

Winter 1990

Performance Improvements for FDDI and CSMA/CD Protocols

David Earl Game
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_etds



Part of the [Computer Sciences Commons](#)

Recommended Citation

Game, David E.. "Performance Improvements for FDDI and CSMA/CD Protocols" (1990). Doctor of Philosophy (PhD), dissertation, Computer Science, Old Dominion University, DOI: 10.25777/d8vc-0z35
https://digitalcommons.odu.edu/computerscience_etds/102

This Dissertation is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

PERFORMANCE IMPROVEMENTS FOR FDDI AND CSMA/CD PROTOCOLS

by

David Earl Game

B.S. June 1973, Massachusetts Institute of Technology

M.S. June 1978, College of William and Mary

A Dissertation submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

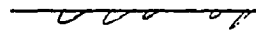
DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

OLD DOMINION UNIVERSITY

December, 1990

Approved by:

 Kurt Maly, Advisor

Copyright by David Earl Game 1990
All Rights Reserved

Abstract

PERFORMANCE IMPROVEMENTS FOR FDDI AND CSMA/CD PROTOCOLS

David Earl Game
Old Dominion University, 1990
Advisor: Kurt Maly

The High-Performance Computing Initiative from the White House Office of Science and Technology Policy has defined 20 major challenges in science and engineering which are dependent on the solutions to a number of high-performance computing problems. One of the major areas of focus of this initiative is the development of gigabit rate networks to be used in environments such as the space station or a National Research and Educational Network (NREN).

The strategy here is to use existing network designs as building blocks for achieving higher rates, with the ultimate goal being a gigabit rate network. Two strategies which contribute to achieving this goal are examined in detail. ¹

FDDI² is a token ring network based on fiber optics capable of a 100 Mbps rate. Both media access (MAC) and physical layer modifications are considered. A method is presented which allows one to determine maximum utilization based on the token-holding timer settings. Simulation results show that employing the second counter-rotating ring in combination with destination removal has a multiplicative effect greater than the effect which either of the factors have individually on performance. Two 100 Mbps rings can handle loads in the range of 400 to 500 Mbps for traffic with a uniform distribution and fixed packet size. Performance is dependent on the number of nodes, improving as the number increases. A wide range of environments are examined to illustrate robustness, and a method of implementation is discussed.

¹This work was supported by CIT grant RF-89-002-01, NASA grant NAG-1-908 and Sun Microsystem grant RF 596043.

²Fiber Distributed Data Interface (FDDI) is an American National Standards Institute standard for token ring networks.

DRAMA³ is a broadband solution to the inherent limitations of an Ethernet network designed to be used in a Metropolitan Area Network (MAN), and is shown to be an effective strategy for extending CSMA/CD⁴ techniques to longer distances, larger numbers of nodes and synchronous traffic. Although implementation requires a large number of transmitter/receivers, a method is proposed which overcomes this problem without affecting the advantages shown. The limitation is one of the availability of bandwidth to a node, not the limitation of the total bandwidth of the network.

³Dynamic Resource Allocation in Metropolitan Areas (DRAMA) is an Ethernet-type network described in Chapter 5 and Appendix B.

⁴Carrier Sense Multiple Access / Collision Detection (CSMA/CD) is a network protocol used by Ethernet, and is discussed in Section 1.4.1.

Acknowledgements

I would like to express my deep appreciation to a number of people for their support during this academic marathon. First of all, I would like to thank my advisor, Kurt Maly, for the enduring patience he has shown, for drive which he instilled in me because many times I think it was indeed his and not mine that kept me going, and for the time he has invested in teaching me what research is all about. I would also like to thank the rest of my committee composed of Mike Overstreet, Ravi Mukkamala, Ed Foudriat, David Doughty and Nick Murray for their constant support; a special thanks to Mike Overstreet for his friendship and kind nature.

To my fellow graduate students Frank Paterra, Sanjay Khanna, LiPing Zhang and Vijay Kale I would like to express my appreciation for your friendship over the years and hopes for success in the pursuit of your degrees.

Without the support of the Webbs, George and Jane, I could never have been able to find time from my job to complete this degree, but the emotional encouragement which I received from them was even more important in bringing this effort to a finish.

To the one person who showed me how exciting learning really is, my high school math teacher, Neil Drummond, I would like to say that you were probably the single greatest influence on my academic life.

I would like to thank my parents, Earl and Peggy Game, for the trust that they have always had in my ability to do the right thing, and for giving me the room to fail in order to learn how to stand on my own.

Last and most important, I would like to dedicate this to my family: my wife, Betsy, and my children, Melissa, Alison and Aaron. Not only have they been an encouragement and a motivation for me, but they have also sacrificed the most by giving up our time together, something that I will never be able to give back. I love each and every one of you, and I promise that I will return to you the same support for fulfilling *your* dreams.

Preface

Throughout the history of mankind, numerous factors have been instrumental in shaping life as we know it. Many of the major advances in improving our way of life are a direct result of advancement in the ability to communicate. For centuries, the only methods of communication were verbal or hand-written symbols. With the invention of the printing press came the opportunity for groups of people to express their ideas in a medium which would not only last for lifetimes, but could be disseminated with great expediency to the masses. New ideas of religion, science, mathematics, politics, philosophy, and expressions of the human spirit in the form of literature and art spread across continents and had a significant effect on the evolution of our civilization.

An examination of recent history reveals the effects of communication on our lives as evidenced by the international move to democracy which citizens in the United States would have never dreamed of seeing in this century. Much of the stimulation for this change was a result of discovering the advantages of a democratic society through communication with the western world. In order to compete in the global markets of today, countries must open their doors to technology and education, but once the door is opened to communication, it is very difficult, if not impossible, to limit the flow of ideas to science alone. Even in societies which are as open to communication as is the United States, mass media such as television, newspapers, radio and movies are a major factor in shaping public opinion.

As all communication in the world becomes more digital in nature, research such as that described in this dissertation will contribute significantly to the way we communicate and the way we educate, as will other advances in hardware such as computing speed and memory capacities. Collectively, these advances provide a potential for disseminating information far beyond anything experienced in the history of man. We must continually evaluate the use of this communications technology, balancing the freedoms inherent therein with the limits of what we find socially, morally and politically acceptable. We are at the crossroads of developing a technology which is likely to have profound consequences which

we are unable to foresee.

This technology also brings with it a responsibility for the computing industry to consider its impact on the society as a whole, for the impact will be similar to that of nuclear energy in the mid-years of the twentieth century. In the past, the weakness of the communications infrastructure has actually proved to be a safeguard from the harmful sharing of private information about people and organizations. In the future, computer scientists, lawyers, politicians, and professionals from many walks of life must address these problems of privacy and ownership rights of information with the hope that we can take full advantage of the technology and yet minimize the potential negative effect on our world community. It is my hope that this research, whether it contributes in a large or small way to the advance of communications, will be used to enhance our quality of life but will not infringe on our basic human rights.

Contents

I	Problem Definition	1
1	Background for Research	2
1.1	Need for Additional Bandwidth	3
1.2	Definition of Network Parameters	5
1.3	Media	7
1.4	Potential of Current Topologies	9
1.4.1	CSMA/CD BUS	12
1.4.2	Token Ring	14
1.4.3	Token Bus	20
1.4.4	Slotted Ring/Bus	21
1.4.5	Star	21
1.5	Summary	22
1.6	Organization of Results	23
II	FDDI	25
2	Influence of TTRT on Performance	26
2.1	Basic Token Ring	28
2.2	Traffic Placement Policy	30
2.3	Timed Token Rotation Protocol	33
2.4	Token Rotation Time	34
2.5	Impact of Token Rotation Time	40

2.5.1	Parameters affecting TTRT	42
2.5.2	TTRT vs Utilization	44
2.5.3	TTRT Impact on Synchronous Traffic	50
3	FDDI at Gigabit Speeds	52
3.1	Limitations of Techniques Presented	52
3.2	Scaling FDDI	53
3.2.1	Parameters and Metrics	54
3.2.2	Results	56
3.3	The Interface Bottleneck	59
3.3.1	Timing Parameters	63
3.3.2	Results	64
3.4	Conclusions	65
4	Destination Removal for FDDI	69
4.1	Related Work	70
4.2	Destination Removal for Token Ring	73
4.2.1	Restrictions on slot reuse	74
4.2.2	Objective	76
4.2.3	Advantages	77
4.3	Analysis	77
4.3.1	Expected Maximum Throughput Increase	77
4.3.2	Overall effectiveness	80
4.4	Simulation	82
4.4.1	Throughput Comparison	86
4.4.2	Long Distances	88
4.4.3	Expansion of Queue Window	91
4.4.4	Delay Characteristics	95

4.4.5	Robustness	97
4.5	Implementation	103
4.6	Conclusions	112
III	DRAMA	114
5	Multichannel CSMA/CD - DRAMA	115
5.1	MAN Requirements	116
5.2	Protocol Analysis	120
5.2.1	Bandwidth Allocation Model	120
5.2.2	Traffic Placement Model	121
5.3	Simulation Results	124
5.3.1	Single Versus Multiband Behavior	124
5.3.2	Size and Scalability	125
5.3.3	Stability	130
5.3.4	Fairness	134
5.3.5	Integration of Voice and Video	136
5.4	Conclusions	143
IV	Conclusions	146
6	Conclusions	147
6.1	FDDI	147
6.2	DRAMA	149
V	Appendix	150
A	FDDI Overview	151
A.1	Design Objectives	151

A.2	Implementation	153
A.2.1	Topology	153
A.2.2	Transmission Medium	153
A.2.3	Signalling	156
A.2.4	Timing	159
B	The DRAMA Protocol	160
B.1	Overview	160
B.2	Objectives	164
B.3	Protocol Description	165
B.3.1	Band Allocation Protocol	165
B.3.2	Traffic Placement Protocol	166
B.3.3	DRAMA Feasibility	170
C	Fiber Optics Technology	172
C.1	Fibers	172
C.2	Transmitters	174
C.3	Receivers	175
C.4	Optoelectrical Integration	176
D	Simple Destination Removal Simulation	177

List of Tables

1.1	Number of Users per Machine	2
1.2	Network Packet Capacity - Electrical Medium	11
2.1	FDDI Timing Specifications	43
2.2	Overhead Delay Parameters for FDDI Scenarios	45
2.3	Maximum Utilization for 125 μ s TTRT	50
3.1	Scaling: Parameter Range for Gigabit Test	55
3.2	Scaling: Gigabit Timing Scenarios	64
4.1	Removal: Curve Interpretation	86
A.1	FDDI: 4/5 Encoding	158

List of Figures

1.1	Ethernet Utilization: HSLDN - small packets	15
1.2	Ethernet Utilization: HSLDN - large packets	16
1.3	Ethernet Utilization: LAN	17
1.4	Ethernet Utilization: HSLDN - longer distance	18
2.1	Token Ring Protocol : Token Capture	31
2.2	Token Ring Protocol : Appending a Message	32
2.3	TTRT timer	36
2.4	FDDI Timing Example	41
2.5	Backbone Predicted Utilization versus TTRT	48
2.6	MAN Predicted Utilization versus TTRT	49
3.1	Scaling: Impact of Nodes on Message Delay	57
3.2	Scaling: Impact of Packet Size on Message Delay	60
3.3	Scaling: Impact of Network Length on Message Delay	61
3.4	Scaling: Impact of Timing Parameters on Gigabit LAN	66
3.5	Scaling: Impact of Timing Parameters on Gigabit MAN	67
3.6	Scaling: Impact of Timing Parameters on Gigabit WAN	68
4.1	Removing Packet at Destination	75
4.2	Removal: Expected Throughput Increase	81
4.3	Unfavorable Destination Address Distribution for Single Ring	84
4.4	Using <i>problem</i> Messages on the Second Ring	85

4.5	Removal: Summary of Results	87
4.6	Removal: Slot Reuse	89
4.7	Removal: Impact of Network Distance - Throughput	92
4.8	Removal: Impact of Network Distance - Slot Use	93
4.9	Removal: Increasing Queue Window to Physical Layer	94
4.10	Removal: Throughput Increase Per Ring	96
4.11	Removal: Access Delay	98
4.12	Removal: Access Delay at 0...100% Load	99
4.13	Removal: Access Delay at 0...70% Load	100
4.14	Dual Counter-rotating Rings with Removal under File Server Loading . . .	104
4.15	Dual Rings without Removal under File Server Loading	105
4.16	Single Ring without Removal under File Server Loading	106
4.17	Single Ring with Removal under File Server Loading	107
4.18	Removal: Packet Structure	109
5.1	DRAMA: Steady State LANG Allocation	122
5.2	DRAMA: Access Delay versus Number of Channels	126
5.3	DRAMA: Access Delay versus Load Varying Number of Nodes	127
5.4	DRAMA: Band Allocation Responsiveness	128
5.5	DRAMA: Band Allocation Fairness	129
5.6	DRAMA: Access Delay Under Scaling Conditions	131
5.7	DRAMA: Access Delay Versus Packet Size (normalized delay)	132
5.8	DRAMA: Network Throughput for Various Placement Policies	133
5.9	DRAMA: LANG Queue Size During Pulsed Load Conditions	135
5.10	DRAMA: Band Allocation Stability	137
5.11	DRAMA: LANG Fairness Varying Tolerance Interval Size	138
5.12	DRAMA: Response Varying Interval Size	139
5.13	DRAMA: LANG Fairness Varying Band Traffic Deviation	140

5.14 DRAMA: Access Delay Varying Periodic Traffic Load	141
5.15 DRAMA: One Channel versus Multichannel DRAMA	142
A.1 FDDI Configuration	154
A.2 FDDI Healing	155
A.3 Clock Jitter	157
B.1 DRAMA: A Network Configuration	161
B.2 DRAMA: Sample Frame	169
B.3 DRAMA: A Possible Network Configuration	171

Part I

Problem Definition

Chapter 1

Background for Research

Despite rapid technological advances which allow us to build faster, more powerful and more reliable computers, machine development never seems to be able to keep pace with man's ability to create new applications requiring additional resources. Over the past 30 years, we have witnessed tremendous increases in central processing unit speeds and memory capacities, accompanied by corresponding decreases in cost. History has shown us that this power increase is always insufficient to outdistance the needs and imaginations of the computing community.

A computer classified as a minicomputer, capable of executing multiple MIPS¹ and supporting many users in recent years, can now be found as the workstation of a single user. An examination of the ratio of users to machines in Table 1.1 indicates a definite trend towards reducing the number of persons using each computer.

ENVIRONMENT	Users/Machines
Mainframe	$n/1$
Personal Computers	$1/1$
Network Computing	$1/n$

Table 1.1: Number of Users per Machine

There are already a number of areas of interest which require as one of the prerequisites a high-bandwidth communications medium. This chapter cites a number of application areas requiring an improvement in communication bandwidth, a specific definition of the problem area investigated (High Speed Long Distance Networks - HSLDNs), a survey of

¹MIPS represents millions of instructions executed per second.

the use of different types of media for HSLDNs, an analysis of the problems of existing network strategies for HSLDNs, and an explanation of how the research comprising this thesis contributes to the solution of this problem.

1.1 Need for Additional Bandwidth

It is difficult to find an application which would not be improved by a better communications infrastructure with higher capacities, faster speeds and better reliability. Some applications require these improvements. The High-Performance Computing Initiative from the White House Office of Science and Technology Policy has defined 20 major challenges in science and engineering which are dependent on the solutions to a number of high-performance computing problems. One major focus of this initiative is the development of gigabit rate networks and a National Research and Educational Network (NREN).

Network Computing is the concept of distributing computational objects across a network in order to smooth computing load and take advantage of special purpose computers such as artificial intelligence machines, simulation machines, and graphics processors in order to operate more efficiently. Much of the computing power which exists is severely underused except during small periods of time. Implementation of these concepts relies heavily on the ability of computer networks to support transport of these objects across the network.

Artificial intelligence and expert systems possess the potential to radically alter the way we perform computing tasks: algorithm development, coding, compiling, interfacing with the operating systems and numerous others. It is difficult if not impossible to find an application which could not be enhanced by expert system support tools. One expects that sophisticated knowledge bases will be of extremely large size and will require large bandwidth networks to support interprocessor communication of this knowledge. As the availability of such systems evolves, the demand for home and business access to this information through public networks will grow, placing an even larger strain on the public

facilities.

One of the most immediate problems demanding increased bandwidths is the lack of support for remotely-generated video images. A graphics processor on a network may be able to generate displays fast enough to support real-time animation on a remote workstation, but tremendous network data transfer rates are required to deliver the images.

Inclusion of video information places a significant drain on the capacity of current networks. Workstation screens have resolutions of approximately 1000x1000 pixels; the screen must be redrawn at least 30 times a second; and typically, a screen can have one of 256 to 1,000,000 colors at each pixel location. Transmission of video images for a workstation such as this with 256 colors would require a data rate of 240 Mbps². Ordinary color television video can be digitized at 100 Mbps and compressed to 45 Mbps [60]. Voice and video traffic are examples of synchronous (periodic) data, one of the major traffic classifications for these networks. The primary characteristic of synchronous traffic is the need to deliver fixed quantities of data at regular, periodic intervals.

We now see the emergence of products such as X-windows which support interactive graphical windows, generating data on one computer but displaying it on another. This is only a first step towards the ability to deliver pictures such as animated images or those created by a video camera. The potential use of capabilities such as these have been shown in commercial products such as the HyperCard by Apple Computers.

Due to recent developments in optics and laser technology, efficient movement of processes and data from one machine to another is not constricted by low communication rates of transmitters and receivers, but by the interface between the applications program and the physical transmitter, which is separated by a layered communications system. Layered system designs such as the Open Systems Interconnection³ (OSI) model incur data transfer, formatting and error checking at most layer interfaces. Although direct memory access

²Notation for data rates is : kbps = kilobits (10^3) per second, Mbps = megabits (10^6) per second, and Gbps = gigabits (10^9) per second.

³OSI is a communications architecture developed by the International Organization for Standardization (ISO)

(DMA) devices help to streamline some of these operations, the implementation of most of these algorithms in software severely limits the speed at which they can execute. The collective overhead required to move the data from the application to the physical layer results in a system which is unable to deliver the data to the physical layer at optical transmitter rates. Therefore, the development of cost-effective fiber devices results in a reverse of the information flow bottleneck between the computer which is generating the information and the medium which is transporting it. Nonetheless, research is progressing to increase both the rate of physical transmission and the ability of the computer to utilize it through more efficient interfaces. The focus of this research emphasizes how to improve the computer/network interface for improved performance by using existing networking strategies rather than designing completely new ones.

1.2 Definition of Network Parameters

In 1990, a typical computer network media would be twisted pair, coaxial cable or fiber optics capable of supporting data in the range of one to 100 Mbps. For standard text and file transfers and configurations of a few nodes, this type of capacity would seem reasonable. However, with the proliferation of computing equipment, the need to communicate between numbers of nodes on the order of magnitude of 100-1000 is realistic and would significantly reduce the proportion of the network capacity allocated to each node. Large numbers of nodes on a network represent the first component of the problem definition. Not only is there a demand of bandwidth resulting from the large numbers of nodes anticipated on networks, but also the types of applications discussed previously compound the situation. In addition, the number of nodes has an impact on other aspects of network performance as will be shown later.

The current cost of transmitter/receiver interfaces for high rate networks makes such interfaces unrealistic for low-cost workstations. Ethernet interfaces are less than \$500 for access to a ten megabits per second network. FDDI⁴ interfaces are currently estimated

⁴Fiber Distributed Data Interface

at \$10,000 for connection to a 100 Mbps network, but Wallach [71] estimates the cost will be down to \$1000 within three to seven years. The connection cost to a one gigabit per second network significantly restricts its applicability to very expensive workstations, mainframes, supercomputers, and/or as a backbone for network interconnection. Although it is not currently economically feasible for workstations to interface directly to a gigabit per second network, the traffic which they generate should be considered in researching this issue because it must be transported over the backbone. High speed network bandwidth represents the second component of the problem definition.

A third significant concern is the inclusion of nodes which are spread over a wide geographical area. Local area networks have been used in configurations in the range of a few kilometers. A logical extension of this usage is to allow communication between nodes separated by larger distances. Networks of larger distance are characterized as metropolitan area networks (MANs) or wide area networks (WANs).

The anticipation of a demand for large scale networks characterized by

- large numbers of nodes (≥ 100),
- much higher data rates (≥ 100 Mbps), and
- greater physical distances (> 5 kilometers),

is the subject of this research. Networks with these characteristics will be referred to as High Speed Long Distance Networks (HSLDN) as compared to WANs which are traditionally much slower (56 kbps). Some of the cases considered in this research incorporate values less than the ranges defined above for two reasons: one for comparison purposes and the other because certain deviations (such as smaller number of nodes in a long distance) remain within the spirit of the definition.

The remainder of the background for this research examines

- which media appear most promising,
- which networking strategies appear extendable, and

- which research areas appear most promising and practical.

1.3 Media

A variety of media is currently in use. Coaxial cable and twisted pair utilize electrical transmission techniques but suffer from low data rates (twisted pair carries less than four megabits per second and coaxial cable <500 Mbps). They also have significant attenuation problems at large distances (five to ten kilometers). Perhaps the most significant problem with these technologies is that the potential for high bandwidths (greater than one gigabit per second) does not appear promising. Factors such as thermal and intermodulation noise provide an inherent limitation.

Microwave transmission can support data rates in excess of hundreds of thousands of bits per second but suffers from a number of deficiencies. The transmitter/receiver must operate in a line-of-sight mode, which severely limits its usage in areas with numerous tall buildings. Second, because the data is transmitted through the air it comes under the regulation of the Federal Communications Commission. There is, therefore, a limitation to the total capacity which can be used by all. This is a significant issue in the area of cellular car phones, for example, because of the limited bandwidth available to automobiles in a localized area. The cost and need for special location of the transmitter/receiver are additional considerations.

Satellite transmission is capable of transmitting data at the HSLDN rates but has two problems. The cost of connection is expensive for transmitter/receivers, although new satellite technology shows promise for drastic reductions in the cost of receiver dishes. Satellite costs are astronomical, and the costs per connection cannot be justified for most computer applications. As in the case of microwaves, satellites use the airwaves as a transmission medium, restricting the total communication bandwidth available.

The most promising technology is that of fiber optics. There are a number of reasons why optical fibers are favorable for the types of networks described. The cable itself is

more expensive than twisted pair or coaxial cable; however, it has many characteristics which make it more attractive and justify the cost. It is also anticipated that the price of copper will not change significantly, and that the price of fiber will be driven below that of copper. In addition, the fibers, transmitters and receivers are still in a state of intense research which has led to increased quality and a reduction in cost. Light emitting diode (LED) transmitters in the range of a few megabits per second only cost about \$10 each, whereas transmitters in the 50 Mbps range cost about \$300. Plastic optical fibers have been developed and are commercially available in the two megabits per second range, with claims that data rates of 100 Mbps will be available at similar costs in the future [65]. These transmitters are relatively inexpensive and research continues to improve the bandwidth/cost ratio. Cables can easily be pulled through tight trunk lines due to the small diameters [9]. Although the cables are not completely immune to sensing, the problems can be minimized with proper containment of the cable [64]. The protected nature of the cable allows the bandwidth to be reused on every cable, resulting in a total bandwidth which is limited only by the number of cables available. In comparison to existing twisted-pair and coaxial cable media, fiber is not only smaller by an order of magnitude, it is also capable of bandwidths larger by orders of magnitude.

Another issue for long-haul networks over ten kilometers is the low attenuation of the signal. In long distance networks, this minimizes the need for repeaters. Tapping a fiber signal, however, does have a negative effect on the power, which will either limit the number of nodes or require repeaters. Most fiber networks today rely on a point-to-point connection that requires each node to repeat the signal to the next station in the network. This has an effect on the type of network topologies which are compatible with optical fibers.

Perhaps the most important advantage of optical fibers is its potential for expansion of bandwidth. Today, most optical fiber systems utilize non-coherent detection. Due to problems of reflection, signal generation and construction of the fiber itself (multi-mode versus single mode), the potential bandwidth has not yet begun to be used. Whereas sophisticated

use of electronic signals allows for detection of changes in frequency, amplitude, phase, or combinations thereof, non-coherent detection does not allow for detection of the information content of the signal except through the presence or absence of photons. Development of similar coherent techniques for optical signals is the subject of much research.

At this time most fiber systems use time-division multiplexing techniques as opposed to frequency or wavelength methods for bandwidth sharing. Current optical fiber data communication systems such as FDDI have the capability of transmitting at rates in the range of 100 Mbps. Laboratory experiments and trial commercial products for communications trunk lines have produced transmissions in excess of one gigabit per second by using Wavelength Division Multiplexing (WDM) techniques [9,70]. Given the availability of FDDI hardware and the predictable drop in cost as a result of its consumer acceptance, one interesting area for research is to investigate methods for extending the network capacity of a network using existing hardware such as FDDI.

Subsequent sections will include a general discussion of the limitations of current standard LAN designs, a review of some of the recently proposed network designs, a summarization of the problems and advantages of each of the designs, and motivation for investigation of a variety of interconnection topologies.

1.4 Potential of Current Topologies

It is relevant to first discuss the effect which the three main parameters described above (number of nodes, data rate and network length) will have on these networks before examining specific topologies. A few articles which consider token-bus or token-ring structures have been written concerning this issue, but the authors come to no definite conclusion [37,46].

The increase in number of nodes has a negative effect on practically all architectures. The token-ring or token-bus architectures suffer greater delay between token arrivals due to

the node delays and token captures on the network. For CSMA⁵ or CSMA/CD⁶ type media access strategies, the number of nodes competing for the network has a negative effect due to the increased probability of collisions as discussed in Section 1.4.1.

Performance characteristics of networks are related to the quantity of information a node sends in a physical transmission. The term used to describe this is a packet of information. If the logical size of the data is different than the physical size, the logical entity is subdivided into packets for transmission on the network. In certain situations the terms packet and frame have different meanings, but they will be used interchangeably here.

The data rate and network length are interrelated as follows. Most local area networks operate with lengths of one to two kilometers and data rates from one to ten megabits per second. These parameters generally limit the number of simultaneous packets on a network to less than one. As the length of a network increases and as the time it takes to place a packet on the network decreases, the number of packets which can be transmitted simultaneously on the network rises. This introduces the problem of how one might best manage these packets to enable fairness of access to the network by all nodes, maximize throughput, and minimize average delay.

Ignoring system overhead, Table 1.2 gives an idea of the relationship among these parameters for specific types of networks. Examination of this table reveals that when dealing with networks characterized by one gigabit per second data rates and network lengths of 100 km, the number of packets to be managed is a design consideration. LAN token ring/bus and CSMA-CD networks are approximated by the second line of the chart, which indicates only a small fraction of a packet is on the network at a time. For example, a network of ten megabits per second capacity, packet length of 2000 bytes and LAN length of one kilometer can only hold .003125 packets at a time. Network designs with capacities such as one gigabit per second, 100 km and packet lengths of 2000 bytes, could contain approximately 31 packets on the ring simultaneously.

⁵Carrier Sense Multiple Access

⁶Carrier Sense Multiple Access / Collision Detect

Data Rate (Mbps)	Net Length (km)	Packet Length (bytes)	Network Capacity (packets)	
10	1	500	.0125000	
		2000	.0031250	
		100000	.0000625	
	10	10	500	.1250000
			2000	.0312500
			100000	.0006250
	100	100	500	1.2500000
			2000	.3125000
			100000	.0062500
100	1	500	.1250000	
		2000	.0312500	
		100000	.0006250	
	10	10	500	1.2500000
			2000	.3125000
			100000	.0062500
	100	100	500	12.5000000
			2000	3.1250000
			100000	.0625000
1000	1	500	1.2500000	
		2000	.3125000	
		100000	.0062500	
	10	10	500	12.5000000
			2000	3.1250000
			100000	.0625000
	100	100	500	125.0000000
			2000	31.2500000
			100000	.6250000

Table 1.2: Network Packet Capacity - Electrical Medium

As the chart shows, the precise number of packets on the network is also dependent on the size of each packet. Increasing the size of the packet just to reduce the complexity of managing the increase in numbers of packets may reduce throughput. If the application only uses a small percentage of a large packet, internal fragmentation is a significant issue. Large packets have the advantage of minimizing the overhead inherent in formatting each packet with source and destination address, packet type, trailers, etc. The disadvantage is that a fixed size large packet does not match the size of the data which the application transmits. The network should be able to support a diverse range of message sizes without an unreasonable amount of wasted bandwidth due to overhead, but internal fragmentation may dictate that larger messages be decomposed into a series of small frames.

The following section describes the impact of these parameters to current network designs.

1.4.1 CSMA/CD BUS

A CSMA/CD bus topology has serious limitations when considered for HSLDN organizations. One of the fundamental results of the analysis of CSMA/CD networks shows the maximum utilization, U_{max} , to be

$$U_{max} = \frac{T}{T + C} \quad (1.1)$$

where

T is the packet transmission time

and

C is the contention interval, defined as

the time required to obtain the channel.

In Ethernet, a contention based protocol, nodes attempt transmission without knowledge of the number of other nodes which are also currently ready to transmit. A packet is transmitted only when a single node transmits. If two or more nodes transmit simulta-

neously, the information collides, resulting in the need to retransmit all of these packets. Ethernet has been designed to allow for early detection of these collisions so as to reduce the lost bandwidth.

For example, if station A begins transmitting at one end of a two kilometer network on a ten megabits per second line, the first bit will take ten μs^7 to propagate to the other end of the network. Station B at the other end of the bus could have begun its transmission just prior to receiving Station A's packet. It will now take an additional ten μs for Station A to detect a collision. Station A has now been transmitting for 20 μs , which would have only allowed the transmission of 200 bits of data. As long as the packet is significantly longer than 200 bits, the amount of bandwidth lost as compared to the bandwidth used on a successful transmission will be small and the maximum utilization will be acceptable. A contention slot is 2τ , where τ is the one-way propagation delay. This 200 bit contention slot is not exactly the value of C in the above formula. Multiple 200 bit slots may be required before the contention is resolved.

Consider the maximum efficiency of an Ethernet network as the number of nodes increases. Due the inherent nature of Ethernet to provide immediate access when no other nodes are trying to use the network, light load conditions are uninteresting. Although minimal collisions occur, utilization of the network is usually low because a single node would not be able to saturate the network for a long period of time. HSLDNs are usually characterized by large numbers of nodes, and the intention here is to extend the analysis of CSMA/CD to HSLDNs.

Stallings [62] shows that with time normalized to the duration of a single packet, the maximum channel utilization U , is

$$U = \frac{1}{1 + 2\tau \frac{(1-A)}{A}} \quad (1.2)$$

where

⁷Timing values will utilize the following standard notation : ms = milliseconds (10^{-3}), μs = microseconds (10^{-6}), and ns = nanoseconds (10^{-9}).

A is the probability that a packet is successfully transmitted.

A is optimized when the node transmits with probability of $1/K$, when there are K of the total number of nodes, N , ready to transmit.

For HSLDNs, the implications are as follows. Figure 1.1 shows the maximum utilization of an Ethernet network where the number of users varies from one to one thousand in a logarithmic scale. The figure is calculated using a fixed bus length of 20 km, a data rate of one gigabit per second and a varying packet size from eight to 64 kilobits. Assuming nodes can determine the number of nodes attempting to transmit, K , maximum utilization indicates an effective delivery of between 100 and 200 Mbps. In order to obtain a utilization of over 50%, packet lengths of approximately 256 kilobits must be used. This is illustrated in Figure 1.2. In the future, packet sizes of this magnitude may be realistic, but the current bottleneck remains the inability of the computer to deliver the load. Large numbers of nodes can demand high bandwidth collectively, but they are incapable individually of utilizing packet sizes on the order of magnitude of one megabit. As a contrast to the HSLDN problem, consider a typical LAN of two kilometers running at Ethernet rates of ten megabits per second. Figure 1.3 shows utilization to approach 100% and provides insight as to why Ethernet works so well in a LAN, but so poorly in a different environment. Figure 1.4 demonstrates even more vividly the impact of large distances on an Ethernet approach. Extending the distance to 100 km requires increasing the packet sizes above one megabits in order to obtain 50% utilization. This result shows that CSMA/CD approaches will not provide high utilization for HSLDNs unless very large packet sizes are used or unless one alters the approach by using techniques such as partitioning the distance (medium) in some adaptive algorithm.

1.4.2 Token Ring

Access to the network by a node in a token ring is obtained by capture of a special token frame which circulates around the ring. A node removes the token, transmits its packet for a specified period of time and places the token back on the ring for capture by downstream

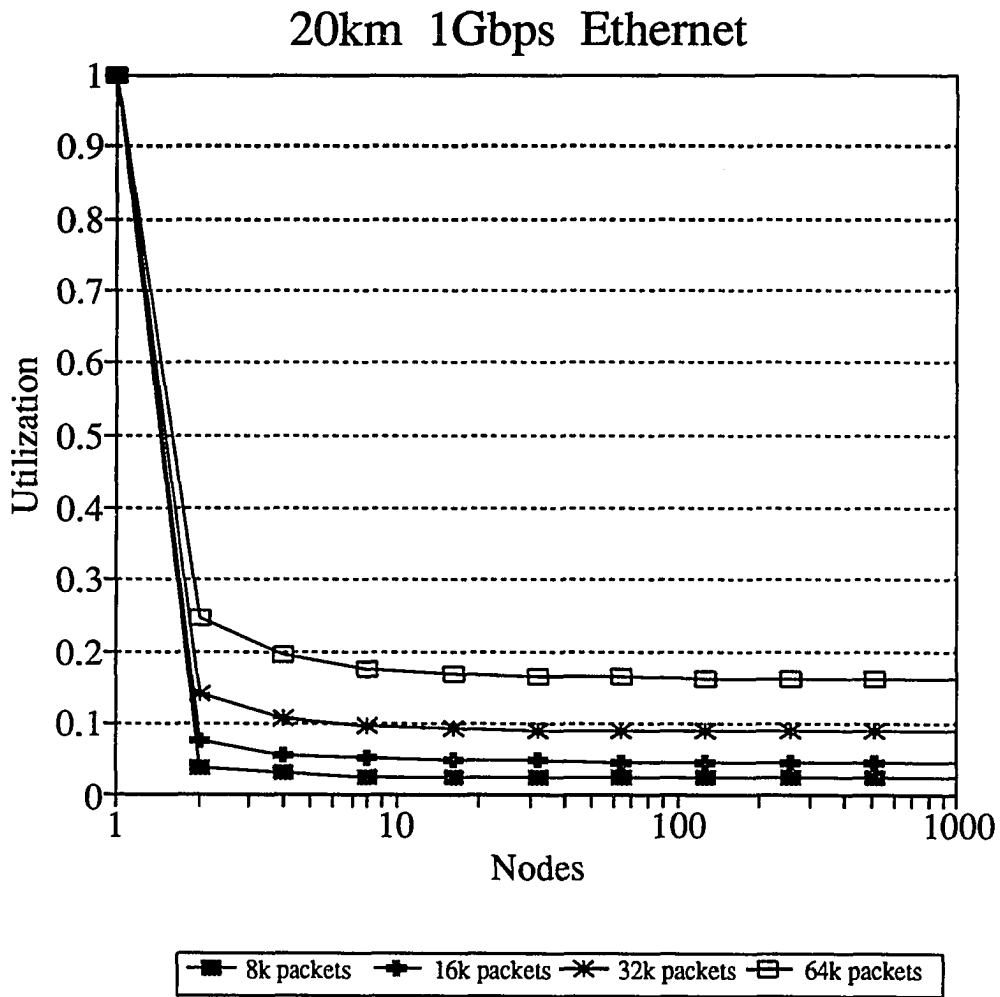


Figure 1.1: Ethernet Utilization: HSLDN - small packets

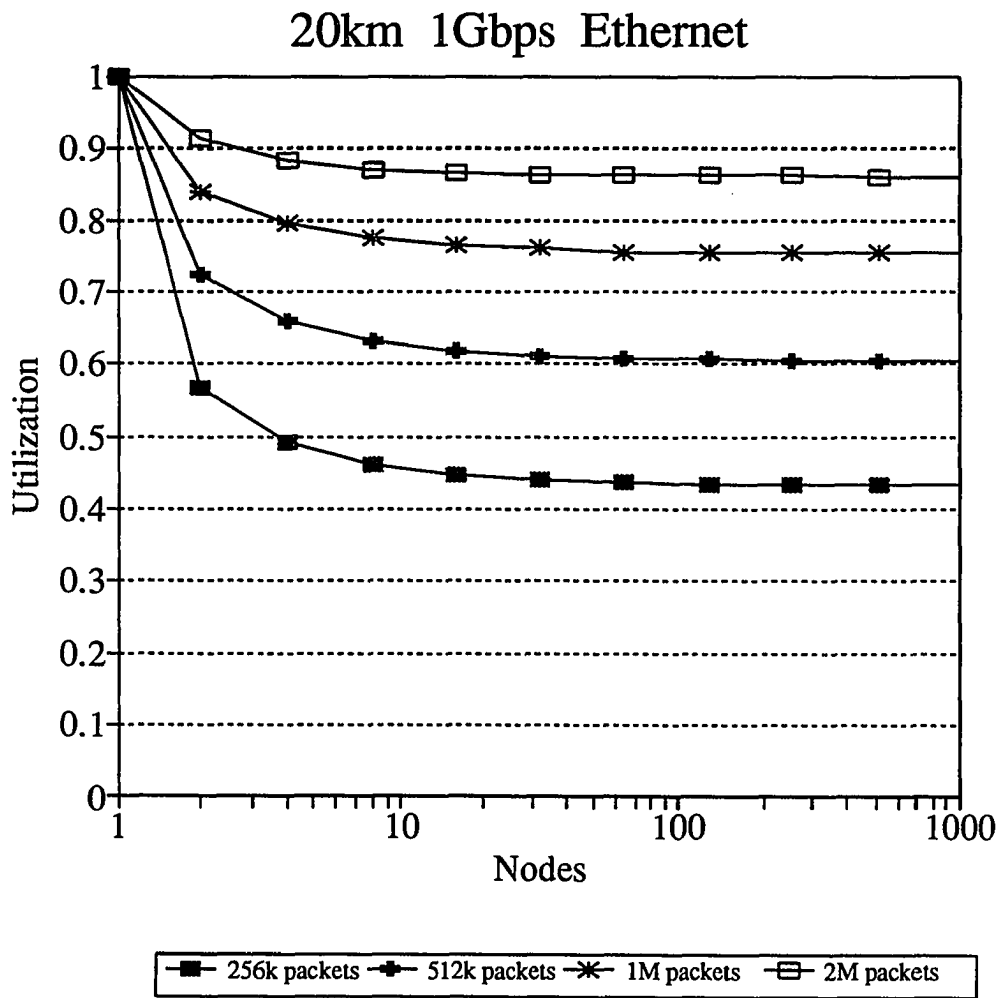


Figure 1.2: Ethernet Utilization: HSLDN - large packets

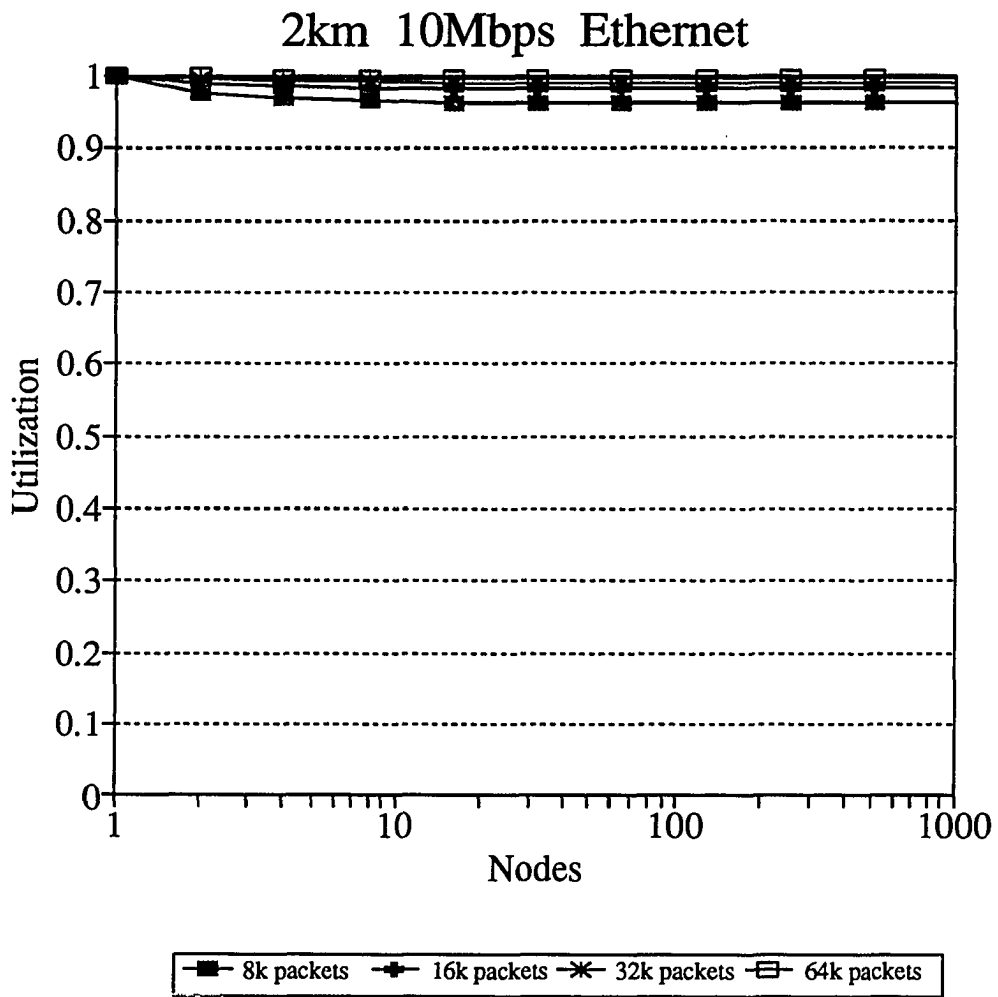


Figure 1.3: Ethernet Utilization: LAN

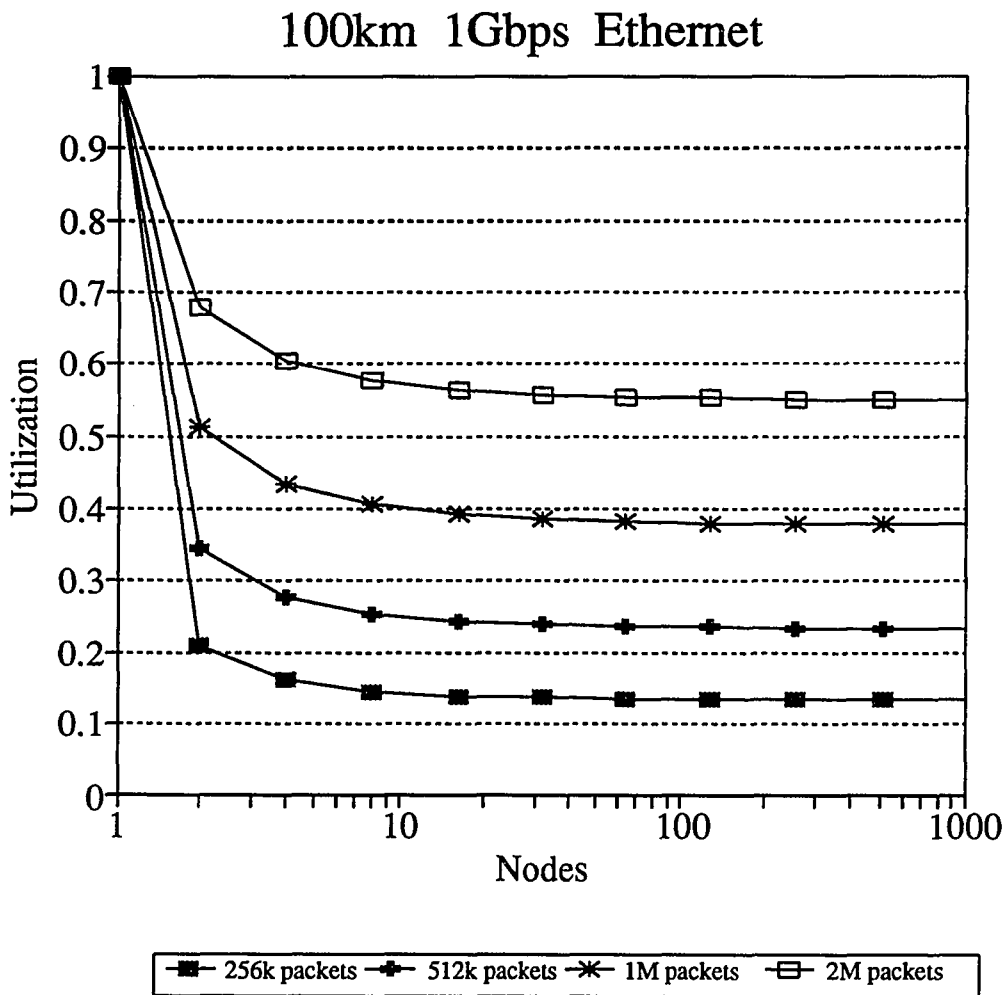


Figure 1.4: Ethernet Utilization: HSLDN - longer distance

nodes. The amount of time allocated to a node for holding the token is a method of allocating bandwidth among the nodes, but for simplicity it is assumed that all nodes hold the token long enough to transmit a single fixed-length packet.

The analysis of throughput of a token ring can be found in Stallings[62]. Utilization is dependent on the time to transmit a packet and on propagation delay. During heavily loaded network conditions where all nodes have queued packets, the utilization can be quantified by the time required to transmit a packet divided by the time to transmit the packet plus the delay

$$U = \frac{1}{1 + \frac{\tau}{N}} \quad (\tau < 1). \quad (1.3)$$

When the packet transmission time exceeds τ , the beginning of the frame, actually the busy-token, returns to the sender before placing a free-token on the network. For example, a two kilobit packet transmitted on a ten megabits per second ring will take 200 μ s. A propagation delay of five μ s/km means that the results hold for a ring up to 40 km.

Retaining this policy, thus requiring the head of the transmitted packet to return before releasing the token, has a detrimental effect for rings which can hold more than one packet. The results for this case are also contained in Stallings [62] and show the utilization to be

$$U = \frac{1}{\tau(1 + \frac{1}{N})} \quad (\tau > 1). \quad (1.4)$$

The application of Equation 1.4 to HSLDNs implies unsuitability when compared with Table 1.2. For $\tau > 1$, the utilization would be low. However, a token ring need not be constrained by the conditions of this equation.

It is not necessary to wait until the busy-token returns to place a free-token on the network. FDDI is an example of a token-ring network which retransmits the token immediately at the end of frame transmission. The process of removing the busy token is explained in more detail in Chapter 2. The utilization of such a network can approach 100%. Once the network fills with packets, the entire network will be in use with the exception of overhead for the token itself and the time required to convert from a receiving to a sending mode.

There is one other matter to consider for ring utilization. This measure only indicates how many packets are on the network as compared to its actual capacity. It does not consider that the data only needs to pass by half of the nodes on the average, but instead travels by all nodes. If the packet was removed at its destination rather than at the sender, the frame capacity of the net could be used for other messages. Acknowledgements could be handled in another way. This does not suggest that the management of this wasted capacity is easy to recapture, but it is mentioned as a point of comparison with other strategies and as a metric for evaluation.

1.4.3 Token Bus

A token bus does not differ significantly from a token ring. In a token ring, media access rights are passed from one node to an adjacent node on the ring. Nodes on a token bus may pass the token to any node on the bus. Therefore, the physical order need not be the logical order. Certainly one could use the physical order of the nodes on the bus as the logical order, and require the last node to pass the token to the first node. This would be the equivalent of the token ring.

The advantage of the token bus as compared to the token ring is the provision for varying the order in which the token is passed to the nodes. There is no inherent reason why a token ring would be prohibited from passing the token in some order other than the physical order of the nodes, but the hardware is typically unidirectional on a token ring, severely limiting the efficiency with which one could deviate from the physical order of the nodes.

Likewise, a fiber optic token bus would normally have the physical and logical order of the nodes the same. Fiber optic connections for HSLDNs must utilize a point-to-point connection due to the high dB losses incurred in tapping the line. For the large numbers of nodes being considered here, the signal loss would be unacceptable. A point-to-point connection is utilized in DQDB⁸[52], an IEEE⁹ slotted bus which runs at a 300 Mbps rate

⁸Dual Queue Dual Bus

⁹Institute of Electrical and Electronics Engineers, Inc.

(150 Mbps per channel). One approach for exploiting the bus structure and the point-to-point connection would be to partition the network into subnets, allowing each subnet to transmit simultaneously. This approach has been suggested for Ethernet-type networks and is the subject of research [12,42] and is not included here.

1.4.4 Slotted Ring/Bus

A slotted ring/bus typically configures the medium to contain a continuous series of fixed-length slots into which nodes place packets and from which they remove them. Slots can be managed by a reservation scheme or a round-robin type of scheduling algorithm, or can access any free slot which passes. In the former case delay is incurred before the node can use the reserved slot, and in the latter case the delay occurs after delivery because the slot must be marked empty by the original transmitter for subsequent use.

Research is ongoing in a collaborative project between Olivetti Research Limited and the University of Cambridge Computer Laboratory in an attempt to build a slotted ring which can execute at gigabit speeds. This network, the Cambridge Backbone Ring, is currently in the design phase, but component testing is providing optimistic results [21]. Simulation results indicate that throughput of about 525 Mbps can be achieved for transmitter rates of 800 Mbps.

1.4.5 Star

Star networks have been used very little in local area networks due to the reliance on the operability of the center node for operation of the network. Depending upon the processing capabilities of the center node, simultaneous transmissions could take place between points on the star. Nodes can be physically arranged in a star, yet function as a token ring. For example, an extremity node could receive the token, transmit the message to the center node which broadcasts to all nodes, and transmit the token to the ring. The center node would then transmit to a node and the process would repeat. If parallel transmissions could take place, better utilization would be possible. Stars provide good potential for exploitation of

simultaneous traffic if sufficient switching complexity is built into the center node.

One major difficulty of stars as they apply to HSLDNs is the great cable distances required to reach the center node. These propagation delays have a negative effect on utilization. In considering the use of fiber optics for building star configuration networks, most current suggestions are to use some type of signal processing hub which is not point-to-point. Typically, some type of wavelength division multiplexing (WDM) technique is employed. Unfortunately, for large numbers of nodes, the problem of tapping or splitting a signal too many times reduces the signal strength to a level which renders it too weak to read properly.

1.5 Summary

The combination of data transfer rate, network distance and packet size indicates that most local area networks only have a single packet on the network at a time. As data transfer rates and network distances increase, the number of simultaneous packets on the network increases dramatically, as summarized in Table 1.2. One of the major issues in moving to the HSLDN environment is the efficient management of these packets. Limb [37] provides analysis to indicate that a slotted ring approach has the greatest potential of these methods.

Linear structures such as the bus and ring have been implemented as broadcast networks. The increased efficiencies which result from reducing the number of nodes to see a specific message have been partially examined by Dobosiewicz [12], Limb [38], and Maly [42]. The approach used in their research was to logically segment the network into smaller links and take advantage of the decrease in propagation delays and parallel transmissions.

Segmentation with hierarchical structures appears to be a logical topology to examine for managing these HSLDNs. The hierarchical tree structure would minimize switch complexity and cable distances. In addition, interconnection schemes such as perfect shuffle [33] have shown promise for high speed feasibility, so it would seem logical that the methods used in parallel computation for interconnection of processors should be examined. Investigation

of different topological organizations was an original subject of interest associated with the research in this paper and still appears to be a promising area, but this subject has had to be deferred due to more pressing problems which manifested themselves during the course of conducting the research.

1.6 Organization of Results

The research which follows is a summary of a variety of approaches which are intended to improve the performance of High Speed Long Distance Networks. Rather than derive a radically different approach to networking, the research presented here is an attempt to suggest subsequent evolutionary directions for two networking strategies either currently in use and/or standards adopted by recognized organizations such as ANSI and IEEE. The intensity of the investigation is not proportionally spread across each of the methods, but is more reflective of my personal contribution to the respective research efforts and the funding facilitating the research.

The next part of the thesis contains the main body of research, which is an investigation of a number of issues related to improving performance of a token ring standard, FDDI. Included is a general description of the network strategy, a discussion of the relationship between performance and the token holding timer algorithm, some conclusions concerning what performance issues will demand the most attention in order to scale FDDI physical transmitter to gigabit rates, and a study of the concept of destination removal in a token ring network. There follow studies of a method originally proposed by Sharrock [58] to extend CSMA/CD protocols to longer distances and higher data rates by using a multi-channel approach. Support is provided for synchronous and asynchronous data and dynamic allocation of bandwidth by partitioning of channels to smaller groups of nodes which are in the same physical proximity. This work is a journal paper submitted to *Computer Networks and ISDN* and represents a collaborative effort of myself, Kurt Maly, Ed Foudriat, Ravi Mukkamala and Michael Overstreet. The last two parts of this paper comprise the conclu-

sions of the thesis, and the paper closes with a few appendices that include an explanation of the basic design of the networks investigated.

Part II

FDDI

Chapter 2

Influence of TTRT on Performance

Network data rates are now available at rates in the 100 Mbps per channel class. The two most prominent of those are competing MAN standards FDDI [63] and DQDB(QPSX) [22], and research is currently progressing in an effort to better understand their performance capabilities and limitations [30,13,69]. Research, however, is going forward to develop networks with better performance, and a national research initiative is underway to develop gigabit per second networks to be employed as a backbone for a national research network [31].

Questions still remain as to the approach which can best suit the requirements of such a network. A national network will be likely to transport synchronous and asynchronous traffic, support large numbers of nodes (at least 100 and likely over 1000), be spread over very large distances (over 1000 kilometers), and is therefore an example of a HSLDN. The impact of considering these types of parameter ranges can be very negative for token rings due to increased token cycle time and for CSMA/CD due to the increased slot times as I have showed in Chapter 1. Although long distance networks transmitting at high data rates may not necessarily employ large numbers of nodes, this research assumes a large numbers of nodes.

Most of the current research in FDDI and DQDB focuses on the characteristics of the network as it is currently defined. Johnson [30] shows that certain timing assumptions

concerning the token-holding-timer algorithm are, in fact, true. Dykeman and Bux [13] show how the timers settings affect the flow of various priority traffic levels used in IEEE 802.4 and FDDI. Valenzano [69] reveals the cyclical access nature of the algorithm. These papers do not address the impact of considering FDDI for other environments such as HSLDNs.

Most public packet-switched and circuit-switched wide area networks such as Telenet, Tymnet, CompuServe and others interface at rates on the order of 56 kbps [61]. Higher data rates are available through dedicated or private leased lines at a T-1 carrier rate of 1.544 Mbps. Integrated Services Digital Network (ISDN) represents the first attempt by public carriers to provide a carrier capable of carrying integrated digital services, but the connection service rate is basically two 64 kbps channels, greatly restricting its applicability. Current applications already demand rates orders of magnitude larger. As networks are developed for gigabit rates, they will be much more expensive, and efficiency will become a much more important factor. A current understanding of performance is important, but consideration for enhancements to increase performance is also demanded.

Increased data rates could be accomplished by focusing on the development of transmitter/receiver devices which are capable of functioning at such high rates, i.e. simply build gigabit rate lasers. A number of the major issues associated with building such devices can be found in the following articles: Hinton provides a classification of photonic devices with an emphasis on building photonic switching networks [26]; Maeda discusses the problems of integrating optical and electrical components [39]; Nakagami describes the tradeoffs and principles of the light sources and detectors currently employed [49]; and Nosu traces the history of coherent lightwave detection [53].

There are, of course, numerous problems associated with such high speed devices other than the transmitter/receivers themselves, such as how to build computers which can process data at the rate of the network and what types of protocols would work best at these rates. One major problem is the integration of electronic components with optical transmitters and

receivers which running at much faster clock rates. The approach taken in the Cambridge Backbone Ring typifies the strategy used to solve this problem. Multiple wide data paths are employed for slower electronic speeds and are time-division multiplexed and buffered to interface with the more narrow, high speed optical component.

Another approach is to examine how current transmitter/receiver technology can be optimized to improve characteristics such as throughput and delay. External parallel channels also show some promise for improved performance [44,42,45] and is the subject of some research. Many architectures have been proposed in an attempt to design more efficient networks. Strategies have included 'train' protocols [67,66,38], hybrid CSMA/CD protocols [15,47,44,42,17], slotted and register insertion rings [25], and numerous others.

In this research I investigate the viability for scaling FDDI, a 100 Mbps token ring protocol, to the type of environment mentioned above by modifying either the physical layer or the MAC level of the architecture. The suggestion for performance enhancement using destination removal of packets is applicable to token ring networks at any rate or distance, and provides a framework for improving other types of networks.

An overview of FDDI is presented in Appendix A. This chapter contains an examination of the impact of the token-holding timer setting on network performance and a method for predicting network performance based on this setting. Chapter 3 investigates some of the problems inherent in scaling the transmitter/receiver to execute at gigabit rates. Lastly, Chapter 4 includes a proposal for modifying the basic design of FDDI to incorporate destination removal into its operation as a means of increasing throughput and reducing delay.

2.1 Basic Token Ring

A token ring network is distinguished by the manner in which transmission rights are granted to nodes on the network. The token packet circulates on the ring, passing by each node. If a node's transmitting queue is empty, it simply continues circulating the token to the next

node. However, if the node has a message to transmit in its queue, it will effectively remove the token by changing some information in the packet. The token itself is not actually removed, but is transmitted in some altered form so that subsequent nodes will not see it as a token and will not attempt to transmit. The modified token continues circulating and is eventually consumed (not retransmitted) when it reaches whatever node is in the process of transmitting (holding the token). The figures which follow explain this process graphically. Terminals represent the nodes, the box in the upper right part of the terminal represents the queue of messages to be sent identified by the destination address, and transmission is in the direction from 5,4, ...,1,5,

Figure 2.1 illustrates how the token is 'removed'. Node four has a message for node two and is waiting for the token before initiating transmission. Figure 2.1.B shows that the token has been modified so that it will not be recognized as such and the message has been placed on the network. Part C shows that the token has been retransmitted on the network and is available for subsequent use.

As the token continues around the ring, subsequent nodes remove the token and append another packet. This scenario is illustrated in Figure 2.2, where node one has a message for node five. In order to send the message, the token is modified, the message for node five is transmitted and a new token is made available to the other nodes. The network becomes filled with messages and *old tokens*. At most one real token is on the network at any instant of time.

The exact time at which the *old tokens* and messages are removed is dependent upon a number of factors: the distance of the network, the length of the packet, the data rate of the network, and which nodes subsequently capture the token for transmission. In the example shown, the old token has not yet encountered a transmitting node and remains on the ring for the moment. The most important factor to note is that all nodes except the node holding the token are forwarding messages. The node holding the token does not forward the incoming message but instead forwards its own message. Incoming messages,

which are either messages which have already been delivered or the same message which this node is transmitting, are absorbed. Eventually, the *old token* will encounter a node which is in the process of transmitting a message and be removed.

Data packets are removed in a slightly different manner. It is important that they be removed by the sender so that a receiver will not accept the message a second time if it recirculates. The tail of the message is removed (modified) at the sender once the address is recognized, whereas the fragments of the headers of these packets are removed as mentioned above [56].

2.2 Traffic Placement Policy

A fundamental element of the design of FDDI is the use of a token-holding timer approach to determine how long a node may hold the token and what type of information may be transmitted during this time frame. A multitude of strategies exist for placement of data on a token ring. There are three primary issues related to traffic placement:

1. which nodes can capture the token,
2. what amount of information can be transmitted, and
3. when the token is retransmitted for subsequent use.

General strategies for capture of the token vary. The simplest technique is to allow each node to capture the token when it arrives. A second method adopted in the IEEE 802.5 token bus protocol provides for prioritization of the token itself, thereby establishing a minimum priority level required of the node desiring to obtain the token. FDDI employs a different strategy developed by Ulm [68], which allows capture of the token based upon local timers which indicate the length of time since the arrival of the last token at that node. This will be further developed in a subsequent section.

The simplest approach for determining the transmission time allotted to each node is to allow each node to transmit a single, fixed-length frame and release the token. Another

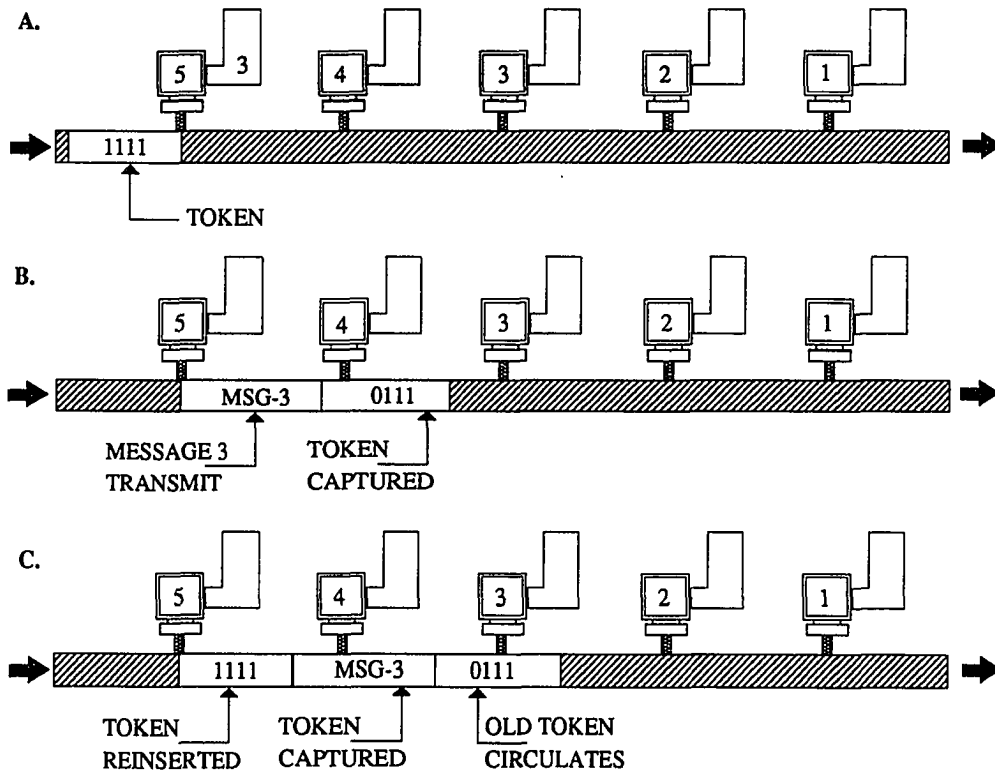


Figure 2.1: Token Ring Protocol : Token Capture

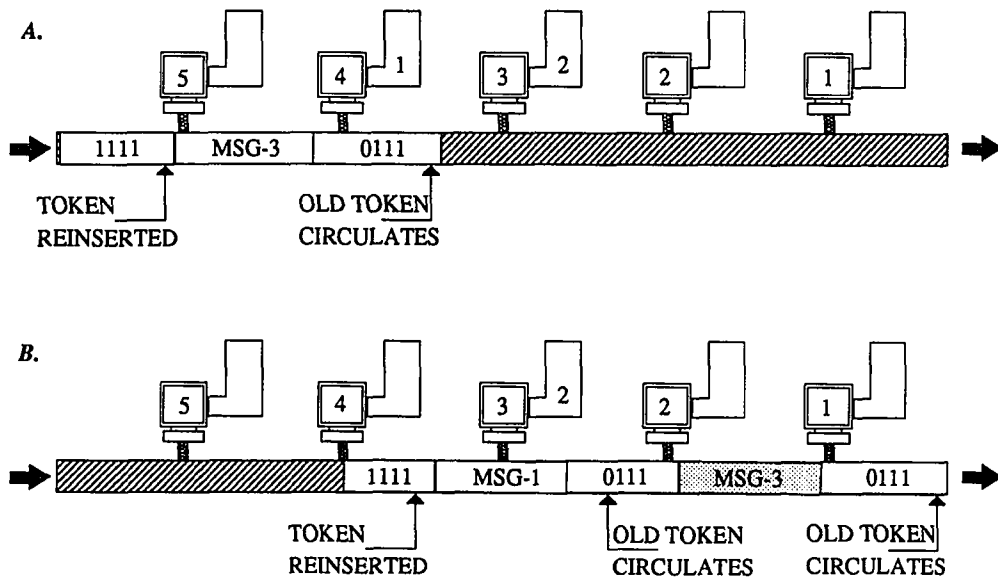


Figure 2.2: Token Ring Protocol : Appending a Message

strategy is to allow for variable length frames. FDDI also uses the local timers to determine the amount of time which a node is allowed to transmit, and it contains specifications for how to use these timers to differentiate between the allowable placement of different traffic classes and different priorities.

The time at which a node replaces the token on the network is also different in FDDI. A typical token ring implementation waits for the first part of the frame to return with the busy (captured) token before transmitting a new one. In rings characterized by a packet length longer than the bit length of the network, this is satisfactory because the head of the frame will return to the sender before transmission is complete. This will not hold true for an FDDI configuration spread over a long distance (100 km). Note that Table 1.2 shows the number of packets for a 100 Mbps network of length 100 km and packet size of 2000 bits to have a capacity for 25 simultaneous frames¹. In order to understand how this affects performance, consider the case where a node can transmit only one frame before releasing the token. Equation 1.4 shows the utilization to be approximately four percent. In contrast, if the policy adopted is to allow the node to retransmit the token upon completion of transmission of the message, Equation 1.3 reveals that a utilization of 50% is achievable for $N = 25$ and 80% for $N = 100$. As a result of this relationship, FDDI replaces the token at the end of transmission of the frame rather than waiting for the head of the token to return.

2.3 Timed Token Rotation Protocol

The Timed Token Rotation Protocol [68] allows FDDI to limit transmission times allowed in various priorities and to assure access to the token within a tolerance interval for synchronous traffic. This is a distributed timing mechanism whereby each node has its own set of timers which are set as a function of the arrival of the token at the node and which determine the amount of time for which the token may be held and the node's data transmitted. It will assist in the understanding of this protocol to consider only synchronous and

¹The original table is expressed in bytes, not bits. The number of packets in bits is $3.125 \times 8 = 25$.

asynchronous traffic as compared to isochronous. Isochronous traffic is implemented only in FDDI-II [6] and a separate mechanism is used for placement of this traffic. Therefore, only synchronous and asynchronous traffic will be included in the explanation in Chapter 2.

2.4 Token Rotation Time

The decentralized access mechanism of a token ring protocol can also place limitations on how long a station has access to the network once it obtains the token. The approach in FDDI is to limit the amount of *time* for which one station can hold the token [68]. Both access to the token and the token-holding restrictions are determined by this algorithm. Each node has a timer which is reset when the token arrives. When the token returns, the node may capture it only for an amount of time which will assure that the token will return to each and every node within a fixed period of time from the arrival of the previous token. The primary motivation for employing this approach to access control is to support synchronous traffic. The method can be shown to place bounds on the access delay to a node and therefore guarantee access to the network on a regular (synchronous) basis. This time period is the *Target Token Rotation Time*, TTRT. The value of this parameter is negotiated amongst all nodes on the network and is essentially the smallest value selected by any node. This defines the maximum amount of time between access to the network and should provide for synchronous traffic.

Although it can be shown that the algorithm guarantees that the token returns within the negotiated time frame on the average [30], it cannot be guaranteed that the node is able to hold the token at all once it returns unless it has a synchronous traffic allocation. This has implications for periodic traffic and maximum throughput and is examined in Section 2.5.

At the time of initialization of the network, the stations compete to define the *guaranteed* interarrival time of the token through the TTRT. The smallest value requested for the TTRT represents the node which requires the fastest token rotation. This becomes the goal for all

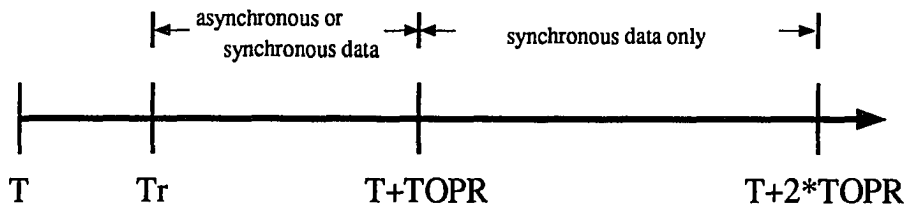
nodes in the system. The negotiated time cannot be guaranteed to be met, but it can be shown that one can guarantee that the token will return within twice the operative TTRT value, TOPR [30]. As a result, a node must negotiate for one-half of the minimum time it requires if access to the network within this time frame is a strict requirement. Johnson[30] also shows that the average time between token arrivals is no more than TTRT.

Each node has a TRT timer which represents the amount of time remaining before the token will be declared late. Let TRT_n represent the timer for node n , $Tarr_n$ the time at which the token last arrived at node n , and $Tnow_n$ the current local time at the node n . Also let TOPR represent the current operating value of the TTRT. Then the following equation defines the TRT at node n .

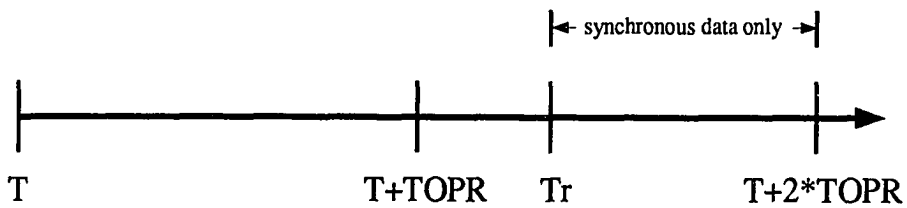
$$TRT_n = TOPR - (Tnow_n - Tarr_n) \quad (2.1)$$

Upon receiving the token, the node will save its current value in THT and reset its token-holding timer to TOPR. If the token arrives back at this node before the TRT timer expires (within the desired limit) it may transmit either synchronous or asynchronous data. The time left on TRT, now in the THT, represents the time available for synchronous transmission. Once the TRT timer expires no asynchronous transmissions may begin, but if a node has already begun, it may continue to completion. A node is not guaranteed time for asynchronous data because the token could not be guaranteed to complete a cycle within TOPR. If the TRT timer does expire before the token arrives, the node may still transmit its synchronous data upon receipt of the token. An example of each of these conditions is shown in Figure 2.3.

In order to maintain the timing restrictions mentioned, there must also be some way to limit synchronous traffic. The value of TOPR is also used to determine the maximum amount of synchronous traffic allowed between tokens. At the time of ring initialization, each node is allocated some percentage of the total available synchronous bandwidth (TOPR). The sum of all allocations for synchronous traffic must not exceed 100% of TOPR in order



Token arrives to allow for both traffic types



Token arrives late - asynchronous not allowed

Figure 2.3: TTRT timer

to guarantee that timing requirements can be met. This allocation requires renegotiation if it is to be altered.

At first glance it appears to be simple to see why the token will return in $2 \times \text{TOPR}$. Once the maximum asynchronous bandwidth, TOPR, is exhausted, the worst possible scenario is that none of the nodes with synchronous bandwidth has received the token yet. After the nodes place their synchronous data, which also has a maximum of TOPR, the token will be delayed $2 \times \text{TOPR}$. This is not a precise assessment of the problem due to the fact that each node has a different timer and that asynchronous traffic may overrun TOPR if a node begins transmission before TRT expires. However, this perspective provides a simple model of what takes place and it is reasonably accurate. A more precise example follows.

Using this simplified model, one more restriction will allow for a stronger statement about the frequency of rotation of the timer. Assume that once the token returns, the node has recognized that the token arrived late and that it refrains from sending any asynchronous traffic until this 'late' time interval has been recaptured. For example, if the first rotation takes twice TOPR and the next rotation uses 25% of TOPR for synchronous transmission, then 75% of the delay has been recaptured.

Time for first token rotation

$$\text{TOPR} \times 2$$

Time for second rotation

$$\text{TOPR} \times .25 \text{ (asynchronous not allowed)}$$

Total time for two rotations

$$\text{TOPR} \times 2.25$$

The token is now only late by $.25 \times \text{TOPR}$. If only synchronous transmission is allowed on the third rotation, as much as 75% of the maximum allotment could be used, and still allow the ring to catch up and resume transmission of asynchronous traffic. It is apparent that under conditions of maximum use of synchronous bandwidth, asynchronous traffic may be locked out. It is possible to say that the maximum delay between token arrivals is $2 \times \text{TOPR}$

and that the average delay is no more than TOPR, a much stronger statement.

Each node has two timers and a flag, which are defined as follows.

1. TRT timer contains the time remaining until the token will be declared late. It is reset to TOPR on either of two conditions: if the token arrives on time, or if the timer expires before the token arrives. The token is not reset upon arrival of a late token in order to accumulate the lateness of the token. Note that if the token arrives late, the timer has already been reset to TOPR.
2. LATE flag is set when TRT expires before token arrival. If TRT expires when LATE is set, the ring must reconfigure. This would indicate that the token has not arrived in $2 \times \text{TOPR}$, indicating ring failure.
3. THT timer contains the time allowed for asynchronous transmission. It is set to the current value of TRT when the token arrives to indicate the time remaining until the TRT timer would have reached TOPR, had it not been reset.

Following is an algorithm which represents how each of the events takes place:

1. Wait for Token

If TRT expires then

If Token is already late ($\text{LATE} = 1$)

Net has failed -> Reconfigure

Else

set $\text{LATE} = 1$

reset TRT to TOPR

continue waiting

2. Upon Token arrival

If Token is late ($\text{LATE} = 1$) then

reset $\text{LATE} = 0$

transmit any synchronous data up to

bandwidth allocation

transmit token

Else (Token is on time)

save value of TRT (early time) in THT

reset TRT to TOPR

while (async data left to transmit) and (THOLD >0)

transmit async data

Transmit any synchronous data up to

bandwidth allocation

Transmit token

3. Repeat waiting

Viewing FDDI from the perspective of a single node provides a reasonable model of the ring operation, but an examination of multiple nodes and their individual timers provides valuable insight into the operation of the ring. Johnson [30] provides a formal proof offering a more precise statement of the timing guarantees of token arrivals. The result of this proof shows that the maximum interarrival time of a token to any node is $2 \times \text{TOPR}$ and makes a more precise statement concerning the conditions required for this to be true.

Consider a simple three node FDDI ring with three nodes N_1 , N_2 and N_3 . Assume that fixed size asynchronous frames are used for all nodes, that transmission of these asynchronous frames requires 25% of TOPR for transmission and that all nodes have a sufficient queue of asynchronous frames to allow the node to transmit an asynchronous frame if the timer permits. Also assume that nodes N_1 and N_2 each have 40% of TOPR allocated for synchronous transmission, and that both will always have the synchronous frames ready for transmission. Note that 20% of the potential synchronous bandwidth is not allocated. Propagation delay between successive nodes on the network will have a value of one. For ease of calculation TOPR will have a negotiated value of 100 units. A global clock will be referenced in addition to the timers at individual nodes. This global clock does not exist in

FDDI operation and is only used for reference.

Consider the scenario depicted in Figure 2.4. At ring startup, the token circulates around the network without any transmission and no one has any data to transmit until the second time around. A microscopic view of the ring operation follows where GC is the time on the Global Clock, N_i represents the TRT timer at node i , H_i represents the THOLD timer at node i and an asterisk (*) at the N_i timer indicates that LATE is set.

Figure 2.4 illustrates how the timers vary and gives some appreciation for the difficulty in understanding the effects of distributing asynchronous bandwidth to the various nodes on the network. Every node is allowed to transmit its synchronous bandwidth, but the asynchronous bandwidth is sporadically available. The traffic placement policy used in this example requires asynchronous data to be transmitted first if bandwidth is available and synchronous transmission last. The FDDI standards document does not define the relative priority given to the two traffic classes if bandwidth is available for both types. It would also seem logical to transmit synchronous data first, but a better case can be made for the policy adopted. If a node has bandwidth available for both, the asynchronous data could be transmitted first with a guarantee that synchronous bandwidth would be available afterwards, but the opposite is not true.

2.5 Impact of Token Rotation Time

Johnson [30] provides analysis concerning the timing requirements of FDDI for both the ideal and the non-ideal (including overhead) cases. For the ideal case, the token can always be guaranteed to return to the station within $2 \times \text{TOPR}$ where TOPR is the current operating value of the TTRT, within which the token should return. For example, if the currently negotiated value of TOPR is $125 \mu\text{s}$, then the token can only be guaranteed to return within $250 \mu\text{s}$. It would appear, then, that one would simply negotiate for one-half of the desired TTRT and that the proper availability of the token could thus be assured. For reasons of maximizing utilization of the network, there is compelling motivation to have a large value

GT	N_1	H_1	N_2	H_2	N_3	H_3	Event
0	100	-	0	-	0	-	Token arrives at node 1 no transmission
1	99	-	100	-	0	-	Token arrives at node 2 no transmission
2	98	-	99	-	100	-	Token arrives at node 3 no transmission
28	75	72	73	-	74	-	n1 sends sync 40 +1
69	34	-	100	32	33	-	n2 sends async 25
94	11	-	75	7	8	-	n2 sends sync 40 +1
135	*70	-	34	-34	*67	-	n3 forwards
136	69	-	33	-	66	-	n1 sends sync 40 +1
177	28	-	*92	-	25	-	n2 sends sync 40 +1
218	*87	-	51	-	*84	-	n3 sends nothing
219	*86	-	50	-	83	-	n1 sends sync 40 +1
260	45	-	100	9	42	-	n2 sends async 25
285	20	-	75	-16	17	-	n2 sends sync 40 +1
326	*79	-	34	-	*76	-	n3 sends nothing
327	*78	-	33	-	75	-	n1 sends sync 40 +1
368	37	-	*92	-	34	-	n2 sends sync 40 +1
409	*96	-	51	-	*93	-	n3 sends nothing
410	*95	-	50	-	92	-	n1 sends sync 40 +1
451	54	-	100	9	51	-	n2 sends async 25
476	29	-	75	-16	26	-	n2 sends sync 40 +1
517	*88	-	34	-	*85	-	n3 forwards
518	*87	-	33	-	84	-	n1 sends sync 40 +1
559	46	-	*92	-	43	-	n2 sends sync 40 +1
600	5	-	51	-	100	2	n3 sends async 25 +1
626	*79	-	25	-	74	-	n1 sends sync 40 +1
667	38	-	*84	-	33	-	n2 sends sync 40 +1
708	*97	-	43	-	*92	-	n3 forwards
709	*96	-	42	-	91	-	n1 sends sync 40 +1
750	55	-	100	1	50	-	n2 sends async 25

Figure 2.4: FDDI Timing Example

of TTRT thus resulting in a tradeoff between the design objective to support synchronous traffic and the need for high utilization.

The following section shows the impact of TTRT on utilization. The research took place concurrently with similar research conducted by Jain [29]². Jain's work consists primarily of simulation results and gives an overview of the effect of TTRT on access delay, response time and throughput; but FDDI timing parameters are not considered. Simulation results are used for certain metrics, and analytical results are used for others; however, simulation and analysis are not used to support the same finding. In contrast, this chapter focuses upon the impact of throughput with a more detailed analysis, and uses simulation to *support* the analytical results.

2.5.1 Parameters affecting TTRT

As cited in [30], the primary components of ring overhead are as follows.

- Total Propagation Delay (D_{prop}) is determined by multiplying the propagation delay for fiber optic media (5085 ns/km) by the length of the network.
- Latency (L_n) occurs at each node and is effectively the delay between the time a bit arrives at a node and departs the node. Therefore, if one examines the round trip of a single bit around the network, the delay is increased by the latency at each node times the number of nodes which are actively using the ring. Here it is assumed that all nodes are accessing the ring. L_{tot} represents the total latency of the network ($L_{tot} = N \times L_n$).
- The number of nodes to capture the token, N_c , increases the delay of the token rotation. In the minimal delay case, no node needs the token and this component is nothing, but no information is transmitted. By focusing upon the process of transmitting a frame at a single node, this overhead becomes apparent. The head of the token arrives at the node and is passed on to the rest of the network while the node waiting to transmit identifies this as the token. Recognition takes place before the tail of the

²Both results were submitted for publication to the same conference, SIGCOMM '90.

L_n	Latency per connection	600 ns
T_t	Token transmission time	880 ns
T_i	Max transmitter idle time	3500 ns

Table 2.1: FDDI Timing Specifications

token is retransmitted, providing the capturing node the opportunity to modify the end of the token, transforming it into a non-token frame, and thereby capturing the token. The delay required to accomplish this is incorporated into Latency, L_n . The node then proceeds to transmit its packet and *retransmit* the token to its neighboring node as explained in Section 2.1 and Figure 2.2.

- As each node captures the token and retransmits it, an additional delay equal to the Token Transmission Time (T_t) will be incurred. Note that the delay from message transmission does not contribute to *overhead* delay.
- The design specifications of FDDI [63] allow for a maximum Transmitter Idle Time (T_i). This represents the time which is required between recognition of the token by the node and beginning of transmission of the frame. As in the previous item, this time is only a factor when nodes are actually capturing the token, but it proves to be an important factor in performance. The design specifications of FDDI refer to this term as LMAX.

Specified maximum times for these delay components can be found in Table 2.1 [63].

Consider three scenarios for FDDI as a basis for evaluating the impact of these parameters:

1. ten nodes separated by a distance of 100 meters (one kilometer total) representing a backbone for interconnecting local area networks;
2. three nodes separated by a distance of ten meters (30 meters total) representing the connection of two mainframe/supercomputers or peripheral equipment which is in a close physical proximity; and

3. 500 nodes each separated by a distance of 100 meters (50 km total) representing a HSLDN or MAN.

Table 2.2 illustrates the delays³ inherent in each of the scenarios. The difference between MAX and MIN TOTAL DELAY is the number of nodes transmitting on a token rotation. For example, given that a network has a specific TOPR value, only a certain amount of time can be spent transmitting data on each token rotation. If that time is assigned to multiple nodes as opposed to a single node, transmission time must be sacrificed in order to complete the token rotation in the same amount of time, TOPR. Therefore, if each node transmits during a single token rotation, overhead is increased and throughput is reduced. MAX TOTAL DELAY assumes each node transmits incurring the higher overhead. MIN TOTAL DELAY assumes as the other extreme that (unrealistically) no node overhead is incurred for token capture and message transmission, but includes the standard node delay ($D_{prop} + L_{tot}$). Factors which would determine the specific value within this range would be primarily the distribution pattern (higher for uniform distribution) and packet size (higher for more small packets). The first example with $L_n = 600ns$ uses the maximum values from the specifications and illustrates the worst case. The second case shows the a more optimistic estimate of the latency, $L_n = 60ns$, and overly optimistic results. One can see that as latency improves below 60 ns, the effect will be significant only in the MIN TOTAL DELAY for mainframe environments.

2.5.2 TTRT vs Utilization

The purpose of this analysis is to develop a tool which is reasonably simple to use and which will be able to predict maximum utilization of an FDDI network. Practically all aspects of the derivation use average values of the random variables with focus given to heavily loaded conditions. Only asynchronous traffic is considered, with justification for ignoring synchronous traffic in the analysis being given in the following section. The end of the analysis will provide results from a simulation of FDDI to illustrate the degree of accuracy

³A one bit delay is equivalent to ten nanoseconds for a 100 Mbps network.

Latency : 600 ns per node

	1.Backbone	2.Mainframes	3.HSLDN
Prop Delay (D_{prop})	5.085 μs	0.1526 μs	2543 μs
Latency (L_{tot})	6 μs	1.8 μs	300 μs
Max Token Trans (T_t)	8.8 μs	2.64 μs	440 μs
Max Trans Idle (T_i)	35 μs	10.5 μs	1750 μs
MAX TOTAL DELAY	54.885 μs	15.0926 μs	5033 μs
MIN TOTAL DELAY	11.085 μs	1.9526 μs	2843 μs

Latency : 60 ns per node

	1.Backbone	2.Mainframes	3.HSLDN
Prop Delay (D_{prop})	5.085 μs	0.1526 μs	2543 μs
Latency (L_{tot})	.6 μs	.18 μs	30 μs
Max Token Trans (T_t)	8.8 μs	2.64 μs	440 μs
Max Trans Idle (T_i)	35 μs	10.5 μs	1750 μs
MAX TOTAL DELAY	49.485 μs	15.0926 μs	4763 μs
MIN TOTAL DELAY	5.685 μs	.3326 μs	2573 μs

Table 2.2: Overhead Delay Parameters for FDDI Scenarios

of these approximations.

All nodes on a FDDI network use the same value of TTRT. If a node does not obtain the token in time to transmit its data and maintain the required timing restraints, it simply forwards the token to the next node. One way of viewing utilization (U) is to represent it as

$$U = \frac{TTRT - TRO}{TTRT} \quad (2.2)$$

where TRO represents the Token Rotation Overhead. During a single round trip of the token, only a certain percentage of the TTRT time can be spent sending data. The rest of the time essentially represents the amount of time required to transmit the token around the ring. Part of TRO is fixed and independent of the network load. All other factors remaining fixed, one can see that for a heavily loaded network, increasing TTRT will increase utilization, but at the expense of delay. Here an estimation for utilization as a function of TTRT is derived.

Using the terms developed in the previous sections, TRO can be expressed as follows

$$TRO = N_c \times (T_t + T_i) + D_{prop} + L_{tot} \quad (2.3)$$

The number of nodes to capture the token on a rotation is dependent upon the available bandwidth for transmission (TTRT-TRO), the packet length, load and the number of packets transmitted by each node which captures the token. All of the variables except N_c are static values; however, for this derivation, separate the components of TRO into two terms as follows, TRO_s representing the component of TRO which is independent of the number of packets transmitted and the dynamic part TRO_d .

$$TRO = TRO_s + TRO_d \quad (2.4)$$

$$TRO_s = D_{prop} + L_{tot} \quad (2.5)$$

$$TRO_d = N_c \times (T_t + T_i) \quad (2.6)$$

The dynamic component is determined by the number of nodes which capture the token. Each time the token is captured, there is a transmitter idle delay and a retransmission of the token. It is possible that a node can transmit multiple packets for a single token capture. If each node capturing the token transmits twice as many messages, the token retransmission (T_t) and transmitter idle delays (T_i) would occur half as often.

Consider the range of values which N_c has with the load uniformly distributed among the nodes. Under low loads, few nodes have messages to transmit. As network load increases, N_c increases, approaching the number of nodes on the network N , then it decreases as the network becomes overloaded. In the last case, as nodes have large queues, one token capture results in a large number of packet transmissions. Eventually, each node holds the token for a period of time which precludes other nodes on the network from capturing the token until its next rotation.

As the queues at each node overload, the utilization actually increases as there are fewer token captures per rotation; however, we are interested in determining the maximum traffic which the network can support without queue buildup. In such an overloaded situation, the network cannot support the traffic levels even though overall utilization may be higher. Therefore, it is assumed that traffic is distributed such that on the average a node only has

a single packet(or less) to transmit per token rotation and that the maximum value of N_c is N . The number of packets which can be transmitted is dependent upon the number of packets which can be transmitted during $TTRT - TRO_s$. N_c would be defined as

$$N_{pt} = N_c = \frac{TTRT - TRO_s}{\frac{P}{R} + T_t + T_i} \quad (2.7)$$

where

- P is the packet length
- R is the transmission rate of the network and
- and N_{pt} is the number of packets transmitted.

As the load increases, the number of nodes capturing the token would have a limit of

$$N_c = N \quad (2.8)$$

and the number of packets transmitted with maximum token captures N_{ptM} would be

$$N_{ptM} = \frac{TTRT - TRO_s - N \times (T_t + T_i)}{\frac{P}{R}} \quad (2.9)$$

Substituting Equation 2.9 into Equation 2.2 , U can be expressed as

$$U = \frac{TTRT - TRO_s - N \times (T_t + T_i)}{TTRT} \quad (2.10)$$

Figures 2.5 and 2.6 illustrate the predicted and real maximum utilization for the backbone and MAN scenarios listed above. This provides a practical basis upon which one can determine how to set the TTRT for a network with a legitimate estimate of its affect on performance. Also note that these two graphs indicate that to a large degree, the setting of the TTRT value has little affect on utilization. As I mentioned previously, the tradeoff in setting the TTRT value is that TTRT should be set as high as possible to minimize overhead and increase throughput; this, however, has the counter-effect of increasing the time between token arrivals and therefore increasing access delay. The graphs indicate that to what degree the TTRT can be set lower without incurring this degradation of performance.

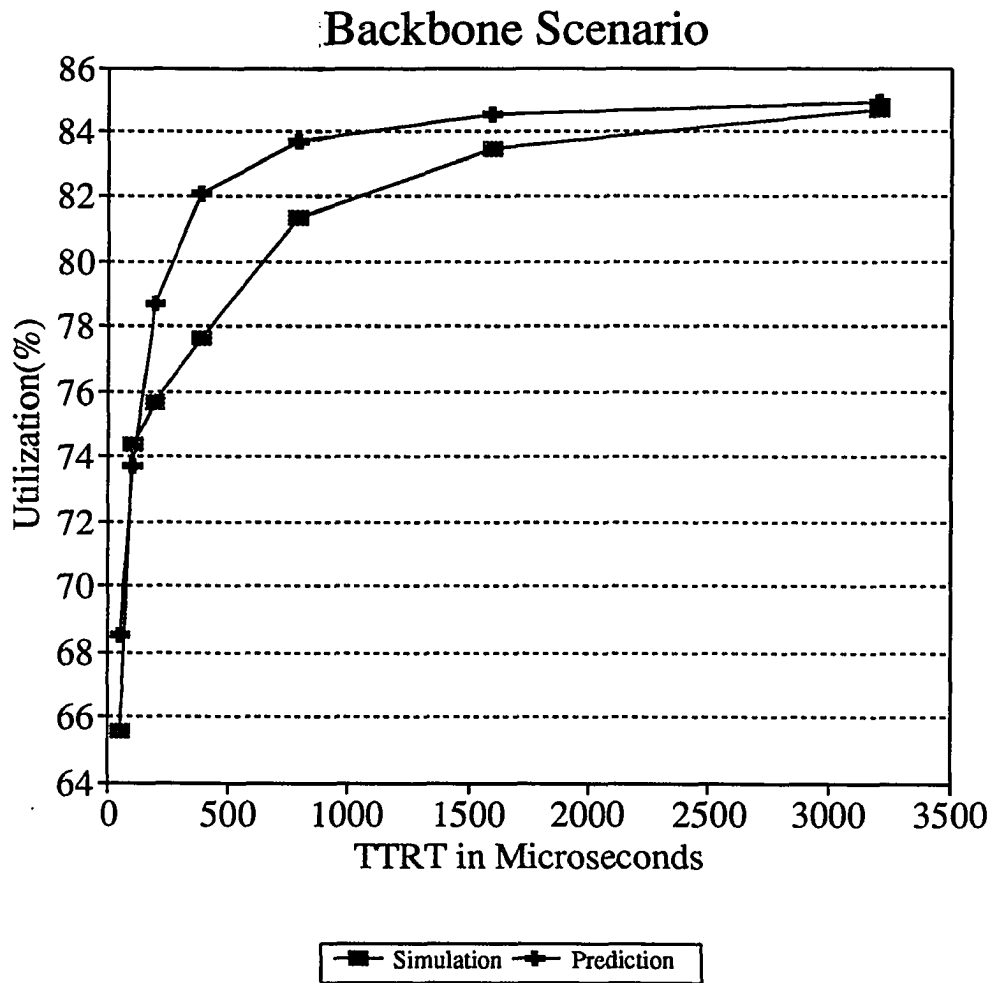


Figure 2.5: Backbone Predicted Utilization versus TTRT

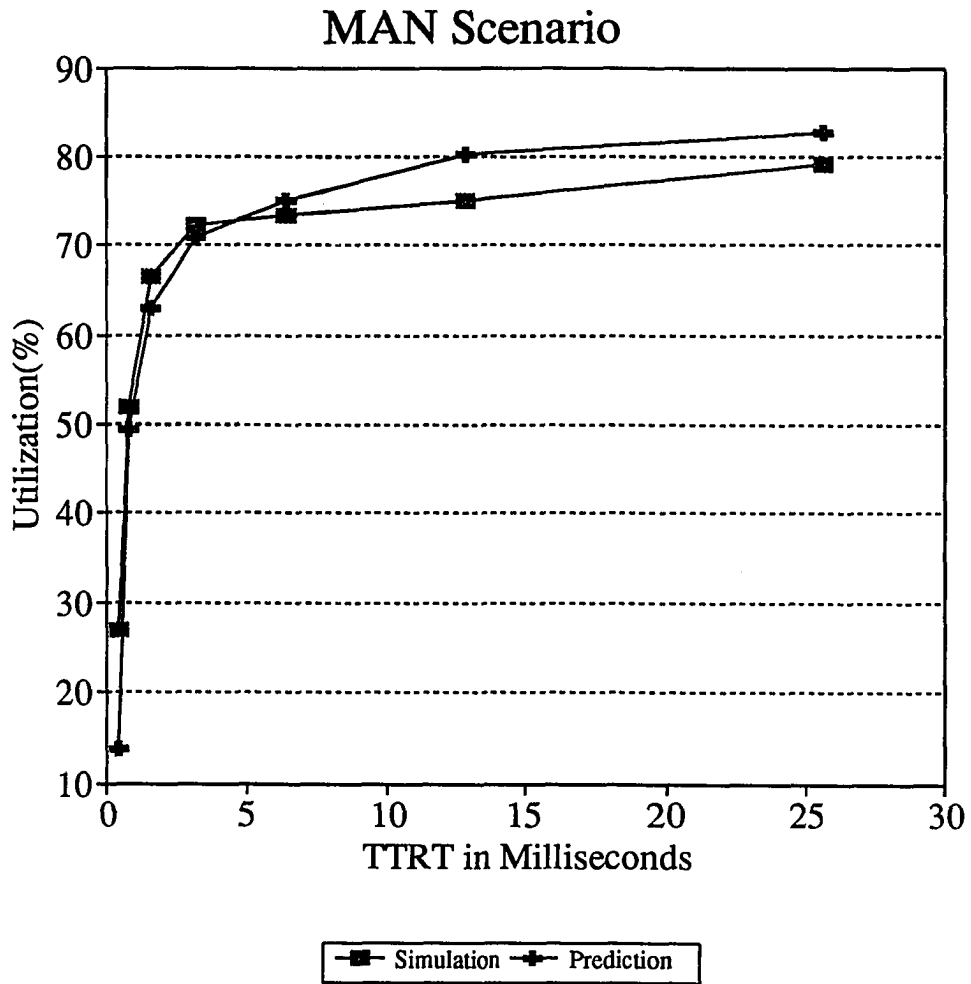


Figure 2.6: MAN Predicted Utilization versus TTRT

	1.Mainframes	2.Backbone	3.HSLDN
MAX TOTAL DELAY	88%	60%	****
MIN TOTAL DELAY	99.9%	95.5%	****

Table 2.3: Maximum Utilization for 125 μ s TTRT

2.5.3 TTRT Impact on Synchronous Traffic

Synchronous traffic has not been included in the previous analysis. Because synchronous traffic by its nature places a uniform load on the system per token rotation, the effect can be explained as follows. Consider the synchronous traffic as an overhead of the token rotation, TRO. After calculating the maximum utilization as described in Equation 2.10, add the percentage of synchronous traffic to the previous utilization value to obtain the true utilization. The only approximation involved is due to the number of token captures which could be higher if, for example, synchronous traffic packets were small and distributed to a large number of nodes. This would further reduce maximum utilization.

Application of the previous results to typical synchronous traffic requirements indicates that FDDI does not support synchronous traffic without significant decrease in utilization. ISDN is a world-wide public telecommunications network which will provide a digital interface capable of delivering a variety of services. A 64 kbps ISDN interface compatibility requires that synchronous traffic must be delivered at the rate of once every 125 μ s. In order to guarantee this arrival rate, either a TTRT of 62.5 μ s must be established, or buffering of packets at the ISDN interface must be accommodated. An average access rate for synchronous traffic can be guaranteed as shown by Johnson [30], so the differences between the speed of access by the node and the ISDN interface can reasonably be handled by buffering. Note that for all applications this trick will not work if gaps in the delivery of data cannot be tolerated. For a conservative estimate, assume that TTRT is 125 μ s, knowing that on the average 125 μ s is attainable. Comparing this with the lower node latency and ignoring packet overhead, the maximum utilization is illustrated in Table 2.3 where the scenarios are ranked according to their ability to support ISDN interfaces.

Table 2.3 indicates that FDDI could not support an acceptable ISDN interface in most configurations and that it would be highly unlikely that synchronous traffic with comparable periodicity could be supported in a long distance (MAN) environment. It is also interesting to note that in the scenario for a backbone, the utilization could drop significantly depending upon the number of nodes which can capture the token on one rotation. This is, of course, also dependent on the packet size being transmitted.

Chapter 3

FDDI at Gigabit Speeds

FDDI [63] is a 100 Mbps fiber optics ring which is commercially available and currently being used primarily as a backbone for internetwork communication. The cost (about \$10,000 per node) is a major factor prohibiting its use in workstations, but this is expected to drop significantly as the product matures. Once that happens, a logical step is to investigate what would be necessary to scale FDDI to use higher transmission rates. Consistent with the OSI philosophy of layering the communications network, FDDI was designed to allow modification of the physical level to support other media and devices. This chapter addresses the subject of scalability of the physical layer. The two primary considerations are the impact of HSLDN environments and the interface of the physical layer speed-up and upper levels of the protocol.

3.1 Limitations of Techniques Presented

If this research could be said to have a single goal, that goal would be to achieve gigabit transmission rates which would function in the HSLDN environment. DRAMA¹ and the destination removal concept for FDDI are two approaches which contribute to reaching this goal. Although each of these are significant improvements to the base technique, neither appear to be able to reach gigabit rates, given the current transmission rates of the base approach.

DRAMA employs a multi-band CSMA/CD. The most commonly used CSMA/CD net-

¹Dynamic Resource Allocation in Metropolitan Areas (DRAMA) is discussed in Chapter 5

work, Ethernet, currently runs in the ten megabit per second range. For DRAMA a LANG² would need the capacity of 100 of these channels to reach gigabit rates. The effectiveness of the approach is shown in Chapter 5, but the practicality of having 100 of these channels is in question due to the cost of such a large number of transmitters and receivers and the complexity of the interface to map traffic onto these channels. Appendix B.3.3, which describes DRAMA, provides a solution to this problem which limits the number of channels for each LANG, yet provides for a total network bandwidth which might approach the gigabit level assuming a sufficient number of LANGs are utilized. If one assumes that the demand for communication bandwidth between computers will continue to grow, and history shows this to be a strong possibility, the DRAMA solution is likely to become less and less applicable.

Application of the technique of destination removal to FDDI, modified to employ a second counter-rotating ring, is shown to produce throughputs of between 400 and 500 Mbps in Chapter 4. Although it alone is insufficient to deliver gigabit rates, a much slower base rate than one gigabit, approximately 250 Mbps, will effectively provide a gigabit network. Nonetheless, scaling of the transmitter is a viable option and is considered separately here.

3.2 Scaling FDDI

Through techniques such as destination removal, parallelism and various methods of topological construction, FDDI can also be used as a building block to increase bandwidth potential. Another possibility is to simply increase the speed of the FDDI transmitter/receiver logic. The issues here are twofold. One concerns the problems inherent in the design of the lasers, which is discussed in Appendix C, and the other is related to the interface with the controller for the network. This section provides results which show how the performance of FDDI would be impacted by simply increasing the speed of the transmitters, emphasizing the delay components considered in the analysis of Chapter 2:

²A LANG is a Local Area Network Group, a set of nodes within a geographical proximity for the purpose of allocation of bandwidth

1. propagation delay;
2. node delay; and
3. time between recognizing the token and transmitting the frame (L_{max}).

FDDI is fundamentally a token ring network. The distinctive characteristics of the network are its use of fiber optics and associated high data rates, a dual counter rotating ring topology, and a token holding timer algorithm to determine the length of time for which a node may hold the token and transmit data. Although FDDI is a dual ring, the second ring is primarily intended to allow for healing in the event of a damaged link [56]. For this reason, only one ring is modeled. This approach to analysis allows for an assessment of scalability independent of the implementation of destination removal with either one or two rings. The inclusion of destination removal would lessen the impact of the parameters, as these results show, but is not part of this study.

The token holding timer algorithm is one whereby each node keeps a local timer as a means of determining how long it can hold the token for transmission. This algorithm is intended to place a bound on access delay for synchronous traffic, and its impact on performance is treated in Chapter 2. In order to minimize this as a factor influencing performance, the TTRT value was set arbitrarily high (20 ms) in these runs. Only asynchronous traffic is considered.

3.2.1 Parameters and Metrics

Extending the rates should improve performance over standard FDDI. Packet transmission times will be proportionally reduced and propagation delays will remain the same. The issue which is addressed here is to determine which factors have the greatest impact on such a network, so that the environment in which it can best be utilized can be better understood. It is reasonable to anticipate that increasing the number of nodes or decreasing the packet size would increase node delays and token captures, both elements which increase network overhead. For example, if the number of nodes is increased, additional node overhead will

will be added to the token cycle, and if the same data is spread to the larger number of nodes, additional token captures will add to the overhead also. Will the implementation of a physical layer capable of gigabit rates be prohibited from reaching the level of performance anticipated due to such factors?

The predominant factors likely to affect performance are number of nodes, length of the network, and packet length. For the simulation, each of the three parameters are tested over a range of three values, each as follows.

<i>Parameters</i>	<i>Range</i>
Nodes	10, 100, 1000
Network Length	1Km, 10Km, 100Km
Packet Length	5K, 10K, 15K

Table 3.1: Scaling: Parameter Range for Gigabit Test

Given the large bandwidth of the network, it is anticipated that large numbers of nodes can be supported. Network length values include LAN, MAN and WAN scenarios, and packet length varies from 5000 to 15000 bits.

The metric used here to evaluate performance is message delay (referenced as Delay in Figures 3.1, 3.2, and 3.3). This is a measure of the time between arrival of a message at the node to the delivery of the last bit of the message at the destination. Other metrics frequently used in network analysis include access delay (time from arrival of the message at the node to the beginning of message transmission), throughput and fairness. Access delay is not graphed because the impact of distance on the network is a concern, and its impact on propagation delays as the distance is lengthened should be included. With the exception of only a few data points, the only cases considered are those in which the network is less than fully loaded, so that throughput is equal to offered load. Throughput is not explicitly graphed, but conclusions are drawn from the delay graphs.

I have shown fairness of FDDI for nodes transmitting only asynchronous traffic in a previous paper [43]. The effect of mixing asynchronous and synchronous traffic remains a question and is not analyzed here, but the example used in Figure 2.4 reveals an unfair

policy by denying node one the opportunity to transmit asynchronous data, while allowing nodes two and three to make multiple transmissions. There is no reason to assume that an increased transmission rate will affect fairness so it too has been ignored here.

3.2.2 Results

For the purpose of comparison, each of the following graphs in Figures 3.1, 3.2 and 3.3 use the same scaling. The horizontal axis represents load on the system in percent of the transmission rate of the network, and the vertical axis represents delay in milliseconds. Results of selected runs from the set of runs described previously are also shown.

Nodes

In a typical token ring network, the number of nodes affects performance in two primary areas. First, there exists a delay introduced on the ring at each node which is using the network. Second, for each node capturing the token on a cycle around the ring, the token arrival to other nodes is delayed by an additional token retransmission. In addition, there is a delay between recognition of the token and transmission of the queued packet, as I explain in Chapter 2. For these reasons, the number of nodes has an adverse effect on performance, but the impact of number of nodes varies as I explain below.

Figure 3.1 shows the effect of varying the number of nodes in four different scenarios. Vertically, the graphs have the same packet size, and horizontally they have the same distance. The top row of graphs more closely represents the network distances of a WAN, 100 km. The extremity values of the ranges for packet length and network length are graphed. Note that in every case, the number of nodes has a negative effect; however, the effect varies with certain combinations of the other parameters.

A comparison of the graphs horizontally shows that if the load is distributed in smaller packets, the number of nodes has a greater effect than it does in situations where the packet size is larger. This can be explained by the fact that the overhead time required for a token capture does have an impact. As the packet size is smaller and thus distributed to more

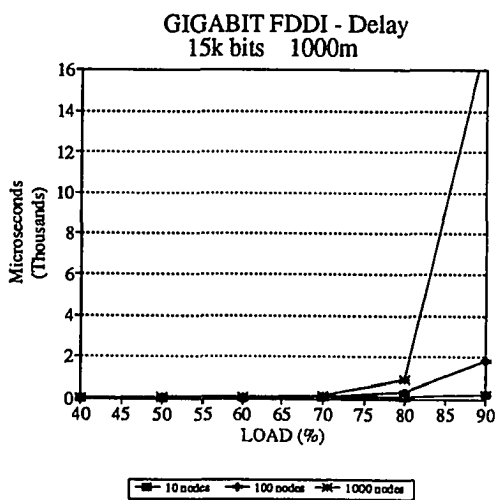
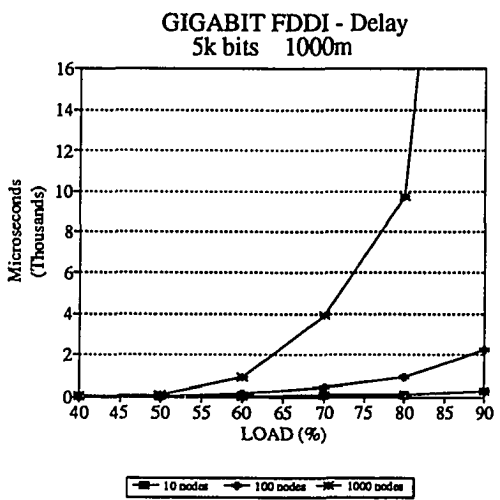
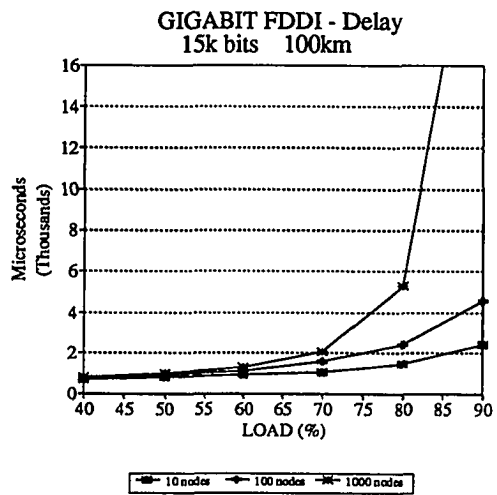
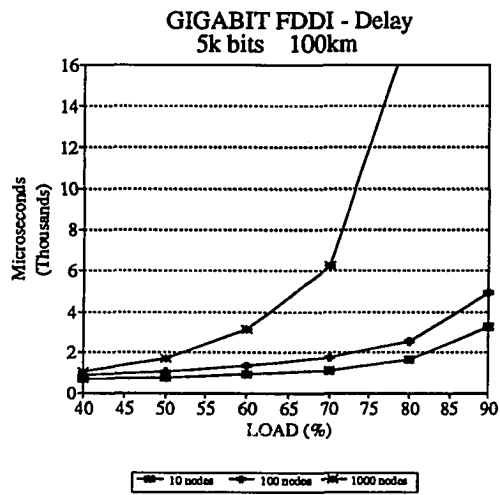


Figure 3.1: Scaling: Impact of Nodes on Message Delay

nodes, additional nodes capture the token on each cycle, introducing additional delays. This is partially a phenomenon of the simulation. The simulation generates packets at the rate specified by the load and packet size. Each time a packet is generated, it is assigned randomly to a node on the network. As a result, the generation of more packets results in a wider range of packet distribution, which in turn results in additional token captures, etc. Other strategies could be used to distribute the packets, but they were not employed in these simulations.

Comparison of the graphs vertically shows the impact which network length has on performance. One would reasonably expect that the increased distance would affect performance solely on the basis of the increased propagation delay. If this is the case, then each graph should have a larger delay in the upper graph, but the amount should be the same for each graph. This is approximately true for the two cases of ten and 100 nodes. Each curve is translated vertically about two milliseconds. For the 1000 node curves, however, this is only true up to a load of about 70%. One can also infer from this difference a few things concerning throughput. The knee in the curves for 1000 nodes indicates that beyond 700 Mbps the delay becomes excessive and that the gigabit transmitter is only permitted to deliver between 700 and 800 Mbps. Note that for ten or 100 nodes, throughput in excess of 900 Mbps is achievable; hence, as opposed to the other two control parameters, network and packet length, the number of nodes appears to be the most difficult of the three to extend.

Packet Length

As I describe in the previous section, packet length and number of nodes, in combination, can have an effect on performance. Figure 3.2 reinforces the previous results. A slightly different set of runs is graphed. All three graphs are made for the longest network, 100 km. The three graphs show scenarios where the number of nodes equals ten, 100 and 1000. Notice that in the first two cases, the effect of packet size is practically insignificant.

However, when the number of nodes increases to 1000, the delay varies significantly. Clearly in this case, the packet size has an effect. The point made is that packet size only has an effect when the number of nodes is expanded. The last case shows that increasing the packet size from 5000 bits to 15000 bits cuts the delay in half for up to 70% and by a more significant quantity for 80%.

Network Length

The last set of graphs in Figure 3.3 shows four scenarios similar to Figure 3.1. In fact, it is the same data, simply presented in a different organization to more clearly see when distance has an impact and when it does not. One would anticipate that the impact of propagation delay is relatively large for the 100 km case and proportionally less for the other two cases. The first row of graphs indicates that although the increased length has a negative effect on the network, delay behaves in a predictable, somewhat proportional manner. In these cases, delay is at or below two milliseconds for all values of distance over the entire range of loads. The second row of graphs makes it unclear whether the number of nodes or packet length is the cause of the problem, but recalling the graphs from Figure 3.2 reinforces the assumption that at 100 km, varying the packet size had almost no effect for up to 100 nodes, and that the number of nodes is the major factor. Once again, the impact of increasing the number of nodes in the second row of graphs as compared to the first row is the most significant parameter.

3.3 The Interface Bottleneck

The design of an efficient system requires a carefully matched set of components. If this part of the design process is ignored, inefficient components will limit the efficient ones. One of the major problems in data communications is the inability of actual networks to achieve data rates comparable to the transmission rate of the network. The sources of overhead limiting this throughput are numerous and protocols are typically the first component to

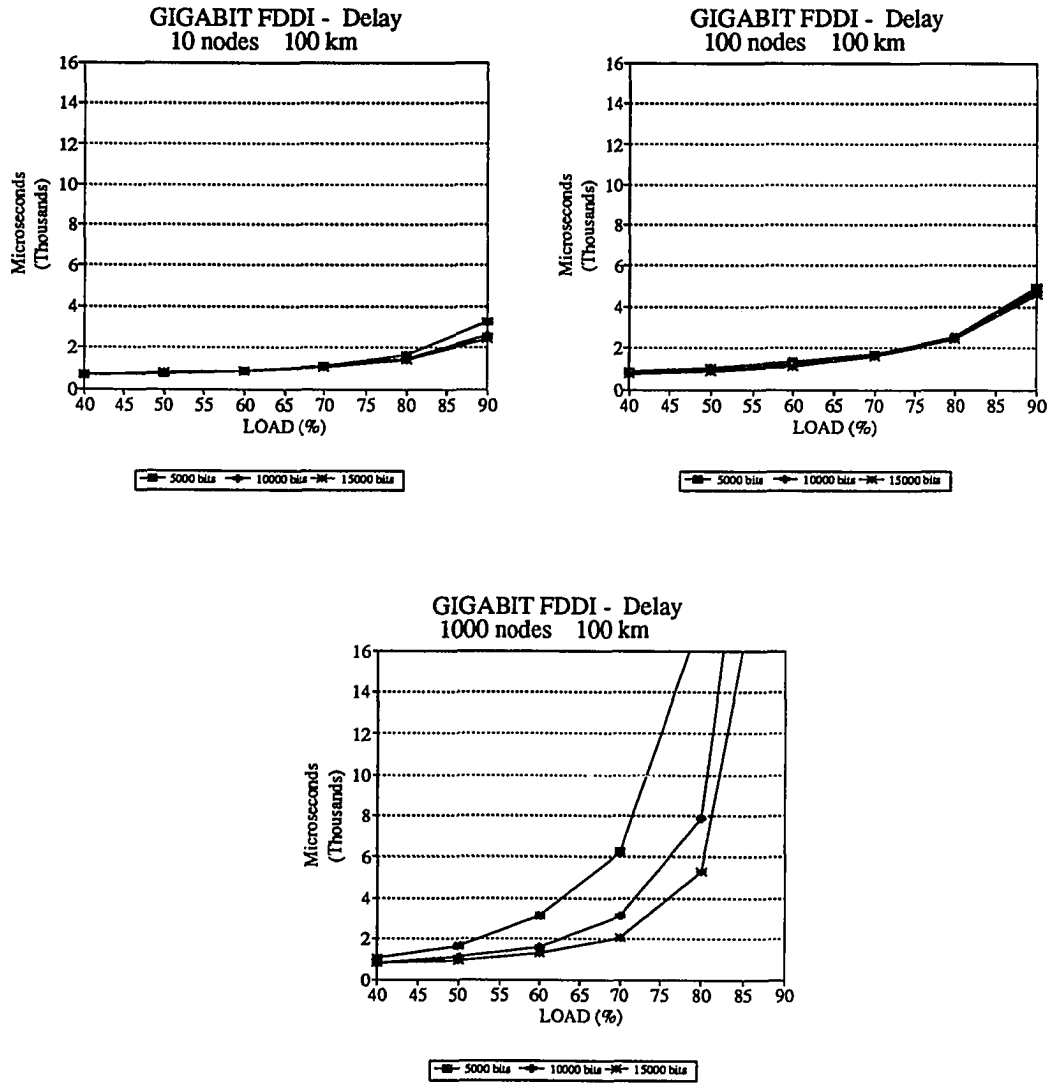


Figure 3.2: Scaling: Impact of Packet Size on Message Delay

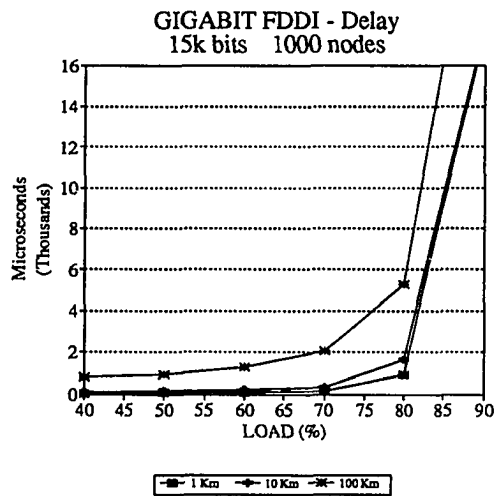
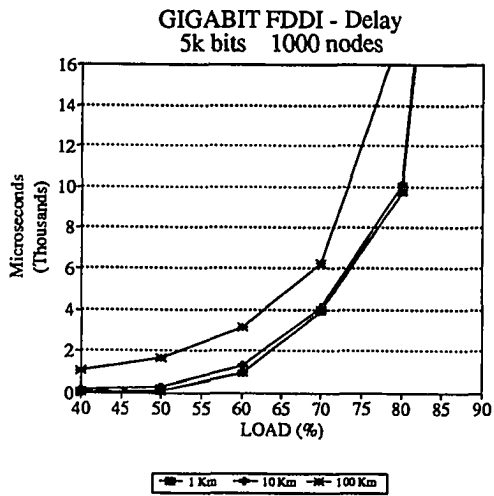
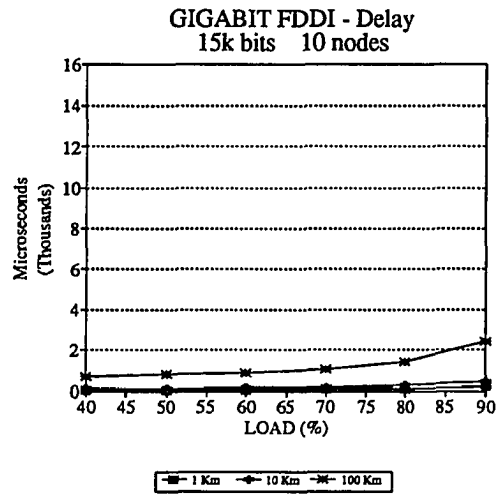
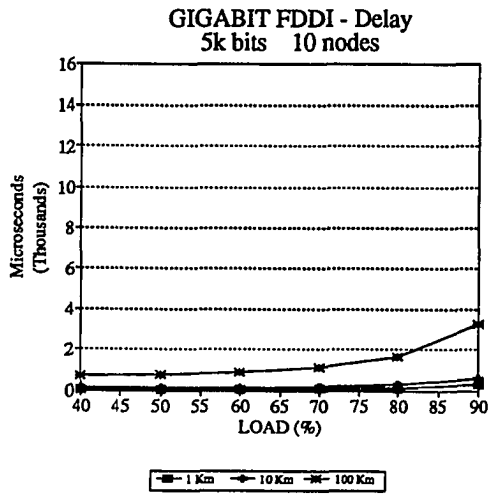


Figure 3.3: Scaling: Impact of Network Length on Message Delay

bear the blame. Although the protocols themselves contribute to some of this overhead, other factors including the operating system and inefficient hardware interfaces appear to be more significant limitations to efficient network utilization. In [8], Clark suggests that it is not necessary to focus on protocols such as the Transmission Control Protocol (TCP) which have been theorized as a major source of the inability of a workstation to deliver data at network rates; rather, the focus should be placed on different processor boards with special memory and controllers for copying data, computing checksum, and other network functions.

It would stand to reason that more attention should be given to streamlining the operating system/network controller interface. The basic paradigm used in viewing this interface is currently that of an input/output device requiring the operating system to move the data from user space to its own buffers, packetize the data, then move the data to the proper input/output ports representing the network controllers. Disk accessing follows this model, but primary memory is interfaced in a much more intimate fashion. In a virtual memory operating system, pages to be swapped in and out are not buffered in main memory by requiring excessive data movement. The use of file servers on a network to serve as remote disks now requires a rethinking of the need to redesign the interface between the processor and the network to treat it more as an extended memory rather than as an input/output device.

Maly and others [41] are currently in the process of investigating a means of improving the performance at the processor/network controller interface which will allow a node to have access to a gigabit network by using parallel FDDI interfaces. Research such as this will provide additional insight into the specific nature of the interface problems and provide for more streamlined interfaces.

In this section, the interface between the physical layer and the link layer is investigated. Another way of viewing this section is as an investigation of the interface between the physical transmitter and the network controller board, when the transmitter is scaled to

a gigabit per second transmitter. In an attempt to determine the sensitivity of previously discussed timing parameters to FDDI running at gigabit rates, this section presents results to establish the need to reduce LMAX³ timing ranges if higher speed transmitters are employed.

3.3.1 Timing Parameters

In Chapter 2 a number of factors contributing to overhead on an FDDI network are discussed. Included in these are:

- propagation delay,
- replication of the token,
- internal delay at the node,
- delay between recognition of the token and transmission of the impending frame.

The impact of propagation delay is shown in the previous section where MANs and WANs are simulated, but the last three of these items are not considered. It is the purpose of this section to establish which, if any, of these timing factors has a significant impact on gigabit rate FDDI performance.

Four curves are generated for each scenario examined. The interpretation of the curves is as follows.

- The *Normal* curve shows FDDI results using all standard specifications for the timing parameters. The only material difference from standard FDDI is the increase in transmission rate to one gigabit per second.
- The *Token = 1* curve eliminates the overhead due to retransmitting the token each time it is captured by setting the token length to one bit. Obviously this is an unrealistic assumption, but by looking at the minimal token length, it will be possible to assess whether or not token passing efficiency is a concern. The reason for setting

³see Section 2.5.1

	LAN	MAN	WAN
nodes	10	100	1000
packet length	5000 bits	10000 bits	15000 bits
network length	1 km	10 km	100 km

Table 3.2: Scaling: Gigabit Timing Scenarios

the token frame length to one bit instead of zero is the result of a timing problem in the simulator itself and roundoff occurring with events essentially taking place at the same time.

- The *NodeDelay* = 0 curve eliminates the processing delay overhead inherent at each node.
- The *LMAX* = 0 curve illustrates what happens when we assume perfect token recognition; that is to say that the impending frame transmission begins at the instant of the token arrival.

Three scenarios are considered here: LAN, MAN and WAN as defined in Table 3.2.

Access delay is considered because of the desire to minimize distance influence from the results as compared to the previous set of runs. Propagation delay cannot be looked at in an orthogonal manner as can the three control parameters used here.

3.3.2 Results

The results in Figures 3.4, 3.5, and 3.6 leave no doubt of the negative impact of the value of *LMAX* on system performance. In every case tested, the impact of eliminating the token and the node delay is barely detectable by viewing the graphs. On the other hand, setting *LMAX* to zero results in a significant change in the access delay curve. At 90% loads, the curves with *LMAX*= 0 show a reduction in delay of anywhere from 50% to 90% of the normal settings. Note that each curve is scaled to the data in the specific graph as opposed to a common scaling. The rationale for this method of scaling is that the comparison is

only with the other curves on that graph rather than with other curves.

3.4 Conclusions

This chapter shows the effect of the number of nodes, network length and packet length for FDDI at gigabit rates. Most of the results show that over the range of parameters examined, delay is on the order of a couple of milliseconds for loads below the 60% level. As load increases above 60%, the delay degrades at different rates depending on the specific case examined. The number of nodes is the one factor which has the greatest effect on performance of the three parameters considered. For numbers of nodes equal to 1000, throughput will be limited to the range of 700 to 800 Mbps, where throughput of over 900 Mbps can be achieved for smaller numbers of nodes. In addition, the number of nodes compounds the problems even more when increased in conjunction with reducing packet sizes. This result further substantiates the significance of the results of destination removal. Destination removal takes advantage of the large number of nodes to offset the negative effect revealed here and to extend dimensions of viability of FDDI to solve the HSLDN problem.

As physical layer transmission rates increase, it is important to understand the significance of matching the interfaces so that the entire communications system functions efficiently. Section 3.3 shows that a reduction in overhead of node delay and token length has essentially no effect on performance, and that the bottleneck which needs improvement is the ability to recognize the token and begin transmission of the frame in a tighter time frame than that defined by LMAX.

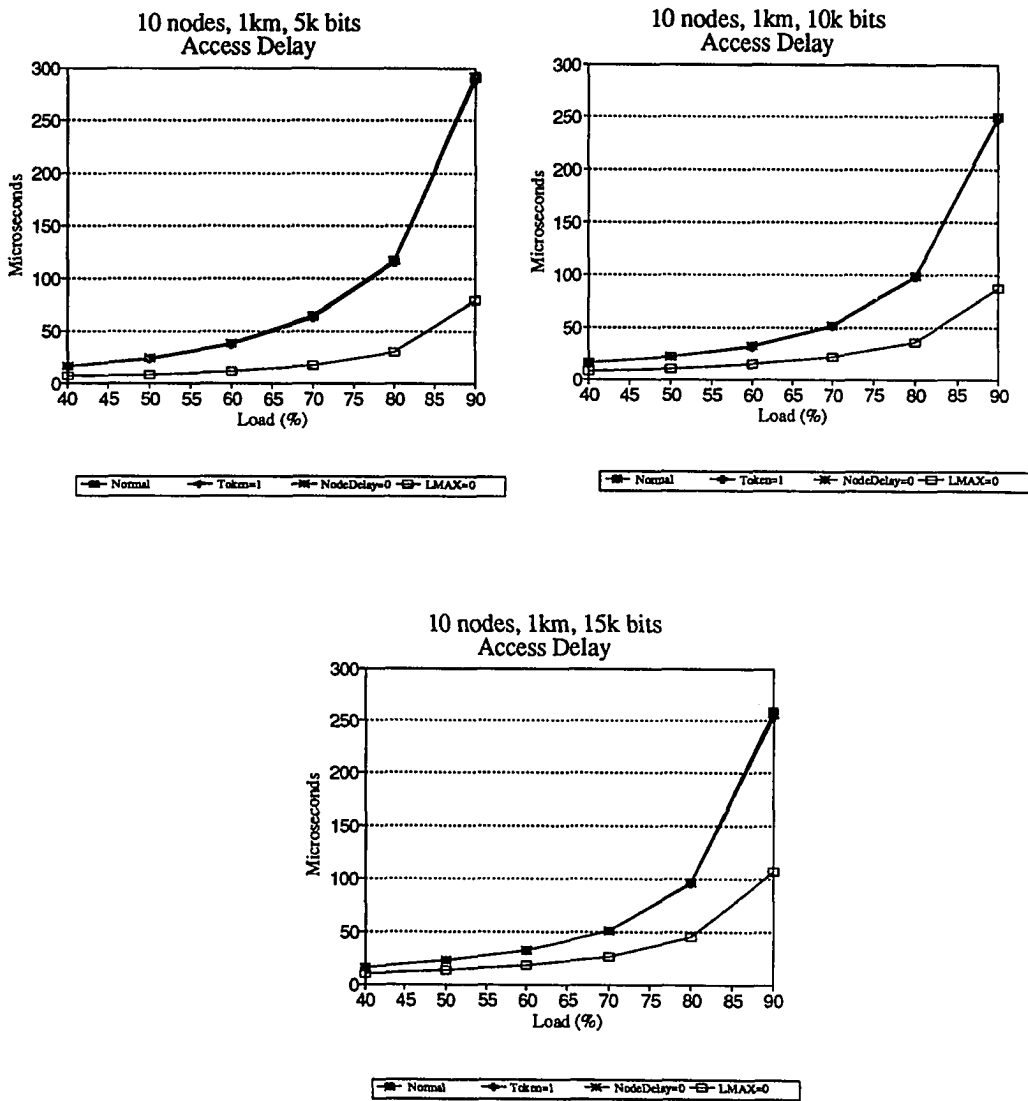


Figure 3.4: Scaling: Impact of Timing Parameters on Gigabit LAN

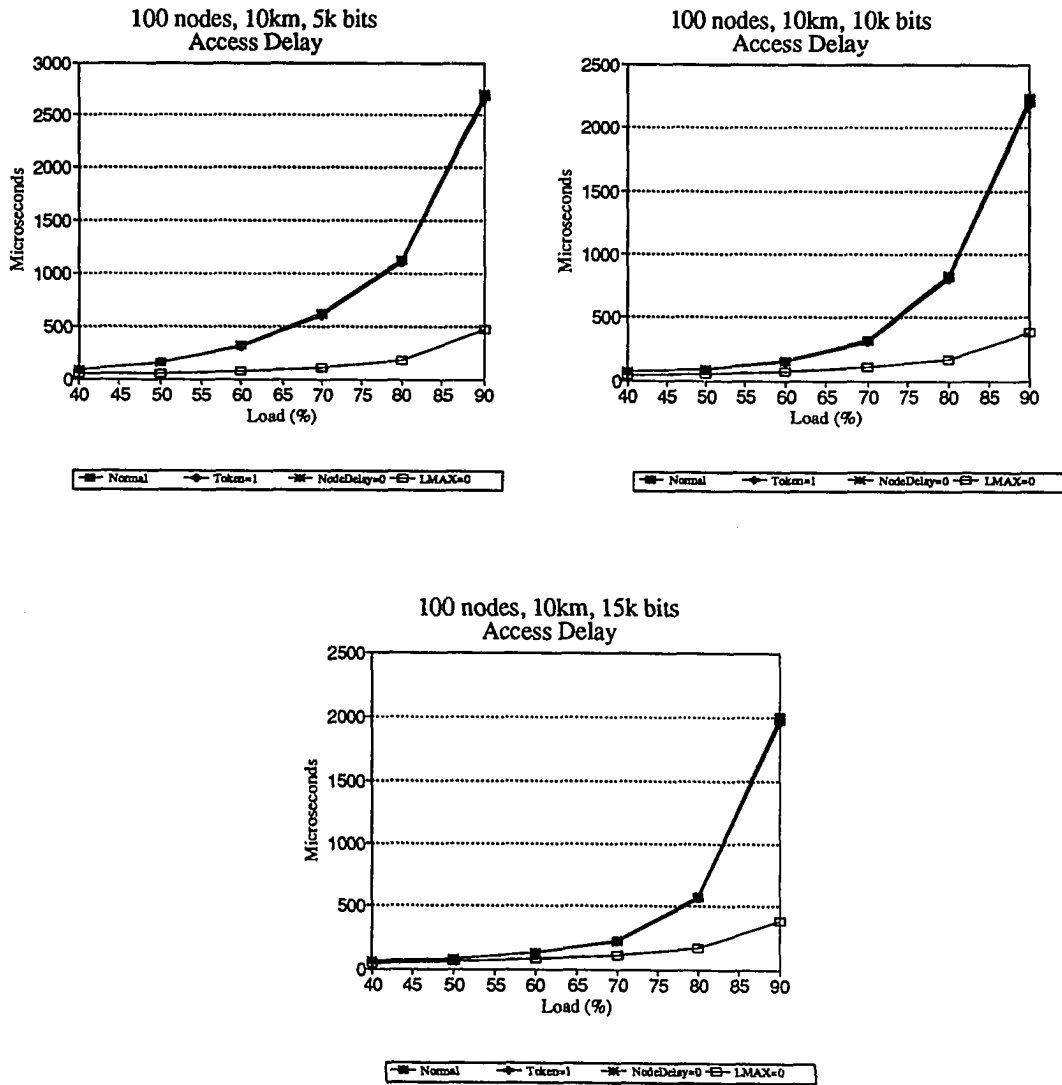


Figure 3.5: Scaling: Impact of Timing Parameters on Gigabit MAN

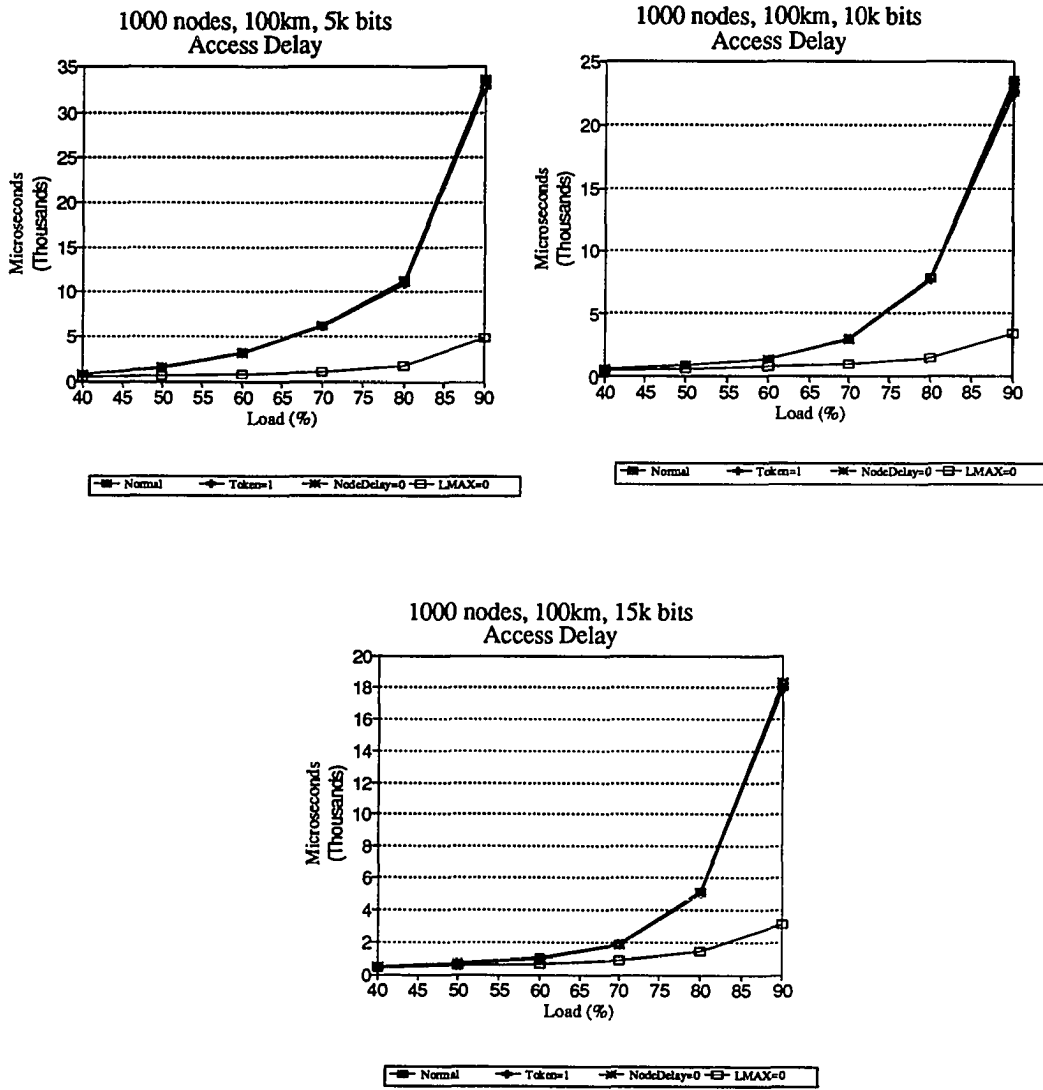


Figure 3.6: Scaling: Impact of Timing Parameters on Gigabit WAN

Chapter 4

Destination Removal for FDDI

Let us consider some of the problems of using FDDI as a means of implementing a HSLDN. For token rings such as FDDI, access delay is increased due to the longer token cycle time for these distances, but the point at which the delays become unacceptable is application dependent. Results in Chapter 3 on FDDI at gigabit rates indicate that the increased number of nodes is the most negative parameter when sizes up to 1000 nodes are considered with timing parameters set at levels defined in the standards documents. The propagation delay also contributes to the degrading performance and presents a significant problem for extensibility of token rings, especially at national network distances.

This chapter presents suggestions for modifying FDDI to enhance performance by utilizing the second counter-rotating ring and destination removal of packets. This does not constitute a proposal to change FDDI as it is currently defined; however, the concepts and results presented are sufficiently compelling to consider a modified FDDI as an approach which is indeed more extensible and as a possible improvement in the efficiency of token ring networks. FDDI is used as the vehicle for illustrating the concepts because it has a second counter-rotating ring and is of current interest.

Related previous research is discussed in the first section in order to put this material in the proper context of its originality. The second section of the paper provides an explanation of how destination removal works in a token ring and points out some of the conditions necessary for its use. The third section presents an analysis which defines an upper bound of sorts on the throughput improvements expected. The fourth section contains simulation

results showing the actual improvements realized and the effects of including a second counter-rotating ring. The fifth section discusses issues for implementation of destination removal, and the conclusions constitute the last section.

4.1 Related Work

A number of techniques including multiple rings, multiple tokens, destination removal of messages, and register insertion have been proposed in an attempt to improve performance on a token ring. FDDI [63] is designed to incorporate two counter-rotating rings, although the secondary ring is not used to transport data simultaneously with the primary ring. The secondary ring is used for fault tolerance and error recovery. Casale [5] is implementing a dual ring optical fiber LAN, Deterministic Access Fiber Network (DAFNE), which incorporates counter-rotating rings and supports simultaneous data traffic on both rings. Published simulation results indicate that maximum throughput increases are approximately double that of a single ring, similar to some of the results presented in this dissertation. The primary focus of DAFNE is to address the hardware design issues associated with simultaneous transmission on both rings by a node, not to optimize throughput.

Kamal [32] shows that increasing the number of tokens lowers delays at low loads, but increases delay at high loads, with no overall increase in network throughput. The use of additional tokens on the *same* ring has the advantage of reducing access delay under certain conditions, but under other conditions, the tokens increase the complexity of network operation with little benefit. Consider a situation where the frame length is much shorter than the length of the network, similar to Figure 2.1. In part C of this figure it is possible to see how another token could be used by node one, if node one's frame is sufficiently short to allow for transmission before arrival of the circulating message, *MSG-3*.

This cannot be guaranteed, however. If a node holds a token while receiving a message currently being transmitted by another node which holds a second token, one of the two nodes must cease transmission so that the other message can be delivered, or it must buffer

the incoming message until it completes transmission of its own message. Other possible scenarios include the arrival of another token while transmitting a message. Under normal use of a token, this would result in the inevitable bunching of tokens if the second token is held until completion of the message transmission; the original token would be reinserted on the ring and the second token also forwarded. Kamal suggests letting tokens interrupt message transmission so that they continue propagating around the ring and using the *old token* for the original token, marking it free at the sender rather than retransmitting the token at the end of the message.

The token management problem is difficult at best. Kamal has given an algorithm for handling these tokens, but he has shown no proof of correctness or indication of robustness of the method. More importantly, the tradeoff of network management complexity for faster access at low loads is not obviously justified. Given that the focus of my research is to increase throughput, this approach does not appear to be advantageous even if the token management problem can be solved.

Dobosiewicz describes a technique by which an Ethernet channel is segmented to minimize the probability of collisions [12] and support multiple simultaneous transmitters, however, the throughput increase is not able to sustain even the basic transmission rate of the channel. Throughput is increased from approximately 60% for standard Ethernet to 70% for the segmentation scheme in the scenario considered.

Concepts related to destination removal as described in this chapter have been employed in an attempt to improve performance in other types of networks, including rings and busses, but each implementation is different in the context of the type of network in which it functions. The Cambridge Ring was one of the earliest slotted rings and was proposed in 1980 [27], followed by a number of additional proposals for optimization of the original design. A slotted ring employs a latency buffer to assure that an integral number of slots exists on the ring; each slot contains a flag marking the slot as full or empty. If a node has data for transmission, it marks the slot as busy and transmits its data. Different methods

exist to release the slot for subsequent use: the source can count the number of slots, anticipate the arrival of the frame, and mark it as empty when it arrives; or the destination node can mark the slot as empty. The latter technique, named Orwell, was discussed by Falconer and Adams in 1985 [16] and marks one of the earliest uses of destination removal. Zafirovic-Vukotic [76] summarizes the various access mechanisms for slotted rings, includes a comparative analysis, and shows that for Orwell, the throughput approaches 240 Mbps for a 140 Mbps channel, representing the ability to accommodate a load approximately 1.7 times the transmission rate. A more recent proposal for the Cambridge Ring, the Cambridge Backbone Ring, will run at 800 Mps and is currently under development [21], but it does not use destination release of the slot.

Register (buffer) insertion rings also have the potential for destination removal, although the message is typically removed by the sender [1,4]. A similar method is suggested by Foudriat [17] for removal of messages at the destination on a CSMA ring network, CSMA/RN. Foudriat shows that CSMA/RN is able to support loads of two times the channel rate of the network.

DQDB, a dual bus MAN which operates at 300 Mbps, has also been considered for destination removal. Slots are generated at one end of the bus and are propagated in one direction to the other end of the bus where they are removed. In the current design of DQDB, a slot can be used only once. Garrett [19] and Rodrigues [55] both consider insertion of erasure nodes on the bus to redefine slots containing data previously delivered so that the slots can be reused. Position of these erasure nodes is crucial to performance and relies on a reasonably static nature of message distribution in order to optimize placement of the erasure nodes. Garrett shows that for a 100 node network, three busses can improve throughput by 40% and that improvement asymptotically increases to a factor of two by increasing the number of erasure nodes.

Xu [74] suggests a destination removal technique for a single ring token ring, and reports the ability to deliver traffic up to 1.5 times the channel rate. The reuse of slots is accom-

plished through Conditional Tokens which only allow use of the frame based on physical position. Xu provides no analysis to show whether this is a minimum, a maximum, or an average case. No indication is provided indicating the relevance of network parameters such as number of nodes, packet length, transmission rate, network length, etc., to the results shown. The contribution of this chapter is not only to provide an analysis showing the expected gains achievable by such a technique, but to also combine the destination removal with counter-rotating rings to significantly exceed the performance of destination removal in token rings and in other types of networks.

The previous work in this area has been able to show throughput increases up to a factor of two times the basic transmission rate of the channel. The fundamental contribution of this work is a technique for token rings to employ destination removal with counter-rotating rings and sustain traffic over 2.5 times the transmission rate for *each* ring; two 100 Mbps rings are able to support a traffic load close to 500 Mbps.

4.2 Destination Removal for Token Ring

The focus of the suggestion for improvement in performance of FDDI in this paper is on recovering the unused packet capacity by removing the packet at the destination and inserting new packets in their place; this technique is commonly referred to as *destination removal*. Destination removal will allow for multiple simultaneous transmitters on the network and increased throughput.

Destination removal works as follows. When a node transmits a message on the network, the message proceeds until it reaches its destination. At that point it is marked as received and the slot containing the message is available for further use. As long as the fixed-length slot is on the ring, it can be used by other nodes. This factor raises two additional questions which are addressed in subsequent sections:

- how long the slot remains on the ring, and
- which nodes (messages) are candidates for reusing the slot.

Figure 2.2 illustrates the process of appending a message to the train of messages. This figure also shows the arrival of *MSG-3* at the receiver. In Figure 2.2.B *MSG-3* has reached the destination node and is now wasting network bandwidth. The shaded section emphasizes that the slot is only delivering an acknowledgment and that the capacity of the network is not being used in an optimal fashion. In this case, the slot could have been used by node three to transmit the message to node two; i.e., two messages could have been delivered instead of one with this packet-slot.

Figure 4.1.B depicts node three removing the message from node five and at the same time inserting its message to node two. The removal of the token at node four and transmission of the message from node four to node one are unaffected by this squeezing of the message from node three to node two into the train of packets. *MSG-2* is once again available for reuse. The squeezed data cannot be longer than the message which it is replacing or it will possibly overwrite trailing messages on the network.

4.2.1 Restrictions on slot reuse

A message transmitted on a standard FDDI ring (without destination removal) would normally travel at least as far as the original sender and could conceivably travel even further if the original sender releases the token before the packet returns. If the time of transmission of the message (t_m) is at least as long as the walk time (w) of the network ($w < t_m$), the message will terminate at the sender. If the time of transmission is much less than the propagation delay ($w > t_m$), the new token will leave the node before the message has made a loop, and the message may not reach the node holding the token for a significant period of time.

When $w < t_m$, the packet arrives at the sender before transmission is complete. Consider this case as it affects destination removal. Assume that node one holds the token and has a message for a node i . When node i receives the slot, it is reusable by any node up to (but not beyond) node one, to send a message to any node between the node using the slot and the original sender of the slot, node one. If an attempt is made to send a message

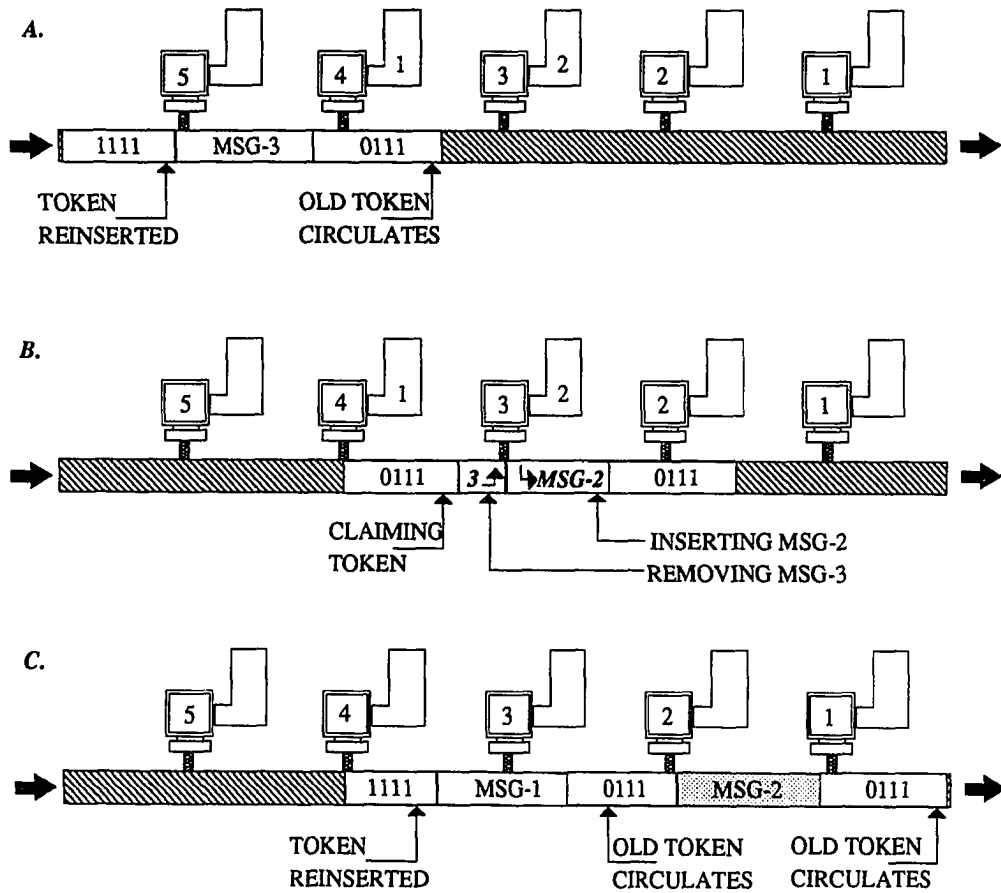


Figure 4.1: Removing Packet at Destination

past node one, the message will be absorbed by node one, the holder of the token. If $w > t_m$, as is the case in Figure 4.1, additional reuse could be made of the packet, but the opportunity for reuse will be terminated at the original sender in this analysis. The proposed method assumes that the original sender will absorb the message and is, therefore, the conservative approach. The optimistic approach would be to attempt to take additional advantage of the slot, going on the assumption that the original sender of the message had already transmitted the token, resulting in a continued traversal of the ring by the slot. This optimistic approach is not considered in this analysis.

Length of packet also presents a problem, as I mention above. Obviously, a message which is inserted into the free slot must of length less than the length of the delivered message. For this analysis, the assumption is made that all packets are of equal length. Fixed packet lengths are a common occurrence in networking strategies. For example, DQDB [52], the primary FDDI competitor for MANs, is slotted and uses fixed-length slots similar to what is proposed here.

4.2.2 Objective

If *destination removal* recaptures the slots containing messages previously delivered, one expects throughput to increase and delay to decrease. The central questions are how much these measures would be affected and how feasible the implementation would be. Recent papers [25,11,17] have used similar techniques to show increases on the order of 1.5 to two times, but these papers have not included any generalizations as to how this technique might apply in an arbitrary case; the simulation results are for specific cases. This analysis allows one to predict the degree to which a method such as destination removal can improve throughput in FDDI or token rings, and to show that a feasible strategy can be developed which does not meet the expectation of Section 4.3, but can approach it.

4.2.3 Advantages

One might assume that the effect of *destination removal* would be simply to provide for an increase in throughput at high loads and have little effect at low loads. However, the method can be shown to improve performance in the following areas.

- Traffic loads at a much higher levels can be supported.
- One of the major problems with token ring networks is the access delay for obtaining the token. These extra slots will reduce average access delay to the network.
- The sensitivity of a token ring network such as FDDI to longer distances and large numbers of nodes will be reduced.

In addition, the fair nature of the underlying basic token ring is preserved because destination removal will only send the message earlier, but never later. The token continues to circulate as in traditional token ring networks. Therefore, this method does not suffer from the fairness problems of most random access strategies such as slotted rings and CSMA/CD.

4.3 Analysis

For this analysis, which focuses on throughput as opposed to delay or other related performance metrics, the assumptions will be that all nodes always have at least one packet in the queue, that destination address space is uniformly distributed among the nodes, and that packets are of fixed length.

4.3.1 Expected Maximum Throughput Increase

If all messages were destined for the neighbor, throughput could be increased by a factor of N^1 , but this is an unlikely scenario. The result derived is a function of the number of nodes, and this function is designated $\mathcal{E}(N)$, the factor by which throughput can be expected to improve under heavily loaded conditions. For example, if utilization is currently 80% and

¹ N represents the total number of nodes in the network.

the expected throughput increase, $\mathcal{E}(N)$, is 1.4, then utilization should be able to reach 112% under the conditions specified in the assumptions above for N nodes.

In order to determine the expected throughput for such a network, consider the traversal of a packet in a single loop around the network in the direction from N down to one. Assume that node N removes the token from the network and transmits a packet. The packet must be destined for one of the nodes $N - 1 \dots 1$. Assume that the packet is destined for node j , where $N > j \geq 1$. Upon receiving the message, either

1. node j has a packet available for transmission to node k where $j > k \geq 1$ or $k = N$ which states that the message can be squeezed into the now available slot and removed before it passes the original sender, or
2. node j has a packet available for transmission to node k , where $N - 1 > k > j$ and it cannot be squeezed without a possibility of being removed by a node which has the token (specifically node N may still be transmitting, so the assumption is that node N is transmitting and that the slot cannot be used).

Define $\mathcal{E}(j)$ to be the factor of expected increase in throughput if the slot has as its destination node j . Let us start with $\mathcal{E}(1)$ and use a recursive derivation

$$\mathcal{E}(1) = \frac{1}{N-1} \times 1 + \frac{N-2}{N-1} \times 0 = \frac{1}{N-1} \quad (4.1)$$

because the expected value of increased throughput is the product of the probability of the message in the head of the queue being destined for node N (the original sender of the slot and the only possible node to which node one can send), $\frac{1}{N-1}$, and the number of messages it can send in the slot, one, plus the product of the probability that the message at the head of the queue is for some other node, $\frac{N-2}{N-1}$, and the number of messages it can send, zero.

The expected increase at node two is composed of three terms:

1. the probability that the message at the head of its queue is for node N , $\frac{1}{N-1}$, times its expected increase (which is one because the slot will be used to send a message to node N and then no longer be used),

2. the probability that the message at the head of the queue is for node one, $\frac{1}{N-1}$, times the expected increase (which is $1 + \mathcal{E}(1)$ because the message is delivered at node one and can be reused again at node one with expected increase $\mathcal{E}(1)$), and
3. the probability that the message at the head of the node is for neither node 1 nor N , $\frac{N-3}{N-1}$, times the expected increase, which is also $\mathcal{E}(1)$ because the slot will proceed to node one available for reuse

Therefore,

$$\mathcal{E}(2) = \frac{1}{N-1} + \frac{1 + \mathcal{E}(1)}{N-1} + \frac{N-3}{N-1} \times \mathcal{E}(1) = \frac{2}{N-1} + \frac{N-2}{(N-1)^2} \quad (4.2)$$

For arbitrary node j , the formula can be generalized to

$$\begin{aligned} \mathcal{E}(j) &= \frac{1}{N-1} + \sum_{i=1}^{j-1} \frac{1}{N-1} \times (1 + \mathcal{E}(i)) + \frac{N-j-1}{N-1} \times \mathcal{E}(j-1) \quad (4.3) \\ &= \frac{j}{N-1} + \frac{N-j}{N-1} \times \mathcal{E}(j-1) + \sum_{i=1}^{j-2} \frac{1}{N-1} \times \mathcal{E}(i) = \frac{1}{N-1} \times (j + (N-j) \times \mathcal{E}(j-1) + \sum_{i=1}^{j-2} \mathcal{E}(i)) \quad (4.4) \end{aligned}$$

for $N > j \geq 3$

and

$$\mathcal{E}(j) = \sum_{i=1}^{N-1} \frac{1}{N-1} \times (1 + \mathcal{E}(i)) \quad (4.5)$$

for $j = N$

The first term in Equation 4.3 represents the expected increase if the message is for node N , the original sender. The second term represents the expected increase if the message is for a node positioned between the current node j and node 1 which is the squeezed message itself plus any expected increase once that message is delivered. The third term assumes that the slot could not be used, so it is passed on to node $j-1$. In the case for $j = N$, the first term is omitted, because it would never send a message to node N , itself.

4.3.2 Overall effectiveness

The following graph shows the increase in throughput expected from a traffic placement strategy as described above. The result of interest in the above derivation is the value of $\mathcal{E}(N)$, which describes the number of expected messages delivered with each packet as it is transmitted from the node holding the token (N). Figure 4.2 shows $\mathcal{E}(N)$ versus N . One can observe that the effect of such a technique is greater as the number of nodes increases.

Recall that this graph shows the factor by which throughput can be increased. Note that for a 100 node problem which operates at 90% maximum utilization, this method increases throughput by a factor of three to 270%. A number of curves are provided. The curve marked *Analysis* is the result of calculating $\mathcal{E}(N)$ for various values of n as derived above. The analytical results can be compared with simulation results in the curve *Simulation*.

This simulation model (included in Appendix D) does not utilize a detailed FDDI model and only models the passing of messages from node to node without timing delays, etc. Here a packet is allowed to be reused an arbitrary number of times, an impractical assumption discussed later. The final two curves in this figure show maximum utilization when one limits the number of times which a slot may be used during a cycle around the network. *Max=2* indicates that the slot may be used twice (reused once). Slot reuse is limited because of the overhead associated with each reuse.

Figure 4.18 illustrates the required frame format to implement destination removal. Although a trailer for each reuse is not necessary, a header for each reuse is necessary. As I mentioned previously, acknowledgements for reuse of the slots are not available, but all headers must be retained in order to allow *downstream* nodes to determine whether or not the slot can be reused without requiring delays at each node to process the header before retransmitting. The two curves labelled *Max=2* and *Max=4* are intended to show what happens when the number of reuses is limited and the amount of overhead per message is fixed. This is explained further in Section 4.5.

Destination Removal Prediction of Benefit

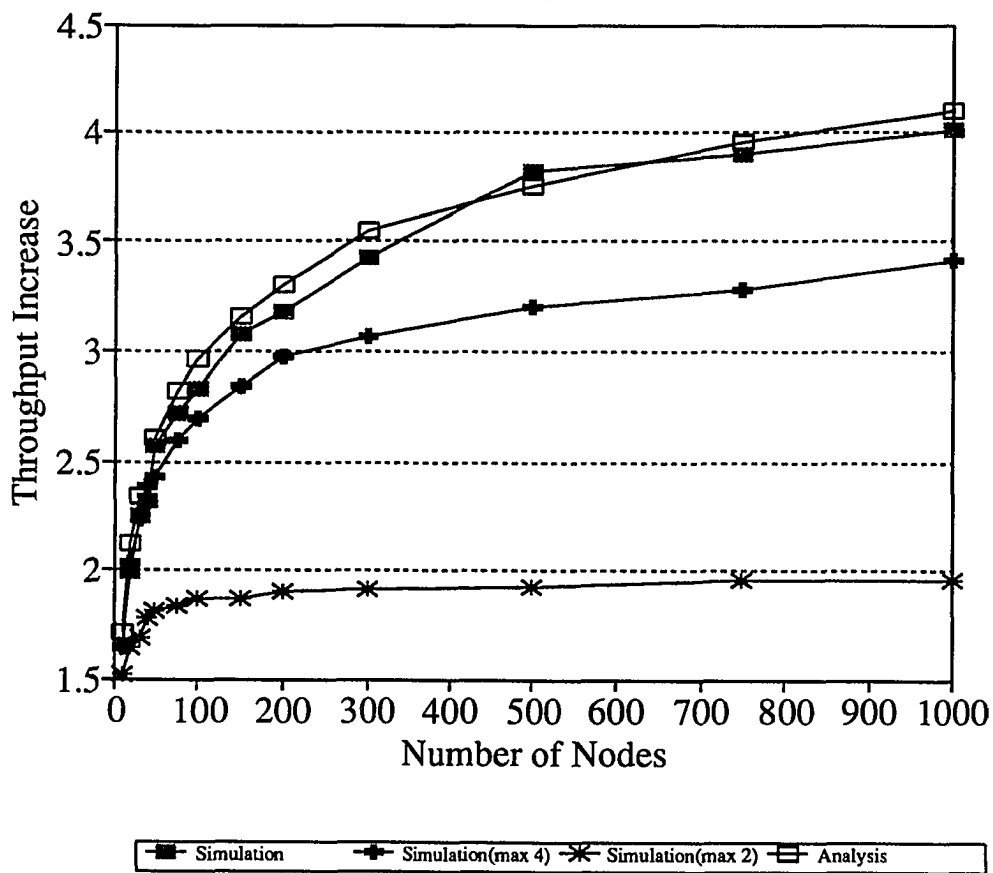


Figure 4.2: Removal: Expected Throughput Increase

4.4 Simulation

I have developed a much more detailed simulation model of FDDI², in order to test performance issues more precisely. A modification was made to the model to allow for the incorporation of destination removal. This simulation model was used to determine how well FDDI might approach the analytical results obtained in the previous section. The primary assumptions used in the analysis of Section 4.3 were that the destination address space was a uniform distribution and that one loop of the ring would provide a sufficient estimate of the throughput. The first assumption is shown here to be a more reasonable one. When other distributions are considered, destination removal produces fairly consistent results. As a test of the robustness of the method under different loading conditions, the file server traffic distribution is also examined in Section 4.4.5.

The assumption concerning the viability of examining one loop of the ring in the analysis turns out to be an optimistic assumption. Consider a single ring under heavy loading, each node sending a message and releasing the token to its neighbor. As the slots move around the ring, it is unlikely that messages which cannot be squeezed into a slot will be able to use succeeding slots.

Figure 4.3.A illustrates the difficulty. In this scenario there are twelve nodes on the ring. The token is rotating in a clockwise direction and node 12 is currently holding the token. Nodes which have messages queued for transmission are shown with the destination of the queued message in parentheses. For example, node 12 has a message for transmission to node five, node one has a message for node ten, etc. The bold arc on the innermost ring indicates the transmission path of the message from node 12 to node five. The remaining arc represents the potential for reuse of the path. Each of the queued messages is unable to use destination removal to insert an additional message. Only nodes five and eight reside on the arc which could accept an additional message, but neither have a suitable destination

²This is *not* the simulation listed in Appendix D. It is a simulation written in Simscript and is too extensive to be listed in this document.

address (the address is beyond node 12).

As the token continues to propagate around the ring, a similar scenario is replayed. When node one captures the token in Figure 4.3.B, the queued messages are still unable to use the insertion arc from node ten to node one. The succeeding token capture at node three still precludes insertion of the messages from nodes five and eight. Hence, the *problem* messages still remain and limit the advantages inherent in destination removal.

These *problem* messages, however, contribute in a positive manner to slots which are moving in the opposite direction. Figure 4.4 shows how a ring rotating in the opposite direction, counter-clockwise, would use these messages. Here the messages from one to ten, eight to five and five to two are all able to be transmitted by the one message cycle. Destination removal is limited primarily by destination addresses which are separated by a large number of nodes; however, messages such as these are also characterized by proximity to the destination for a ring rotating in the opposite direction. The hypothesis is that the dual counter-rotating nature of the second ring in FDDI could be used to take advantage of this property.

In an attempt to determine the actual impact of destination removal, a number of simulation runs were made using the following parameters.

- The number of nodes was the major parameter of investigation and was examined for values of up to 500.
- FDDI parameters for node latency, transmission speed, token size, packet overhead, etc., were taken from the standards document [63].
- Packet sizes were constant at ten kilobits.
- Network distances of up to 5000 km are considered, although the majority of the results are for 100 km distances.
- Token rotation time was set at 40 ms in most cases (higher for distances in excess of 1000 km) in order to avoid having low token rotation times artificially limit utilization

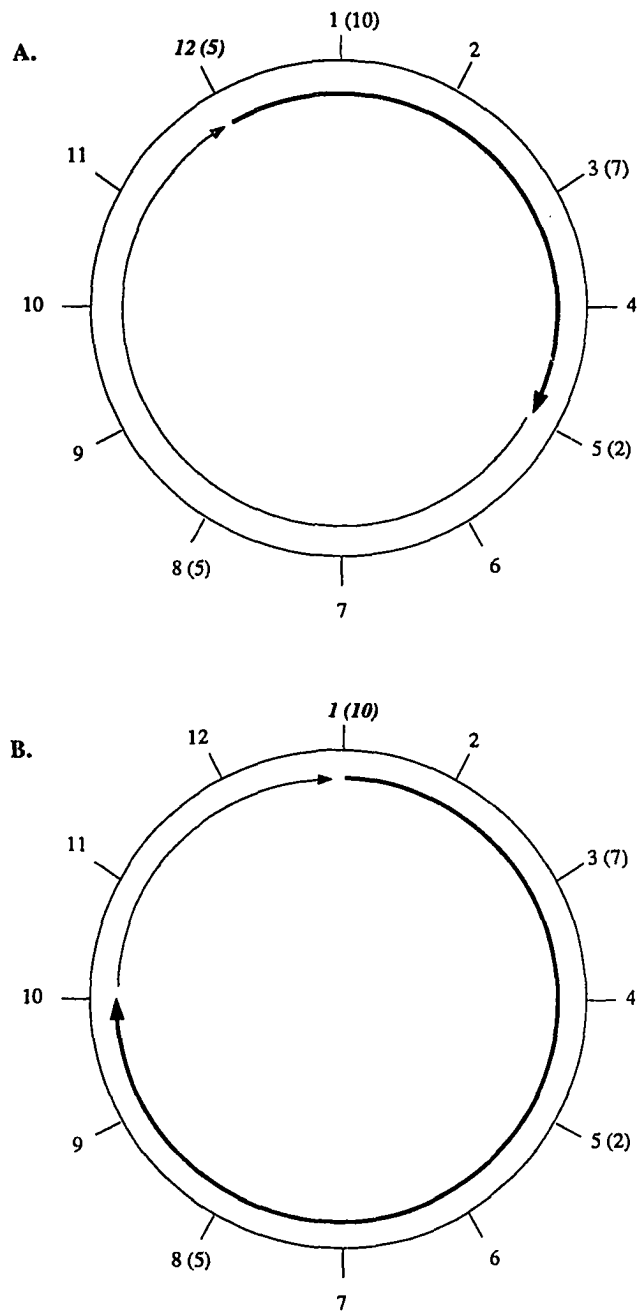


Figure 4.3: Unfavorable Destination Address Distribution for Single Ring

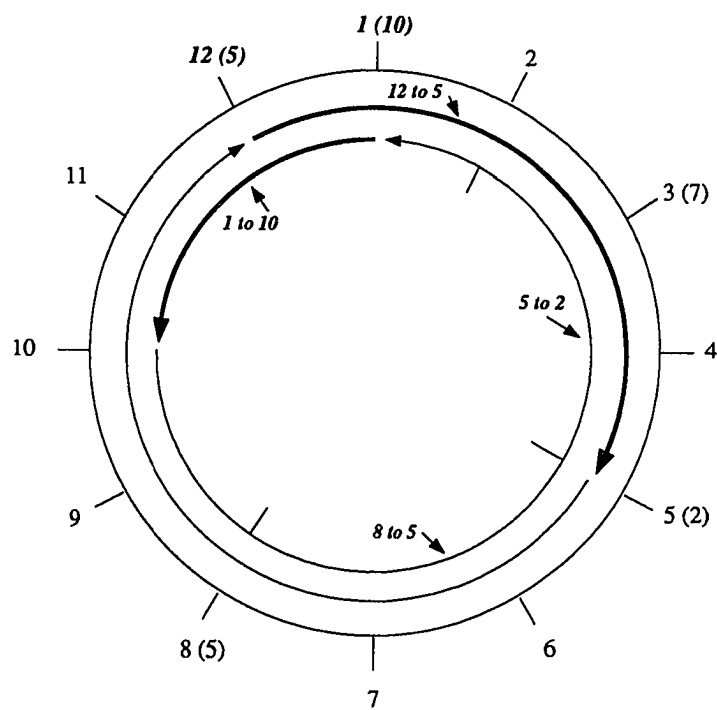


Figure 4.4: Using *problem* Messages on the Second Ring

as I explained in Chapter 2.

- Load was set at levels to assure that the network was overloaded when determining maximum throughput. These levels range from 100% to 900% depending on the scenario tested, and were adjusted with preliminary runs to assure that the network reached maximum throughput without unnecessary excessive loading.

4.4.1 Throughput Comparison

LABEL	LOAD	EXPLANATION
1RwoRMV	100%	1 Ring without Removal
1RwRMV	300%	1 Ring with Removal
2RwoRMV	200%	2 Rings without Removal
2*1RwRMV	N/A	Doubling 1 Ring with Removal
2RwRMV	600%	2 Counter-rotating Rings with Removal

Table 4.1: Removal: Curve Interpretation

The results are summarized in Figure 4.5, which contains five curves. The model associated with each curve is shown in Table 4.1. The first two curves (*1RwoRMV*, *1RwRMV*) illustrate the performance of a single ring FDDI, both with and without destination removal. Three different dual-ring scenarios are also included. These curves show no removal at all (*2RwoRMV*), use of destination removal using counter-rotating rings (*2RwRMV*), and the use of destination with rings which rotate in the same direction (*2*1RwRMV*). The case of two rings with no removal (*2*1RwRMV*) provides a means of comparing the improvement of the counter-rotating characteristic. The values for *2*1RwRMV* are not simulated results, rather they are calculated by doubling the curve *1RwRMV*.

The first curve (*1RwoRMV*) without destination removal goes slightly downward as the number of nodes increases due to the negative impact of numbers of nodes and distance³ on utilization. The second curve (*1RwRMV*) shows that throughput is about 120% for less than 50 nodes, rising to 150% at 100 nodes and 165% for 500 nodes. Even though this increase approaches the performance of two rings which do not use removal (*2RwoRMV*)

³In this graph, each node is separated by 100 meters; therefore, as the number of nodes increase, so does the total distance of the network. A 500 node network is 50 km.

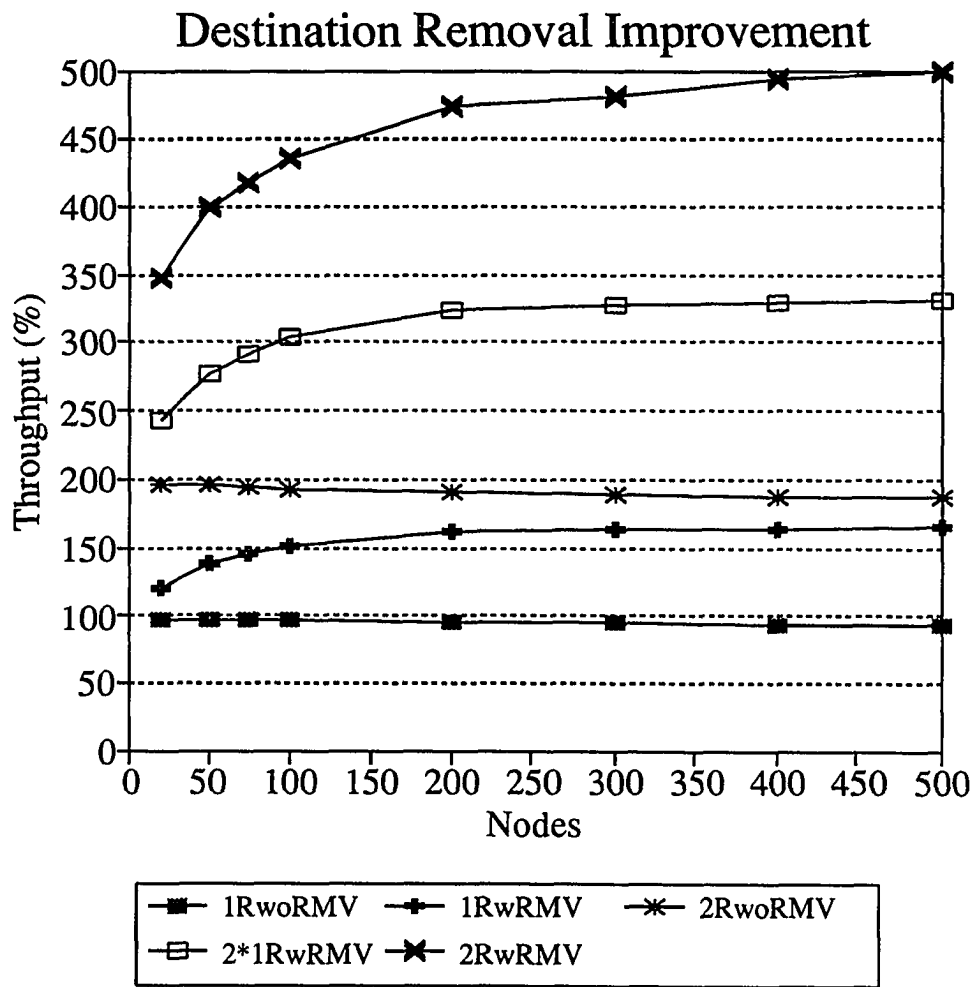


Figure 4.5: Removal: Summary of Results

for large numbers of nodes, it is still far below that predicted in the optimistic case of the analysis.

For large numbers of nodes, a single ring destination removal network approaches the performance of a dual ring FDDI network without destination removal. The fourth curve ($2*1RwRMV$) indicates that the anticipated throughput of two rings using removal is approximately 330%. In comparison to $2*1RwRMV$, the fifth curve ($2RwRMV$) reveals the major result of this part of the research, the effect of the counter-rotating nature of the rings which uses the *problem* messages of a single ring as an advantage. Performance improvement for even small numbers of nodes is on the scale of 100 Mbps of additional throughput over what would have been anticipated by employing two rings. The combination of destination removal and counter-rotating rings is capable of approaching throughput levels of 500% (500 Mbps).

Figure 4.6 provides a means of comparing the simulation results more directly with the previous analytical results. The derivation shows the expected number of uses of the slot as does this figure. Each time a message is placed on the ring as a result of a token capture, the number of times the message is used prior to reaching the original sender is counted (including once for the original message). The graph reveals that each slot in a single ring is used between 1.5 and 1.7 times for more than 100 nodes; however, each slot in a dual counter-rotating ring is used between 2.4 and 2.6 times for a comparable number of nodes. Once again, it emphasizes the advantage of the counter-rotating nature of the ring. In addition, the figure shows the degree to which the network can approach the results predicted by the analysis.

4.4.2 Long Distances

The basic definition of the problem investigated by this research includes, but is not limited to, long distances. In an attempt to determine applicability of destination removal for distances suitable for the National Research and Educational Network (NREN), the design must be able to function in distances over 100 km; therefore, distances of 1000 and 5000 km

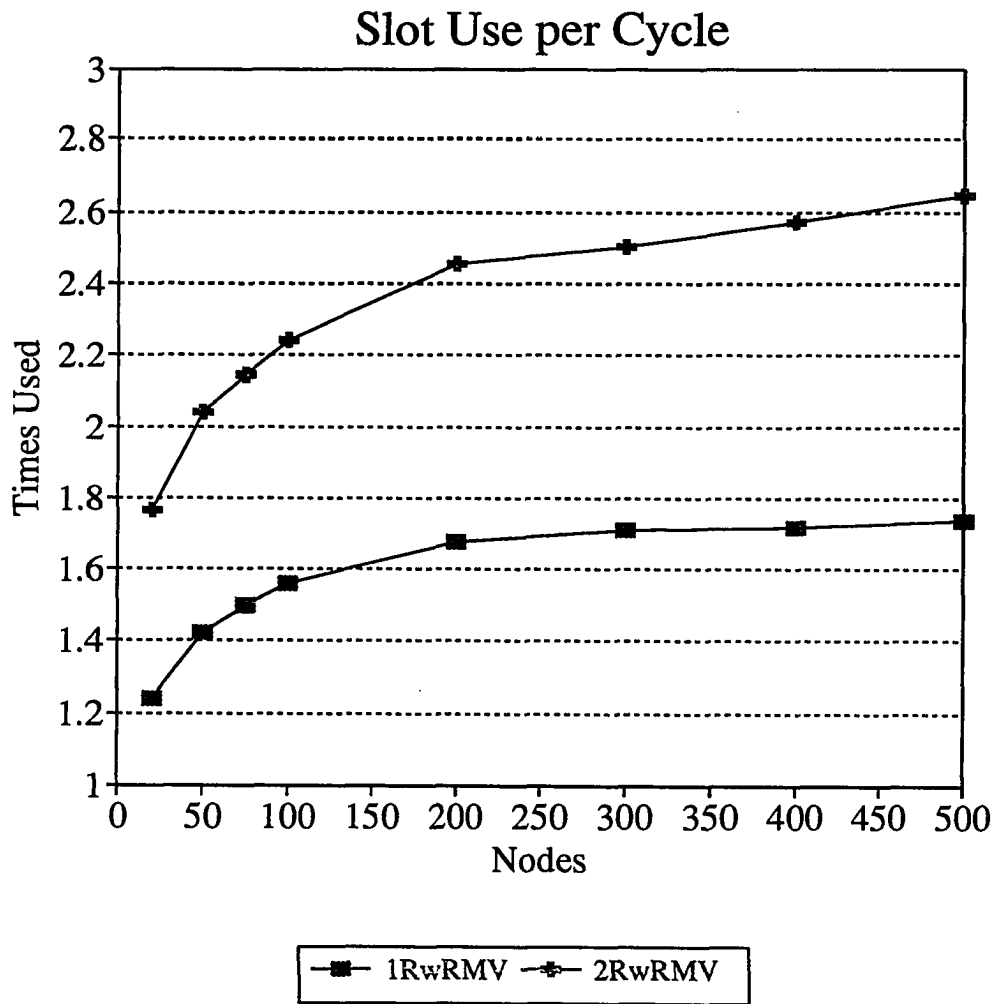


Figure 4.6: Removal: Slot Reuse

are evaluated. Figures 4.7 and 4.8 contain results similar to the last two figures with two basic differences. First, the distance is the same for all data points on a graph; internode distance varies with each data point on a curve so that total network distance remains the same. Second, due to the long distances, the TTRT settings are adjusted as annotated to assure that the basic ring operation is not choked by low timer limits.

Figure 4.7 illustrates the impact on throughput of distances sufficiently large to cover the entire country. The two graphs at the top of the figure tend to imply that destination removal suffers to some degree at longer distances. The graph at the bottom of the figure dispels this assumption. The difference between the graph in the upper right and the one at the bottom of the figure differ only in the setting of the TTRT value, and it is consistent with the analysis of Chapter 2 which shows how setting TTRT too low can limit system utilization. By raising the TTRT setting in the lower graph to 500 ms, utilization was increased 50 Mbps for the *2RwRMV* case and to a lesser degree for the other cases. Cases such as this indicate the significance of the relationship between TTRT setting and utilization.

Recall that the results of the Analysis section predict the factor by which throughput increases, not the number of bits per second. Figure 4.8 parallels Figure 4.6, but considers longer, fixed distances (all points have the same total network distance in Figure 4.6, but they have the same internodal distance in Figure 4.8). The results allow one to conclude that the second counter-rotating ring has a benefit for increasing throughput which is comparable to the benefit for shorter distances of ten to 100 kilometers which represent the distances used in other results shown.

Another question raised by this figure is why slot reuse decreases in the 5000 km example with a higher TTRT when compared to the other 5000 km scenario, yet the throughput increases. It is not exactly clear what this interaction represents, but it is likely the result of an interaction between TTRT setting and propagation delay and should be the subject of future investigation. Nonetheless, the increase in throughput is similar to the previous

results.

4.4.3 Expansion of Queue Window

As I mentioned previously, one would expect that the optimal case for destination removal would occur if all messages were destined for the immediate neighbor (right or left for the dual ring scenario). If a node had the ability to examine its queue of waiting messages to find the 'best-fit' message, one would expect the performance to approach the optimal case, N times the normal throughput of a single ring. This performance would clearly be impossible to implement, but a feasible implementation of the idea would be to consider limiting the queue from which the physical layer of the network could choose for transmission. Also note that the proposal here is to use some message other than the head of the queue only when reusing slots. This would maintain a degree of fairness and prevent a message from being barred from the network because a large distance is needed to travel to the destination.

Figure 4.9 shows the results of a simulation model which allows the physical layer to choose from a limited number of messages at the head of the queue of waiting messages, as opposed to always selecting from the head of the queue or selecting from the entire queue. The network parameters such as distance, packet length, etc. are the same as in Figure 4.5. The graph shows an anticipated diminishing return on the size of the queue from which messages would optimally be selected. The curves show a dual counter-rotating ring using destination removal where the queue size is varied from the head of the queue only, to a queue of the first ten elements.

The bottom curve shows that a throughput of 500 Mbps (250 Mbps per channel) can be achieved by destination removal as originally defined, employing the dual counter-rotating ring. If the node is designed so that the decision process always uses the message with the minimum distance to destination of the two at the head of the queue ($QSize\ 2$), throughput increases between ten and 40 Mbps can be achieved where the number of nodes ranges from 20 to 500. For each additional message considered, the increase in throughput becomes smaller. Consideration of the first ten messages ($QSize\ 10$) as compared to the first two

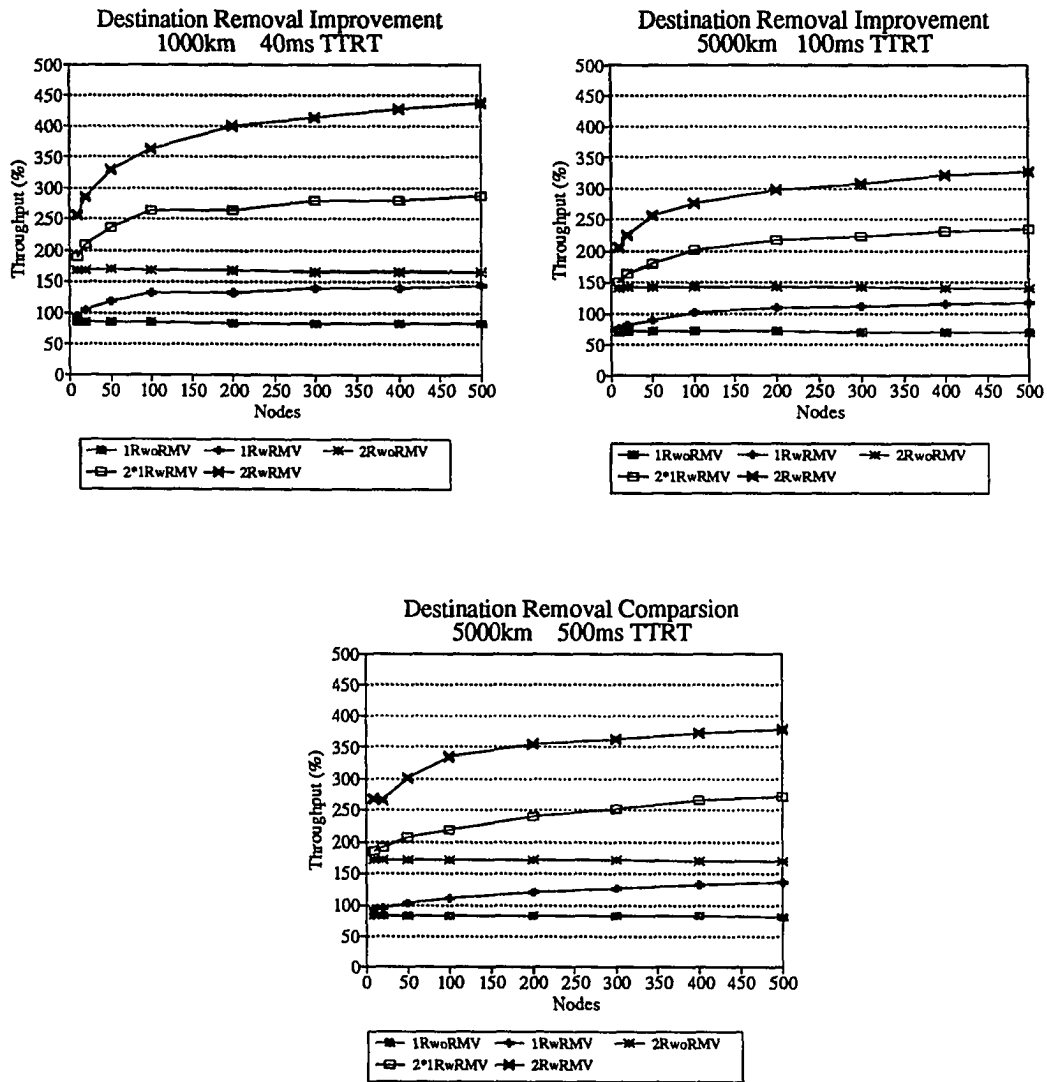


Figure 4.7: Removal: Impact of Network Distance - Throughput

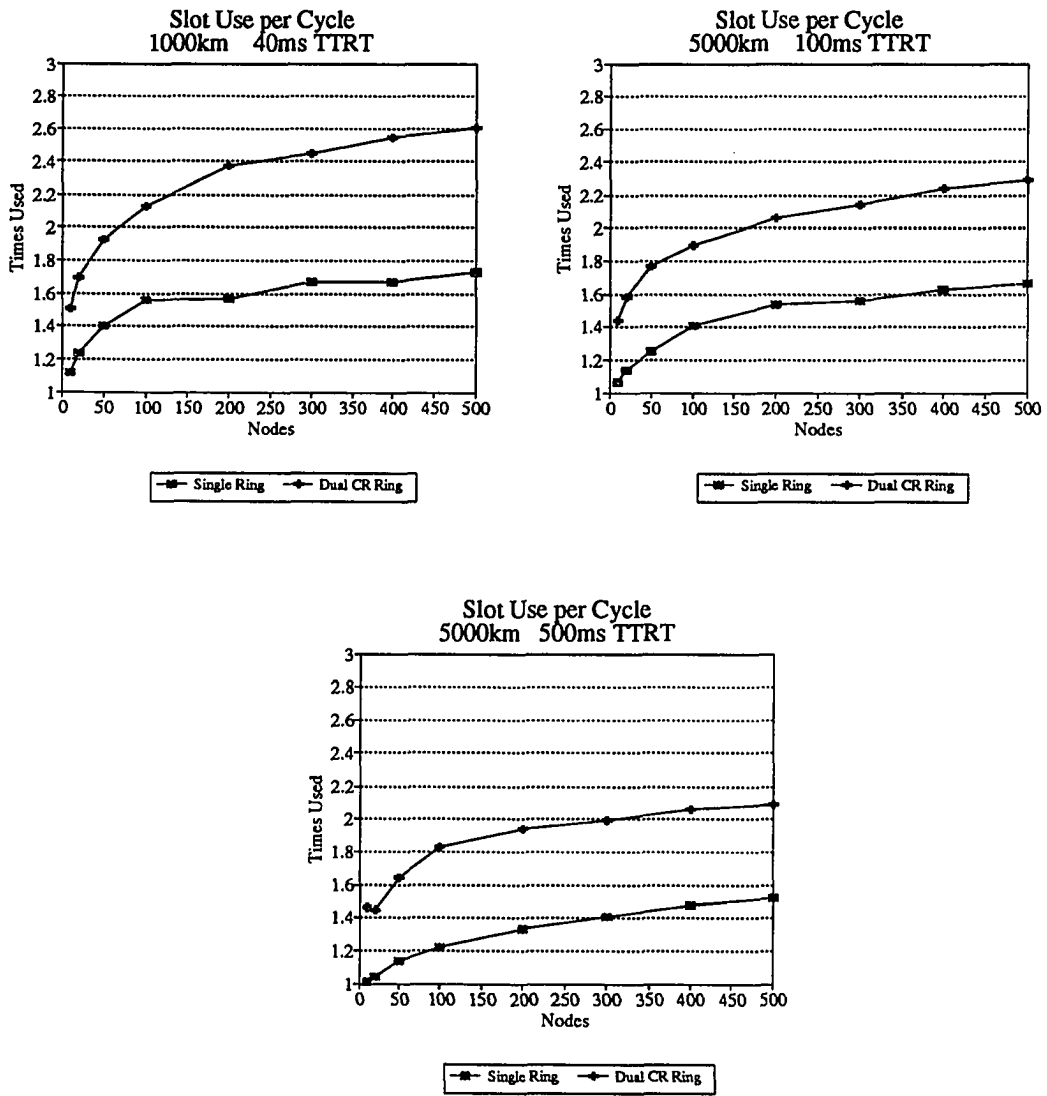


Figure 4.8: Removal: Impact of Network Distance - Slot Use

Throughput Per Ring - Dual Ring with destination removal

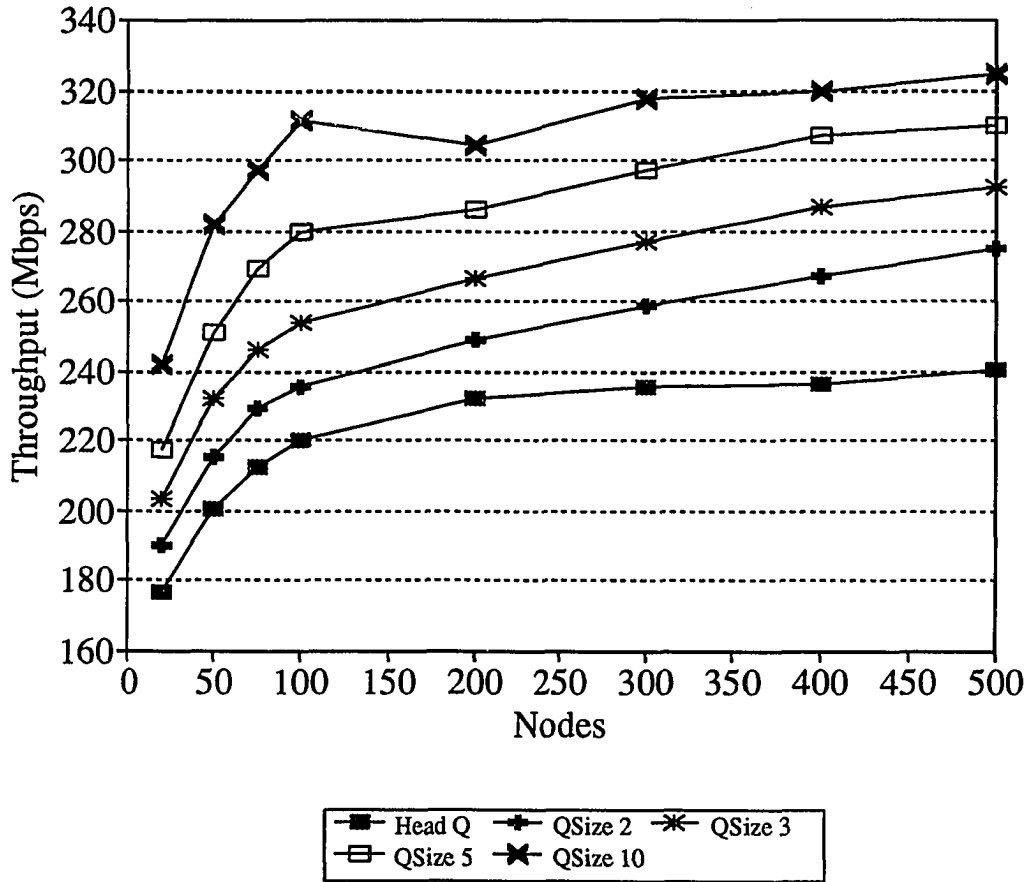


Figure 4.9: Removal: Increasing Queue Window to Physical Layer

messages ($QSize\ 2$) only produces an additional increase in throughput of between 50 and 80 Mbps.

The complexity of an interface such as this must be considered relative to the benefits gained; however, Figure 4.9 indicates that throughput can approach 650 Mbps from two 100 Mbps interfaces under heavily loaded conditions, assuming that such an interface can be engineered and economically produced.

Figure 4.10 compares the factor by which throughput increases for three techniques. The two curves labelled $1RwRMV$ and $2RwRMV$ are the single and dual counter-rotating ring versions of destination removal, and the values are the same results seen previously. The curve labelled $2RwRMVQ$ shows the effect of assuming that a node has an infinite queue and can remove the most appropriate message. This curve approximates very closely the original expectations of the analysis.

4.4.4 Delay Characteristics

One of the original advantages of destination removal cited in this chapter is the ability to reduce access delay at low loads in addition to the increase in throughput at high loads which has been shown. In order to substantiate this claim, a scenario of 200 nodes, network length of 20 km (each node separated by 100 meters), packet length of ten kilobits, and TTRT setting of 40 ms is considered. Consistent with most of the results presented in this section, dual and single rings in combination removal and no removal strategies are considered. These delay graphs also provide a basis upon which one can determine the rationale behind the setting of TTRT values in these tests.

Each of the figures containing access delay results are displayed with different x-axis and y-axis scaling, but all values are millisecond quantities. The first graph, Figure 4.11, is basically representative of the throughput characteristics developed in the previous section. Access delay for the dual counter-rotating ring curve, $2RwRMV$, shows that the delay is not only relatively flat, but that access delay is below one millisecond for loads up to 400%. As expected, each of the four curves have extremely long delays as the load approaches the

Dual Ring FDDI Increase of Destination Removal

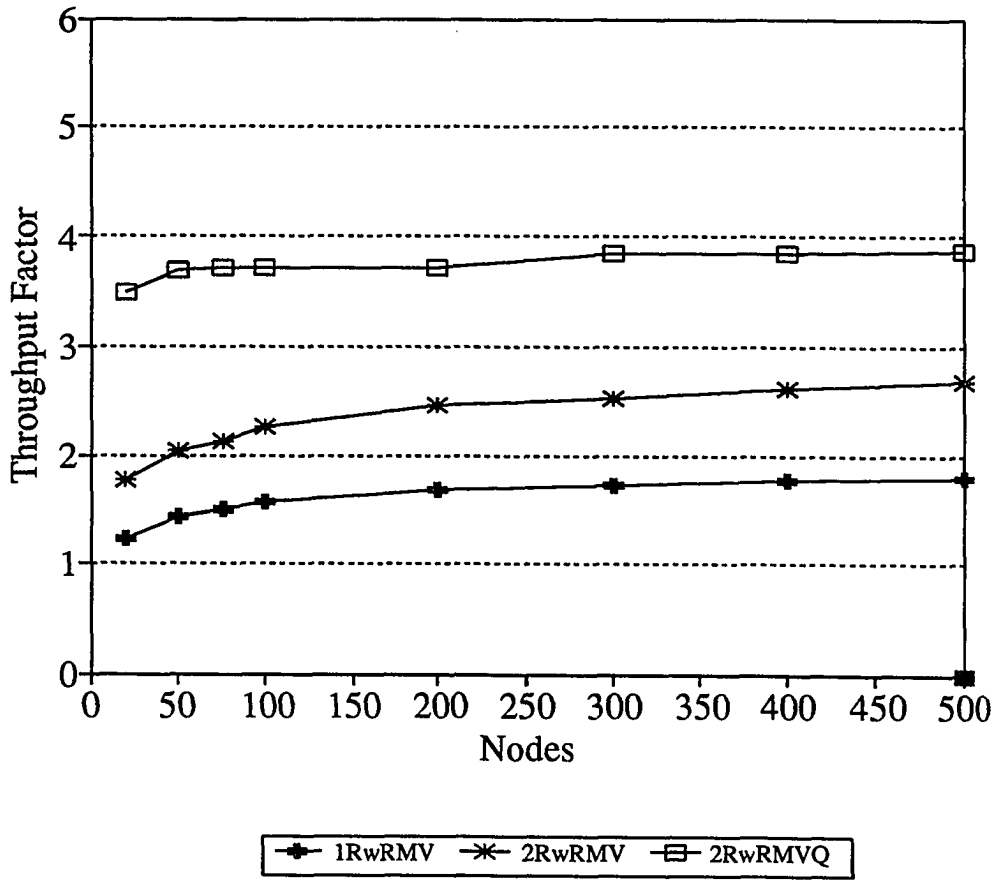


Figure 4.10: Removal: Throughput Increase Per Ring

throughput limit from Figure 4.5.

The second graph, Figure 4.12 shows performance over reasonable loads for a standard FDDI or token ring network, 0...100%. In this range, even destination removal on a single ring is significantly reduced. A 100% load, the two dual ring examples exhibit similar performance. The value of access delay for *2RwoRMV* is 0.1 ms, and for *2RwoRMV* the delay is 0.07 ms. The effect of destination removal has not yet become significant over these ranges; whereas, the effect of the second ring has reduced delay as expected. In order to see the differences at even lower loads, a third graph, Figure 4.13 is included.

4.4.5 Robustness

In order to assure viability of this technique, I examine its performance under different load conditions. As I stated previously, one of the important factors in the analysis of this method is the distribution of traffic. On one extreme, messages passed to an immediate neighbor would result in the slot being used by each and every node, an unrealistic assumption. The analysis and simulation results assume a uniform distribution of the source and destination address space. A more reasonable deviation from a uniform address space is to consider what happens in a file server/client environment.

Traffic Definition

A typical server/workstation ratio is $\frac{1}{20}$, a single file server for each 20 nodes. It is also reasonable to expect that the clients are located in a physical location randomly distributed relative to the server being accessed. Factors contributing to the probability of this distribution include the spread of user home accounts to various files servers, the location of various programs and system files spread across all file servers, and the ability of a user to log into any machine on the network. In a long distance network, the latter assumption would be less likely; however, the basic throughput *increase* of destination removal does not depend on the network distance (see Figures 4.7 and 4.8).

For a load L , expressed as the fraction of the network bandwidth, packet length P , and

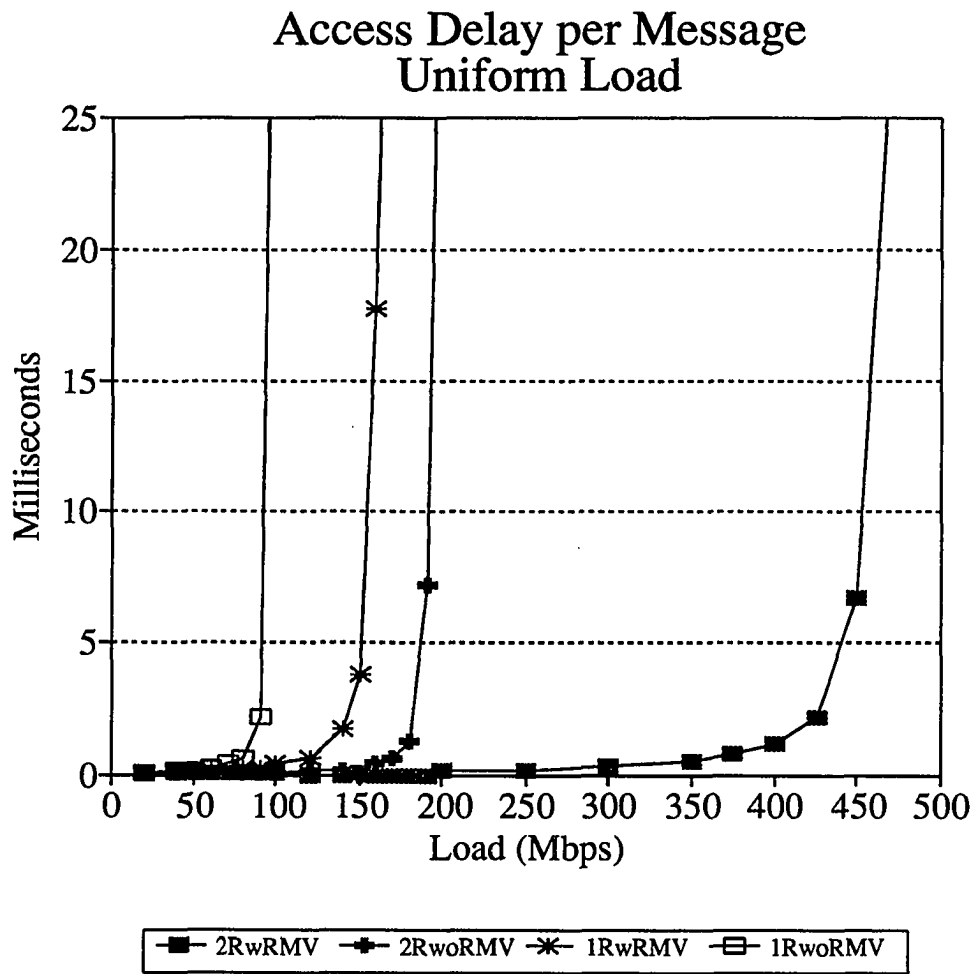


Figure 4.11: Removal: Access Delay

Access Delay per Message Uniform Load

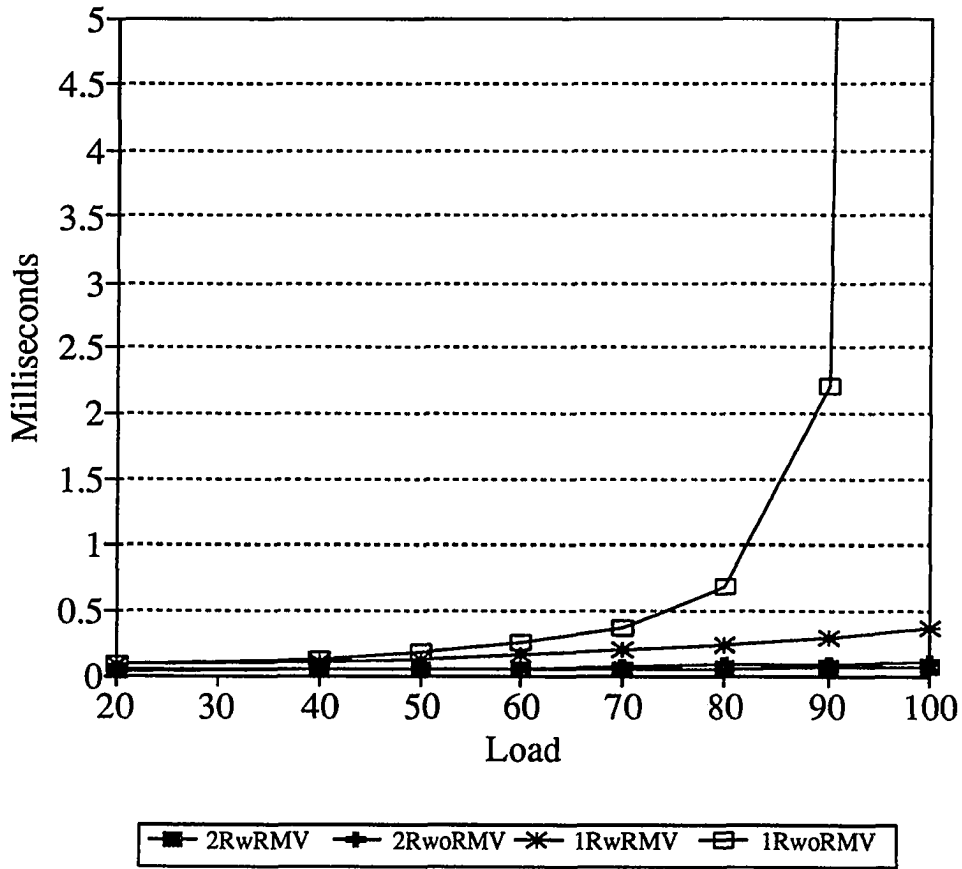


Figure 4.12: Removal: Access Delay at 0...100% Load

Access Delay per Message Uniform Load

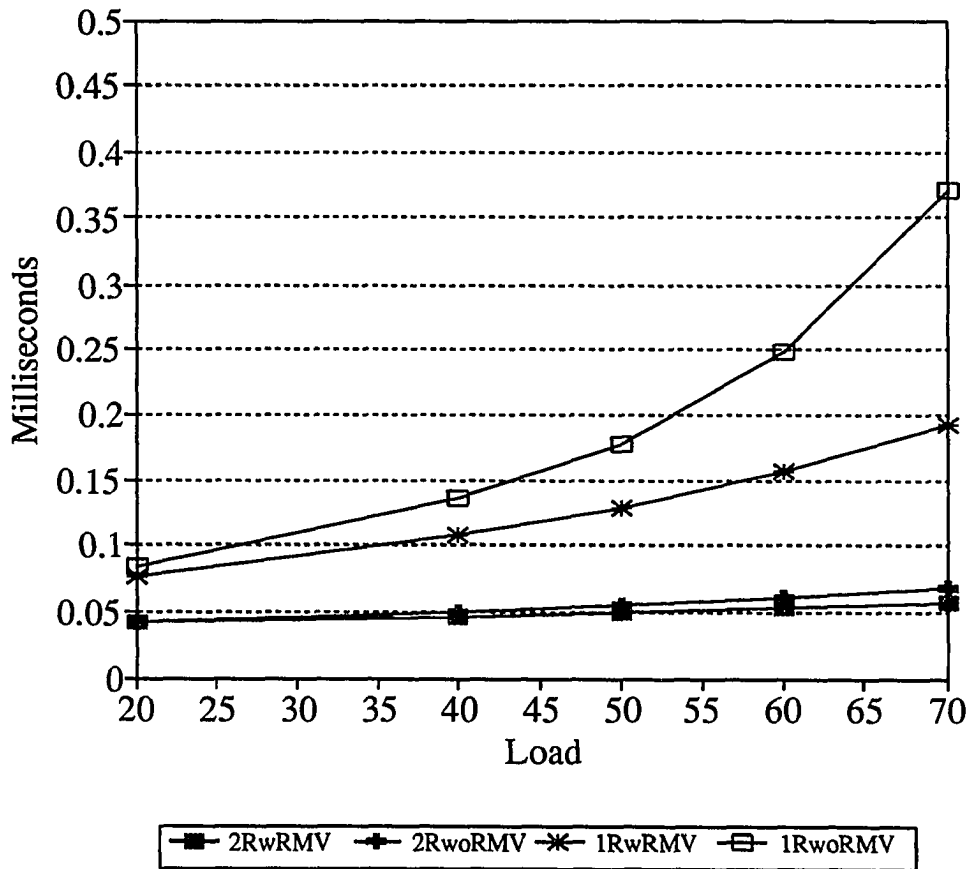


Figure 4.13: Removal: Access Delay at 0...70% Load

transmission rate R , the arrival rate of packets to the collective network is, λ .

$$\lambda = L \times \frac{R}{P}$$

The distribution of the traffic arriving at rate λ is defined here in terms of two discrete probability functions which define how source and destination addresses are determined:

- $s(i)$, a function representing the probability that a message originates at node i , and
- $d_j(i)$, a function representing the probability that a message which originates at node j is destined for node i .

A uniform address space for N nodes is defined as follows.

$$s(i) = \frac{1}{N}$$

and

$$d_j(i) = \begin{cases} \frac{1}{N-1} & i \neq j \\ 0 & i = j \end{cases}$$

For the file-server scenario used here, it is assumed that file servers have a greater likelihood of originating messages because of the low interaction between clients and the fact that for each request from the client for disk access from the server, the server has a message in return. Therefore, the file-server carries a fraction c of the load. The rest of the traffic, $1 - c$, is evenly distributed among all non-server nodes.

Assume that there is one server for every r nodes and that r divides evenly into N^4 , which simplifies the notation. The source address space $s(i)$ is defined here.

$$s(i) = \begin{cases} \frac{c}{N/r} & i \bmod r = 0 \\ \frac{1-c}{N-N/r} & i \bmod r \neq 0 \end{cases}$$

The destination address space is assumed to be a mirror image of the source address space. Messages originating at non-server nodes are equally likely to need any server,

⁴There are N/r servers.

but the probability of the destination being the server is dependent upon the value of c . Messages originating at server nodes are equally likely to be destined for any other node on the network. Therefore, the destination address space $d_j(i)$ is defined first for servers and next for non-servers.

$$d_j(i)_{j \bmod r = 0(\text{servers})} = \begin{cases} \frac{1}{N-1} & i \neq j \\ 0 & i = j \end{cases}$$

and

$$d_j(i)_{j \bmod r \neq 0(\text{non-servers})} = \begin{cases} 0 & i = j \\ \frac{1-c}{N-1-N/r} & i \bmod r \neq 0 \\ \frac{c}{N/r} & i \bmod r = 0 \end{cases}$$

Results

The scenario here considers a network of 200 nodes, spread over a distance of 20 km. Ten of the nodes are file servers distributed uniformly around the network. Packets are a constant size, 10000 bits, representing a typical page size virtual memory operating system or a disk sector but not reflecting packets for short commands. Network transmission rate is the standard FDDI rate, 100 Mbps.

The performance of FDDI under the functions described above is included for server traffic concentrations of 33%, 50% and 75% and for all basic network strategies in Table 4.1 except $2*1RwRMV$. For the purpose of comparison, results of a uniform distribution are also included. Figures 4.14, 4.15, 4.17 and 4.16 summarize these results. Note that the graphs are not presented in terms of utilization but in terms of access delay. Nonetheless, they illustrate the robustness of this strategy; one can determine from the graphs approximately where throughput is limited because delay becomes excessive.

Figure 4.14 compares the various loads for dual counter-rotating rings employing destination removal and shows that a uniform distribution actually falls in the range of the three concentrations studied. All three concentrations are able to support 350 Mbps throughput with access delays of one to two milliseconds. Delivery to the destination requires approximately an additional 0.1 ms. At 400 Mbps, access delays are still at or below five milliseconds for all except the 75% file server loading. The 75% loading case represents an

extreme and unlikely case. It would be more reasonable to expect that the traffic generated for the file servers would be proportional to the traffic for all of its nodes, 50%. The specific traffic distribution, however, always varies with the application and network. Even though delays are becoming significant, the throughput peaks at approximately the same values for all three loads, at approximately 500 Mbps.

Figures 4.15 and 4.16 show that without removal, there is little difference in the delays. This is a predictable result given that the standard token ring takes no advantage of, nor is a victim of any disadvantages in the address space. Figure 4.17 shows once again that the method behaves similarly for both file server loading and a uniform distribution. Here, lack of a second counter-rotating ring dampens the effect of destination removal. The primary result shows that despite the appearance of a simple address space such as a uniform distribution, reasonable perturbations of this address space result in fairly stable performance.

4.5 Implementation

A fundamental requirement of any design is that it should constitute a feasible solution; that is, one which can be implemented. Here the method of destination removal is shown to be such a solution. A method for recognizing the availability of the slot and using the slot is presented.

In order to recognize the availability of the slot and use it, the following constraints must be met.

- Either the original receiver must be able to recognize the destination address and mark the packet as delivered, or subsequent nodes on the ring must independently be able to determine the fact from the source and destination addresses. The method used here is to mark the packet as delivered by the receiver.
- The marking of the packet as delivered must be accomplished so that other nodes can determine that the packet has already reached the destination and is available for

Message Access Delay Varying File Server Load Concentration

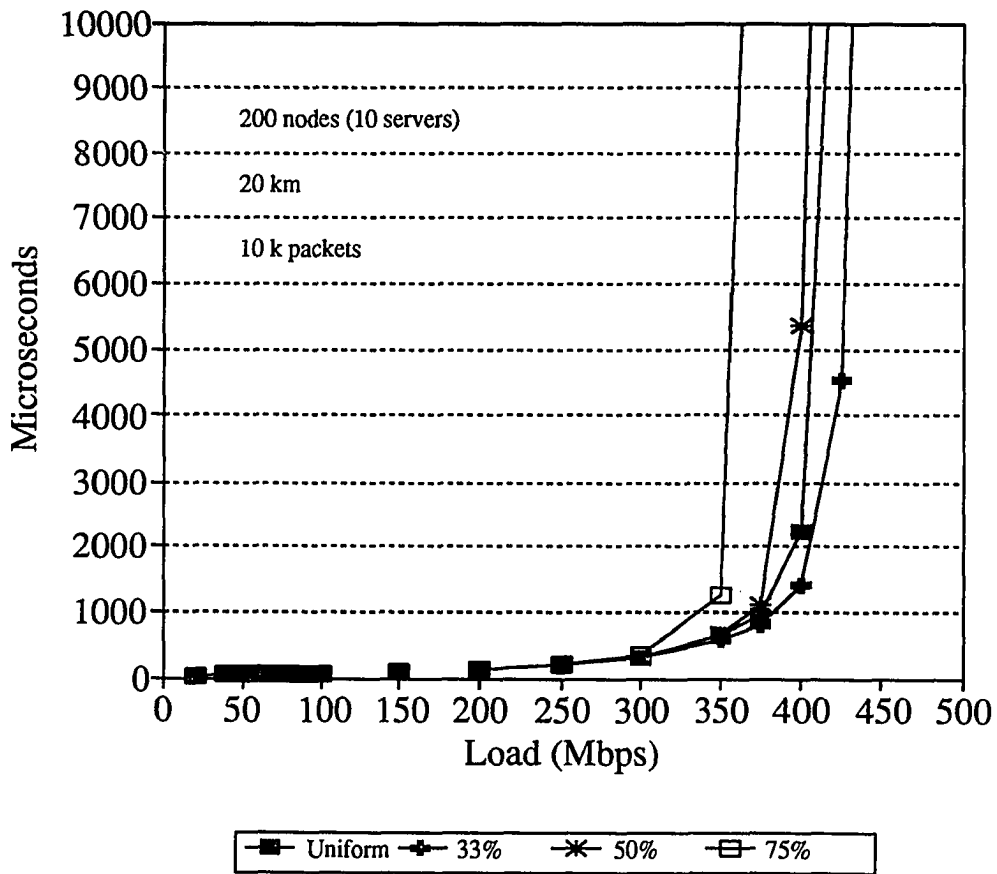


Figure 4.14: Dual Counter-rotating Rings with Removal under File Server Loading

Message Access Delay Varying File Server Load Concentration

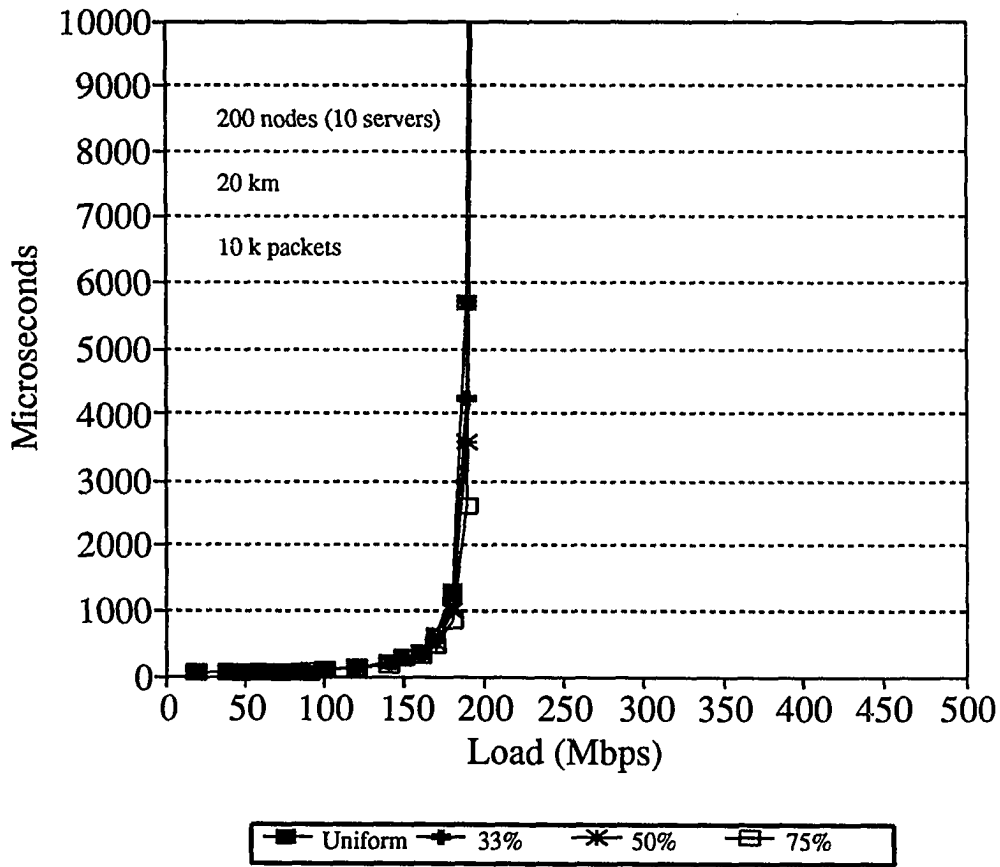


Figure 4.15: Dual Rings without Removal under File Server Loading

Message Access Delay Varying File Server Load Concentrations

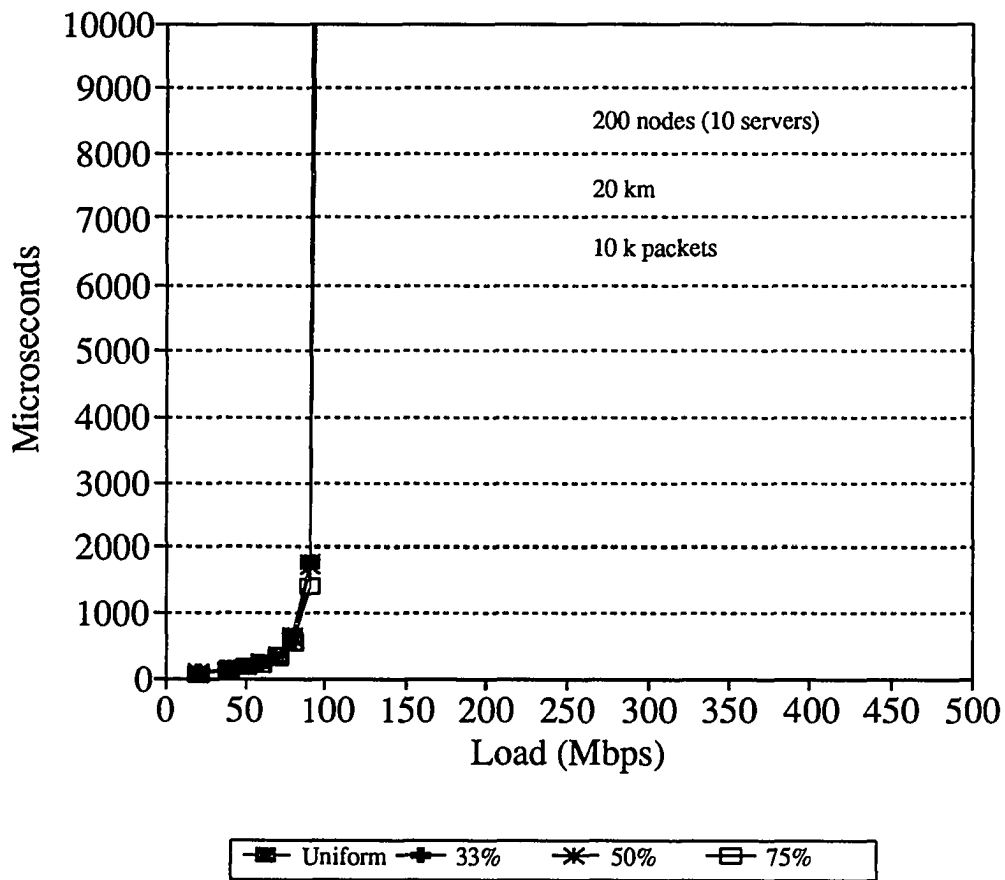


Figure 4.16: Single Ring without Removal under File Server Loading

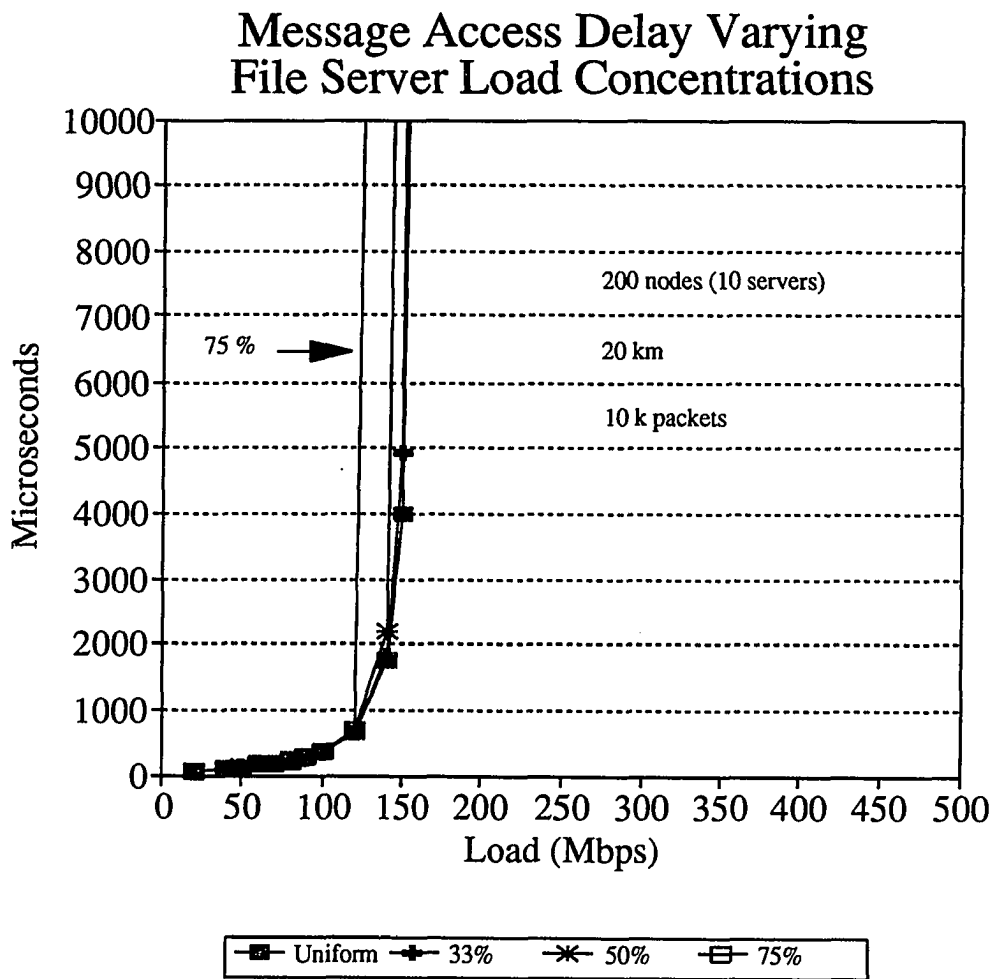


Figure 4.17: Single Ring with Removal under File Server Loading

reuse

- Once the slot is recognized as a reusable slot, the node must be able to determine whether the packet it has queued for transmission can utilize the slot.
- The processing required should not add unreasonable delay to the ring operation.

Placing the addresses and certain *delivered* flags in the header allows for interpretation of the source and destination information prior to receiving the data component of the frame. The original receiver simply looks at the destination address and if that address matches its own address, the original receiver marks the message as received by using the *delivered* flag, shown in Figure 4.18. As mentioned previously, this status could be recomputed at each node, but it would be simpler for each node receiving a message to compare for its own address while all other nodes check for a flag, as opposed to calculating for any arbitrary combination of source and destination addresses at every node. Placement of this flag after the header makes it possible for other nodes on the ring to determine the frame's availability before the arrival of the data component.

Trailers would typically contain two important components: the cyclic redundancy code (CRC) check field, and a flag or series of flags which indicate the status of reception. The CRC code is used by the receiver of the message to determine validity of the transmission. The CRC flag is also used by the sender as an indication that the data propagated itself around the ring without error. The additional flags would be used by the sender of the data to distinguish conditions such as proper transmission without reception or proper transmission with reception. For example, a receiving node could have a buffer overflow problem and simply not be in a position to store the record when it arrives.

If the trailer is reused as the data field is reused, the result is that the receiver is still capable of using the CRC to determine whether errors exist in the transmission, although it is not able to convey acceptance conditions back to the sender. Under the same condition the sender cannot use the trailer for any of the conditions previously mentioned. The consequence of losing the trailer can be neutralized by using the upper level protocols to

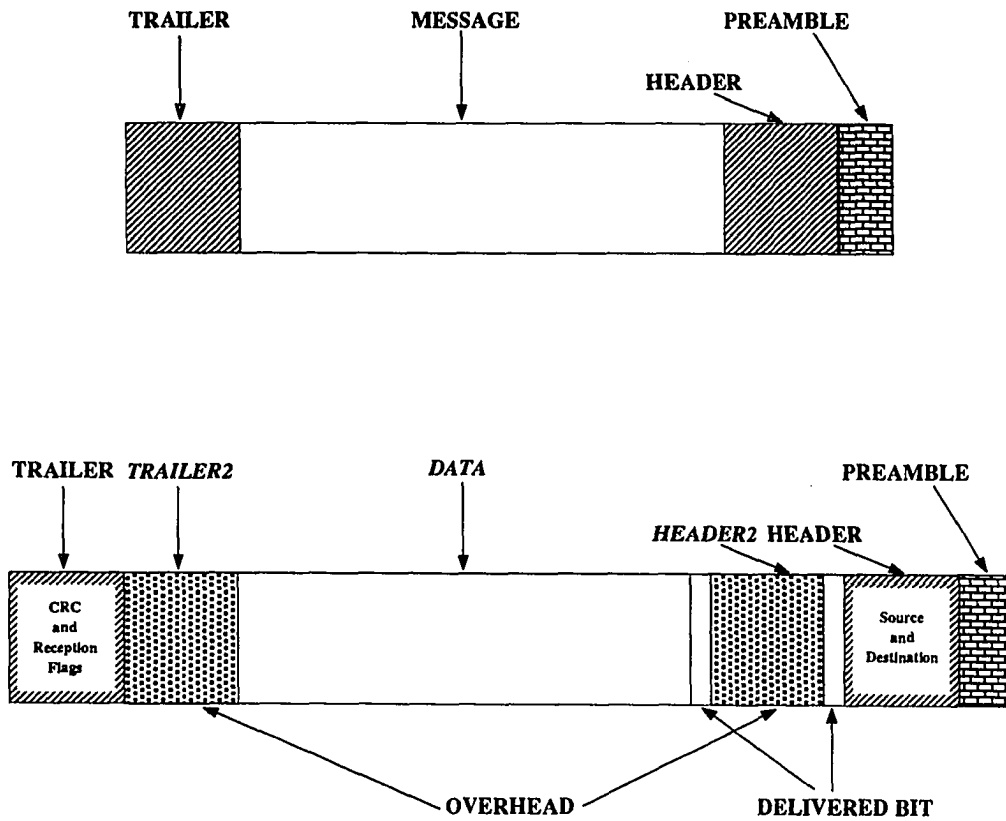


Figure 4.18: Removal: Packet Structure

implement a buffered technique such as go-back-n and by requiring acknowledgements to be piggy-backed or sent in separate frames, preferably the former. Buffered protocols are shown to significantly improve performance in networks spanning large distances, such as those defined in a HSLDN [62]. Piggy-backed acknowledgments require practically no overhead (a few bits) and provide the opportunity to minimize the effect of loss of acknowledgement.

Although all trailers except the first (outermost) one can be reused by subsequent slots, headers present a special problem. If a header is reused, a delay is incurred at each node to provide for a header modification at the receiver before retransmission. Although header reuse is a possible technique and is employed in many network designs, it is not considered as an option here because of my interest in large numbers of nodes. In order to accomplish the header reuse, a delay equivalent to the size of the header would be incurred at every node. Table 2.2 gives an estimate of the impact. This table shows that a latency, L_{tot} , of hundreds of microseconds would be incurred assuming a header size of 60 bits (600 ns) and 500 nodes.

Given these considerations, it is proposed in Figure 4.18 to leave the header/trailer pair of the original message and insert additional pairs for the reuse of the data component. Figure 4.18 illustrates one such pair designated as *HEADER2* and *TRAILER2*. This overhead expense is limited by the number of header/trailer pairs reserved. Figure 4.6 shows that the average number of uses of a slot for the same set of data used in the destination removal results of Figure 4.5 to be less than three for the dual ring. Therefore, this number is determined at network configuration time by the number of nodes on the network, and once the slot is used this number of times, it cannot be reused again. A dynamic structuring of the packet can also be employed to reduce the overhead during light loads and increase the number of reuses as loads increase, similar to the process used to renegotiate *TTRT* timer values.

One of the major results in Chapter 3 shows that *LMAX*, the time between recognition of the token and transmission of data, plays a significant role in performance aspects of

FDDI. In order to achieve a successful implementation of the method, the time between recognition and beginning of transmission of the data reusing the slot should be significantly lower than the value of $LMAX$. Such a recognition technique should not be confused with the recognition of a token because the approach to determining accessibility to FDDI is based on a timer, where the reuse strategy is based on the presence or absence of a flag showing a preceding delivery of the packet and on the determination of whether the packet to be inserted can be received before returning to the original sender of the slot. When FDDI receives a token, communication goes up to the Media Access Control (MAC) level to determine whether there is sufficient time available for transmission and to set timer limitations on how long the token can be held: all of these operations are done at the MAC level. Destination removal requires a much simpler strategy to determine availability for reuse. It can be compared to the strategy in DQDB which incorporates a busy bit in the field of each slot and only incurs a one bit delay at each node to examine, set the bit, and begin transmitting the data.

Another issue related to the reusability of the slot is the relationship between the length of the message and the length of the slot to be used. All of the research presented here assumes that there is a fixed slot length, contrary to the current definition of FDDI. This approach is a common strategy and is the same strategy adopted in DQDB. Use of a fixed slot length eliminates this problem. The degree to which internal fragmentation affects the performance is not considered here because of its dependence upon specific traffic patterns, but represents a reasonable extension of this research.

The last issue is to determine how the node which reuses the slot can determine whether or not the message is destined for a node which lies between itself and the original sender of the frame in a time period which allows the node reusing the slot to capture the header and transmit the data before the frame goes by (or without unreasonable delay). For example, if there are 20 nodes, node five originally transmitted the frame, and the frame now arrives at node 17 available for reuse, node 17 can send a message if it is destined for

nodes 18, ..., 20, 1, ..., 5. If standard addressing methods are used, the relationship between address and position is likely to be arbitrary and to require a table lookup mechanism. A simple modification to the address which appends a local node position number on the end of the current address format, would allow for a node to solve this problem in the following manner by using this position number.

Let s, r represent the position of the original sending node and the node attempting to reuse the slot respectively, and let d represent the position of the destination of the new message.

If $s > r$, then $r < d < s$ must hold,
else $r < d < N$ or $d < N$ must hold.

Each of these relational operations can be implemented with sequential comparators such that the determination can be made as the addresses arrive. The time between recognition of the usefulness of the slot and beginning transmission would be negligible.

Given the original conditions I stated as necessary for a successful implementation of destination removal, it is clear that feasible solutions exist for each of the requirements.

4.6 Conclusions

This chapter shows how to gain a multiplicative effect in throughput for token rings by employing a combination of a second counter-rotating ring in concert with the use of destination removal. An analysis of expected throughput increase is developed and compared with that which can actually be achieved. The performance results are determined by simulation; the results of which are as follows.

- Evidence has been given to show that 100 Mbps rings can deliver up to 250 Mbps on each ring assuming a uniform destination address space. Throughput is enhanced at both MAN and WAN network lengths.
- Access delay can be significantly reduced in a single ring with the use of destination removal at low loads (less than 100%). When employing destination removal on

a second ring, the effect of the second ring is more significant than the effect of destination removal up to about 150% load, at which point the effect of destination removal becomes significant.

- The concept of destination removal is dependent upon the destination address space. A uniform distribution was used in the analysis and most of the simulations for this work; however, other destination address spaces were considered, resulting in throughput comparable to the uniform address space and indicating robustness of the destination removal technique.

Feasible suggestions for implementation of the technique are also presented.

The degree to which this technique can be successfully employed is dependent upon a number of factors including number of nodes, network length, distribution of packet size, distribution of message destination address space, and maximum access delay requirements. Representative cases for MAN and WAN scenarios have been examined, but the main consideration was not given to extremely long distance networks. Instead, because of the dependency of destination removal on large numbers of nodes and the specific distribution of destination address space, the primary focus is on these two parameters.

Part III
DRAMA

Chapter 5

Multichannel CSMA/CD - DRAMA

¹ A network design by Sharrock [58] overcomes many of the problems inherent in CSMA/CD networks when data rates and distances are extended. The system, Dynamic Resource Allocation in Metropolitan Areas (DRAMA), is based on broadband technology and allows for allocation of bandwidth among clusters of nodes in the total network. DRAMA incorporates both band allocation and traffic placement protocols. Its band allocation algorithm is shown to be fair, stable and responsive to dynamic load conditions. Primary design objectives of DRAMA include:

- extend the size of CSMA/CD networks to distances much greater than two kilometers without any loss in speed and capacity,
- integrate synchronous traffic (real time voice or video) and asynchronous traffic (file transfers, mail messages, etc.),
- handles diverse loads and momentary traffic overloads,
- be fair,
- (re)allocate resources in a near optimal manner.

¹Chapter 5 and Appendix B are extracted from a paper submitted for publication to Computer Networks and ISDN. The majority of the material enclosed was written by the author of this thesis; however, all authors listed in original paper collaborated in the research and writing [40]

With DRAMA, CSMA/CD can be extended from low load and medium speed (ten megabits per second) to high load and high speed networks (greater than 100 Mbps) and operating distances can be extended to MAN distances. Results indicate that the design objectives above are achievable. The investigation is performed on two levels. On one level, the band allocation strategy shows that resources can successfully be allocated to that part of the network possessing the greatest demand. On the second level, a partitioning of the network is examined in detail to determine performance in terms of traditional metrics of delay, throughput, and fairness. Partitioning is determined by the organizing the nodes in physical proximity into separate Local Area Network Groups (LANGs). Bands are assigned to LANGs, and any node in the LANG may access the band according to DRAMA traffic placement policy. Performance within a LANG constitutes the majority of the results presented here. These properties are established through simulation studies.

5.1 MAN Requirements

Given all of the design objectives of DRAMA, it is important to understand that the most significant of these is the extension of CSMA/CD to a larger geographical area in order to function in a metropolitan area network (MAN). The primary characteristics that distinguish a MAN from a LAN and which must be addressed for an acceptable MAN solution are:

- increased data rates (greater than 100 Mbps),
- greater node count (from 100 to 1000),
- integration of synchronous and asynchronous traffic,
- dynamic allocation of bandwidth as local demands change,
- larger geographical distances (from two to 200 km).

These characteristics are a subset of the definition of a HSLDN.

Increased data rates are required because of the increased demands of each workstation (video, file transfers, etc.) and the increased number of workstations that will be sharing the bandwidth. Local area networks based on CSMA/CD, such as Ethernet [10,48] have throughput limitations which depend upon the ratio of propagation delay to packet transmission time. Increasing the data rate has a negative impact on throughput unless the frames are made very large which results in internal fragmentation inefficiencies, or unless propagation delays are decreased, which is contrary to the desirable characteristic of large geographical extent. Ring networks have been designed to support these high data rates but they suffer from higher access delays at low loads. An increased number of nodes accentuates the delay problems of ring networks and increases the number of contention intervals required to obtain access to CSMA/CD networks.

Other protocols have been proposed to increase throughput, but they suffer from a conflict with one of the other desirable MAN properties. Lee shows that Hubnet [36] is able to achieve acceptable performance up to 87% of the capacity for data traffic. Thus it extends CSMA/CD type protocols to the 100 Mbps range but it cannot handle synchronous traffic because of its unbounded access delay. Several proposed multichannel LAN protocols extend the original CSMA/CD scheme to improve performance [7,45,73]; but these protocols either do not integrate different traffic types, do not expand the size of the LAN, or cannot allocate resources dynamically.

As video, voice and other synchronous traffic demands are placed on the network, a way must be found to integrate synchronous and asynchronous data and to respond to dynamic load changes. Video traffic, which must be transmitted at specified rates, places large load increments on the system. Analog video circuits require up to ten MHz, and digitized video transmission requires 1.5 to 140 Mbps, depending on the quality and type of application [37,75]. A typical voice call may last for several minutes, generating four to 64 kbps of data. Propagation delay, from voice sample generation at the source to actually hearing the voice sample at the receiver, should be less than a few hundred milliseconds [46]. However,

typically more than 50% of the voice circuit is silent [20]. Studies show that data traffic is bursty, may tolerate widely varying delays, and that the distribution of packet sizes is bimodal [35,59]. Considerable effort has been made to adapt protocols to integrate different types of traffic such as data, voice, and video both within and outside of ISDN [34].

Most high bandwidth systems provide a framing mechanism for handling integrated traffic. FDDI [63] provides for synchronous traffic by controlling token circulation rates. While this scheme is effective in most cases, extreme variations in synchronous traffic rates may require a renegotiation of the token rotation time to change the apportionment of synchronous/asynchronous bandwidth. FDDI-II and DQDB both use a 125 μ s frame with internal slots reserved for synchronous traffic. This structure allows easy connection to ISDN networks for coupling to wide area systems. DRAMA provides an equivalent scheme equally applicable to ISDN coupling. In circumstances such as silence in a voice call or an unchanged video image, synchronous bandwidth may go unused. DRAMA is tailored to easily recover the unused synchronous traffic bandwidth in each frame cycle to make it available for asynchronous traffic.

Synchronous and asynchronous traffic should be managed through dynamic bandwidth allocation. Both FDDI-II and QPSX use strictly time division multiplexing (TDM) whereas DRAMA uses both TDM and frequency division multiplexing (FDM). In FDM broadband systems such as [24,50], the broadband frequency spectrum is statically partitioned by user group and/or by traffic class. For example, CableNet [54], Sytek's LocalNet [14], and Mitre CableNet [18] partition the bandwidth into fixed bands for particular applications of specific groups of users: some bands are permanently reserved for analog video channels, some are reserved for TDM for a set of closely located users, and some bands are dedicated to specific functions such as process control. This approach might be reasonable if the traffic mix were predictable and fairly constant, but the bandwidth requirements of diverse traffic classes such as voice, data, and video can fluctuate widely over short periods of time.

Finally, the network should be able to function in a wide geographical area. Physical proximity of computing neighbors should not be a prohibitive restriction of the network. A CSMA/CD network such as Ethernet can achieve ten megabits per second over a distance of about two km for 100 users if active repeaters are used. An example is given in Section 1.4.1 to show how maximum throughput can be reduced from over 80% to less than 20% if the distance is increased to 100 km, ignoring all of the problems of signal loss over such distances. Other systems, such as the Token Bus [66], or Token Ring [18,34] have better throughput performance but greater access delay.

Initial performance studies for FDDI [13] and DQDB [51] concentrated on access delay but did not consider algorithms for negotiating allocation of bandwidth to nodes based on their needs. More recent studies have shown that this is a problem for DQDB. Hahne and Huang reveal how access to a DQDB network is a function of the position of the node on the bus [23,28]. Resource arbitration frames are defined in the FDDI Station Management proposal standard in order to support different synchronous requirements and bursts of traffic between two nodes. Some studies of fairness and access have been done to determine FDDI support for the real-time requirements of the space station [30]; however, these studies have not included the wide variety of voice, video, and data loads that generally occur.

The DRAMA protocol, introduced in [57,58], produces a gatewayless network that covers a metropolitan area and allows for dynamic sharing of bandwidth among both different types of traffic and different clusters of nodes. An overview of the protocol is presented and results from two detailed simulation studies evaluating important performance aspects of DRAMA are presented. Appendix B contains a DRAMA overview along with a discussion of how DRAMA accommodates the important attributes of any metropolitan area protocol. Section 5.2 presents the metrics used to evaluate DRAMA's performance, and Section 5.3 summarizes the results of the simulation studies. Section 5.4 describes the performance characteristics of DRAMA that are demonstrated from the protocol structure and the simulation studies.

5.2 Protocol Analysis

The DRAMA protocol was evaluated using simulation modeling. Two basic simulation models implemented in SIMSCRIPT II.5 were used to evaluate DRAMA thoroughly at the traffic placement level and at the bandwidth allocation level.

5.2.1 Bandwidth Allocation Model

The bandwidth allocation model is a simulation of the upper levels of the DRAMA protocol. It allows for evaluation of the algorithms for allocation of bands to LANGs [44]. The metrics used to evaluate the band allocation strategies include:

1. *Band fairness* is the average difference between individual LANG utilization and overall network utilization. If n is the number of LANGs, ρ_i the utilization of the i^{th} LANG, and $\bar{\rho}$ is the average network utilization, then band fairness is defined as:

$$Fairness = \frac{\sum_{i=1}^n |\rho_i - \bar{\rho}|}{n} \quad (5.1)$$

2. *Responsiveness* is the LANG's average excess deviation from the utilization tolerance interval. If τ denotes the utilization tolerance, then

$$Responsiveness = \frac{\sum_{i=1}^n \max(0, |\rho_i - \bar{\rho}| - \tau)}{n} \quad (5.2)$$

3. *Recovery* measures the average time required to recover from dramatic changes in load on an individual LANG, resulting from either large increases or drops in traffic. For example, introduction of a new video signal could cause a significant reallocation of bandwidth.

Maly describes details of the band allocation algorithms in [44]. The general approach is to reallocate bands as LANG loads fluctuate; LANGs with higher loads are usually allocated additional bands. As defined in the fairness metric above, the utilization of each LANG's resources should be as close to the average for the entire network. When a LANG's bands

become relatively underutilized, some of its bands are forfeited to other LANGs whose utilization is above average. In the simulation, loads within a LANG were parametrically and randomly varied as well (parameter λ) in order to model a more realistic environment. Figure 5.1 illustrates the concept of a tolerance interval around the average network utilization. In this diagram, all LANGs are within the tolerance range and except under special conditions described in [44] would neither acquire nor release bands.

In order to provide immediate availability of bands for LANGs that suddenly experience a surge of traffic, some bands are reserved in a global pool. We refer to the reserve as the *steady pool*, the minimum number of bands that should be kept in the global pool if at all possible. If several bands simultaneously receive large load increases, the global pool can quickly be used up; however, LANGs will eventually release bands to the global pool until the steady pool value is again reached. Each LANG has a minimal number of bands reserved for its use exclusively regardless of load conditions in other LANGs.

The granularity of the simulation is at the level of adjusting a utilization statistic (in effect, the amount of traffic) on the bands assigned to individual LANGs. That is, individual messages are not generated but instead gross load is adjusted randomly at each LANG as a function of time. This level of granularity is appropriate for understanding the band allocation protocol.

5.2.2 Traffic Placement Model

The traffic placement model simulates the operation of a single LANG with a fixed number of bands subjected to randomly varying synchronous and asynchronous traffic at the message level. It provides a model to study the lower level protocol, the traffic placement. The LANG model does not address issues of band allocation, but it does facilitate understanding of issues such as the validation of the band allocation approach.

The main objectives of the traffic placement model experiments are:

- to implement the DRAMA protocol at the message (circuit and packet) level as it would operate within a single LANG,

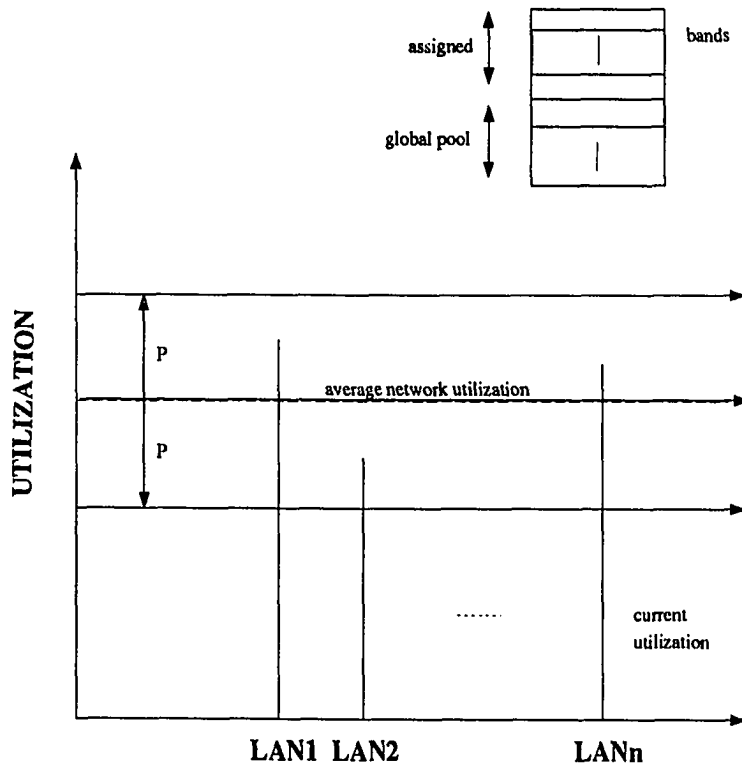


Figure 5.1: DRAMA: Steady State LANG Allocation

- to compare several traffic placement policies as noted above, and
- to measure performance under a varying traffic mixes and loads.

The study modeled a LANG as a group of nodes that can communicate with nodes across the network. By restricting the simulation to a single LANG, the execution time required for each run is considerably reduced. The results for a single LANG are easily generalized to the multiple LANG case since each LANG places messages independently.

Studies using the band allocation model demonstrate that bands can be reallocated rapidly while keeping the band utilization at every LANG from varying more than a small percentage from the total network average [44]. Therefore for the traffic placement model, the number of bands is held constant in order to study the effects of traffic placement on the network. The interest in demonstrating network performance under heavy integrated traffic load conditions justifies this restriction. Additional information related to the model itself can be found in [42].

We use the following performance metrics for system evaluation:

1. *Access delay*: Because network access delay is a critical factor in protocol performance, it was measured in several experiments. The access delay does not include the transmission and propagation delays. δ_i denotes the average access delay at node i .
2. *Fairness*: The network access delay for each of the n nodes should be independent of that node's position in the network and should be close to the average access delay E of the network. Each node's deviation from the network average is measured by the degree of fairness D_f , where

$$D_f = \sqrt{\frac{\sum_{i=1}^n (\delta_i - E)^2}{n}} \quad (5.3)$$

- δ_i is the mean access delay at the i th node,
- E is the mean access delay in the LANG, and
- n is the number of nodes in the LANG.

Ideally, D_f should be zero. Note that this equation is used to measure fairness within a single LANG. Fairness among LANGs uses the band allocation model, equation (1).

3. *Throughput*: In the studies, throughput is the percent of the network capacity taken up with successful traffic. Throughput of a LANG is a function of offered load measured as a percent of network capacity. In these studies offered load ranged from to 200%.
4. *Recovery time*: We were interested in the response of a single LANG to impulse traffic. To this end, in some experiments a sudden dramatic impulse of traffic was generated in order to determine how much time it would take for the system to return the message queue sizes to within five percent of the prepulse load. Ten percent of the nodes were given a burst of additional data traffic to build up these queues.

5.3 Simulation Results

In this paper, the suitability of DRAMA for MAN systems is demonstrated by reporting on simulation experiments utilizing both the band allocation model [44] and the traffic placement model [42]. First the results of a single versus multichannel structure are presented. Next DRAMA behavior as a function of network parameters such as the number of bands, the number of LANGs, and nodes per LANG is discussed. Then the stability of DRAMA at high loads and *momentary overloads* is considered. Fourth, the fairness issues are discussed. Finally, we examine DRAMA's capability to handle varying degrees of integrated traffic is examined.

5.3.1 Single Versus Multiband Behavior

Figure 5.2 illustrates the dramatic advantage of using a multiband rather than a single band structure, even with a fixed network capacity. Here the number of bands assigned to a LANG is changed, although the total bandwidth available to the LANG is held constant. The access delay for integrated traffic is improved when the LANG has at least five bands. Ten or more bands provide additional but reduced degree of improvement. Delay is reduced

because synchronous traffic is confined to only a few bands and the data traffic can select any of the free bands. Thus the probability of collisions in the CSMA/CD protocol is significantly reduced over that of a single band.

5.3.2 Size and Scalability

From Figure 5.3, we observe that traffic within a LANG is unaffected by the number of nodes assigned to it as long as the effective total length of the LANG remains unchanged. Thus most of the size and scalability issues are related to the performance of the upper level, that is, to the band allocation protocol. Figures 5.4 and 5.5 indicate how the number of bands affects performance as measured both by fairness and responsiveness; data are at increments of ten percent. Here λ is the message arrival rate and is used to vary network load. Variation in pool size and in the magnitude of changes in traffic caused no any significant changes in performance. Figures 5.4 and 5.5 also show that the curves tend to flatten out at about 50 bands, or about ten bands per LANG (the number of LANGs is five). Thus it is reasonable to conclude that for good performance *the average number of bands per LANG should be at least ten*. With more than ten bands per LANG improvement in performance is negligible.

Figure 5.6 addresses an additional scalability issue; it shows the effect of increasing the bandwidth, the number of bands and the number of nodes simultaneously. Scaling up from ten bands with 100 nodes to 50 and 500 respectively, improves performance. The increase in available bands allows further separation of synchronous and asynchronous traffic so that some band is usually idle whenever a data packet is to be transmitted. At the upper node and band count, little deterioration in access delay occurs up to 80% of offered load. Most other protocols fail to reach this level of performance and most get worse as the node count increases.

These results indicate size and scalability characteristics that are excellent for MAN systems. Increasing the number of nodes, bands, and overall bandwidth generally has a positive effect on DRAMA's ability to allocate bands to LANGs fairly and to decrease

DATA DELAY

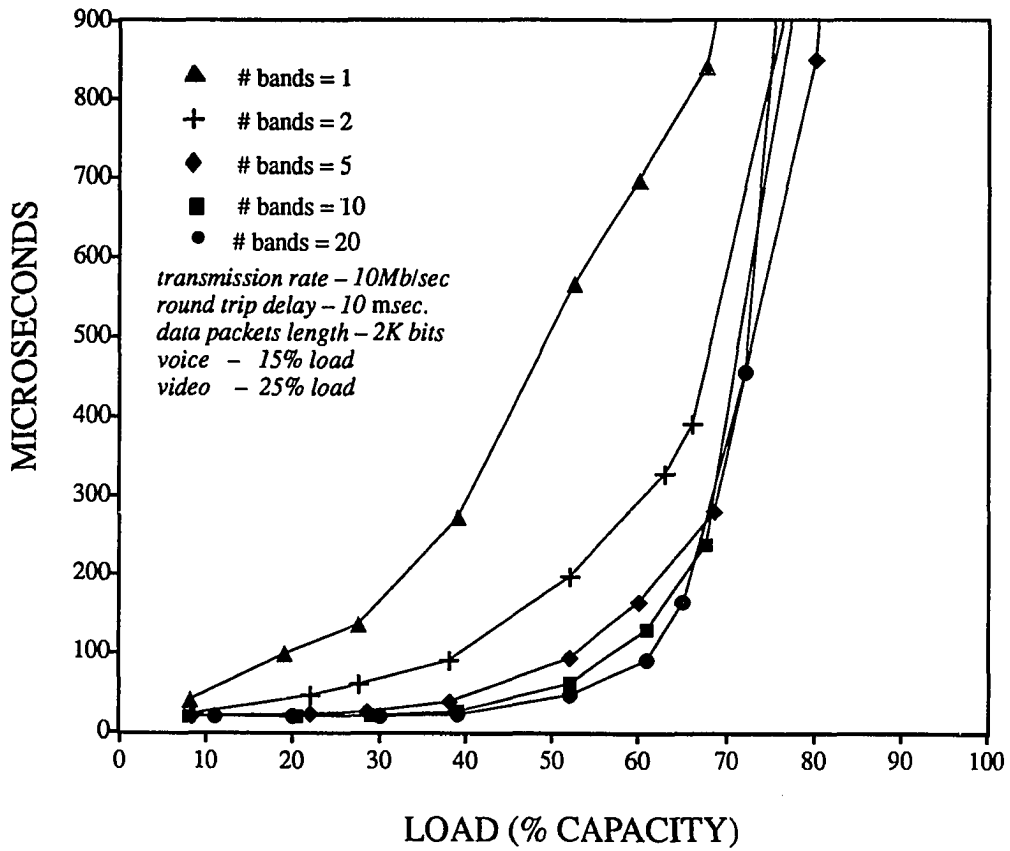


Figure 5.2: DRAMA: Access Delay versus Number of Channels

DATA DELAY

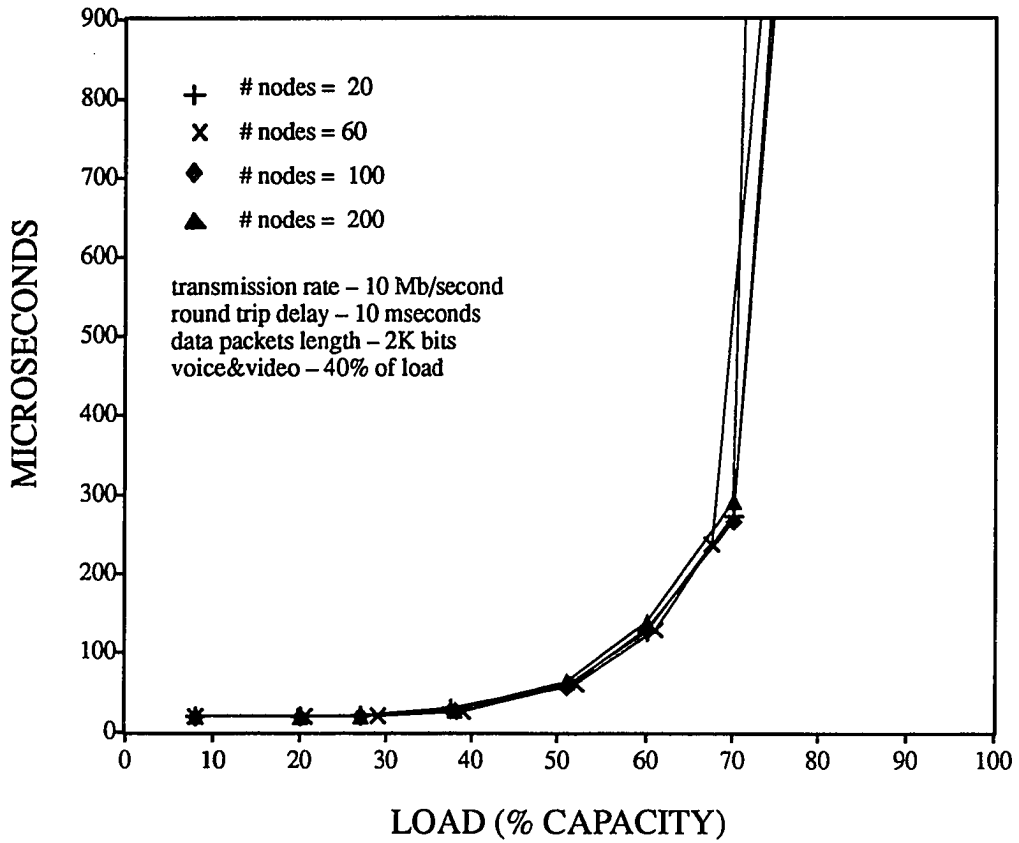


Figure 5.3: DRAMA: Access Delay versus Load Varying Number of Nodes

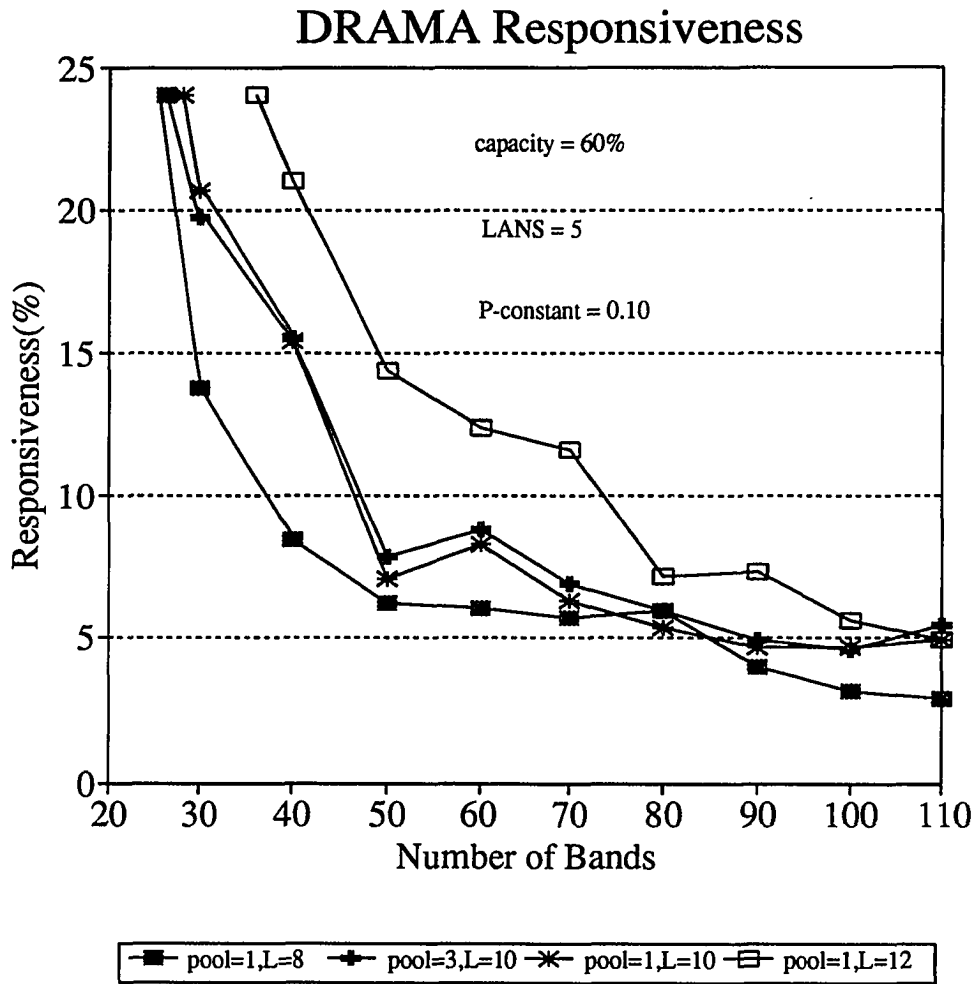


Figure 5.4: DRAMA: Band Allocation Responsiveness

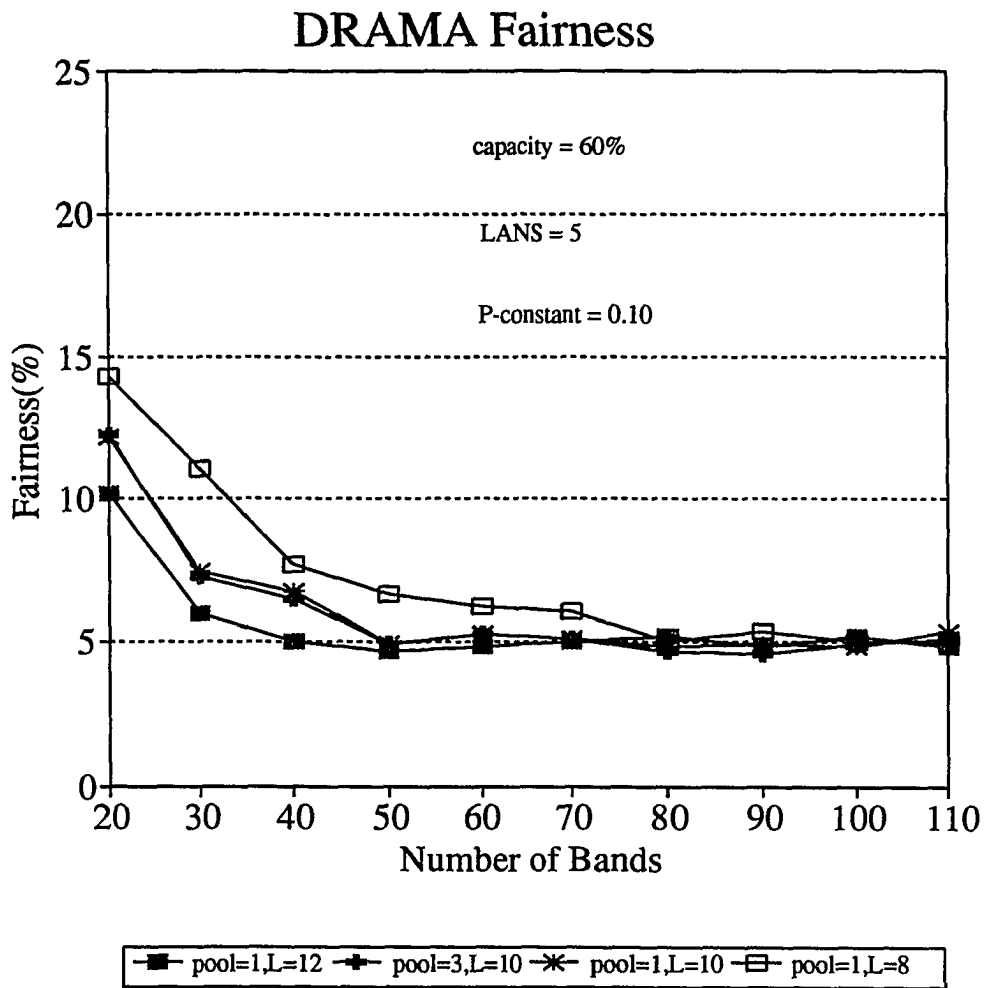


Figure 5.5: DRAMA: Band Allocation Fairness

delay under proportional loading conditions.

Another scalability factor relates to the effect of increasing the size of packets within a frame. Figure 5.7 shows the tolerance of DRAMA for increased packet size. As long as the size of the packet remains small compared to the size of the frame, packet delay remains essentially unchanged. In this case the frame size was chosen to be ten kilobytes. When packet sizes approach 30 to 40% of frame size, changes in access delay become more evident. This graph has been normalized to reflect the fact that eight kilobits of data need only one eight kilobit packet but need eight one kilobit packets.

5.3.3 Stability

Stability, the ability to adjust to new load levels and recover from overloads, is required at both protocol levels. In evaluating the protocols, two primary issues are considered. First, system behavior at high loads is important at the traffic placement level since some protocols have the degrade at high load conditions. Second, the system's response to major changes in traffic patterns is important. More specifically, how a system responds to a dramatic impulse of traffic load at both protocol levels is a concern. At the upper level, can bands be reallocated quickly? At the lower level, will performance recover if bands cannot be supplied and the traffic placement policy is forced to handle an overload?

For message traffic, access delay at high loads clearly becomes large; the question is whether throughput is maintained. Figure 5.8 demonstrates the efficiency of the DRAMA traffic placement protocol. Up to about 75% of capacity, practically all traffic on the net is message traffic with a negligible (less than one percent) amount of noise. At about a 90% load the message traffic levels off and stays constant, independent of the amount of traffic offered. The last data point we measured is with an offered load of 185% of capacity. The amount of noise, i.e., collision, is about five percent and the additional capacity wasted is between five percent and 15% depending upon the placement policy. This result occurs in all traffic placement policies considered, showing another aspect of stability. Backoff seems to perform better under these high load conditions.

DATA DELAY

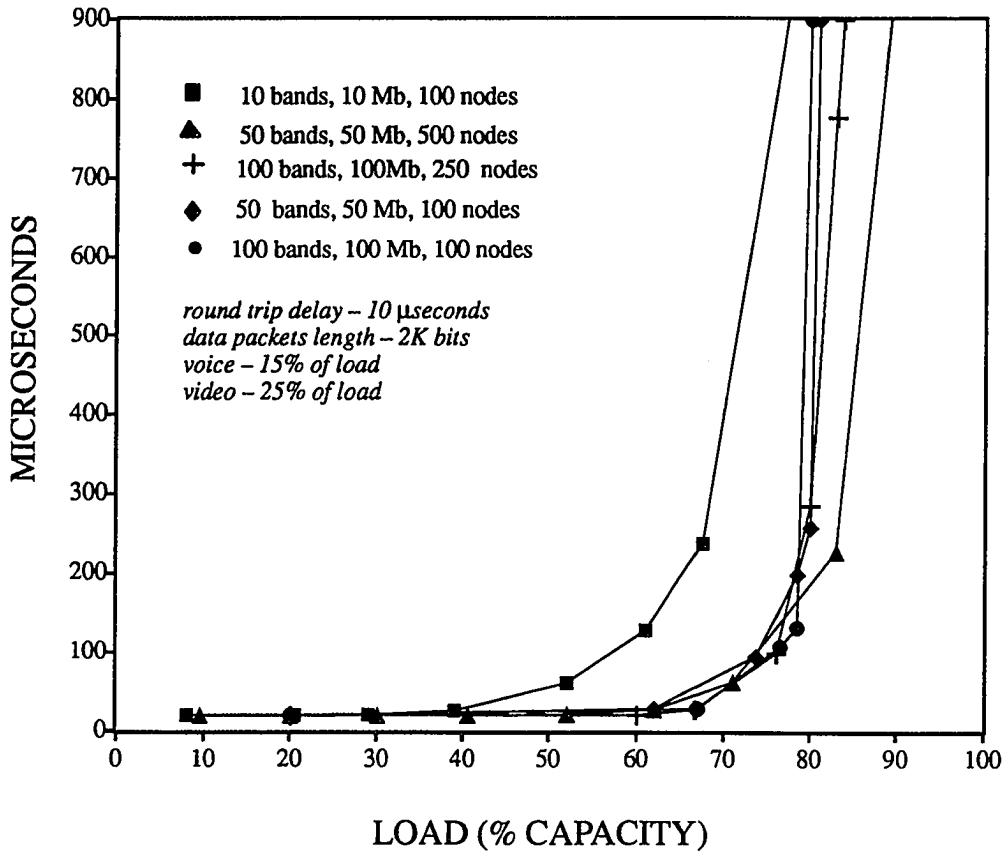


Figure 5.6: DRAMA: Access Delay Under Scaling Conditions

DATA DELAY

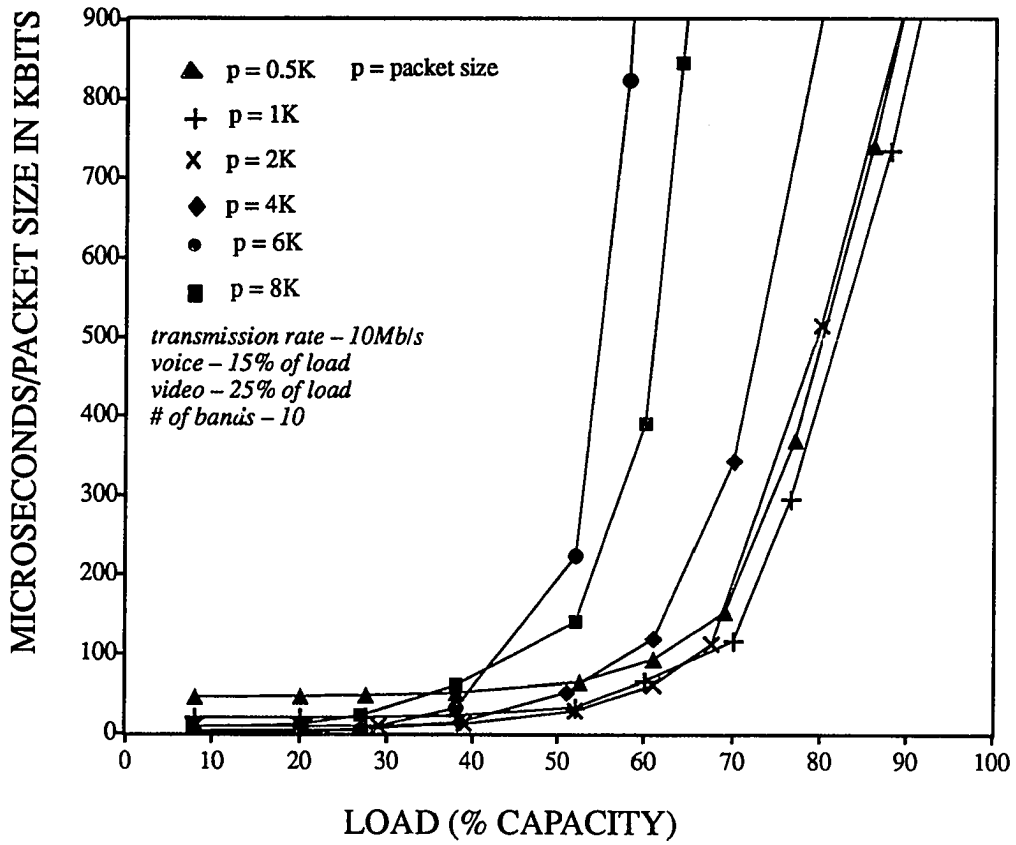


Figure 5.7: DRAMA: Access Delay Versus Packet Size (normalized delay)

THROUGHPUT

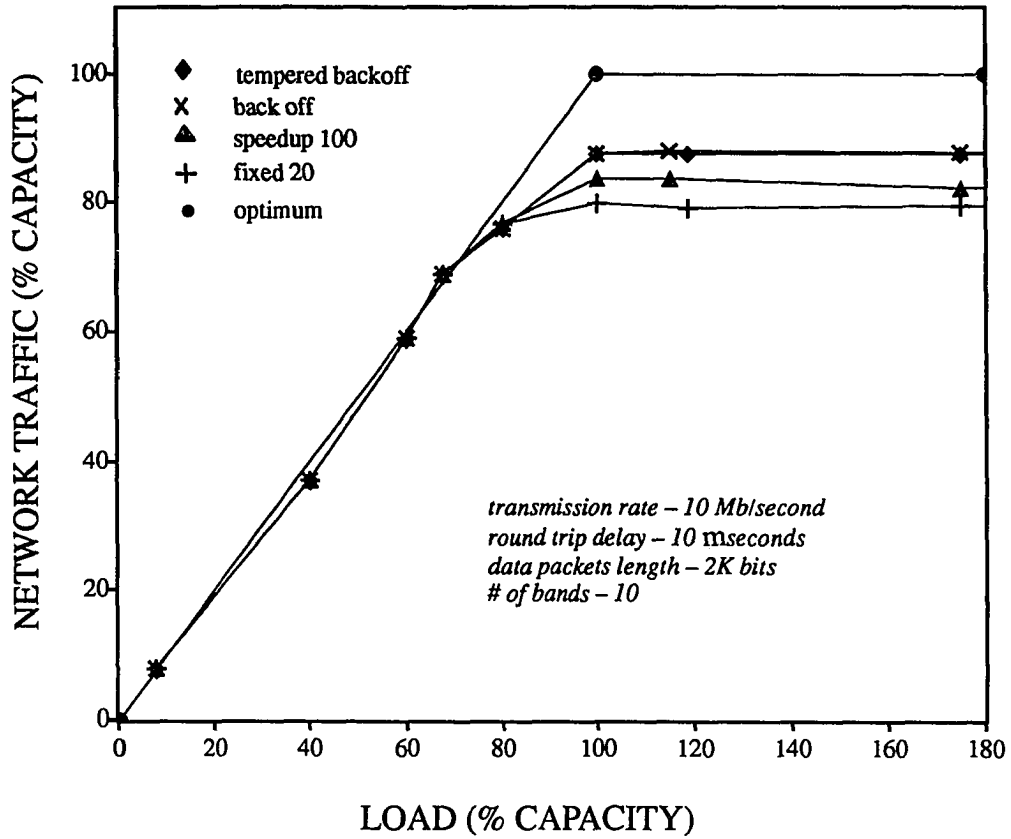


Figure 5.8: DRAMA: Network Throughput for Various Placement Policies

Figure 5.9 examines the effect of a traffic impulse on queue size in a LANG; specifically, how long it takes for the LANG's queue size to return to prepulse levels. In this test the comparison is made using three conditions:

1. a uniform load of 80% of the network capacity;
2. a uniform load of 90% of the network capacity; and
3. a uniform load of 80% *plus* an impulse starting at two seconds that would be equivalent to the difference between the traffic load of cases one and two over the 12 second run. The total load is the same as for case two. For the impulse conditions, three traffic placement schemes are compared.

In the third case, the queue sizes return to prepulse levels after about eight seconds. Also note that the traffic pulse was placed on only ten percent of the nodes in order to exaggerate the effect.

Finally, Figure 5.10 illustrates the upper level protocol stability and the near optimality of the algorithms (points in the graph are at intervals of 33 ms). The lowest line in the graph represents the most fairness any algorithm could achieve. Even here we have small deviations from zero because the number of bands allocated to LANGS must be integers [44]. The horizontal line in Figure 5.10 gives the average deviation of our algorithm from this optimal solution over the total simulation time. The third graph shows how quickly the system returns to a stable state after a strong disturbance. We ran experiments with symmetrically opposite changes by both suddenly dropping and suddenly increasing the load of the net; the results were equivalent in both cases. The conclusion drawn from this figure is that *not only is the band protocol near optimal but it is also stable.*

5.3.4 Fairness

This section deals with two definitions of fairness: one as it applies to the allocation of bands to LANGs (LANG Fairness) and the other as it applies to equivalent access to the LANG's resources within a LANG (Node Fairness). Figure 5.11 shows that LANG fairness

PULSE LOADING

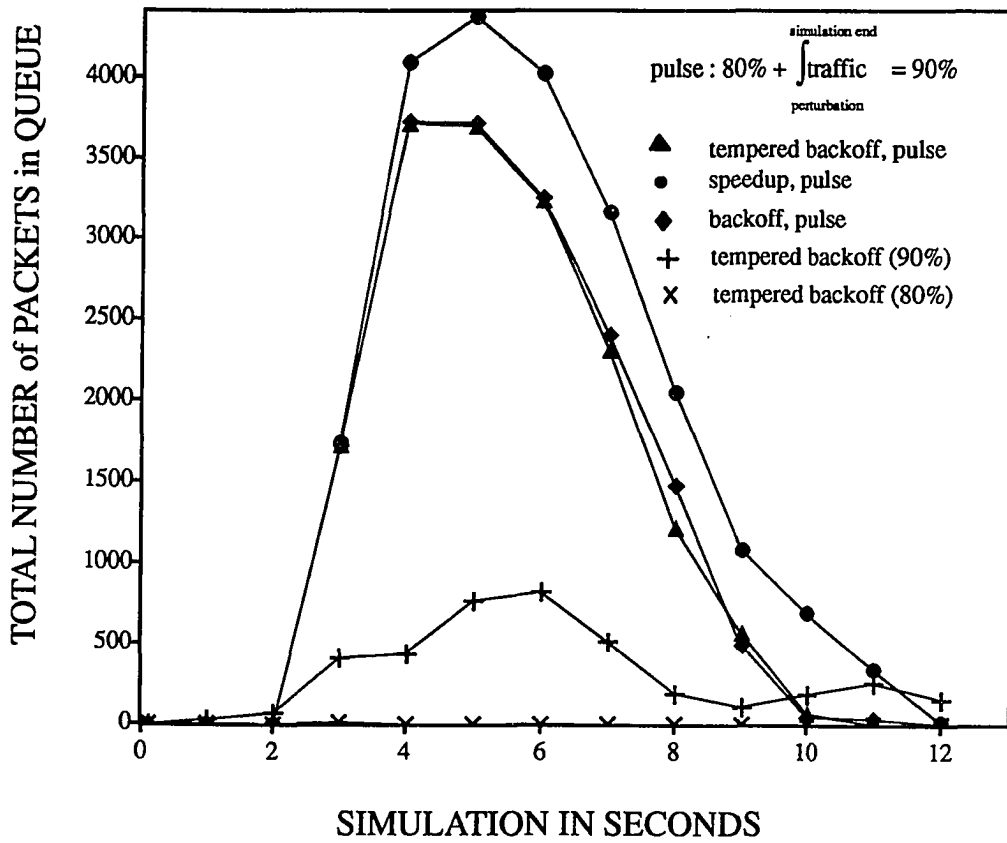


Figure 5.9: DRAMA: LANG Queue Size During Pulsed Load Conditions

is relatively insensitive to both pool size and the deviation of load from one band to the next. Figure 5.11 shows that increasing the acceptable range of utilization deviation (P) among the bands decreases LANG fairness (numerically it gets larger) because LANGs will hold bands longer as P increases. Figure 5.12 illustrates a tradeoff: as fairness is increased, response degrades because access is faster when bands are underutilized. This degradation is most likely to occur if the bands are not being reassigned to LANGs with overutilized bands. Figure 5.13 measures fairness versus load and shows it also to be reasonably insensitive to the load deviation.

5.3.5 Integration of Voice and Video

Node fairness is demonstrated through reference to Figures 5.3, 5.5, 5.6 and 5.13 (see discussion in the next section), which all show that access delay in DRAMA is relatively insensitive to many of the network parameters as long as load remains constant. It indeed shows that the traffic placement protocol described above shares bandwidth efficiently between traffic types without wasting capacity.

The curves of Figure 5.15 depict the average access delay of a data package under various network conditions. As expected, the delay is high even for low loads in a 50 Mbps channel without synchronous traffic. Here, the collision slot time is a large percentage of the time needed to send a packet, so collisions become very costly. The ten megabits per second channel with synchronous traffic uses a framed Ethernet, where framing is necessary to provide guaranteed access for the synchronous information. The cost of incorporating synchronous traffic is significant, since the framing creates a point, located immediately after the voice/video frame terminates, where the probability of collisions is high. In contrast, DRAMA effectively separates the synchronous effects caused by framing so that data packets generally have immediate access to at least some bands. In addition, DRAMA provides for efficient recovery of the portion of the frame not used for synchronous messages.

Figure 5.15 further supports the conclusion reached in Section 5.3.1: The use of multiple bands with framing for synchronous traffic is the best method for handling integrated traffic

DRAMA Stability Band Allocation

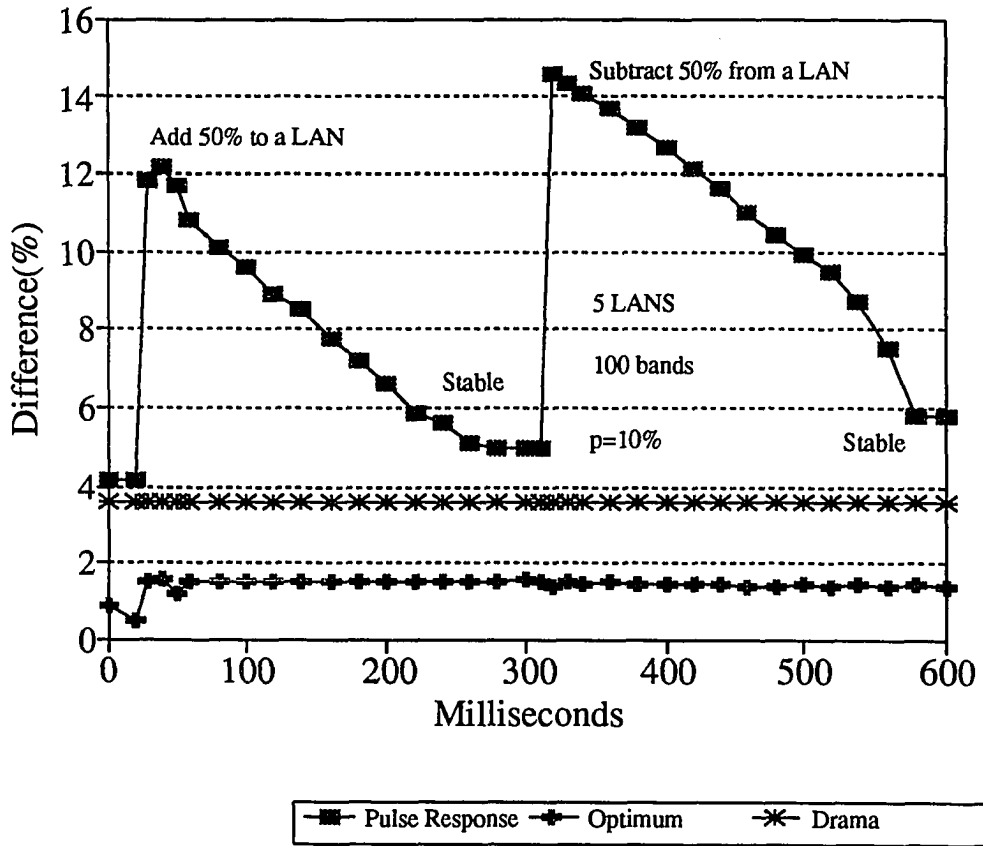


Figure 5.10: DRAMA: Band Allocation Stability

DRAMA Fairness Varying Tolerance

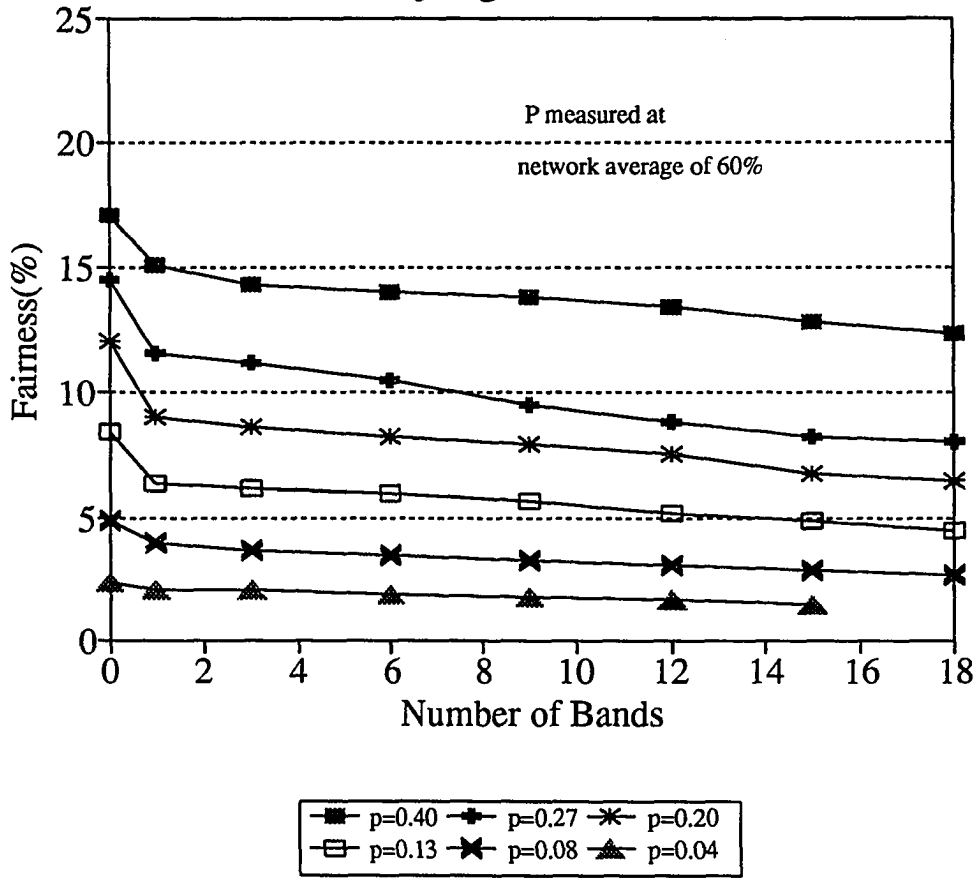


Figure 5.11: DRAMA: LANG Fairness Varying Tolerance Interval Size

DRAMA Responsiveness Varying Interval Size

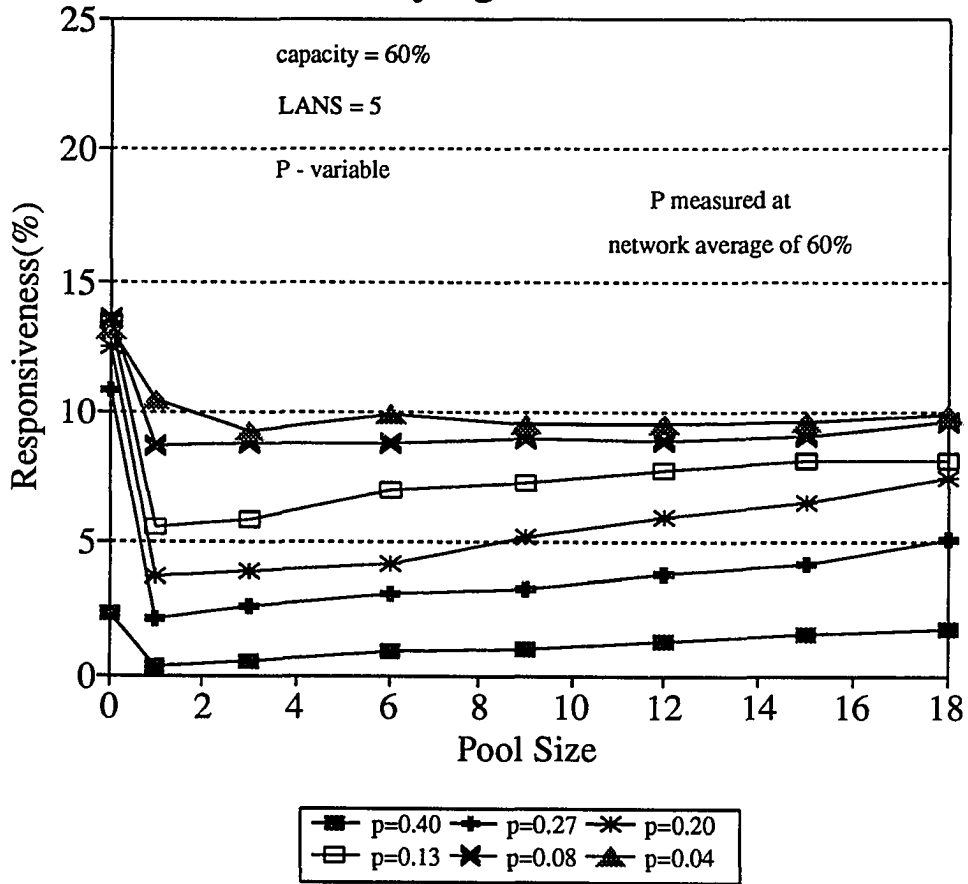


Figure 5.12: DRAMA: Response Varying Interval Size

DRAMA Fairness

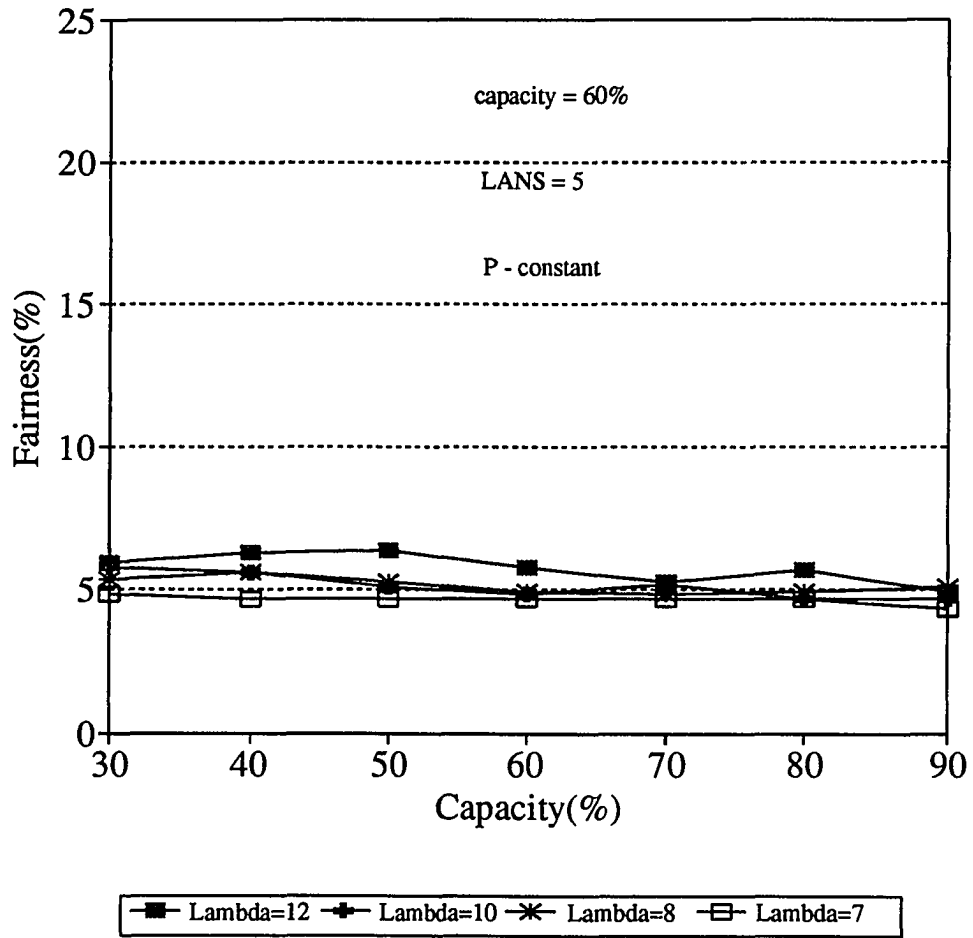


Figure 5.13: DRAMA: LANG Fairness Varying Band Traffic Deviation

DATA DELAY

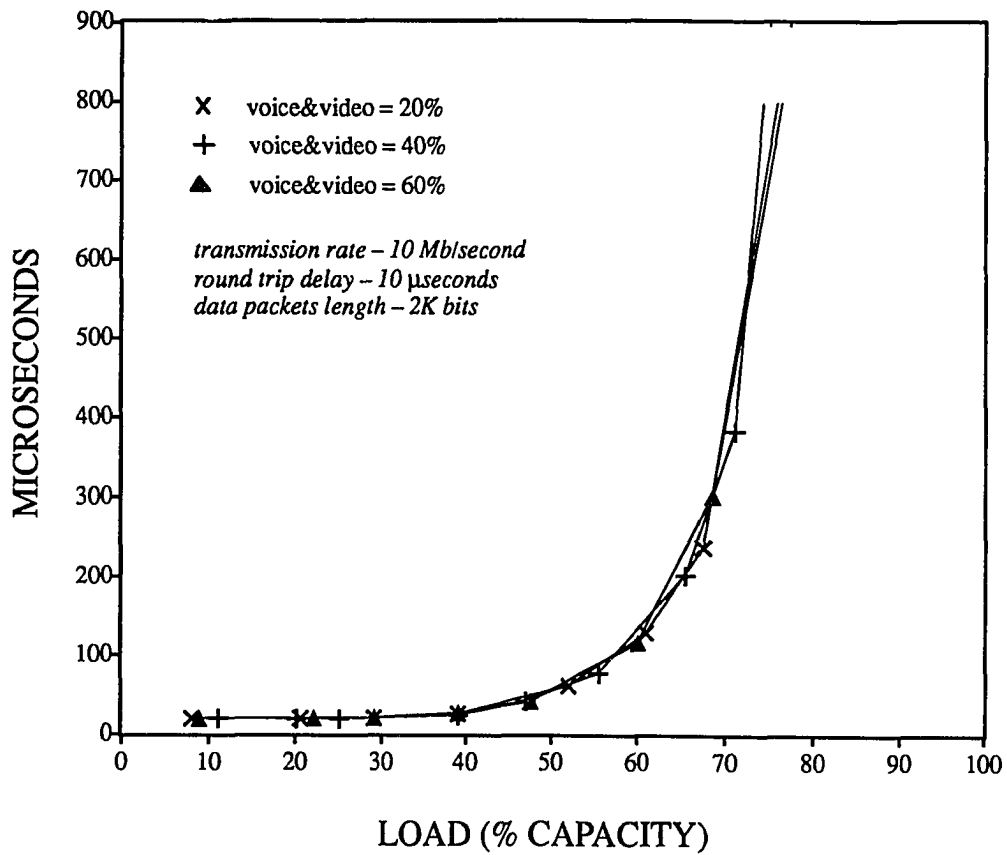


Figure 5.14: DRAMA: Access Delay Varying Periodic Traffic Load

DATA DELAY

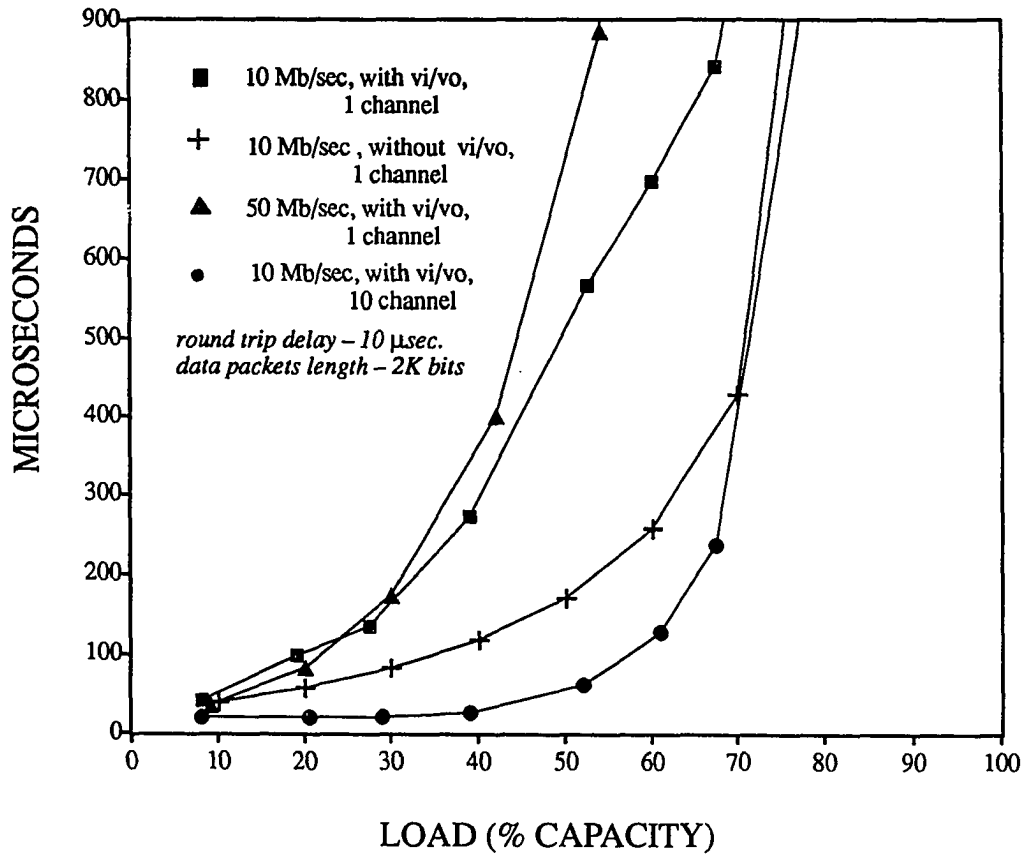


Figure 5.15: DRAMA: One Channel versus Multichannel DRAMA

in a CSMA/CD environment.

5.4 Conclusions

The DRAMA protocol extends the advantages of LANs in such a way that the protocol can be used in metropolitan areas with a large number of users. One of the key concepts behind the protocol is the use of broadband or fiber technology and the splitting of a large channel (around 500 Mbps) into many small bands. Similarly, all nodes in the network are grouped into LANGs based on their proximity. Bands are then assigned to individual LANGs on a mutually exclusive basis; each LANG transmits only the bands it owns but receives on all bands. Since DRAMA is based on CSMA/CD, a most crucial performance factor for grouping is the round trip delay. By using a mutually exclusive band assignment, the round trip delay from one for the entire net to one for the constituent LANGs is reduced.

For this protocol to succeed a mechanism must exist whereby the band assignment can be changed automatically as LANGs exhibit different traffic utilizations and mixes. This protocol allows for load balancing and also reduces the amount of noise due to collisions because in CSMA/CD the ratio of noise and load is a nonlinear function.

Maly [44] introduced an heuristic algorithm to solve the band assignment problem in a distributed fashion so that band assignment is handled on the order of ten milliseconds; that is, it will take on the order of ten milliseconds for a LANG to acquire or release a band. The algorithm's objective is to provide a "fair" assignment where "fair" means that all nodes have the same expected access delay to the net independent from location and time.

The analysis of the DRAMA protocol reported here and in other papers by Maly [44,42] indicates that the protocol has several important features. The most impressive characteristic is its extreme flexibility. A single metropolitan area network, using the DRAMA protocol can support:

- LANGs with widely varying number of nodes,

- LANGs spread across a wide geographic area,
- Dramatic fluctuations of load, and
- Widely varying mixtures of traffic types.

The protocol provides this flexibility since:

- A network can quickly rebalance loads on the order of 30 ms by reallocating bands among the LANGs.
- The protocol effectively integrates voice, video, and data on a CSMA/CD network.
- The traffic placement policies work well with dynamic resource allocation since the number of bands with synchronous traffic is kept to a minimum.
- The network is stable even at very heavy loads and with momentary overloads at some nodes.
- The protocol combines low access delay at low load (typical of CSMA/CD protocols) and acceptable performance at higher load normally associated with token ring type protocols.

These studies suggest that, even with a limited bandwidth (say ten megabits per second), use of a multiband network rather than a single band (say ten one megabit per second bands versus one ten megabits per second band) can significantly lower average access delay (though transmission time for large packets on a one megabit per second band will be longer). In addition, the multiband approach allows integration of synchronous and asynchronous traffic. These studies provide additional encouraging performance information on the DRAMA protocol; both *backoff* and *tempered backoff* allow satisfactory traffic placement.

A weakness of the design is that each node must have as many transmitter/receiver pairs as there are bands. To overcome this shortcoming, a design alternative in which each

node has typically from three to five signaling devices per node has been proposed. This would limit the bandwidth available to each node to be below the total bandwidth of the network. For example, if there were 100 ten megabits per second channels used to create a gigabit network, each node would only be able to use 50 Mbps at any time. At the current bandwidth demands of a typical workstation, this is not a problem; however, as those demands increase, the cost of additional transmitter/receiver pairs and the complexity of the network interface to manage these physical access points is likely to become prohibitive.

Lastly, we have examined the relationship between the policies for global allocation of bands; local sharing of bands within a LANG has also been examined; and these policies have been found to be compatible; Within reasonable bounds, improving one does not adversely affect the other. The pool size of bands need not be large. The number of bands within a LANG is not crucial but methods for determining the number of bands to assigned to a LANG have been suggested.

Part IV

Conclusions

Chapter 6

Conclusions

6.1 FDDI

FDDI is a token ring network based on fiber optics which is capable of a 100 Mbps rate. Traffic placement is determined primarily by a decentralized token-holding timer algorithm dependent largely on the setting of the Target Token Rotation Time (TTRT) timer. Two rings are defined in the standards document, but the second ring is employed solely for the purpose of fault-tolerance.

- A method has been developed which allows one to accurately determine the effect which the TTRT setting has on the utilization of a specific network configuration. The general result is that utilization is relatively unaffected as TTRT is reduced, but only to a point; then the utilization drops dramatically. A tradeoff between fast access for synchronous traffic and overall throughput exists; setting the TTRT extremely low guarantees fast access but reduces utilization. Reduction of the TTRT value to a point close to the drop in the utilization curve allows one to set the timer to accommodate synchronous traffic with the fastest guaranteed access, yet avoid deterioration of network performance.
- Use of an ISDN interface with FDDI in a HSLDN environment has been shown to be impractical due to the excessive propagation and node delays.
- Scalability of the transmitter at the physical (PHY) layer to achieve gigabit rates has been shown to be a viable approach. Simulation results also show the following points.

- The large number of nodes is the the single HSLDN factor having the worst impact on performance at gigabit rates. An increased number of nodes results in additional node delays, and a distribution of the data over a larger number of nodes which in turn leads to more token captures and more overhead. Packet length can also have a negative effect for the same reasons.
- Consideration should be given to a new MAC design in which the node is able to reduce the time delay between token recognition and frame transmission (LMAX). Reduction of LMAX was considered in a number of scenarios with significant reduction in delay characteristics for all cases examined.
- The most significant result of this body of research is the multiplicative effect in throughput achievable in token rings such as FDDI by employing two techniques:
 - destination removal and
 - dual counter-rotating rings.

An analysis of destination removal is presented which provides an an upper bound of sorts on the throughput increase which can be achieved. Use of the second *counter-rotating* ring is the critical factor in approaching the predicted increases. For a large number of nodes (greater than 100), two 100 Mbps rings are able to support throughput rates between 400 and 500 Mbps. In addition, a number of issues related to destination removal are presented.

- The technique is dependent upon the destination address space, so a uniformly distributed destination address space was employed for the analysis, in an attempt to use an average case. A file server scenario is also presented with results indicating that the method is reasonably robust and not dependent on a uniform destination address space.
- Destination removal uses the large numbers of nodes to improve performance and offset other negative effects of the large numbers of nodes.

- Destination removal is a feasible approach. Suggestions for implementation of the technique are given in sufficient detail to justify the feasibility conclusion.
- Destination removal also contributes to improved delay performance.
- Representative cases for MAN and WAN scenarios have been used to show that the method functions well in extremely long distances (1000 and 5000 km).

The most important requirements for use of this technique are fixed packet size and a large number of nodes. The method does not deteriorate for small numbers of nodes, rather it results in less improvement.

6.2 DRAMA

DRAMA is a broadband solution to the extension of Ethernet networks from a LAN to a MAN environment which overcomes inherent limitations of larger geographical distances and faster data rates. Groups of nodes are partitioned by geographic locality into LANGs. Network channel capacity is divided into bands which are dynamically allocated to LANGs. Each band employs a CSMA/CD strategy within a regular slot structure which provides for asynchronous and synchronous traffic.

- A strategy is provided which can quickly reallocate bands from LANGs with under-utilized bands to LANGs with excessive traffic loads.
- Varying synchronous traffic loads does not affect the ability of the network to use the remaining channel capacity for asynchronous data.
- Large fluctuations in the load on the networks can readily be absorbed.

Although implementation requires a large number of transmitter/receivers, a method is proposed which overcomes this problem without affecting the advantages shown.

Part V
Appendix

Appendix A

FDDI Overview

A.1 Design Objectives

One of the major considerations in the design of FDDI is the desire for applicability in a number of problem areas. Special consideration was given to using FDDI as a backbone for interconnection of LANS, as a network to support MAN and LAN environments, and for backend communication which is to interconnect mainframes and peripherals. Some of the features of the design may not be fully appreciated without understanding the variety of intended applications.

FDDI was also designed for compatibility with successive designs. The proposal for a second version, FDDI-II, is already in existence. Here the research deals exclusively with FDDI. Compatibility in the context of FDDI means that FDDI functions with a subset of the features of FDDI-II. FDDI nodes would not function in an FDDI-II network if all of the features of FDDI-II were being used; however, if FDDI-II was to execute using only its FDDI features, an FDDI node could be placed in an FDDI-II network.

FDDI utilizes low-cost fiber optics technology. Fiber technology shows the greatest potential for extending the current bandwidth limitations. The design is not intended to limit expandability in any of the main characteristics of a HSLDN. Faster transmitter/receivers could be substituted to a point (discussed in Chapter 3). In addition, up to 1000 nodes could be supported at a distance up to 100 km. Each of these factors could be increased at a tradeoff of reducing the other.

Nodes may be separated by a maximum distance of two kilometers between adjacent

nodes with a total separation of 100 km, so the network uses a distributed timing mechanism and distributed recovery algorithm. A degree of fault tolerance should be accommodated in the event of node failure, network failure due to improper configuration, or catastrophic failure such as a cable cut.

FDDI supports four classes of traffic and allows for prioritization. The four types of traffic are isochronous, synchronous, restricted asynchronous, and unrestricted asynchronous. Isochronous traffic must be delivered in at regular intervals. The difference between isochronous traffic and synchronous traffic is that synchronous traffic has a tolerance interval, but isochronous must be delivered at precisely the same time for each frame.

Asynchronous traffic is classified as either restricted or nonrestricted. The restricted data is provided to allow for monopolization of the asynchronous bandwidth for limited periods of time. This has applicability in backend operation where peripherals need a large bandwidth for limited periods of time, e.g., a disk drive communicating with a mainframe. An upper bound exists for the amount of time in which two nodes may engage in restricted mode.

The four traffic classes have been listed in the order of their respective priorities. In addition, a finer granularity of priority may be utilized for nonrestricted traffic. Due to the fact that synchronous traffic can be viewed in an orthogonal manner compared to asynchronous traffic, the synchronous traffic is not of primary concern in this research although it is addressed.

In order to support the various traffic requirements and fairness of access by all nodes, a distributed scheduling policy developed by Ulm [68] is employed.

In summary, FDDI is already designed to be extendable to a MAN. It supports the same set of parameters described in Chapter 5 for a MAN, including support for synchronous and asynchronous traffic. FDDI is a relatively unexplored network and the focus here is to study performance and explore areas for performance enhancement.

A.2 Implementation

A.2.1 Topology

FDDI [63,56] is composed of counter-rotating dual rings. Each node is either a class A node or a class B node. Class A nodes may function as standard FDDI nodes or as wiring concentrators as seen in Figure A.1. Class B nodes are connected to the ring through a wiring concentrator class A node and are only connected to one of the rings. Class A nodes are not required to be a wiring concentrator, but may be. Class A rings are dual ported and are connected to both rings, having the potential to transmit and receive on each ring. Each of the rings transmit at 100 Mbps. The dual rings provide a fault tolerant capability for FDDI. Figure A.1 gives an example of class A and B nodes connected in a network. Figure A.2 shows how this network would heal itself in the event of the cutting of a cable connecting two nodes; however, a second similar cut would divide the nodes into two sets which would not be able to communicate with one another.

Communication between nodes is point-to-point. For class A nodes the cable is a dual fiber cable providing the inbound signal from ring one and delivering the outbound signal on ring two through one connection and the outbound signal for ring one and the inbound signal on ring two through the other connection. Class B nodes will have a single dual fiber cable for input and output on the same ring.

A.2.2 Transmission Medium

As mentioned previously, FDDI is an optical fiber based technology. One of the design objectives was to develop a low-cost fiber network, so cost of components was a factor in the elements chosen. The medium chosen was a multimode fiber. LEDs (Light Emitting Diode) are used as transmitters because they are both cheaper and safer than lasers. If a curious user happens to cut a fiber and look into it, an LED causes no harm but a laser can cause retinal damage. PIN photodiodes are used as the optical receiver [3].

Due to the current stage of development of fiber optic detection, coherent detection of

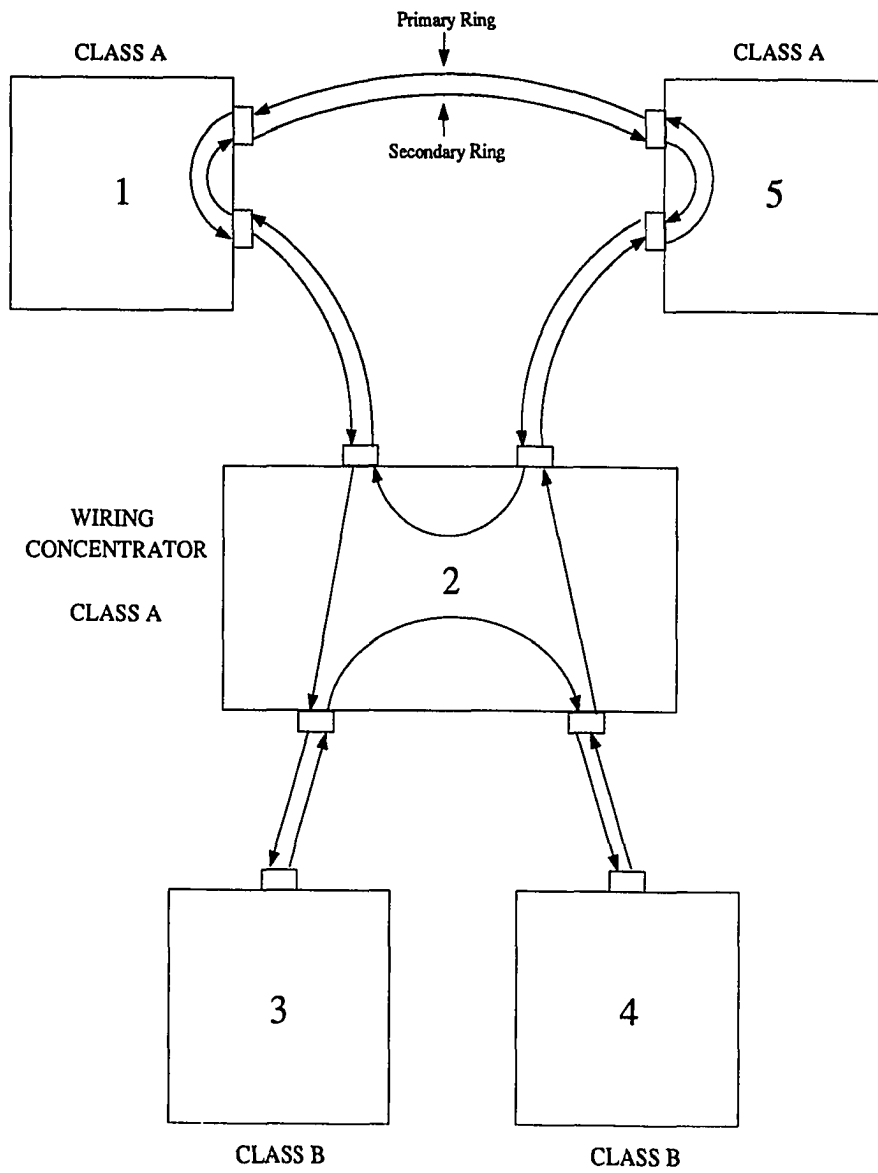


Figure A.1: FDDI Configuration

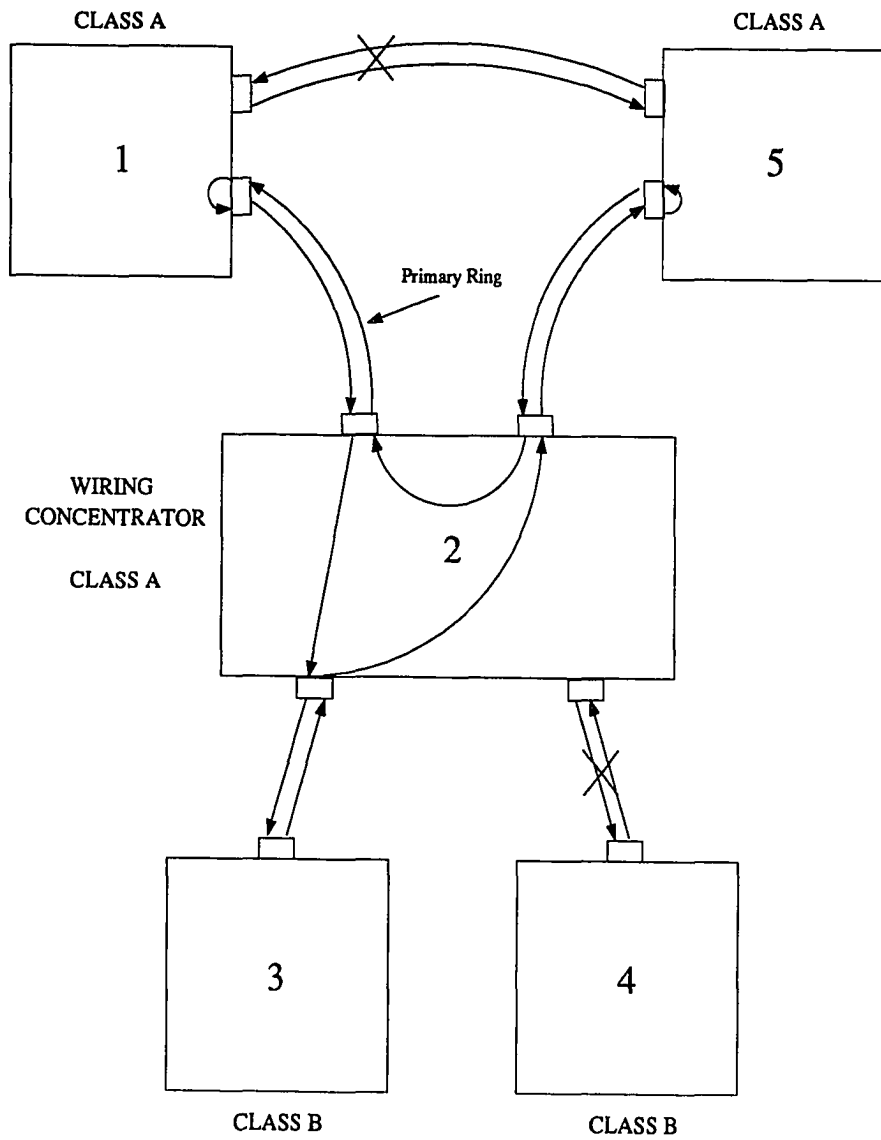


Figure A.2: FDDI Healing

the signal is much more expensive and only applicable in certain types of configurations. Non-coherent transmission means that the detector looks only for the presence or absence of light, rather than examining more subtle aspects of the signal such as the frequency or wavelength of the signal. Error rates for the choice of components to implement a fiber optics medium are on the order of 2.5×10^{-10} .

A.2.3 Signalling

Given the non-coherent detection scheme, one would classify this as a digital, as opposed to analog, signalling technique. Numerous digital signalling techniques exist [62]. One differentiating feature of these signalling techniques is the ability to transmit timing information in the data itself. This is not without a cost. Manchester encoding techniques, for example, have a transition from low to high or high to low during each bit time. In addition to obtaining the value of zero or one, the receiver also can be assured of proper synchronization with the transmitter. If zero is represented by the absence of light and one by the presence of light, a sufficiently long sequence of consecutive zeros or ones could allow for timing differences between transmitter and receiver to go undetected. Figure A.3 gives an example of how data can be misinterpreted if the clocks of the transmitter and receiver drift too much; this phenomenon is called jitter. Note that the Manchester signal requires twice the bandwidth of a Non-Return-Zero (NRZ) signal which changes at most once per bit time.

FDDI uses a 4/5 encoding technique. For each sequence of four bits to be transmitted, an extra overhead bit is included. In order to provide some timing information (but with less overhead than Manchester), only those patterns which will guarantee a signal change once every three bits are used. Whereas Manchester encoding has a 50% overhead for timing, 4/5 has only a 20% overhead. By examining Table A.1 one can find those bit sequences which do not have this run length characteristic. The valid bit sequences are used to represent data and framing information. Even though FDDI transmits at a rate of 125 Mbps, the stated transmission rate is 100 Mbps due to this overhead.

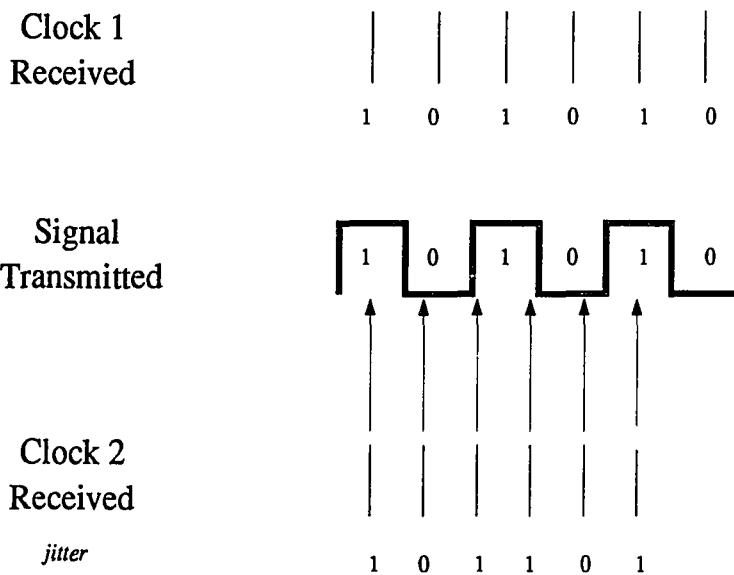


Figure A.3: Clock Jitter

FDDI 4/5 Symbols

Control and Framing Characters	
BINARY	USE
11111	IDLE
11000	SD Sequence
10001	SD Sequence
00111	Terminate Data
01101	RESET
11001	SET
00000	QUIET
00100	HALT

Illegal Characters
BINARY
00001
00010
00011
00101
00110
01000
01100
10000

Data Characters	
BINARY	USE
11110	0 data byte
01001	1 data byte
10100	2 data byte
10101	3 data byte
01010	4 data byte
01011	5 data byte
01110	6 data byte
01111	7 data byte
10010	8 data byte
10011	9 data byte
10110	A data byte
10111	B data byte
11010	C data byte
11011	D data byte
11100	E data byte
11101	F data byte

Table A.1: FDDI: 4/5 Encoding

A.2.4 Timing

As mentioned previously, timing for the received signal is derived from the 4/5 encoding of information. No explicit timing signal exists, so the clocks of individual nodes can vary. As a result of the use of point-to-point node connections, a node does not need to worry about transmitting at the exact same rate as received. Instead the transmitter retimes the signal each time it departs a node. If a node transmits at a slower rate than it receives, then it is inevitable that some bufferring of the received data is necessary. FDDI has resolved this problem by requiring all clocks be within a tolerance of 0.005%, and also requiring the buffers at each node to absorb the difference between the received and transmitted signal rates by an elasticity buffer. The maximum frame length and variance between clocks requires a maximum elasticity buffer length of ten bits at each node.

Appendix B

The DRAMA Protocol

B.1 Overview

DRAMA addresses the basic issues listed above for MANs. The fundamental design employs a two level structure. At the upper level (i.e. MAN level), bands are allocated to groups of nodes. The nodes are clustered by distance and function into local area network groups (LANGs). This type of clustering is an important characteristic required for good performance of DRAMA. At the lower level (i.e. LANG level), each group of nodes must decide how to utilize the bands assigned to it. This two level approach allows for accommodation of the various traffic conditions the MAN might experience.

For logical purposes, assume that a single broadband cable connects all the nodes. The topology of the network is such that no closed loop exists in the cable layout and a direct cable path exists between any two nodes in the network as shown in Figure B.1. The cable bandwidth is frequency divided into two sets of bands:

- those dedicated to particular LANGs; and
- those in a global pool of bands that may be acquired by any of the LANGs.

For each LANG, requesting, acquiring, or releasing a band depends on the current distribution and amount of traffic within that LANG relative to the current traffic within the entire network. In the DRAMA system, the mechanism to control the assignment of bands to a LANG is called the *band allocation protocol*.

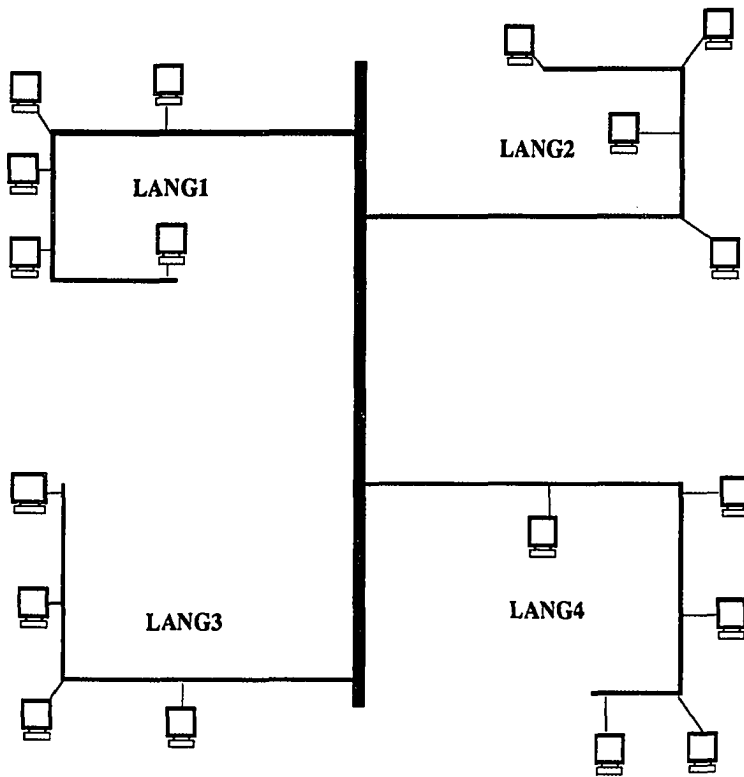


Figure B.1: DRAMA: A Network Configuration

In MAN systems, static partitioning according to average requirements will often waste idle bandwidth and at other times will be insufficient to satisfy a LANG's traffic requirements even if bandwidth elsewhere in the network is not in use. DRAMA uses the band allocation protocol to allow each LANG to acquire a greater amount of bandwidth than would be possible if the bands were statically partitioned among the LANGs, thus using the load diversity and traffic variations as an advantage rather than a hindrance to the design.

Another desirable characteristic of a MAN system is that all nodes be able to communicate with one another in a uniform way regardless of distance. That is, the network should not distinguish between messages destined for local nodes and distant nodes. DRAMA provides uniform communication by allowing both inter-LANG and intra-LANG traffic on each band and by using the integrated voice/data protocol proposed in [58] on each voice/data band. The only additional delay between more distant nodes should be because of the inherently longer propagation delay between them, not because of any extra switching or connection setup time. This is true in DRAMA.

The lower level structure in DRAMA is based on a LANG. Within a LANG the traffic placement policy controls access to resources and is based upon CSMA/CD. This protocol is chosen for two reasons: it has a history as a reliable network design and it provides exceptional performance when the geographical distance between nodes is small. Groups of nodes are typically clustered among various locations of a company within a particular city, installations on a large ship or military base, or among different departments in a university.

Generally, CSMA/CD is not suitable between nodes separated by more than two kilometers because the collision interval is directly proportional to the propagation delay between the most distant nodes. See Figure 1.2. This problem is circumvented by restricting transmission privileges on any band to exactly one LANG at a time, while allowing all LANGs to receive *all* transmissions. In this way, the CSMA/CD-based protocol can be used over the entire set of LANGs with the same efficiency as in a single LANG. Contention intervals

are short, and are determined by the length of the LANG.

Equation 1.2 from the previous analysis is affected in the following manner. The utilization of the network is strongly dependent upon the value of the term in the denominator of Equation 1.2, $2\tau\frac{(1-A)}{A}$. The value of this term must be less than one in order to achieve utilizations of 50% or more under loaded conditions. If the distance increases from a value of two to 100 km, the value of τ and the term itself increases proportionally by a factor of 50. Consider a LAN which has a value of 0.2 for the term. Maximum utilization would drop from 83% to 16%.

In assigning traffic using CSMA/CD, the amount of non-message traffic caused by collisions is a nonlinear function of load. Bands which are heavily loaded have low utilization due to the high number of collisions. By providing more bands to the heavily loaded LANGs, the number of collisions is reduced and overall throughput is increased. The increase in collisions in lightly loaded LANGs, from which bands are removed, is more than offset by the decrease in collisions in heavily loaded LANGs. Thus it is important that the upper level protocol of DRAMA reallocate bands among LANGS in a manner that balances utilization of all allocated bands.

In summary, for DRAMA:

- the *increase in bandwidth* is achieved by increasing the number of bands or the bandwidth of each band,
- partitioning the nodes into groups (LANGs) that contend only on a subset of the bands with nodes within a geographical proximity allows for *geographical expansion* and a *large number of nodes*,
- the band allocation protocol is a mechanism for *dynamic allocation of bandwidth*,
- *integration of synchronous and asynchronous traffic* is facilitated by the traffic placement protocol proposed in [58] and is further explained in Section B.3.2.

B.2 Objectives

Objectives for the DRAMA protocol are:

Robustness: Node failures should not cause the transmission protocol or the band sharing mechanism to fail, nor should failure of one local network cause other local networks to fail. With the exception of node chattering, CSMA/CD bus architectures are insensitive to individual node failures.

Availability: Each LANG should have a guaranteed minimum bandwidth available at least equal to a “normal LANG” requirement, between one and ten megabits per second. Global allocation schemes should adapt to each LANGs varying load requirements.

Feasibility: The overall design should be practical and economical.

Fairness: The band sharing policy should adapt to traffic conditions and should minimize the likelihood that a LANG retains poorly utilized bands, particularly when bands are more urgently required by another LANG. In addition, nodes within a LANG should have equal access to the channels assigned to that LANG.

Adaptability: The reallocation strategy should guarantee that bands are reassigned quickly to overloaded LANGs and that varied traffic patterns can be accommodated without unusual delay or overhead.

Efficiency: Network throughput should be maintained even under high load conditions.

Stability: The network should not degrade under bursts of extremely heavy traffic, and should return to a normal traffic pattern within a reasonable time.

Robustness is achieved through the LANG recovery procedures described in [58,57]. Providing bands dedicated to each LANG accomplishes the desired availability. Details of design feasibility are provided in Section B.3.3. Finally, the fairness, adaptability, efficiency, and stability aspects of DRAMA are demonstrated the results presented in Section 5.3.

In this paper, the performance of DRAMA for these objectives is analyzed by examining the upper level band allocation protocol, the lower level traffic placement policy, and their integration. The algorithm to allocate bandwidth is crucial to the overall performance. At the upper level, the reassignment of bands in a dynamic environment is investigated. The analysis of the global bandwidth allocation scheme emphasizes how LANGs share available bandwidth. At the lower level, each node uses a traffic placement policy to determine which of the bands assigned to it should be used for synchronous and asynchronous traffic. By studying the protocol in detail at each of these two levels, we evaluate its operation under varying conditions.

B.3 Protocol Description

We now discuss the band allocation protocol (upper level) and the traffic placement protocol (lower level) of DRAMA.

B.3.1 Band Allocation Protocol

In DRAMA the bandwidth is frequency divided into $M+1$ bands. One band is reserved for a *slotted band control channel* used by all the LANGs to coordinate band sharing; the remaining M fixed size bands are available for voice, data and video transmissions. The M bands are partitioned into a set of *dedicated bands* and a set of *available* bands. Dedicated bands guarantee that no LANG “starves” and that each LANG’s performance is at least that of a normal, solitary LANG using a baseband cable.

Available bands are either assigned to a LANG or are in a global pool. Bands in the global pool are shared via a dynamic, fully distributed, band sharing policy. It allows each LANG to obtain global bands based on its current needs, the current needs of other LANGs, and the current availability of global bands. When a band is acquired by a LANG, that LANG has *exclusive* transmission rights on that band as long as the band remains allocated to it, although each node listens to all bands for packets addressed to itself. In theory, to implement DRAMA each node would require a receiver for each band and enough

frequency-agile transmitters to service the number of bands upon which it must transmit simultaneously. In practice, receivers can be shared by a node as shown in Figure B.1. Upon receiving data for its nodes, the receiver retransmits the information on an independent system shared by all its nodes. Each LANG may have a receiver set to handle all bands for its nodes. Therefore, each node needs a receiver to detect when the band it wishes to transmit on is occupied and up to the time that the collision detection time has expired. In all probability a node may need fewer receivers than transmitters and in most cases a single frequency-agile receiver is sufficient.

The LANGs coordinate the request, acquisition, and release of global bands via the LANG's slotted band-control channel in which each LANG owns a particular slot that is used for these functions. The slotted scheme allows the band sharing policy to be collision free and guarantees that bandwidth reallocation can take place in a fixed amount of time. Specifically, if the current number of available global bands is at least equal to the demand for additional global bands, then all acquisitions are accomplished in one band control cycle. Even if this is not the case, if the bands allocated to a LANG are insufficient for its current needs and a band can be released by another LANG, then a request, release, and reallocation occur within at most one band control cycle (although the reallocation will go to the band most in need of additional resources).

Integrated traffic is transmitted using a CSMA/CD-based protocol originally developed in [58] and discussed in the next subsection. Intranet and internet traffic may coexist on any band. In fact, the transmission protocol makes no distinction between the two types of traffic; the primary idea is to allow each source, from a transmission protocol point of view, to transmit to any destination as if the destination were in the source's LANG.

B.3.2 Traffic Placement Protocol

In this section an expanded integrated synchronous/asynchronous transmission protocol originally proposed in [58] is described. In this protocol, the fraction of a band's capacity allotted to each traffic type depends on the current voice/video load and data load. Time

on each band is slotted into frames. Each band's frames are delimited by "frame-begin" markers. For reliability and robustness the frame marker also contains the identifier of the LANG to which the band is currently allocated. At any time, exactly one node in the LANG is responsible for broadcasting the frame-begin markers for a particular band, although the identity of the node may vary over time. For clarity, we refer to this node as a *band leader*; note that a LANG may have multiple band leaders if it holds multiple bands. The band leader of the LANG's *dedicated* band is also the current LANG leader, transmitting the LANG's band-control slot in the band sharing scheme.

Each frame is partitioned into a *synchronous region* and a *data region*. The boundary between the two regions will vary from frame to frame depending on the amount of voice and video data. The synchronous region can consume the entire frame if necessary; likewise the data region can consume an entire frame in the absence of synchronous traffic. Call setups are treated as data packets.

The synchronous region provides one virtual circuit for each established (one way) call. This means that a two-way call uses two bands, one for each direction of the call. Silent periods, which comprise roughly 60% of an average voice conversation [2], are not transmitted. One varying-size slot is allocated in the synchronous region to each (one way) call. The slot contains the digitized voice data followed by control information called the *control byte*. Video traffic is treated in an identical fashion; the amount of video data sent in any frame depends on the data compression techniques, if any, which are employed. The control byte informs the other nodes whether the virtual circuit will terminate after this frame or will be continued. The slots for the different circuits are contiguous and precede any data transmitted in the frame. Each circuit has a *position number*, i , in the sequence of slots and begins its transmission after $i - 1$ previous slots have been sensed in this frame via their control bytes. The position number of a circuit decreases whenever a circuit with a lower position number terminates.

Figure B.2 contains a frame in which three calls (C_A, C_D, C_E) signal that they will con-

tinue; two calls (C_B, C_C) signal that they will terminate; one call (C_D) has a silent period and one call (X) has been successfully established. Data packets ($D_1, D_2, \text{call-set-up}_X$) are placed according to CSMA/CD access.

Note that once synchronous traffic has been placed on a band, the band cannot be released since reassignment is considered to be too difficult. In order to provide for cooperation between the global band allocation and traffic placement protocols, some effort must be made to assure that bands can be released when it becomes necessary to reassign the band to another LANG. This cooperation is enhanced if synchronous traffic within a LANG is placed to minimize the number of bands with synchronous traffic. In contrast, data traffic should be placed on as many bands as possible to minimize collision and delay.

A number of traffic placement schemes, investigated to assist in minimizing delay for data traffic, are summarized here (see [42] for more details):

Fixed: In the event of a collision, this policy delays a constant amount of time before retransmitting. It assumes no knowledge of the state of the network and provides a baseline of comparison with those strategies that do attempt to use knowledge of the state of the network.

Backoff: This method, which is substantially the same as standard binary backoff, waits longer between retransmissions if repeated collisions occur. Traffic status (busy/not busy) is determined by the number of collisions experienced.

Speedup: Based upon the size of its own message queue, a node decides how it should attempt to place traffic. Delay before transmission is inversely proportional to the queue size. Somewhat opposite to backoff, this is a selfish policy in which any busy nodes tries to use as much of the available bandwidth for itself as possible.

Tempered Backoff: Given the intuition that speedup would be preferable in light loads and backoff in heavy loads, a combination of the two, which we call tempered backoff is also investigated. If a node is experiencing fewer collisions, tempered backoff will

DRAMA Frame Structure

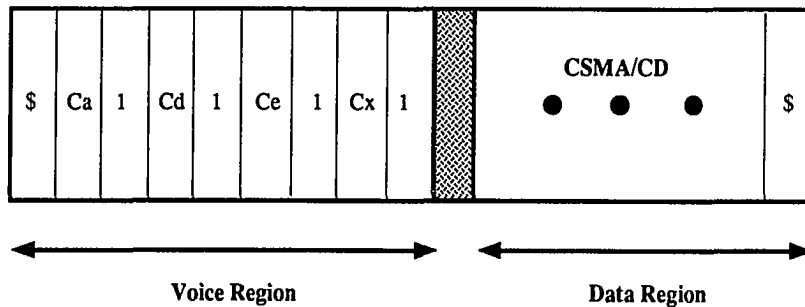
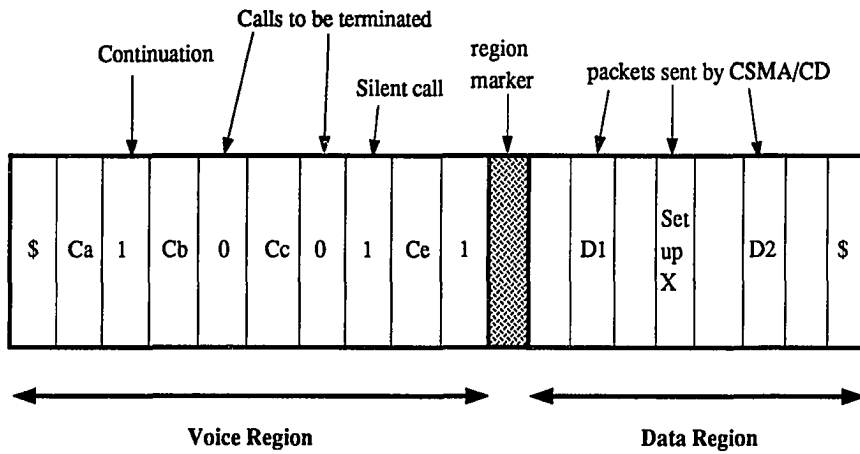


Figure B.2: DRAMA: Sample Frame

approximate speedup. As collisions increase, the delay incorporated into the formula for backoff will quickly dominate the term for speedup and will approximate a backoff approach.

B.3.3 DRAMA Feasibility

In the original DRAMA system [58] each node needed as many receivers as there were channels and enough tunable transmitters to service the channels assigned to its LANG. This is not economically feasible. Figure B.3 presents a possible solution to this cost problem by showing how nodes in a LANG can share receivers and how the number of transmitters per node can be limited. The network sketched has 100 Mbps total capacity, divided into 100 channels. The LANG illustrated has ten bands assigned to it. Each node needs at least one tunable transmitter/receiver pair to be able to determine whether a selected band is free and to detect a possible collision on that band after it has begun transmitting. Nodes can have more than one tunable transmitter if they need to send information on more than one channel simultaneously.

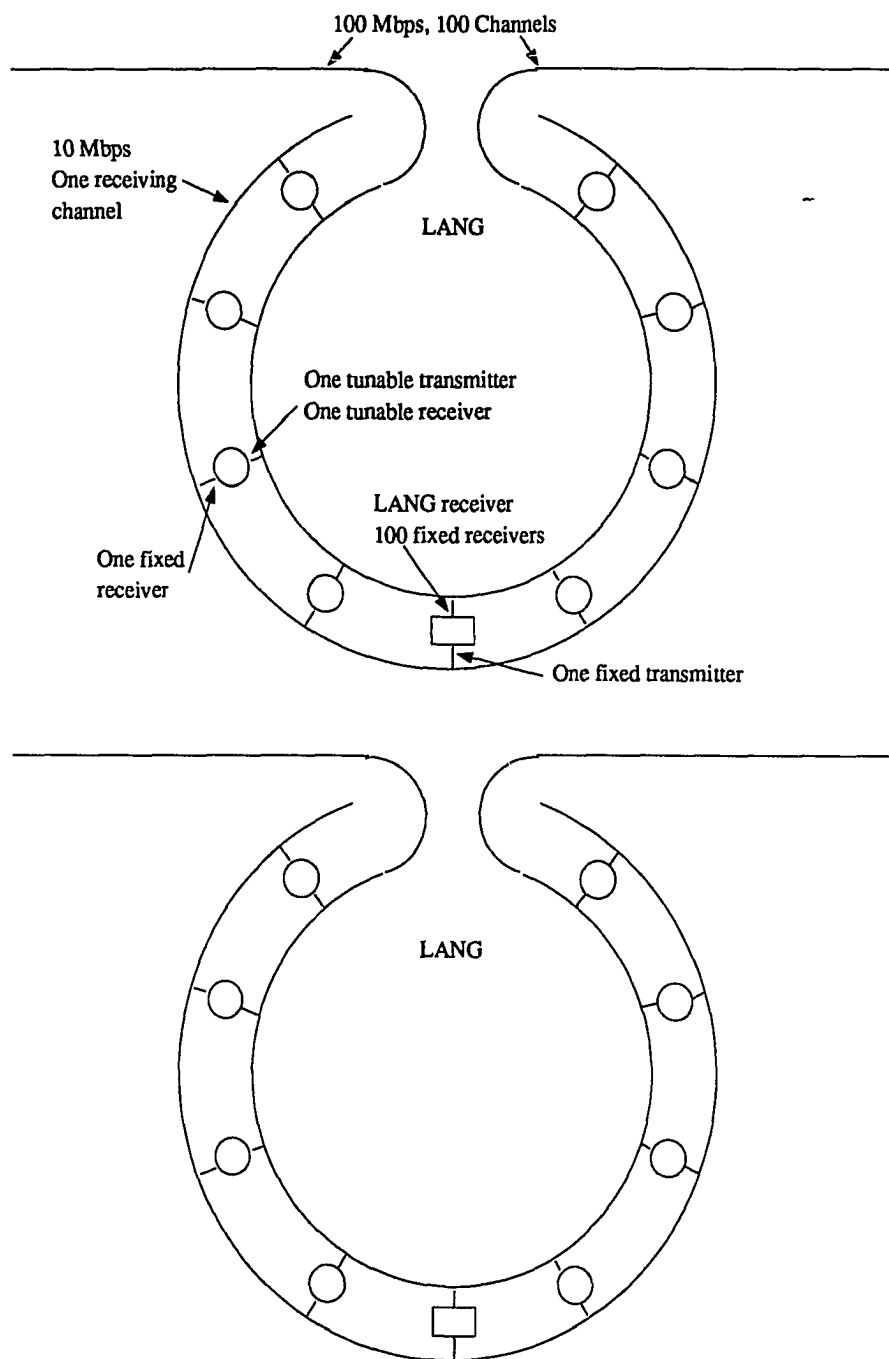


Figure B.3: DRAMA: A Possible Network Configuration

Appendix C

Fiber Optics Technology

Light has long held researchers in fascination, but its applicability in the area of communications has only received serious attention within the last 30 years. As with any communications system, there must be three important components

- a transmitter,
- a receiver, and
- a medium to carry the signal.

This appendix contains a summary of each of these components in a fiber optics system. It also contains a section describing the current state of research in the integration of optical and electrical components which will likely follow the evolution of semiconductor scaling from small scale to VLSI¹[3,39,49,72].

C.1 Fibers

It is impossible to live on this planet without being confronted with the knowledge that light travels in free space, without a guide. One might reasonably ask why a guide such as fiber is required at all. As with microwave transmission, light can be directed in the form of a beam into free space. In the right application it might even be a suitable solution, but it also contains a number of disadvantageous properties. Under ideal conditions, the signal can be transmitted over distances of 150 km, but more typical distances would be in

¹Very Large Scale Integration

the range of a few kilometers. The primary categories contributing to the attenuation of the signal are absorption and scattering. Absorption is caused by elements in the air such as water vapor or CO_2 . Scattering is caused by larger particles in the atmosphere such as smoke or fog.

Experiments concerning sending light in a guide are recorded to have occurred as early as 1870 when Tyndall showed that light could be guided within a water jet. Research into the development of waveguides initially (1950s) produced guides which were marked by extremely high attenuation, 1000 dB/km, but today they have attenuation less than one dB/km. This makes them extremely attractive for long distance communication.

The basic principle of a fiber is that it is composed of two concentric materials, an inner and outer core. The differences in refractive indices of the two materials is such that light of certain wavelengths will experience total internal reflection when injected into the inner core; none of the light escapes from the fiber. This holds unless the fiber is exposed to severe bending.

Most fibers are composed of glass but some are created from plastic. Plastic fibers are less expensive to manufacture and are more sensitive to attenuation. Transmission losses are in the ten dB/km range. Typically they are not made entirely of plastic, but have a silica core with plastic cladding. The area of application for plastic fibers is in short distance, low bandwidth requirements. Plastic fiber systems are commercially available at two Mbps [65]. Glass fibers are made from a variety of glass materials with different refractive indices. Glass attenuation has approached 0.2 dB/km and is capable of much higher bandwidths. Commercial glass fiber systems are available in the one Gbps range.

Attenuation of the signal is the result of a number of properties of the fiber including non-homogeneity of the material, impurity of the material and changes in vibrational states of the lattice of the material. Some of these attenuation factors are a function of the wavelength of the light propagating down the fiber. Windows of good performance exists

around 850, 1300 and 1500 nm² and most systems use one of these wavelengths.

In addition to material type, fibers are classified according to the mode(s) of light which can be transmitted in the fiber. When light is injected into the fiber, the sources of the light do not all enter at the same angle. A more shallow path translates into a shorter distance to the destination. This means that when light enters the fiber, various modes of propagation will reach the destination at different times. This concept is called dispersion and is a limiting factor in the bandwidth capacity of the fiber. Dispersion is also a function of wavelength and a consideration in the selection of fiber/transmitter/receiver. By limiting the diameter of the inner core to the size of the wavelength being transmitted, single mode transmission can be obtained. Single mode transmission has much greater potential for bandwidth expansion but requires more precise transmitters in order to place the light into the inner core. A third fiber type, graded index fiber, is a compromise of both. As contrasted with the previous two types, these fibers have a gradual rather than abrupt change in the refractive indices of the inner and outer core. This minimizes the effect of dispersion caused by the various modes.

C.2 Transmitters

The light emitting source used in a fiber optics system must deliver a beam of light sufficiently concentrated to inject the power level required for proper detection by the receiver. It must also be able to function at the required bandwidth. Two types of light sources are used for this purpose, light emitting diodes (LEDs) and injection laser diodes (ILDs).

LEDs are less expensive, more reliable and easier to drive but are particularly inefficient at applying the power of the generated light into the fiber. They require a large surface area to generate the light necessary to compensate for the misdirected light. Larger surface areas increase parasitic capacitance which in turn restricts the potential bandwidth. For this reason surface-emitting diodes are considered inferior to edge-emitting diodes. Edge-emitting diodes provide a more directed beam of light and are considered suitable for single

²The abbreviation nm represents nanometers 10⁻⁹

mode transmission.

Lasers clearly will be required for coherent transmission techniques. They have higher bandwidth and can transmit in a much narrower linewidth than LEDs. Linewidth represents the ability of the source to generate a signal at a specific wavelength. A narrower linewidth indicates less signal strength at neighboring wavelengths. Lasers possess a higher power output and a narrower beam than LEDs and are more efficient at delivering the required power to the fiber. The two primary lasers used are Fabry-Perot (FP) and dynamic single mode (DSM). One of the DSM laser diodes, a distributed feedback laser (DFB), has a sufficiently narrow linewidth to support data rates of 1.6 Gbps in practical systems.

C.3 Receivers

The receiver of a fiber optics system detects the presence of light. Reception of photons turns the device on and produces a power gain. The two types of receivers typically used for this purpose are avalanche photodiodes (APDs) and PIN photodiodes. Receivers are sensitive to the wavelength of the light which it is sensing. Different materials are used depending upon the wavelength.

The greatest advantage of APDs is the high gain they exhibit. APDs work well in wavelength ranges up to 1600 nm. This matches well with the performance characteristics of the glass fibers being used. Cut-off frequencies for APDs are typically in the one to two gigahertz range.

PIN photodiodes have the advantage of higher cut-off frequencies, above ten gigahertz. The problem with them is that they exhibit no gain over the received signal. In order to utilize a PIN, it must be coupled with a field-effect transistor (FET) amplifier. In order to minimize the parasitic capacitance of interconnection of these two elements, the monolithic integration of these two devices has been investigated. This is an example of integration of electrical and optical devices.

C.4 Optoelectrical Integration

As mentioned at the end of the previous section, integration of the electrical and optical components is a significant issue to be solved in order to move to data rates significantly above the one to two gigabits per second range. The concept of parasitic capacitance is an unavoidable problem when connecting discrete components. Capacitance, the ability of a device to retain a charge, is a negative effect in switching devices because it limits the rate at which the device can change state. Monolithic integration of these devices onto the same substrate reduces this capacitance and increases the bandwidth.

Maeda [39] gives examples of some of the devices which have been integrated since 1972. These include lasers and FETs, photodiodes and FETs, transmitter/receiver pairs, and laser arrays. When mounting these electrical and optical devices on the same substrate, one of the problems is matching the substrate material with both devices. For example, GaAs has proven very successful in the integration of electrical devices. When applied to optical devices, the wavelength is short, 800 nm. This wavelength produces large attenuation and dispersion in the fiber when compared with longer wavelengths. Other substrates do provide for longer wavelengths and are the subject of much investigation in this area.

Appendix D

Simple Destination Removal Simulation

A number of simulations were employed in the development of this research. In Chapter 4 two simulations are discussed. The simulation used to produce Figure 4.2 is included here to include the level of complexity of the simulation. This Pascal model, as compared to the primary Simscript model employed for all of the other results in the thesis, is only concerned with the number of times a slot can be reused as it circulates around the ring.

```
program test_remove(input,output);
var i,j,iter,messages,k,l,n : integer;
    done:boolean;
    infl,fl : text;
    infn,fn : string[20];
    rsp:char;
    maxmess,thismess: integer;

function getran(x:integer):integer;
var z:integer;
begin
    (* GENERATE RANDOM INTEGER <> x *)
    (* BETWEEN 0 AND n - # NODES      *)
```

```

    z:= random(n);
    while z=x do
        begin
            z:= random(n)
        end;
    getran :=z;
end;

begin
    writeln('Enter value of max uses of message');
    readln(maxmess);
    writeln('Enter outputfile name');
    readln(fn);
    writeln('Enter inputfilename');
    readln(infn);
    assign(fl,fn);
    assign(infl,infn);
    rewrite(fl);
    reset(infl);
repeat
    messages :=0;
    writeln('Enter number of nodes');
    repeat readln(infl,n); until n>1;
    writeln('Enter number of iterations');
    repeat readln(infl,iter); until iter > 1;
    randomize; (* INITIIALIZE GENERATOR *)
    for i:= 1 to iter do

```

```

begin  (* STARTING AT NODE n-1 *)
      (* TIMES SLOT REUSED *)
      thismess:=0;
      repeat
        (* GENERATE DESTINATION *)
        j:= getran(n-1);
        until j<>0;
      while (j<>0) and (thismess<maxmess) do
      begin
        messages:=messages + 1;
        thismess:=thismess+1;
        done:=false;
        (* VISIT NODES LOOKING FOR A *)
        (* MESSAGE WHICH WILL FIT *)
        repeat
          (* k IS DESTINATION OF MESSAGE *)
          (* QUEUED AT CURRENT NODE *)
          k:= getran(j);
          if k>j then
            begin (* MESSAGE DOES NOT FIT *)
              j:=j-1; (* TRY NEXT NODE *)
              done := j=0;
            end
          else
            begin (* MESSAGE FITS *)
              j:=k; done:=true
            end;
          end;
        end;
      end;

```

```
        (* UNTIL END OF CYCLE REACHED OR *)
        (* A MESSAGE IS INSERTED          *)
    until done;
    end; (* END CYCLE or MAXIMUM REUSE REACHED *)
end; (* NEXT ITERATION (CYCLE) *)
writeln('iterations = ',iter);
writeln('messages = ',messages);
writeln(fl,n,' ',iter,' ',messages);
writeln('Any More (Y/N)'); readln(infl,rsp);
until (rsp='N') or (rsp='n');
end.
```

Bibliography

- [1] L. Bhuyan, D. Ghosal, and Q. Yang. Approximate analysis of single and multiple ring networks. *IEEE Transactions on Computers*, 38(7):1027–1040, July 1989.
- [2] P.T. Brady. A technique for investigating on-off patterns of speech. *Bell System Technical Journal*, 44, January 1965.
- [3] W.E. Burr. The FDDI optical data link. *IEEE Communications*, 24(5):18–23, May 1986.
- [4] W. Bux and M. Schlatter. An approximate method for performance analysis of buffer insertion rings. *IEEE Transactions on Communications*, COM-31:50–55, 1983.
- [5] S. Casale, V. Catania, A. Faro, N. Parchenkov, and L. Vita. Design and performance evaluation of an optical fiber LAN with double token rings. *Computer Communications*, 12(3):159–166, June 1989.
- [6] Liam Casey. Channel allocation in fddi ii. April 1986.
- [7] I. Chlamtac and O. Gynz. Performance models of asynchronous multitrunk hyperchannel networks. *IEEE Transaction on Computers*, C-36, no.2:138–146, Feb. 1987.
- [8] David D. Clark, Van Jacobson, John Romkey, and Howard Salwen. An analysis of TCP processing overhead. *IEEE Communications Magazine*, 23–29, June 1989.
- [9] P. Cochrane and M. Brain. Future optical fiber transmission technology and networks. *IEEE Communications*, 45–60, November 1988.

- [10] Digital, Intel, and Xerox. *The Ethernet, a local area network. Data link layer and physical layer specifications*. Technical Report, Digital, Intel and Xerox, Sept. 30.
- [11] Dobosiewicz and Gburzynski. Ethernet with segmented carrier. *Computer Networkin Symposium*, 72-78, April 1988.
- [12] W. Dobosiewicz and P. Gburzynski. Ethernet with segmented carrier. *Computer Networking Symposium*, 72-78, April 1988.
- [13] D. Dykeman and W. Bux. Analysis and tuning of the FDDI media access control protocol. *IEEE Transactions on Communications*, 6(6):997-1010, July 1988.
- [14] G. Ennis and P. Filice. Overview of a broad-band local area network protocol architecture. *IEEE Journal on Selected Areas in Communications*, SAC-1, No.5:832-841, November 1983.
- [15] G.M. Exley and L.F. Merakos. Throughput - delay performance of interconnected CSMA local area networks. *IEEE Journal on Selected Area in Communications*, 1380-1390, Dec. 1987.
- [16] R. Falconer and L. Adams. Orwell: a protocol for an integrated services local network. *British Telecom. Technical Journal*, 3(2):85-91, April 1985.
- [17] E.C. Foudriat, K. Maly, C.M. Overstreet, S. Khanna, and F. Paterra. *A Carrier Sensed Multiple Access Protocol for High Data Rate Rings*. Technical Report, Old Dominion University, February 1990.
- [18] T.B. Fowler and S.F. Holmgren. A wideband cable bus local area network. *COMP-CON*, 405-414, Fall 1982.
- [19] M. Garrett and S.-Q. Liu. A study of slot reuse in dual bus multiple access networks. *Proceedings of IEEE Infocom 90*, 617-629, June 1990.

- [20] T.A. Gonsalves and F.A. Tobagi. Comparative performance of voice/video/data local area networks. *IEEE Journal on Selected Areas in Communications*, SAC-7, No.5:657–669, June 1989.
- [21] D. Greaves, D. Lioupis, and A. Hopper. The cambridge backbone ring. *Proceedings of IEEE Infocom 90*, 8–14, June 1990.
- [22] IEEE 802.6 Working Group. Proposed standard: distributed queue dual bus MAN. August 7, 1989.
- [23] E. Hahne, A. Choudhury, and N. Maxemchuk. Improving the fairness of distributed-queue-dual-bus networks. *IEEE Infocom 90*, 175–184, June 1990.
- [24] M. Hatamian and E.G. Bowen. Homenet: a broadband voice/data/video network on a CATV system. *AT&T Technical Journal*, 64:347–367, February 1985.
- [25] W. Hilal and M.T. Liu. Analysis and simulation of the register-insertion protocol. *Proceedings, Computer Network Symposium*, 1982.
- [26] H.S. Hinton. Architectural considerations for photonic switching networks. *IEEE Transactions on Communications*, 6(7):1209–1226, August 1988.
- [27] A. Hopper. The cambridge ring - a local network. *Advanced Techniques for Microprocessor Systems*, 67–71, 1980.
- [28] E. Huang and L. Merakos. On the access fairness of the DQDB MAN protocol. *Proceedings of Phoenix Conference on Computers and Communication 90*, 556–559, March 1990.
- [29] R. Jain. Performance analysis of FDDI token ring networks: effect of parameters and guidelines for setting TTRT. *SIGCOMM '90*, 264–275, 1990.
- [30] M.J. Johnson. Proof that timing requirements of the FDDI token ring protocol are satisfied. *IEEE Transactions on Communications*, 35(6):620–625, June 1987.

- [31] R.E. Kahn. A national network: today's reality, tomorrow's vision, part 2. *EDUCOM Bulletin*, 14-21, Summer/Fall 1988.
- [32] A. Kamal. On the use of multiple tokens on ring networks. *Proceedings of IEEE INFOCOM '90*, 15-22, June 1990.
- [33] M.J. Karol. Optical interconnection using shufflenet multihop networks in multi-connected ring topologies. *SIGCOMM Symposium*, 25-33, August 1988.
- [34] D. Karvelas and A. Leon-Garcia. Performance of integrated packet voice/data token-passing rings. *IEEE Journal on Selected Areas in Communications*, SAC-4, No 6:823-832, Sept. 1986.
- [35] H. Kume. *Ethernet Evaluation in Fuji Xerox*. Technical Report, Software Tech. Center, Fuji Xerox Co. Ltd., Japan, September 1985.
- [36] E.S. Lee, P.I.P. Boulton, and B.W. Thompson. HUBNET performance measurement. *IEEE Journal on Selected Areas in Communication*, 6(6):1025-1032, July 1988.
- [37] J.O. Limb. Performance of local area networks at high speed. *IEEE Communications*, 22(1):41-45, August 1984.
- [38] J.O. Limb and C. Flores. Description of fasnet - a unidirectional local-area communications network. *Bell System Technical Journal*, 61(7):1413-1440, September 1982.
- [39] M. Maeda and H. Nakano. Integrated optoelectronics for optical transmission systems. *IEEE Communications*, 26(5):45-51, May 1988.
- [40] K. Maly, E. Foudriat, D. Game, R. Mukkamala, and C.M. Overstreet. Dynamic allocation of bandwidth in a multi-channel metropolitan area network. *submitted to Computer Networks and ISDN*, September 1989.
- [41] K. Maly, E. Foudriat, C. M. Overstreet, and R. Mukkamala. Private communication. 1990.

- [42] K. Maly, E.C. Foudriat, D. Game, R. Mukkamala, and C.M. Overstreet. Traffic placement policies for a multi-band network. *SIGCOMM 89*, August 1989.
- [43] K. Maly, D. Game, L. Zhang, E. Foudriat, and S. Khanna. *Fairness Problems at the MAC Level for High-Speed Networks*. Technical Report, Old Dominion University, February 1990.
- [44] K. Maly, C. M. Overstreet, X. Qiu, and D. Tang. Dynamic resource allocation in a metropolitan area network. *SIGCOMM Symposium*, 13–24, August 1988.
- [45] M.A. Marsan and D. Roffinella. Multichannel local area network protocols. *IEEE Journal on Selected Areas in Communications*, 885–897, Nov. 1983.
- [46] N.F. Maxemchuk. Twelve random access strategies for fiber optic networks. *IEEE Transactions on Communications*, 36(8):942–950, August 1988.
- [47] N.F. Maxemchuk. A variation on CSMA/CD that yields movable TDM slots in integrated voice/data local networks. *Bell Systems Technical Journal*, 61, no 7:1527–1550, September 1982.
- [48] R.M. Metcalfe and D.R. Boggs. Ethernet: distributed packet switching for local computer networks. *Commun. ACM*, 19, July 1976.
- [49] T. Nakagami and T. Sakurai. Optical and optoelectronic devices for optical fiber transmission systems. *IEEE Communications*, 26(1):28–33, January 1988.
- [50] A.N. Netravali and Z.L. Budrikis. A broadband local area network. *Bell System Technical Journal*, 64, no. 10:2449–2465, December 1985.
- [51] R.M. Newman, Z.L. Budrikis, and J.L. Jullett. The QPSX man. *IEEE Communications Magazine*, 26(4):20–28, April 1988.
- [52] R.M. Newman, Z.L. Budrikis, and J.L. Jullett. The QPSX man. *IEEE Communications Magazine*, 26(4):20–28, April 1988.

- [53] K. Nosu. Advanced coherent lightwave technologies. *IEEE Communications*, 26(2):15–21, February 1988.
- [54] M.W. Rahm. Cablenet: a user perspective. *COMPCON*, 337–342, Fall 1982.
- [55] M. Rodrigues. Erasure node: performance improvement for the IEEE 802.6 MAN. *Proceedings of IEEE Infocom 90*, 636–643, June 1990.
- [56] F. Ross. FDDI - a tutorial. *IEEE Communications*, 24(5):10–17, May 1986.
- [57] S. Sharrock, K. Maly, S. Ghanta, and H. Du. A broadband integrated voice/data/video network of multiple LANs with dynamic bandwidth partitioning. *INFOCOM '87*, 417–425, March 1987.
- [58] S. Sharrock, K. Maly, S. Ghanta, and H. Du. A framed, movable-boundary protocol for integrated voice/data in a LAN. *SIGCOMM '86*, 9 Pages, August 1986.
- [59] J. F. Shock and J. Hupp. Measured performance of an Ethernet local network. *Communications of the ACM*, 711–721, December 1980.
- [60] P.W. Shumate. Optical fibers reach into homes. *IEEE Spectrum*, 43–47, Feb. 1989.
- [61] W. Stallings. *Business Data Communications*. MacMillan Publishing Co., 1990.
- [62] W. Stallings. *Data and Computer Communications*. MacMillan Publishing Co., 1988.
- [63] Draft Proposed American National Standard. FDDI token ring media access control (MAC) ASC X3T9.5 rev. 10. February 28, 1986.
- [64] M. Stephens. Plastic optical-fiber LAN unveiled. *Infoworld*, 11, January 16, 1989.
- [65] M. Stephens. Security of fiber-optic networks in doubt. *Infoworld*, 11, January 9, 1989.
- [66] F.A. Tobagi, F. Borgonovo, and L. Fratta. Expressnet: a high-performance integrated-services local area network. *IEEE Journal on Selected Areas in Comm.*, SAC-1, no. 5:898–913, November 1983.

- [67] F.A. Tobagi and M. Fine. Performance of unidirectional broadcast local area networks: expressnet and fasnet. *IEEE Journal on Selected Areas in Comm.*, SAC-1, no. 5:913–925, November 1983.
- [68] J. Ulm. A timed token protocol for local area network and its performance characteristics. *IEEE*, Feb. 1982.
- [69] A. Valenzano, P. Montuschi, and L. Ciminiera. Some properties of timed token medium access protocols. *IEEE Transactions on Software Engineering*, 16(8):858–869, August 1990.
- [70] S. Wagner and H. Kobrinski. WDM applications in broadband telecommunications networks. *IEEE Communications*, 22–30, March 1989.
- [71] S.L. Wallach. FDDI tutorial: LAN industry gets another standard. *LAN Magazine*, 44–47, March 1987.
- [72] J. Wilson and J.F.B. Hawkes. *Optoelectronics: An Introduction*. Prentice Hall, 1983.
- [73] P.C. Wong and T.S. Yum. Design and analysis of a contention - based lookahead reservation protocol on a multichannel local area network. *IEEE Transactions on Communications*, 36, no.2:234–238, Feb. 1988.
- [74] M. Xu and J. Herzog. Concurrent token ring protocol. *Proceedings of IEEE INFOCOM '88*, 145–154, March 1988.
- [75] H. Yasuda, F. Kanaya, and H. Kawanishi. 1.5444-mbits/s transmission on tv signals by interframe coding system. *IEEE Transactions on Communications*, 1175–1180, October 1976.
- [76] M. Zafirovic-Vukotic, I.G. Niemegeers, and D.S. Valk. Performance analysis of slotted ring protocols in HSLANs. *IEEE Transactions on Communications*, 6(6):1011–1024, July 1988.

Autobiographical Statement

David Earl Game

I was born on June 27, 1952 in Newport News, Va. Except for the time which I spent in Massachusetts as an undergraduate from 1970 to 1973, I have resided in Newport News for my entire life. I have a wife, Betsy, and three children, Melissa, Alison and Aaron, ages eleven, seven and three respectively.

I have two previous degrees as follows:

- B.S., Mathematics, Massachusetts Institute of Technology, May 1973.
- M.S., Applied Science, College of William and Mary, May 1978.

Following are a list of my published articles:

- *Traffic Placement Policies for a Multi-band Network*, Maly, Foudriat, Game, Mukkamala, Overstreet, SIGCOMM 89, August 1989.
- *Performance Analysis for High-Speed Protocols*, Game, Maly, 21st Annual Pittsburg Conference on Modelling and Simulation, University of Pittsburg, May 4-5, 1990.
- *Improvements for the Next Generation of FDDI*, Game, Maly, accepted for Phoenix Conference on Computers and Communication, March 1991.

When I was a graduate student at the College of William and Mary (1977), I was supported through a research assistantship at ICASE, NASA Langley, Hampton, Va. Upon graduating from William and Mary in 1978, I was hired as an Instructor of Computer Science at Christopher Newport College in Newport News, Va. I have subsequently been promoted to an Assistant Professor and have served for a two year period as Chairman of the Department of Computer Science, where I continue to be a member of the faculty. Funding for the research done in this thesis was through a research assistanship from NASA Langley, Sun Microsystems and the Center for Innovative Technology (CIT).