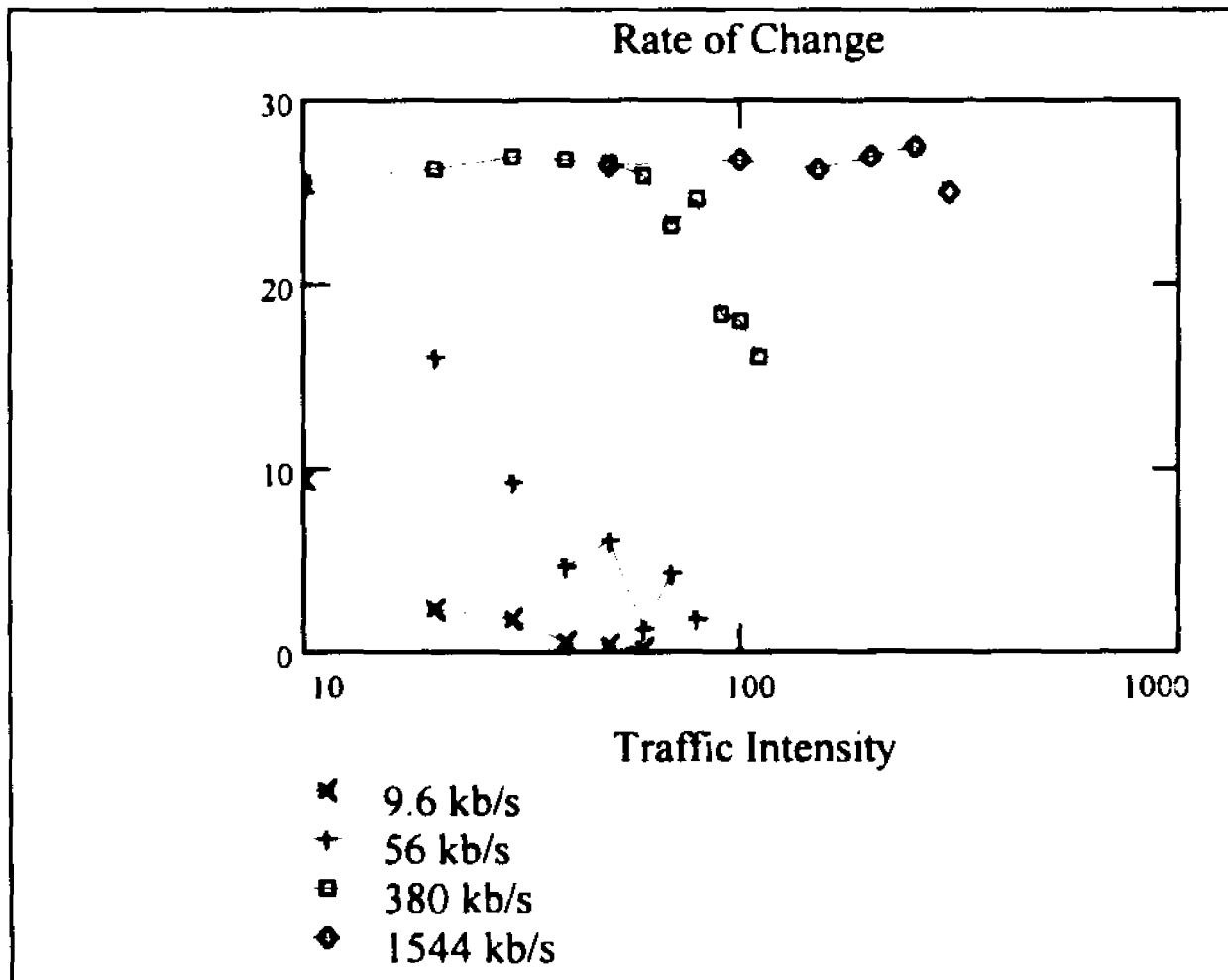


Figure 6 64 Rate of change of throughput (Model 2 - full duplex)

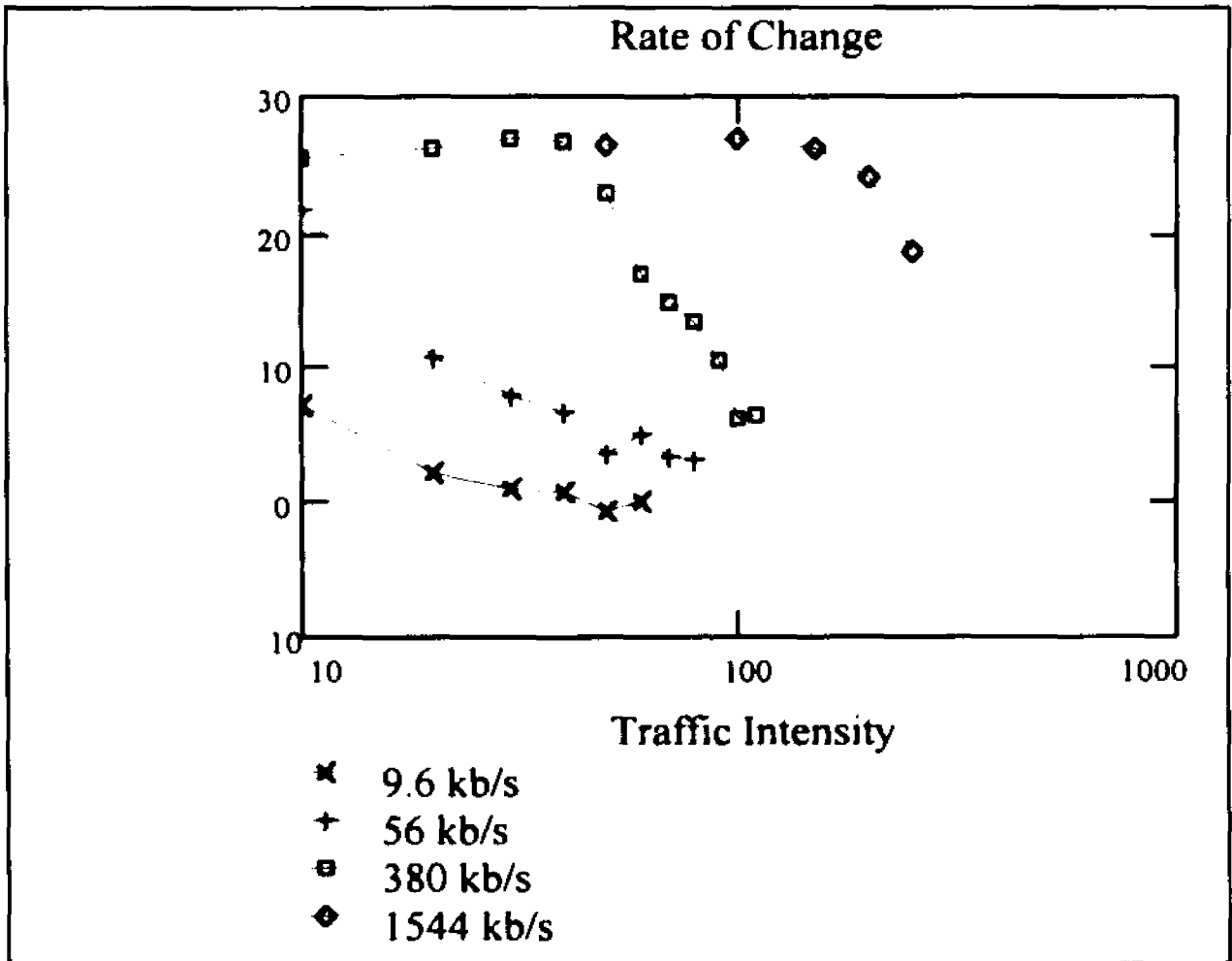


Figure 6 65 Rate of change of throughput (Model 2 - link failure)

Link Speed	Full Duplex		Half Duplex		Link Failure	
	Traf. Int.	Throughp.	Traf. Int.	Throughp.	Traf. Int.	Throughp.
9.6	--	--	--	--	--	--
56	20	116	10	59	10	59
380	250	1500	150	885	150	895
1544	1000	6081	600	3606	650	3613

Table 6. 2. Points of separation for Model One

Link Speed	Full Duplex		Link Failure	
	Traffic Intensity	Throughput	Traffic Intensity	Throughput
9.6	--	--	--	--
56	--	--	--	--
380	50	1324	30	789
1544	250	6709	150	3979

Table 6. 3 Points of separation for Model Two

From these figures, with the help of rate of change of delay times, the tables 6.2 and 6.3 are extracted for points of separation for the two simulation models. Once the point of separation is identified for a model with particular link properties, the equation of the line that approximates the throughput before the point of separation can be found by interpolating all the data points up to and including the point of separation. There are many interpolation methods, of which least square method is one. It is simple, straightforward and effective. It states that given points $x_1, \dots, x_n$ with corresponding points $y_1, \dots, y_n$, a function $y = mx + b$ can be found by solving the simultaneous equations:

$$\left(\sum_{k=1}^n x_k^2 \right) m + \left(\sum_{k=1}^n x_k \right) b = \sum_{k=1}^n y_k x_k$$

and,

$$\left(\sum_{k=1}^n x_k \right) m + nb = \sum_{k=1}^n y_k$$

with the total error for n points:⁹

$$e = \sum_{k=1}^n \left| ax_k + b - y_k \right|.$$

⁹W.Cheney, D. Kincaid, *Numerical Mathematics and Computing*, Brooks/Cole Publishing, Monterey, California, 1980, pp 267-269.

Link Speed (kb/s)	<i>Throughput = m_1 * Traffic Intensity + b_1 (kb/s)</i>					
	Half Duplex		Full Duplex		Link Failure	
	m_1	b_1	m_1	b_1	m_1	b_1
9.6	--	--	--	--	--	--
56	5.75	1.1	5.75	1.1	5.76	1.0
380	5.92	-2.7	6.02	-6.7	5.98	-4.9
1544	6.04	-5.9	6.07	-12.9	6.04	-5.2

Table 6. 4. Traffic intensity coefficients of throughput for Model One

Link Speed (kb/s)	<i>Throughput = m_1 * Traffic Intensity + b_1 (kb/s)</i>			
	Full Duplex		Link Failure	
	m_1	b_1	m_1	b_1
9.6	--	--	--	--
56	--	--	--	--
380	26.51	-3.3	26.22	-0.4
1544	26.77	-10.5	26.54	3.6

Table 6. 5: Traffic intensity coefficients of throughput for Model Two

Substituting Traffic Intensity for x and throughput for y , and solving for coefficients m_1 and b_1 , we get the results as shown in Table 6.4 for the first simulation model and Table 6.5 for the second simulation model. It must be noted that the magnitude for total error is dependent upon the number of points that is being used in the interpolation. Those entries that do not have any coefficients indicate that the point of separation is not observed for that network with that link speed and property.

From these two tables it can be observed that the slope m_1 approaches six for the first simulation model and twenty-six for the second simulation model. Since a link failure has the same effect as removing that link from the network, these results suggest that the relationship between traffic intensity and throughput before the point of separation is independent of the links of a network but dependent on the number of virtual circuits, which, in turn, depends on the number of nodes in the network.

These results imply that in a given a network with a certain node and virtual circuit configuration, the rate of increase of throughput, prior to the point of separation, is constant regardless of the link speeds or link properties. Furthermore, the throughput prior to the point of separation can be approximated by the following linear function:

$$\text{Throughput} = m_1 * \text{Traffic Intensity} + b_1 \quad (6-1)$$

Since the observations made above are only valid for throughput levels before the point of separation, the next step is to examine the relationship of throughput at the point of separation and link speeds. Using Table 6.2 for the first model and Table 6.3 for the second model, and using least squares methods, with link speeds substituted for x and throughput substituted for y , we get an equation in the form of :

$$\text{Throughput} = m_2 * \text{Link speed} + b_2 \quad (6-2)$$

This shows that the throughput level at the point of separation has a direct relationship with the link speeds. The coefficients for slope m_2 and intercept b_2 for the different link properties of the first and the second models and are shown in Tables 6.6 and 6.7 respectively

Consequently, this allows us to find the throughput at the point of separation given the speed of the links. It must be noted that for small link speeds the results are negative. This implies that for those speeds, any resulting throughput is after the point of separation and the throughput does not increase linearly with increased traffic.

Link Property	<i>Throughput = m_2 * link speed + b_2 (kb/s)</i>	
	<i>m_2</i>	<i>b_2</i>
Full Duplex	3.989	-66.9
Half Duplex	2.372	-48.7
Link Failure	2.374	-44.8

Table 6. 6. Link speed coefficients of throughput for Model One

Link Property	<i>Throughput = m_2 * link speed + b_2 (kb/s)</i>	
	<i>m_2</i>	<i>b_2</i>
Full Duplex	4.404	-90.4
Link Failure	2.74	-252

Table 6. 7 Link speed coefficients of throughput for Model Two

As a consequence of the two previous findings, we can set the two equations equal to each other and get the following equality:

$$m_1 * \text{traffic intensity} + b_1 = m_2 * \text{link speed} + b_2$$

when simplified, it yields:

$$\text{traffic intensity} = \frac{m_2}{m_1} * \text{link speed} + \frac{b_2 - b_1}{m_1}$$

This equation approximates the traffic intensity at the point of separation, given the speeds of the network links. It is important to note that this is an approximation since previous interpolations have introduced some error into the equations, and it depends on the topology, link properties (full duplex, half duplex or failure) and the number of the virtual circuits setup within a network.

Replacing m_2 / m_1 with m_3 and $(b_2 - b_1) / m_1$ with b_3 , Tables 6.8 and 6.9 summarize the coefficients for the two simulation models at the point of separation using the equation

$$\text{Traffic Intensity} = m_3 * \text{Link Speed} + b_3 \quad (6-3)$$

Link Speed (kb/s)	<i>Traffic Intensity</i> = $m_3 \cdot \text{Link Speed} + b_3$					
	Full Duplex		Half Duplex		Link Failure	
	m_3	b_3	m_3	b_3	m_3	b_3
9.6	--	--	--	--	--	
56	0.6847	1.1	0.4062	-0.4	0.4057	1.6
380	0.6532	-0.2	0.3952	-1.0	0.3906	0.5
1544	0.6486	-1.2	0.3869	-1.5	0.3866	0.4

Table 6. 8. Traffic Intensity coefficients at point of separation for Model One

Link Speed (kb/s)	<i>Traffic Intensity</i> = $m_3 \cdot \text{Link Speed} + b_3$			
	Full Duplex		Link Failure	
	m_3	b_3	m_3	b_3
9.6	--	--	--	--
56	--	--	--	--
380	0.1661	-3.5	0.1045	-9.6
1544	0.1645	-3.8	0.1032	-9.4

Table 6.9 Traffic Intensity coefficients at point of separation for the Model Two

An examination of the coefficients for a particular link property show that m_1 , regardless of the link speeds, converges. This demonstrates that a linear relationship exists between the link speed and the traffic intensity at which the point of separation occurs. These results, when extended to other link speeds, approximations of traffic intensity and throughput are found for the point of separation, which is a useful tool in network design and planning.

We also observe that for the second model, the difference of values of m_1 (which determines the location of point of separation) between the full duplex and link failure is smaller than for the first model. This indicates that a link failure in a smaller network has a bigger impact on throughput than a larger network since in the larger network a) there are more nodes that are not affected by the failure and b) there are more alternate paths for rerouted messages. Hence the larger networks are more robust and can tolerate link failure better.

6.3.2 Throughput Saturation

We examine the part of the throughput curve after the point of separation. As discussed earlier, after the point of separation the rate of change of throughput starts decreasing until it reaches approximately zero. Once it reaches zero, the throughput is in its saturation level, not changing even if the amount of traffic is increased. The point where the rate of change of throughput reaches zero, is called the point of saturation. In this section we will examine the behavior of throughput between point of separation and point of saturation.

Observing the natural system of electromagnetism, we see that the last part of a magnetization curve of a ferro-magnetic material follows a similar pattern as the network throughput curve. As the force of the magnetic field increases the magnetism of the material also increases approaching an asymptote, after which any increase in the magnetic field has no effect upon the increase of magnetization of the material. The magnetization curve differs for materials with different magnetic properties. The magnetization curve for a ferro-magnetic material can be approximated by a curve that has the equation:

$$T = \frac{p * M}{q + M} \quad (6-4)$$

where T is the magnetization of the material and M is the force of the magnetic field with constants p and q .¹⁰

To find these constants, we can rewrite this equation as :

$$T + q = p * \frac{M}{T} \quad (6-5)$$

This linear relationship has its y-intercept as $-q$ and its slope as $1/p$. Furthermore we observe that a y-axis asymptote for equation 6-4 occurs at p . For the throughput curve between the point of separation and point of saturation, we substitute throughput for magnetization of material (T) and traffic intensity for magnetic field (M). From this substitution, we get the following equation.

$$\text{Throughput} + q = p * \frac{\text{Traffic Intensity}}{\text{Throughput}} \quad (6-6)$$

which approximates the maximum throughput a given configuration can reach.

However, before we attempt to find the values for p and q , the point of saturation of throughput must be located. To achieve this, we again use the rate of change curves. The point of saturation occurs at a point where the rate of change of throughput and the rate of change of delay times both approach zero. For the figures that do not satisfy both

¹⁰ Cullwick E.G , The Fundamentals of Electro-Magnetism (Cambridge. The University Press, 1966), p 197

conditions, due to lack of data points, the last measured point is considered the point of saturation and is a source of error especially for 380 kb/s and 1544 kb/s links. Hence Figures 6.53 - 6.60 will produce the following table

Link speed (kb/s)	Traffic Intensity	Throughput
9.6	40	57.6
56	160	364.2
380	600	2028.0
1544	1200	6719.3

Table 6.10: Point of saturation for full duplex links (Model One)

Tables 6.11 and 6.12 show the points of saturation for different link properties for the first and the second simulation model respectively. Combining Tables 6.2 and 6.11, and 6.3 and 6.12 we delineate the regions of throughput between point of separation and point of saturation for the first and second simulation models. To determine the values for p and q , we examine the curves resulting from plotting Traffic Intensity vs (Traffic Intensity / Throughput) and interpolate the slope and the y-intercept of the resulting graphs. Figures 6.66 - 6.71 plot these curves and the resulting approximated and actual throughput curves for link speeds of 56, 380 and 1544 kb/s for the first simulation model with full duplex links.

Link Speed (kb/s)	Full Duplex		Half Duplex		Link Failure	
	Traff Int	Throughp.	Traff. Int	Throughp.	Traff. Int.	Throughp
9.6	40	57.6	40	30.0	40	40.1
56	160	364.2	110	173.9	140	237.2
380	600	2028.0	600	1244.2	700	1619.1
1544	1200	6719.3	900	4363.1	1050	5424.7

Table 6.11: Points of saturation for Model One

Link Speed (kb/s)	Full Duplex		Link Failure	
	Traffic Intensity	Throughput	Traffic Intensity	Throughput
9.6	50	160.1	60	122.6
56	80	690.0	80	627.8
380	150	3048.7	150	2348.9
1544	300	7958.7	250	6107.5

Table 6.12: Points of saturation for Model Two

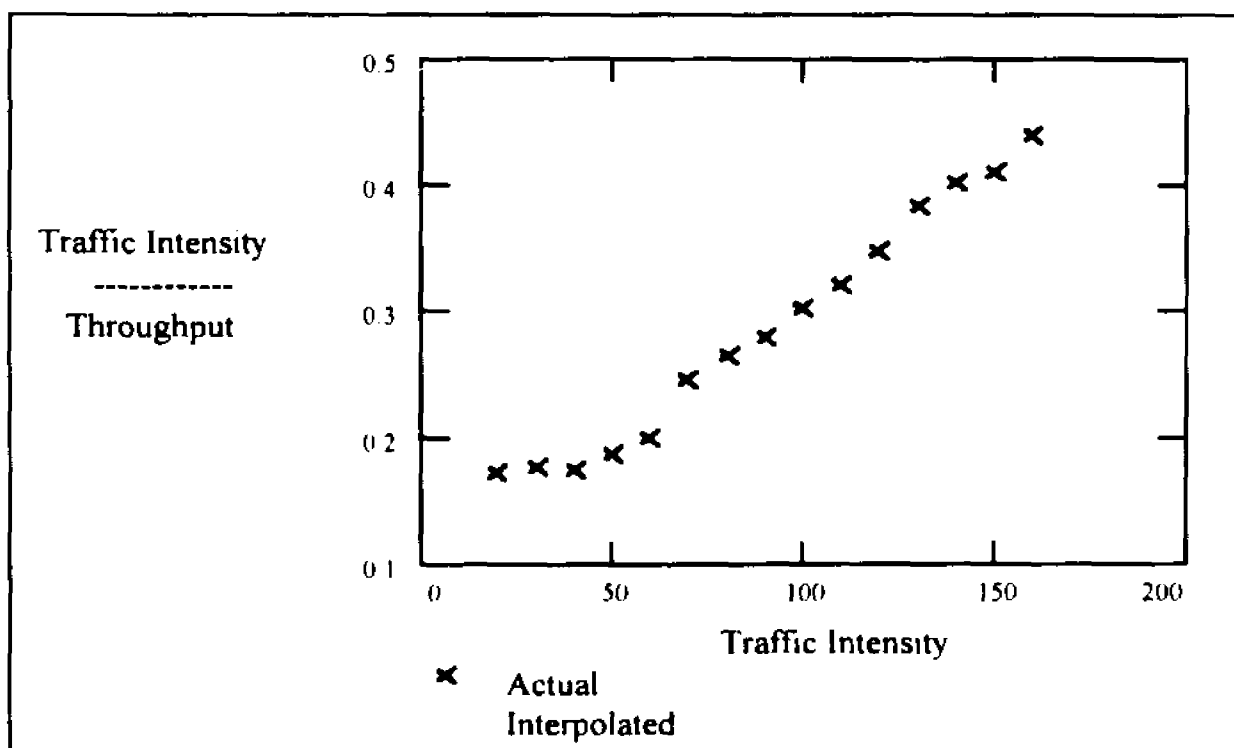


Figure 6.66 Throughput Interpolation for 56kb/s full duplex links - Model One

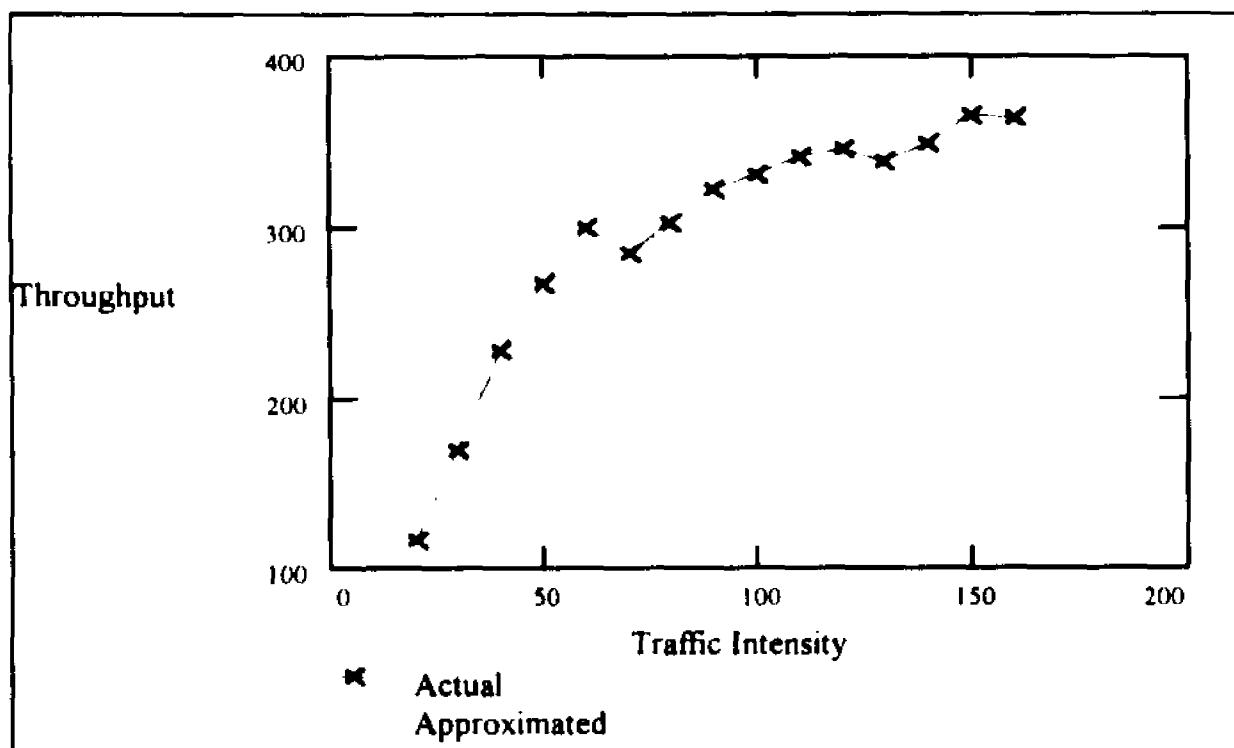


Figure 6.67 Throughput Approximation for 56kb/s full duplex links - Model One

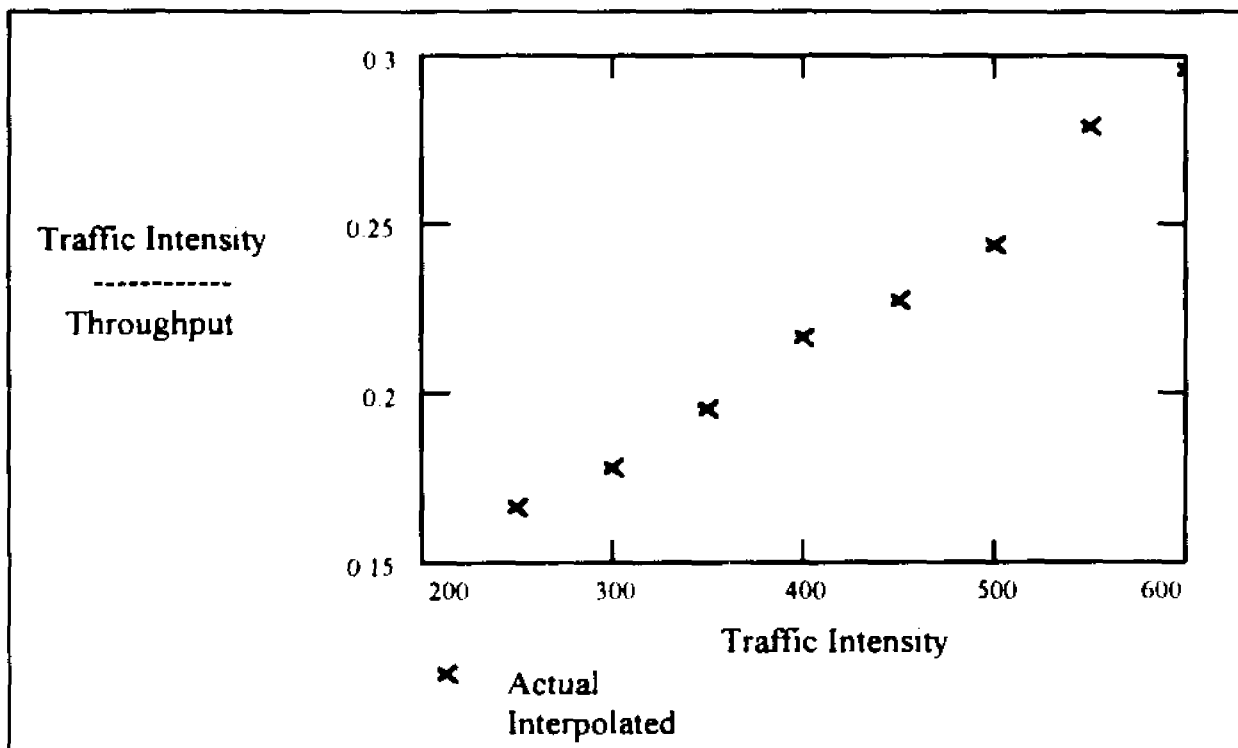


Figure 6.68. Throughput Interpolation for 380 kb/s full duplex links - Model One

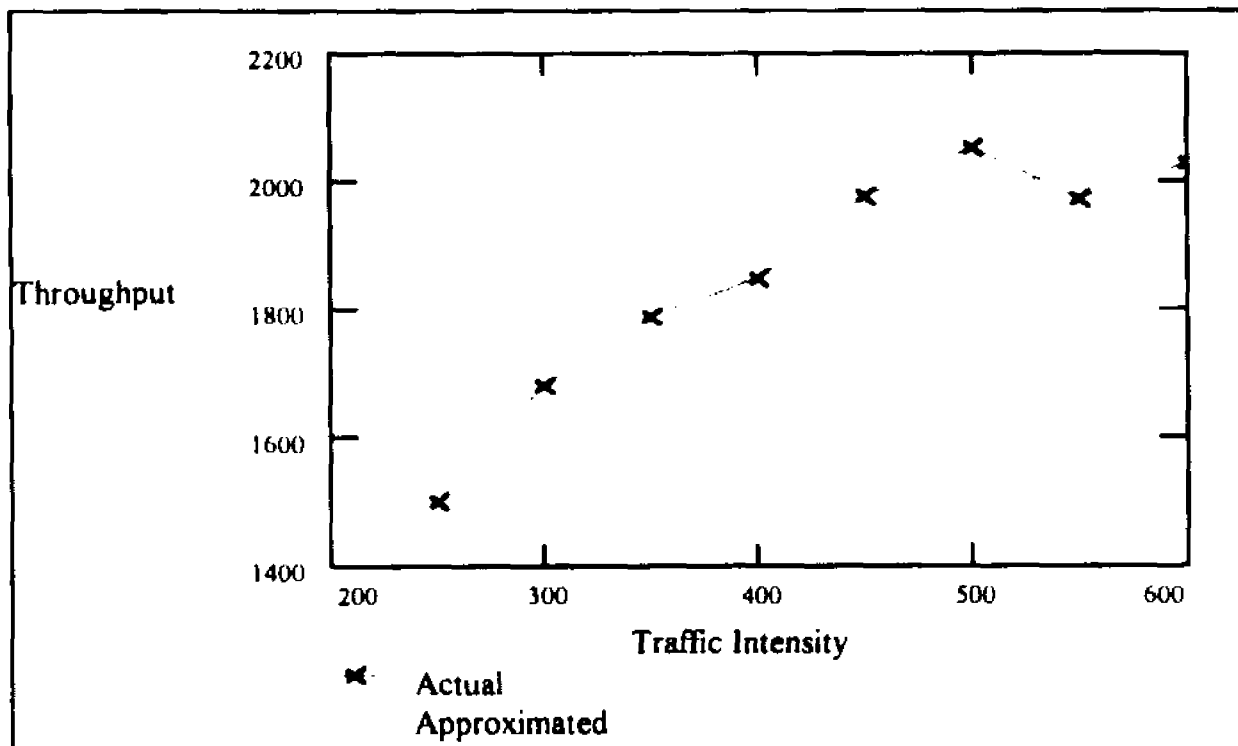


Figure 6.69. Throughput Approximation for 380kb/s full duplex links - Model One

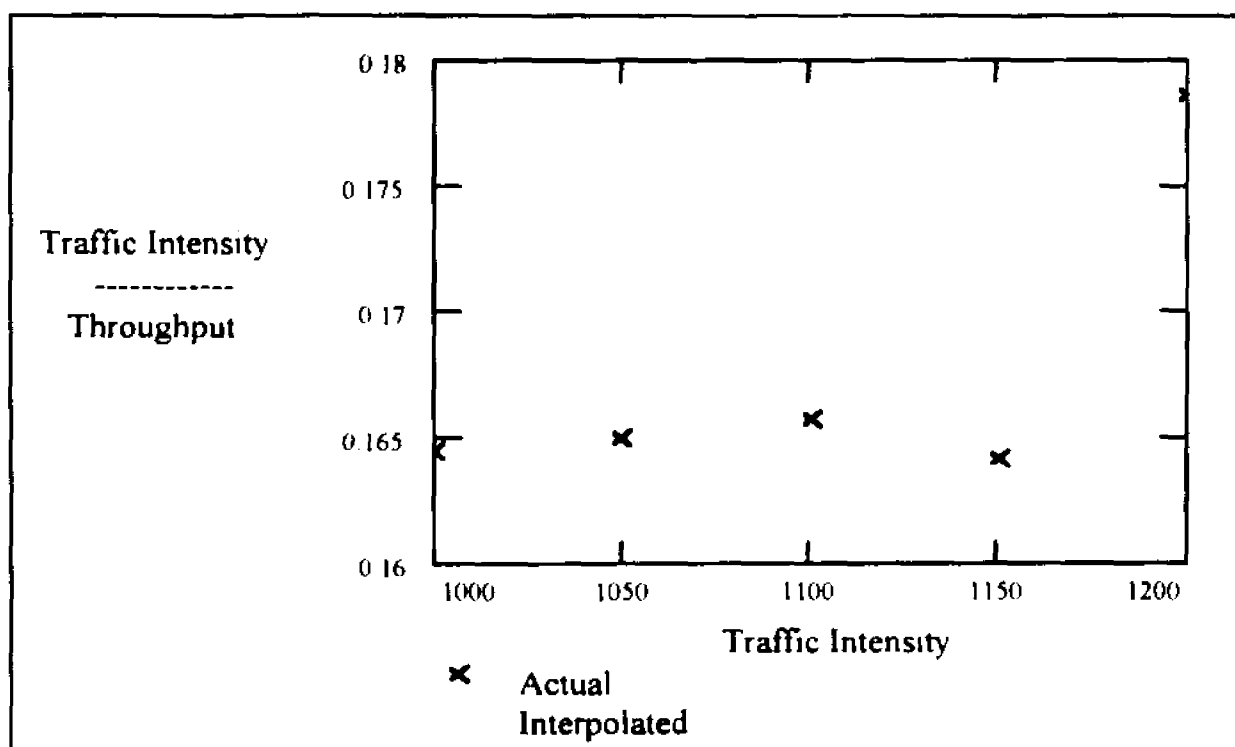


Figure 6.70: Throughput Interpolation for 1544 kb/s full duplex links - Model One

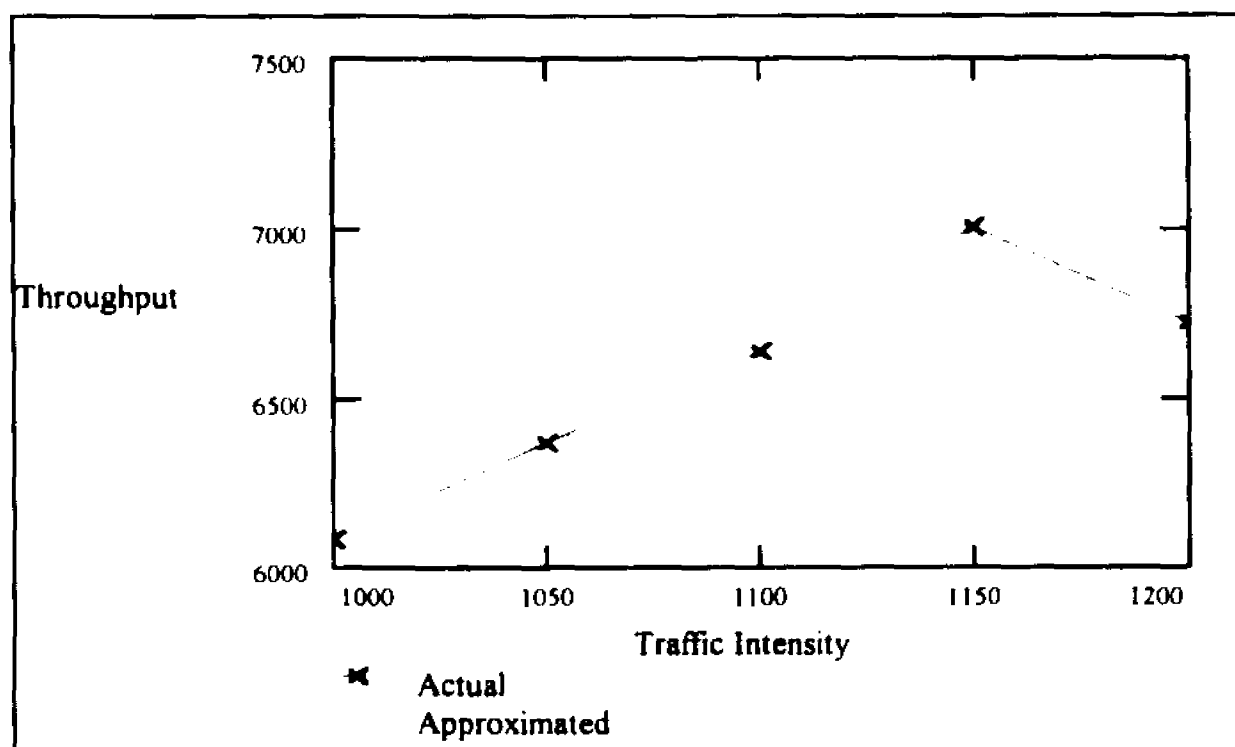


Figure 6.71: Throughput Approximation for 1544 kb/s full duplex links - Model One

Link Speed (kb/s)	$\text{Throughput} = \frac{p \cdot \text{Traffic Intensity}}{q + \text{Traffic Intensity}} \quad (\text{kb/s})$					
	Full Duplex		Half Duplex		Link Failure	
	p	q	p	q	p	q
9.6	62.8	0.04	32.17	0.07	45.46	-0.10
56	482.1	0.1	195.42	0.05	287.76	0.09
380	2687.3	0.1	1407.5	0.05	1980.3	0.1
1544	18210	0.1	8573.8	0.09	14528	0.13

Table 6.13. Throughput coefficients for point of saturation for Model One

Link Property	$\text{Throughput} = \frac{(m_s \cdot \text{link speed} + b_s) \cdot \text{Traffic Intensity}}{\text{Traffic Intensity}}$	
	m_s	b_s
Full Duplex	6.983	40.2
Half Duplex	3.724	-8.1
Link Failure	5.224	-4.7

Table 6.14: Rate of change of p coefficients for Model One

Link Speed (kb/s)	$Throughput = \frac{p * Traffic\ Intensity}{q * Traffic\ Intensity}$			
	Full Duplex		Link Failure	
	p	q	p	q
9.6	178.4	0.03	132.2	0.03
56	898.5	0.03	843.9	0.04
380	--	--	4278.1	0.03
1544	--	--	--	--

Table 6.15: Throughput coefficients for point of saturation for Model Two

Link Property	$\text{Throughput} = \frac{(m_s \cdot \text{link speed} + b_s) \cdot \text{Traffic Intensity}}{\text{Traffic Intensity}}$	
	m_s	b_s
Full Duplex	15.519	29.4
Link Failure	10.974	121.3

Table 6.16: Rate of change of p coefficients for Model Two

Tables 6.13 and 6.15 summarize the values of p and q for the two simulation models.

A closer examination of the coefficients of p in those tables indicates that there is a linear relationship between the coefficients and the data rate of the links. Using the least squares method, we find the slope m_1 and the y-intercept b_1 as shown in tables 6.14 and 6.16. From these tables, given the link speed, we approximate p for the throughput after the point of separation and the maximum throughput for the model. To simplify the approximations, the q coefficient is eliminated since it is a constant. Hence, the throughput of the network between the point of separation and the point of saturation has a linear relationship with the data rates of the links in the network. This relationship is

$$\text{Throughput} = \frac{(m_1 * \text{link speed} + b_1) * \text{Traffic Intensity}}{\text{Traffic Intensity}} \quad (6-7)$$

or

$$\text{Throughput} = \frac{(m_1 * \text{link speed} + b_1)}{\text{Traffic Intensity}} \quad (6-8)$$

Using the results from Table 6.13 and 6.14, we extrapolate the throughput curve for each model and find the maximum throughput for each model. Combining these curves with the previous curves for point of saturation, we approximate the complete throughput curve for a particular network model with given link properties and data rate for any traffic intensity. To do this for a given model with given link properties, we use the following algorithm:

1. **Input** Traffic Intensity and Link speed
2. **If** Traffic Intensity $\leq m_i \cdot \text{Link Speed} + b_i$
then
 - 2a. Throughput := $m_i \cdot \text{Traffic Intensity} + b_i$**else**
 - 2b. Throughput := $(m_i \cdot \text{Link Speed} + b_i) + (m_s \cdot \text{Link Speed} + b_s) / \text{Traffic Intensity}$

Using this algorithm, an approximated throughput curve for a model - given the link speeds and properties (full duplex, half duplex or link failure) - is produced, such as the one in Figure 6.72 for 56kb/s links with link failure

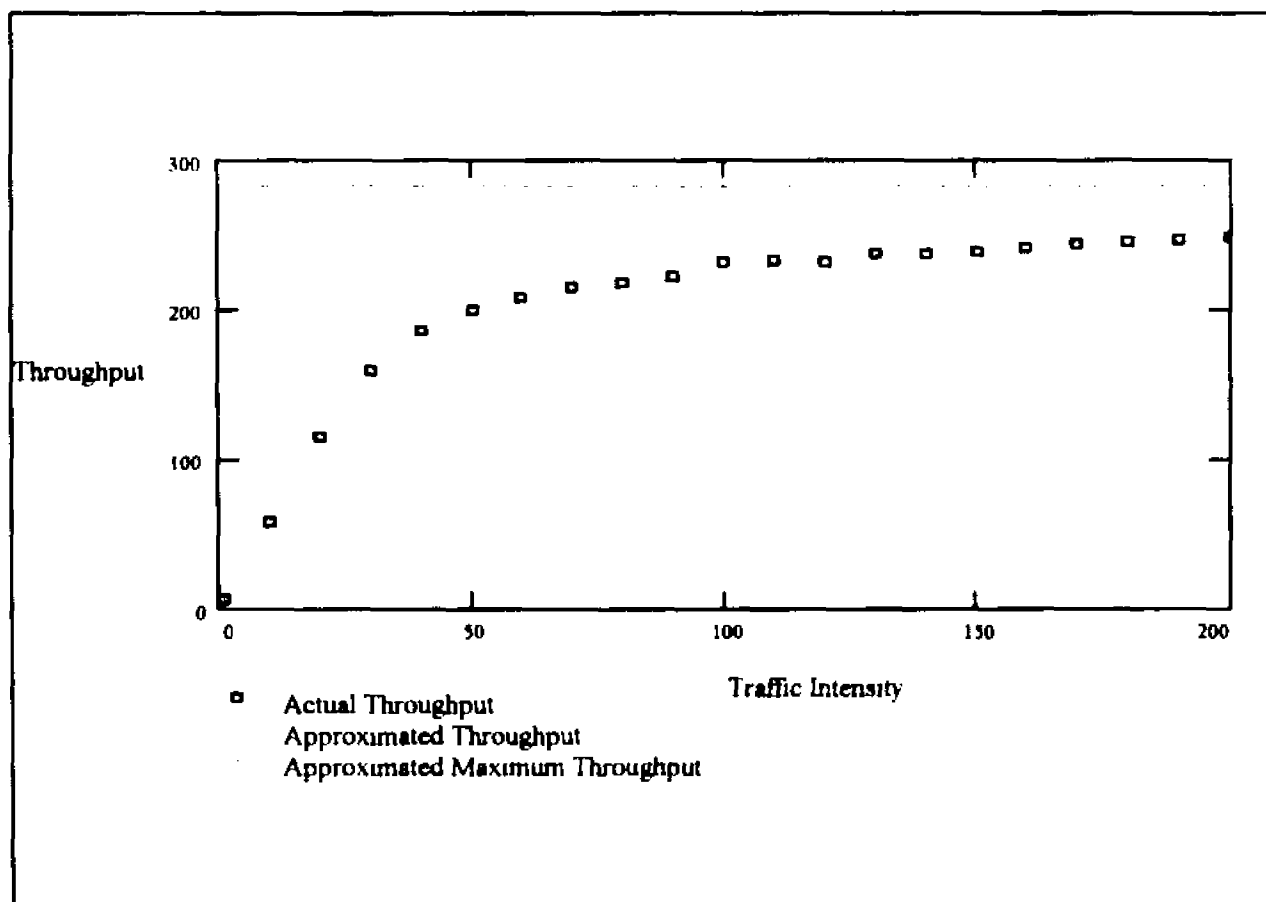


Figure 6.72: Throughput Approximation Curve for 56 kb/s with link failure (Model One)

6.3.2 Other Factors That Affect Performance

The approximated throughput is less than the theoretical maximum which is sum of the capacities of all the links in a network. The difference between the theoretical maximum and the maximum achievable throughput is due to the topology of a network. Since the simulation models are not fully connected, a message encounters delays and bottlenecks at the nodes along its route. This causes the throughput to be lower than for a network where the messages do not encounter any delays between the source and destination nodes. The degree of connectivity of a network, as defined in section 3.1.3.1, affect the delay times which consequently impact the throughput of the network. The throughput is maximized and delays are minimized in a fully connected network, where each node in the network is directly connected to every other node. However, for large number of nodes, this becomes economically not feasible.

Different degrees of network connectivity yield different shapes of throughput curves which may be considered a network topology's signature. In our studies, this is evident in the throughput curves for the first simulation model where the actual throughput drops off at a certain point even though the traffic intensity increases. This is due to the lower number of messages delivered to the destination at higher traffic intensities. The increase in number of messages in all virtual circuits causes the transmission queues in highly used links to grow, thus decreasing the number of messages delivered per unit time. This reduces the overall throughput of the system. As the traffic further intensifies, rerouting the messages through those links that have lighter loads temporarily increases the

throughput until those links also become saturated. The throughput increase that takes place with higher traffic intensities diminishes as evidenced by the curves in chapter five.

Furthermore, those nodes that have single link to the rest of the network experience congestion at smaller traffic intensities than those nodes that have more than one link, providing further proof that better connectivity reduces congestion and leads to higher throughput and better delay times.

VII. CONCLUSIONS

7.1 Summary of Findings

From the results of the simulation studies it is concluded that throughput, delay times and Quality of Service (QoS) are related to each other very closely. For example, (QoS) deteriorates very rapidly as the delay times increase and delay times increase as throughput increases. In virtual circuit switching, the traffic intensity depends on the number of virtual circuits, which in turn depends on the number of nodes in a network. As the traffic intensity between a set of nodes increases, the throughput of the system increases linearly with the approximating equation $y = mx + b$ where m and b are constants and x is traffic intensity and y is throughput. This increase in throughput, has little or no effect on the delay times and QoS. During this increase delay times remain low, and QoS acceptable.

As the traffic intensity further increases, the throughput ceases to increase linearly. The delay times start to increase and QoS degrades. These changes in the performance parameters of the network coincide at a point which is called the point of separation. The results show that there is a linear relationship between the link speeds and the traffic intensity at the point of separation. After the point of separation, the throughput of the network approaches a maximum which is called the point of saturation. The throughput curve between the points of separation and saturation may be approximated by a hyperbolic equation $y = \frac{px}{q + x}$, where p and q are constants and x is traffic intensity and y is throughput. The constant p is an asymptote which represents the maximum

throughput for a network. It also has a linear relationship to the link speeds. As the throughput approaches the point of saturation, the delay times also approach a maximum level. We then conclude that for a network with given properties (topology, nodes, links etc.), the overall throughput as well as the message delays approach a maximum as the traffic intensity increases

Another major factor affecting the performance of a network is its topology. The extent of connectivity of the nodes dictates the shape of the throughput curve. For example, different topologies of networks produce traceable features (i.e. peaks and valleys) in the throughput curves

In conclusion, our results have shown that results from one experiment leads to the prediction of results for other experiments with similar input parameters. Specifically, the point of separation is predicted for a network with a given topology for different link speeds. Additionally, partial results from the experiments approximate the saturation point for all sets of throughput curves at different link rates. These results have also shown that when designing a network, there are factors which must be considered very closely. These factors are the number of nodes, the number of links, the properties of the links, the link speeds and the topology or the connectivity of the network. Consequently, the points of separation and saturation in conjunction with the factors mentioned above establish a parameter list which is used to predict the behavior of a network

7.2 Further Research Areas

Further research can be conducted in the following areas:

- The effect of different topologies on the performance of a network
- The effect of partitioned traffic patterns on the performance of a network
- The effect of different routing algorithms and metrics on the performance of a network
- The effect of blocking on the performance of a network
- The effect of using various link speeds on the performance of a network
- The effect of using links other than point-to-point links on the performance of a network.

APPENDIX

Queuing Theory Concepts

The following formulae assume that the arrival times are distributed according to Poisson distribution for a single server where

- a. the probability of the occurrence of a success or failure is constant
- b. the probability of a success or failure is independent of what has previously happened
- c. the probability of an occurrence or a change during time interval t is approximately proportional to t , provided that t is small.
- d. the probability of more than one occurrence or change during any small interval t is approximately zero when compared to the probability of single change during that interval

Probability that zero units are in the system when the rate of arrival λ and service μ given:

$$P_0(t) = 1 - \frac{\lambda}{\mu}$$

Probability that n units in queue plus the one being serviced

$$P_n(t) = \left(\frac{\lambda}{\mu}\right)^n P_0(t)$$

Expected number being serviced plus waiting:

$$E_n = \frac{\lambda}{\mu - \lambda}$$

Expected number in queue - length of waiting line not including those being serviced:

$$E_w = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

Expected waiting time in line

$$E_t = \frac{E_w}{\lambda}$$

Expected time in system

$$E_s = E_t + \frac{1}{\mu}$$

Probability that the number in the queue plus the number being serviced is greater than n .

$$P(N > n) = \left(\frac{\lambda}{\mu} \right)^{n+1}$$

REFERENCES

- S. V. Ahamed, Intelligent Networks Encyclopedia of Telecommunications, New York Academic Press, 1989.
- S. V. Ahamed, V. B. Lawrence, Transmission Aspects in ISDN, ISDN, DECNET and SNA Communications, Cambridge: Harvard University, 1990.
- V. Ahuja, Design and Analysis of Computer Communications Networks, New York: McGraw Hill, 1982.
- A. O. Allen, Probability, Statistics and Queueing Theory with Computer Science Applications, New York: Academic Press 1978.
- C. Antonelli, The Diffusion of Advanced Telecommunications in Developing Countries, Paris: OECD Publications, 1991.
- E. Ayanoglu, J. Wand, Priority Statistical Multiplexer Design for SNA/SDLC Access to a Virtual Circuit Packet Network, AT&T Technical Journal, November-December 1988 pp. 24-28.
- P. Bratley, B. L. Fox, L. Schrage, A Guide to Simulation, Heidelberg: Springer-Verlag, 1983.
- E. W. Cheney, Numerical Mathematics and Computing, New York: Brooks/Cole Publishing Company, 1980.
- H. Cravis, Communications Network Analysis, Lexington Books, 1981.
- E. G. Cullwick, The Fundamentals of Electro-Magnetism, Cambridge: Van Nostrand Company Inc 1967.
- J. J. Degan, G. W. R. Luderer, A. R. Vaidya, Fast Packet Technology for Future Switches, AT&T Technical Journal, March-April 1989 pp. 36 - 58.
- R. Freeman, Telecommunication System Engineering. Analog and Digital Network Design, New York: John Wiley and Sons, 1980.
- Heiber, An Overview of Universal Information Services: Concepts and Technologies of Future Networks, AT&T Technical Journal, March-April 1989 pp. 5-13.

T.J. Herr, T.J. Plevyak, ISDN in the USA: The Opportunity Begins: Networks for the 1990s, New York: John Wiley and Sons, 1988.

P. G. Hoel, S.C. Port, C. J. Stone, Introduction to Statistical Theory, Boston: Houghton Mifflin Company, 1971.

R. V. Hogg, A. T. Craig, Introduction to Mathematical Statistics Fourth Edition, New York: Macmillan Publishing Co 1978

R. Hueber, Integrated Broadband Communications: The Programme RACE, Networks for the 1990s, New York: John Wiley and Sons, 1988.

S. L. S. Jacoby, J. S. Kowalik, Mathematical Modelling with Computers, Englewood Cliffs, New Jersey: Prentice Hall Inc. 1980

R. Jain, Congestion Control in Computer Networks: Issues and Trends, IEEE Network Magazine, May 1990 pp 24-30

H. Kobayashi, Modeling and Analysis: An Introduction to System Performance Evaluation Methodology, Reading: Addison Wesley Publishing Company, 1978

R. Lal Sharma, P.J.T. de Sousa, A.D. Ingle, Network Systems Modelling Analysis and Design, New York: Van Nostrand Reinhold, 1982.

I. Mitrani, Modelling of Computer and Communications Systems, London: Cambridge University Press, 1987.

J. Panico, Queueing Theory, New York: Prentice Hall, 1969.

M. Rodrigues, Evaluating Performance of High Speed Multiaccess Networks, IEEE Network Magazine, May 1990 pp. 36-40.

H. Rosner, Packet Switching, New York: Van Nostrand Reinhold, 1982

M. Schwartz, Telecommunication Networks Protocols, Modeling and Analysis, New York: Addison Wesley, 1987.

A. Tanenbaum, Computer Networks, New York: Prentice Hall, 1988.

Telecommunication Network Based Services, Paris: OECD Publications, 1989

The Usage of International Data Networks in Europe, Paris: OECD Publications, 1979