AD-A283 243

ATION PAGE

Form Approved
OMB No. 0704-0188

average 1 hour per response, including the time for reviewing instructions, searching existing data sources, ng the collection of information. Send comments regarding this burden estimate or any other aspect of this to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

DATE	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Performance modeling and analysis of Parallel Processing and Low Earth Orbit Satellite Communications Systems

5. FUNDING NUMBERS

6. AUTHOR(S)
Dr. Nathaniel J Davis, IV, Chairman

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

8. PERFORMING ORGANIZATION REPORT NUMBER
94-027D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STREET
WRIGHT-PATTERSON AFB OH 45433-7765

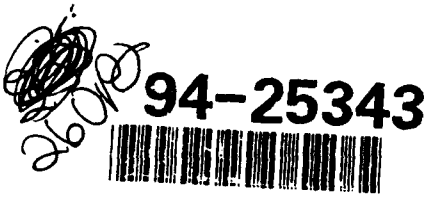
10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distribution Unlimited
MICHEAL M. BRICKER, SMSgt, USAF
Cheif Administration

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)



14. SUBJECT TERMS	15. NUMBER OF PAGES 237
	16. PRICE CODE

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

NSN 7540-01-280-5500

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Block 12b. Distribution Code.

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

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

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

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

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

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

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

94-027D

PERFORMANCE MODELING AND ANALYSIS OF PARALLEL PROCESSING AND LOW EARTH ORBIT SATELLITE COMMUNICATION SYSTEMS

by

Richard A. Raines

Dr. Nathaniel J. Davis, IV, Chairman

Electrical Engineering

(ABSTRACT)

This dissertation presents unique and valuable insight into the analysis of packet-switched data communication systems. The research described in this dissertation examines performance characteristics of two types of packet-switched data communication systems. The first system to be analyzed operates in a parallel processing environment where cooperating processors independently perform assigned tasks. In this environment, the packet delay performance is dominated by queuing delays. The second type of system examined operates in a low earth orbit (LEO) satellite communications network environment. In this type of network, delay performance is affected by both queuing and propagation effects.

The objectives of this research are to study the effects of queuing and propagation on the average packet delay, the number of buffers required to implement the networks that interconnect the parallel processors, and the satellite resource utilization rates. For both types of communication systems, mathematical metamodels [Agr85] are developed to capture the effects on packet delay caused by incremental changes in network dependent parameters.

Part I of this research performs average packet delay and buffer cost comparisons of the augmented shuffle exchange network (ASEN) and the multistage cube (MSC) network. It is shown that the packet delay associated with the ASEN is between 20 and 25 percent lower than that of a similar sized MSC network. In addition to the delay benefits of the ASEN, network implementation cost savings for the ASEN are shown to be 9 to 16 percent lower than the MSC.

Innovative mathematical design tools are developed and applied to the parallel processing interconnection network environment. These tools are used for predictive modeling of packet delay given network dependent parameters. The simple and concise models are shown to have predictive accuracy within 1 percent of the observed simulation delay results.

Part II of this research focuses on LEO satellite system communications. Six different constellations, providing whole-earth coverage are modeled and analyzed. The number of satellites within the constellations examined range from 36 to 77. The analysis of these LEO satellite systems consists of examining the packet delay characteristics of these dynamic systems as well how resource requests to the satellites are distributed. It is shown that when packet delay is the only design criteria, the differences in delay between the 36-satellite system and the 77-satellite are minimal and do not warrant the use the 77-satellite system over the 36-satellite system. The satellite resource utilization analysis captures the resource request load balancing characteristics of the systems. From a load balancing perspective, the 54-satellite system yields the best performance while the 36-satellite system the worst. A third unique aspect of the research presented in Part II is the application of metamodeling to the LEO satellite system environment. The metamodels developed reduce a complex 8-factor packet delay representation to simple, yet accurate, 2 and 3-factor relationships. These metamodel delay relationships are shown to have a predicted versus observed packet delay "best case" accuracy of 8 and 4 percent for the 3 and 2-factor models, respectively. Predicted versus observed packet delays are typically within 20 percent of agreement.

This research makes two contributions to the state-of-the-art knowledge in packet-switched communications system analysis. First, the metamodeling of interconnection networks and LEO systems are first of their kind. These models can allow for reduced simulation trials and more expeditious design decision making. The second contribution is the development of an integrated LEO satellite system model not seen in previously published research. This model can be used to further advance research in the LEO satellite system environment.

PERFORMANCE MODELING AND ANALYSIS OF PARALLEL PROCESSING AND LOW EARTH ORBIT SATELLITE COMMUNICATION SYSTEMS

by

Richard A. Raines

Dr. Nathaniel J. Davis, IV, Chairman

Electrical Engineering

(ABSTRACT)

This dissertation presents unique and valuable insight into the analysis of packet-switched data communication systems. The research described in this dissertation examines performance characteristics of two types of packet-switched data communication systems. The first system to be analyzed operates in a parallel processing environment where cooperating processors independently perform assigned tasks. In this environment, the packet delay performance is dominated by queuing delays. The second type of system examined operates in a low earth orbit (LEO) satellite communications network environment. In this type of network, delay performance is affected by both queuing and propagation effects.

The objectives of this research are to study the effects of queuing and propagation on the average packet delay, the number of buffers required to implement the networks that interconnect the parallel processors, and the satellite resource utilization rates. For both types of communication systems, mathematical metamodels [Agr85] are developed to capture the effects on packet delay caused by incremental changes in network dependent parameters.

Part I of this research performs average packet delay and buffer cost comparisons of the augmented shuffle exchange network (ASEN) and the multistage cube (MSC) network. It is shown that the packet delay associated with the ASEN is between 20 and 25 percent lower than that of a similar sized MSC network. In addition to the delay benefits of the ASEN, network implementation cost savings for the ASEN are shown to be 9 to 16 percent lower than the MSC.

Innovative mathematical design tools are developed and applied to the parallel processing interconnection network environment. These tools are used for predictive modeling of packet delay given network dependent parameters. The simple and concise models are shown to have predictive accuracy within 1 percent of the observed simulation delay results.

Part II of this research focuses on LEO satellite system communications. Six different constellations, providing whole-earth coverage are modeled and analyzed. The number of satellites within the constellations examined range from 36 to 77. The analysis of these LEO satellite systems consists of examining the packet delay characteristics of these dynamic systems as well how resource requests to the satellites are distributed. It is shown that when packet delay is the only design criteria, the differences in delay between the 36-satellite system and the 77-satellite are minimal and do not warrant the use the 77-satellite system over the 36-satellite system. The satellite resource utilization analysis captures the resource request load balancing characteristics of the systems. From a load balancing perspective, the 54-satellite system yields the best performance while the 36-satellite system the worst. A third unique aspect of the research presented in Part II is the application of metamodeling to the LEO satellite system environment. The metamodels developed reduce a complex 8-factor packet delay representation to simple, yet accurate, 2 and 3-factor relationships. These metamodel delay relationships are shown to have a predicted versus observed packet delay "best case" accuracy of 8 and 4 percent for the 3 and 2-factor models, respectively. Predicted versus observed packet delays are typically within 20 percent of agreement.

This research makes two contributions to the state-of-the-art knowledge in packet-switched communications system analysis. First, the metamodeling of interconnection networks and LEO systems are first of their kind. These models can allow for reduced simulation trials and more expeditious design decision making. The second contribution is the development of an integrated LEO satellite system model not seen in previously published research. This model can be used to further advance research in the LEO satellite system environment.

PERFORMANCE MODELING AND ANALYSIS OF PARALLEL PROCESSING AND LOW EARTH ORBIT SATELLITE COMMUNICATION SYSTEMS

by

Richard A. Raines

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

APPROVED:

Dr. N. J. Davis IV, Chairman

Dr. C. W. Bostian

Dr. T. Pratt

Dr. S. F. Midkiff

Dr. M. Abrams

June 1994

Blacksburg, Virginia

PERFORMANCE MODELING AND ANALYSIS OF PARALLEL PROCESSING AND LOW EARTH ORBIT SATELLITE COMMUNICATION SYSTEMS

by

Richard A. Raines

Dr. Nathaniel J. Davis, IV, Chairman

Electrical Engineering

(ABSTRACT)

This dissertation presents unique and valuable insight into the analysis of packet-switched data communication systems. The research described in this dissertation examines performance characteristics of two types of packet-switched data communication systems. The first system to be analyzed operates in a parallel processing environment where cooperating processors independently perform assigned tasks. In this environment, the packet delay performance is dominated by queuing delays. The second type of system examined operates in a low earth orbit (LEO) satellite communications network environment. In this type of network, delay performance is affected by both queuing and propagation effects.

The objectives of this research are to study the effects of queuing and propagation on the average packet delay, the number of buffers required to implement the networks that interconnect the parallel processors, and the satellite resource utilization rates. For both types of communication systems, mathematical metamodels [Agr85] are developed to capture the effects on packet delay caused by incremental changes in network dependent parameters.

Part I of this research performs average packet delay and buffer cost comparisons of the augmented shuffle exchange network (ASEN) and the multistage cube (MSC) network. It is shown that the packet delay associated with the ASEN is between 20 and 25 percent lower than that of a similar sized MSC network. In addition to the delay benefits of the ASEN, network implementation cost savings for the ASEN are shown to be 9 to 16 percent lower than the MSC.

Innovative mathematical design tools are developed and applied to the parallel processing interconnection network environment. These tools are used for predictive modeling of packet delay given network dependent parameters. The simple and concise models are shown to have predictive accuracy within 1 percent of the observed simulation delay results.

Part II of this research focuses on LEO satellite system communications. Six different constellations, providing whole-earth coverage are modeled and analyzed. The number of satellites within the constellations examined range from 36 to 77. The analysis of these LEO satellite systems consists of examining the packet delay characteristics of these dynamic systems as well how resource requests to the satellites are distributed. It is shown that when packet delay is the only design criteria, the differences in delay between the 36-satellite system and the 77-satellite are minimal and do not warrant the use the 77-satellite system over the 36-satellite system. The satellite resource utilization analysis captures the resource request load balancing characteristics of the systems. From a load balancing perspective, the 54-satellite system yields the best performance while the 36-satellite system the worst. A third unique aspect of the research presented in Part II is the application of metamodeling to the LEO satellite system environment. The metamodels developed reduce a complex 8-factor packet delay representation to simple, yet accurate, 2 and 3-factor relationships. These metamodel delay relationships are shown to have a predicted versus observed packet delay "best case" accuracy of 8 and 4 percent for the 3 and 2-factor models, respectively. Predicted versus observed packet delays are typically within 20 percent of agreement.

This research makes two contributions to the state-of-the-art knowledge in packet-switched communications system analysis. First, the metamodeling of interconnection networks and LEO systems are first of their kind. These models can allow for reduced simulation trials and more expeditious design decision making. The second contribution is the development of an integrated LEO satellite system model not seen in previously published research. This model can be used to further advance research in the LEO satellite system environment.

ACKNOWLEDGMENTS

This dissertation culminates a three year effort and completion of yet another chapter in my life. These three years in Virginia have seen a period of physical recovery, spiritual growth, and academic learning. During this time, many individuals have made direct impacts on my life and have assisted me in getting to where I am now.

First, I thank Dr. Nathaniel J. Davis, IV for accepting me as his Ph.D. student. His efforts in guiding this research, enduring the many questions I had, as well as the numerous revisions of technical papers and this dissertation are greatly appreciated. I also thank you Nat for being more than an advisor, for being a friend. I value this friendship as well as the working relationship we have established over the years.

I also thank my committee members, Drs. C. W. Bostian, T. Pratt, S. F. Midkiff, and M. Abrams for their knowledge and guidance during this effort. Each was always available for my questions and concerns. Special thanks go to Dr. Tim Pratt for teaching me satellite communications, setting me straight when needed, and always having the time to answer my many questions.

I cannot begin to thank my family enough for their love and support during these past three years. I promised Helen in 1987 that a Masters degree was my last, but broke that promise when accepting this assignment. Without her love and support, I know I could not have made it. Thank you Garrett and Kelley for understanding when I could not spend as much time with you as I would have liked to.

Lastly, and by no means least, I thank the good Lord for the many blessings he has bestowed upon me and my family. Through Jesus Christ, all things are possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This research examines performance characteristics of two types of communication systems. The first communication system operates in a parallel processing environment where queueing delays are the dominant factors affecting the performance of the system. The second communication system to be analyzed is a low earth orbit (LEO) satellite communications network. In this environment, signal propagation delays play an important role in the performance of the system. Analyses of both systems are conducted to determine average packet delay times associated with the respective networks in transmitting and receiving messages. Additionally, system cost metrics, to include allocated memory space, are evaluated for the interconnection networks. The modeling and analysis of the LEO systems also examines the satellite resource requests. Mathematical metamodels [Agr85], which characterize the system delay performance, are developed to aid the network analyst in determining optimal network designs for a chosen application.

The work to be presented in this dissertation effort is discussed in two parts. Part I examines network communication performance of interconnection networks for parallel processing systems. These types of networks are designed and developed to support independent processors whose numbers range from tens to thousands. The network supporting these processors is composed of switching elements used to route messages within the network and communication links (wires) between appropriate switches. Representative parallel processing systems are packaged in electronic cabinets of size on the order of a few square feet. Because of the close proximity of the processors, propagation times along the communication links are in the range of a few nanoseconds (10^{-9} seconds). Therefore, these systems are predominantly affected by the

delays incurred at the network switching elements as a message waits to be transmitted to the next switching element along its path from source to destination.

Part II of this effort analyzes the system performance of low earth orbit (LEO) satellite communication systems. These systems are composed of multiple satellites cooperatively working together to provide global communications for data applications. In an orbital system such as the ones to be examined by this effort, propagation delays, as well as queueing delays, play important roles in the overall system delay characteristics. With orbital altitudes on the order of a few hundred nautical miles, the minimum one-way propagation delay a message can expect to encounter is a few milliseconds (10^{-3} seconds). Queueing delays are encountered because of the unique nature of low earth orbiting system architectures. Proposed LEO systems [Mot90, LoC91] are envisioned to have on-board processing capabilities not available in their predecessor systems, the geostationary (GEO) systems. These processing capabilities are required to route messages from the originating site, through the multiple satellites within the constellation and finally to their predetermined destinations.

The motivation for this research results from system design goals to establish and provide reliable, efficient, and cost-effective communications in a real-time environment. Reliable communications can be achieved by designing the system hardware and software to have large mean-time-between-failures (MTBF) and also by providing redundant system hardware that can be utilized in the event of hardware failures within the system. Efficient communications result from the proper allocation and application of system resources (switches, memory, etc.) along with routing algorithm software that provides minimum paths (in terms of shortest distance or least congested) through the system from source to destination. Cost-effective communications deal with the design of the system so that an optimal (based on application) amount of system resources (switches, memory, satellites) are built into the system so that utilization rates are high, but also allow for growth in terms of scalability and increased functionality.

Part I of this research effort is composed of four chapters. Chapter 2 presents an introduction and background on the motivating factors that have led computer architects to develop parallel processing systems. Also included in Chapter 2 are the research goals of Part I. An overview of parallel processing systems is discussed in Chapter 3. This overview includes system classification taxonomies, interconnection network definitions and examples, switching methodologies, and representative systems implemented to take advantage of parallel processing capabilities. The modeling and analysis of interconnection networks for parallel processing is the topic of Chapter 4. This analysis includes the development of mathematical metamodels which characterize the delay performance of various network architectures. These metamodels are then applied to the Augmented Shuffle Exchange Network (ASEN), whose packet-switched model is developed for analysis and comparison with another interconnection network, the Multistage Cube (MSC) network. Part I concludes with Chapter 5, which contains a summary of the research efforts.

Part II, the modeling and analysis of low earth orbit satellite communication systems, consists of six chapters. Chapter 6 provides an introduction to the research and also motivating factors of the research. Chapter 7 discusses satellite communication parameters and the role each plays in establishing real-time communications. Previously published research related to satellite communication systems is the topic of Chapter 8. This chapter focuses on the research results obtained from efforts in satellite channel modeling and analysis, constellation modeling and analysis, and crosslink architectures of multiple satellite systems and how each of these apply to low earth orbit satellite systems. The modeling and analysis of low earth orbit satellite systems is addressed in Chapter 9. This discussion includes the choice of modeling tool, network operating assumptions, model development, and model validation. Chapter 10 presents the results of the modeling efforts to include delay performance characteristics and satellite resource utilization rates. Mathematical models are also developed to characterize the system delay. A summary of

research contained in Part II is discussed in Chapter 11. Concluding remarks are presented in Chapter 12.

PART I

MODELING AND PERFORMANCE ANALYSIS OF INTERCONNECTION NETWORKS FOR PARALLEL PROCESSING

CHAPTER 2

INTRODUCTION

2.1 Background

Increasing the performance of one's self or a human-made object has always been, and in all likelihood will always be, a goal which humanity strives to achieve. Performing tasks more efficiently and with greater speed have compelled humanity since the advent of fire. This innate drive has led modern day computer architects to examine ways of meeting societal processing needs.

Processing needs of society fall into two categories: real-time, and those tasks that do not require immediate results. Processing tasks such as medical diagnosis, weather prediction, speech recognition, entertainment, and defense systems require that computational results be obtained almost immediately after the start of the application. These types of tasks require an inordinate amount of data and analysis to arrive at a solution. Because of the computational complexity of the algorithms required to perform these tasks, real-time processing is, in most cases, not feasible in a uniprocessor environment.

Traditional uniprocessor systems perform tasks sequentially. Although Very Large Scale Integration (VLSI) techniques and overlapping and pipelining of instructions have reduced the execution time of single instructions, the computational requirements and complexities of certain large tasks exceed the capabilities of the fastest uniprocessor machines [Hay88]. Because of these

limitations, architects have developed parallel processing systems, composed of many cooperating processors, to meet the real-time processing requirements of contemporary applications.

Many concerns face the parallel processing systems designer. One of the major concerns lies in how to interconnect the multiple processors. This is not an easy decision since the designer must consider the application and the economics associated with determining the number of processors to be interconnected. Numerous interconnection schemes have been proposed to aid the designer in choosing an interconnection network [Fen81, AdS87]. Each of these schemes have common goals. First, to have the communication time between processors in the network be substantially less than the processing time required by an individual processor to execute the instruction or set of instructions. Second, the network should be robust to failures within the network. Third, the network should allow for expansion in the number of processors without degradation in required performance.

Additional design considerations must be taken into account when developing a parallel processing system interconnection network. The switching method in which packets (messages) traverse the network must be determined. Four basic choices exist: circuit switching (resources are dedicated to a given message for path setup and transmission); packet switching (resources are dynamically allocated and released to and from a packet as requested); virtual cut through (packet headers are examined and forwarded to the next channel along the transmission path before the receipt of the entire packet); and wormhole routing (similar to virtual cut through with the exception that blocked packets remain in the network instead of being buffered as in virtual cut through).

The network control strategy is another design factor that must be considered. By this, the method in which individual switches within the network are controlled must be determined. Two basic strategies exist: centralized control where all network switches are controlled by a single source and distributed control where individual switches have the control logic necessary to

determine appropriate settings. As with the choice of switching methodology, design tradeoffs exist in choosing the type of control strategy to use for the network.

The ability of a network to perform in the presence of hardware failures is of utmost importance in a parallel processing environment. Designers look to fault tolerant network architectures for performance robustness in faulty environments. This robustness is accomplished, in most cases, by the addition of redundant hardware which remains idle until faults are detected.

To facilitate the design and development of interconnection networks, system designers have turned to modeling a proposed system through simulation. Simulation is necessitated by the complexity of analytical approaches in determining the performance of large scale parallel processing systems. The emergence of commercial software simulation packages provides to the network architect a low-cost and time-efficient method for system modeling. In addition, simulation provides supplemental information that can be used to validate analytical models when they are available. Simulation allows for a wide range of comparative analyzes which are normally not possible through analytic modeling.

2.2 Research Goals

The focus of this research is two-fold. First, to extend the body of knowledge relative to interconnection network modeling by analyzing the performance of the Augmented Shuffle Exchange Network [KuR87, KuR89] relative to the multistage cube network. The second aspect of the research to be discussed in Part I is to develop and apply mathematical models that can be used to predict the performance of a given interconnection network.

One of the figures of merit that this investigation concentrates on is the average delay incurred by a message as it traverses the network. By analyzing the delay experienced by messages entered into the network, for various network loading, determinations can be made as to the desirability of one network over another. Additional design insight is gained by analyzing the implementation costs of each of the networks. The cost parameters considered in this investigation

result from examining queue lengths of network switches for light to heavy network loading. Switch queue lengths are chosen so that the probability of overflow by any switch queue within the network is less than one percent.

Another important figure of merit of this research keys on accurately predicting the delay performance of a chosen network. While analytic models exist for classic queuing theory models (e.g., M/M/1, M/G/1), predictive models for massively parallel systems do not. Simple and concise regression models are developed to gauge analytically the relative performance of one network versus another. These models prove to be beneficial in allowing the network designer to reduce the amount of simulations necessary in analyzing a system's performance over a wide range of network sizes.

2.3 Summary

This chapter has presented an overview of the processing problems that presently face the computer architect. As a result of encountering these obstacles, designers have proposed and developed multiprocessor systems to meet the real-time processing demands of today's applications. In these multiprocessor systems, designers have proposed and implemented various strategies for interconnecting individual processors. Part I of this investigation analyzes the performance characteristics of the Augmented Shuffle Exchange Network (ASEN) relative to the multistage cube (MSC) network.

Chapter 3 presents an overview of parallel processing systems. Classification methodologies for parallel systems are discussed along with advantages and disadvantages of a particular taxonomy. This discussion is followed by an examination of interconnection networks which includes categorization, primary switching elements, control structures, and fault tolerant capabilities. A synopsis of switching methods for communication systems as applied to interconnection networks follows the presentation on interconnection networks. Chapter 3 concludes with a review of four contemporary parallel processing systems. The discussion of these

systems relates system characteristics to interconnection network type and switching methodology used to implement interprocessor communications.

Chapter 4 presents the results of two independent studies on the performance of interconnection networks for parallel processing. The first study develops mathematical relationships for predicting the delay of three types of interconnection networks. The second study analyzes the performance characteristics of the ASEN and then compares them to the MSC. Mathematical relationships developed in the first study are extended and then applied to the ASEN. Chapter 4 opens with the formulation of mathematical models for four interconnection network topologies (ASEN, MSC, single stage cube (hypercube), and Illiac IV mesh-type). These models are used predict the network packet delay associated with a particular network. A review of the previous work of Kumar [KuR87, KuR89] in the modeling and analysis of ASEN in a circuit switched environment is discussed next. Model derivations for the ASEN and the MSC follow. These derivations include discussions of network operating assumptions, network simulations, and the validation of simulation models. Comparisons of network packets delay and associated network costs are analyzed and compared for the ASEN and the MSC. The latter sections of Chapter 4 present the application of derived predictive models to the ASEN.

Chapter 5 concludes Part I of this dissertation effort with a summary of the work presented. Recommendations for extensions to this research are also discussed.

CHAPTER 3

AN OVERVIEW OF PARALLEL PROCESSING SYSTEMS

3.1 Introduction

This chapter presents the basic characteristics of parallel processing systems. This was written to familiarize the reader with contemporary concurrent processing systems and the architectural issues which face the design engineer. Section 3.2 briefly discusses classification techniques used to categorize parallel processing systems. Section 3.3 describes interconnection networks, their functions, classifications, and fault-tolerance capabilities. Switching methodologies for parallel processing systems are presented in Section 3.4. Chapter 3 concludes with a survey of contemporary parallel processing systems presented in Section 3.5.

3.2 Classification of Parallel Processing Systems

One of the major issues that has plagued computer architects lies in how to accurately classify a computer system. Many taxonomies [Fly66, Fen72, Hän77] have been formulated to provide descriptive tools at the global level. Each taxonomy provides categorization insight, but fails to provide a complete classification tool. Feng's taxonomy [Fen72] attempts to compare computer systems by computing their degree of parallelism. The *degree of parallelism* represents the maximum number of bits per unit time that the system can process. The key disadvantage of Feng's approach is its lack of information related to the overall system structure. Händler's taxonomy [Hän77] is an extension of Feng's in that the degree of pipelining as well as the degree of parallelism are identified. The *degree of pipelining* describes the system's ability to decompose a process into distinct processes which may be executed in an overlapped manner. As with Feng's taxonomy, Händler's lacks in providing overall system representation. The taxonomy of Flynn [Fly66] provides the system structure the other two taxonomies do not but does not provide the

parallelism detail given by Feng and Händler. Flynn's taxonomy provides a compact and simple approach for classifying systems. Because of this simplicity, it is probably the most universally accepted and used of the taxonomies referenced above.

Michael Flynn, in the *Proceedings of the IEEE*, 1966, proposed a method for classifying computer systems based on the number of instruction and data streams associated with a system. A *stream* refers to a sequence of items, either instructions or data, that are executed or operated upon by a processor contained in the machine. Flynn proposed that any computer system could be classified into one of four categories, based on his concept of streams. These categories are:

- Single Instruction stream - Single Data stream (SISD)
- Single Instruction stream - Multiple Data stream (SIMD)
- Multiple Instruction stream - Single Data stream (MISD)
- Multiple Instruction stream - Multiple Data stream (MIMD)

The four classification categories of Flynn are shown in Figure 3.1.

The SISD machine represents the traditional von Neumann architecture. This architecture is characterized by a single processor which uses a single instruction stream and a single data stream for its operation. The MISD machine is characterized by multiple instruction streams, supplied by multiple processor, which operate on a single data stream. In theory, a MISD machine's multiple processors operate on a single stream of data. At present, no true MISD machine exists. Both the SIMD and MIMD machines are considered to be parallel processing systems. A SIMD machine, such as the Illiac IV [BaB68], is characterized by a single instruction stream. This single stream is spawned to each processor for operation on data local to the processor. The processors of a SIMD machine operate in "lock-step" (sequential) instruction execution. Each processor in the system is required to report completion of the instruction

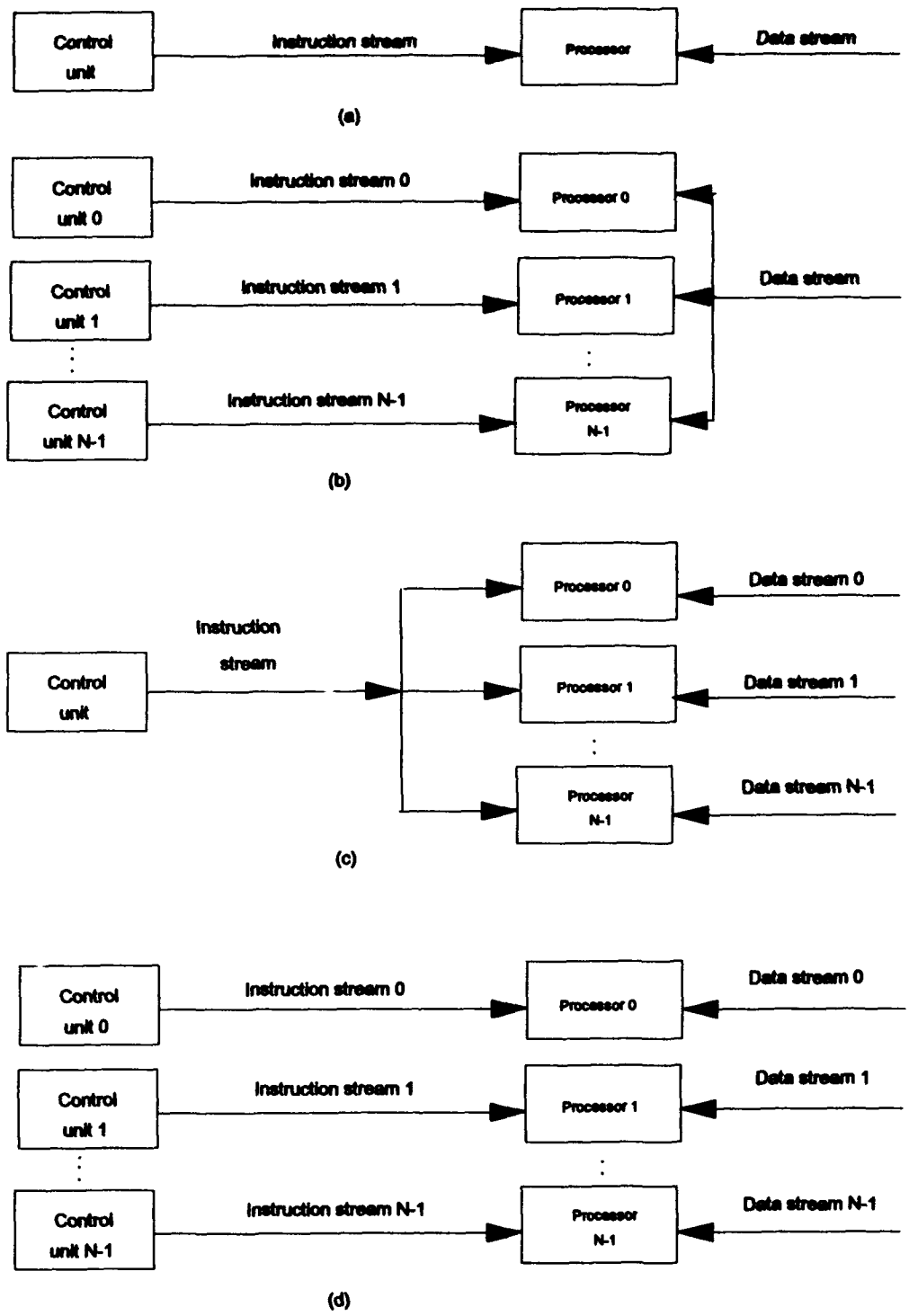


Figure 3.1. Flynn's taxonomy. (a) SISD; (b) MISD; (c) SIMD; (d) MIMD.

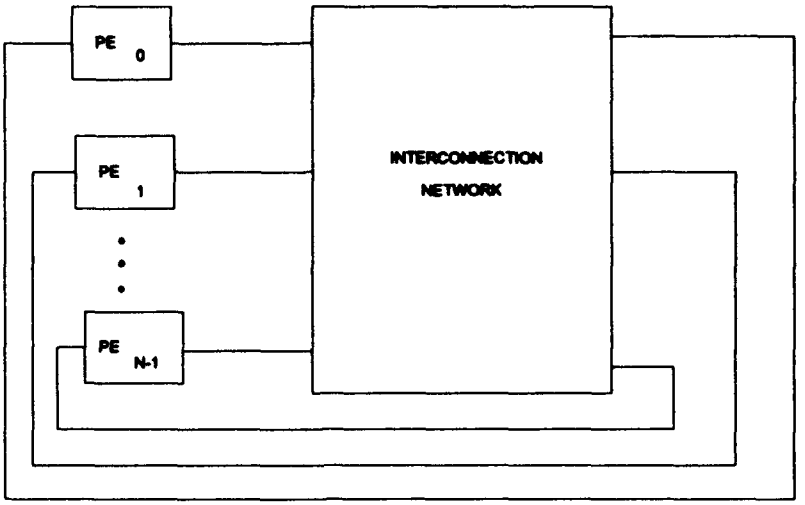
execution before the next instruction is issued by the central controller. An MIMD machine, such as the Intel Hypercube [Int86], is truly parallel in nature. Each MIMD processor is assigned an instruction stream thus allowing for concurrent operation and execution of instructions.

3.3 Interconnection Networks

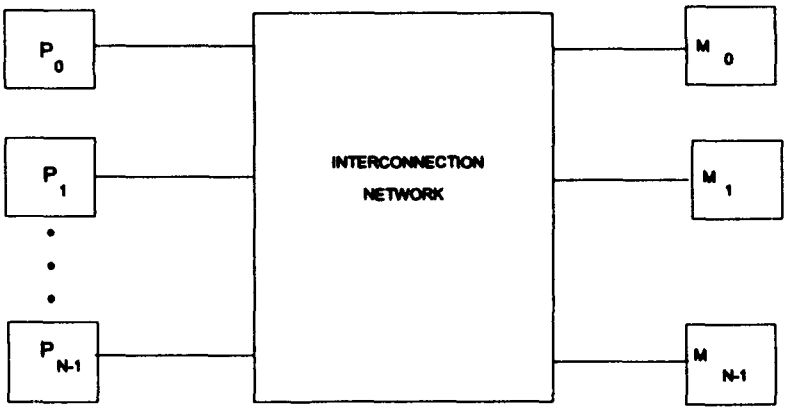
When two or more processors are to communicate with one another, communication connections must be established. How well these processors communicate with one another depends on the topology of the network and the interprocessor communication switching methodology. This section presents an overview of the different topologies of interconnection networks.

From a network point of view, parallel processing systems can be grouped into one of two architectural categories: processor-to-memory (P-M) or processing element-to-processing element (PE-PE, where a PE is a processor memory pair). Processor-to-memory architectures use bi-directional networks to connect processors to memory modules. Processor-to-memory architectures are characterized by heavy network loading which results from interprocessor communications and memory accesses across the network. In a PE-to-PE architecture, the network is uni-directional and provides inter-PE communications only. The PE-to-PE architecture differs from the P-M architecture in that no commonly accessible memory modules exist. As a result, the network loading is less in a PE-to-PE system than in a P-M system. Figure 3.2 shows the PE-to-PE and P-M architectures.

Interconnection networks can range from simple and inexpensive to complex and cost prohibitive. The simplest method (logically) for interconnecting multiple processors is the ring structure. The complexity of the ring, O(n), is proportional to the number of processors in the ring.



(a)



(b)

Figure 3.2. Parallel processing system network viewpoint. (a) PE-to-PE system architecture; (b) Processor-to-Memory system architecture.

The ring interconnection network forms a closed loop by connecting neighboring processors in a uni- or bi-directional ring. The communication time in the ring is a function of the number of processors in the ring. As a result, computer systems implementing a ring topology are severely

limited in the number of processors that can be connected. Therefore, the ring interconnection is feasibly applied only when the number of processors to connect together is small. On the other end of the complexity-cost spectrum is the crossbar switch shown in Figure 3.3. The crossbar switch has n inputs and n outputs. An n-by-n crossbar switch provides full-connectivity between the n inputs and the n outputs. The benefits of full-connectivity must be paid for in hardware complexity and cost. The high cost of a crossbar switch results from a circuit complexity which is proportional to the square of the number of processors and memory devices connected to its input and output ports.

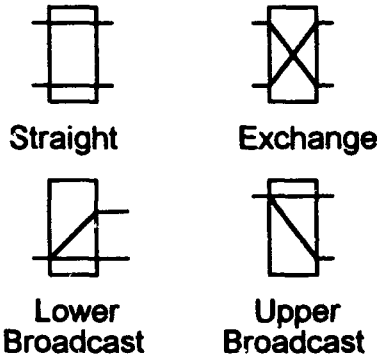


Figure 3.3. Four settings of a 2-by-2 crossbar switch.

To overcome the ring's inherent scalability restrictions and the crossbar hardware complexity costs, network designers have researched and developed two classes of interconnection networks which allow for large numbers of processors while, at the same time, achieve relatively low interprocessor communication times. The two classes of interconnection networks are the single-stage (direct) and the multistage (indirect) interconnection networks. Single-stage networks use point-to-point links to connect processing elements. Multistage networks consider the network to be a separate entity comprised of switching elements (i.e., crossbar switches of relatively small

size (32-by-32 or smaller)). Processing elements are connected to the inputs and the outputs of the network. The inputs and outputs of the network are the input and output ports of the switching elements.

3.3.1 Single-Stage Networks Numerous research efforts [BaB68, Sto71, Sie90, HsY87, RaD88, Sco92] have focused on the development and analysis of single-stage or direct networks. These networks have the ability to reconfigure their interconnection links based on the implementation of their interconnection function. Two of the most popular single-stage networks analyzed are the Mesh and the Cube networks. The forerunner of the Mesh networks is the Illiac IV machine designed in the late 1960s and early 1970s at the University of Illinois. The Illiac IV is a SIMD machine housing 64 processing elements (PEs). Each PE operates in lock-step. Information passing is performed via nearest-neighbor communications. Figure 3.4 shows a 16-PE Illiac network. Each box shown in Figure 3.4 represents a 5-by-5 crossbar switch. Four inputs and outputs of the crossbar switch are used to connect to the north, south, east, and west neighbors of a particular PE. The remaining input/output pair of the crossbar is used for communications with the PE local (associated) to the switch. The numbers on each box indicate the logical address of the PE associated with a particular switch. The interconnection of logical PE addresses are determined by the four functions described in [BaB68].

The second single-stage network predominantly analyzed is the Cube network [Sie77]. This interconnection network's name is derived from the physical interconnection pattern of the processing elements. The size of a Cube network is usually given by its *dimension*. The dimension m, is $\log_2(N)$, where N is the number of PEs in the system . As an example, consider the 8-PE Cube shown in Figure 3.5. The PEs are located on the vertices of the Cube. Processing elements are interconnected via crossbar switches. For an m-dimensional Cube, $(m+1)-by-(m+1)$ crossbar switches are used. As discussed for the Illiac IV network, one input/output pair in the switch is used by the local PE and the other m pairs used to connect to m

neighbors. Determination of neighboring PE addresses is achieved by performing the $Cube_k$ function. The $Cube_k$ interconnection function pairs PEs whose addresses differ only in the k^{th} bit position (e.g., $Cube_2$ pairs PEs 6 (110) and 2 (010) shown in Figure 3.5).

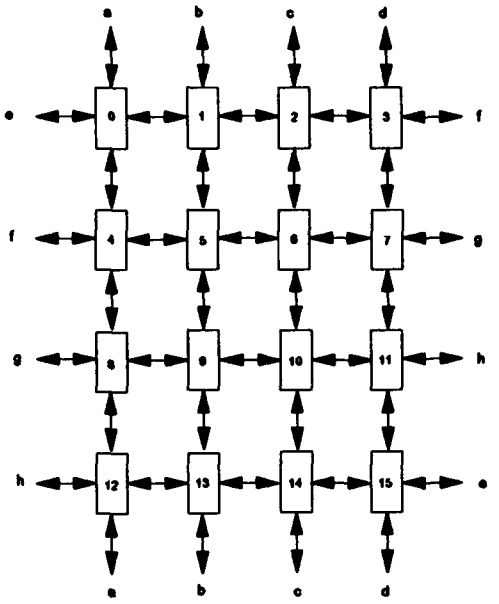


Figure 3.4. The Illiac IV mesh network, N = 16 processors.

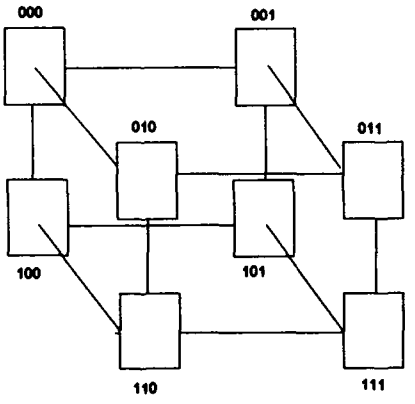


Figure 3.5. The Cube interconnection network, m = 3, N = 8.

3.3.2 Multistage Networks Multistage networks can be described by three characterizing features: the *switching element*, the *network topology*, and the *control structure* [HwB84].

3.3.2.1 Switching Element The *switching element* most commonly used in multistage networks is the crossbar switch. As mentioned above, the crossbar allows for all possible permutations between its input and output ports. A multistage network with N PEs will contain at least $\log_p(N)$ stages, where p is the size of the crossbar switch. Each stage of the network will consist of N/p crossbar switches. As an illustrative example, consider the 8-PE Multistage Cube (MSC) network [McS81, Sie90] shown in Figure 3.6. Each stage of the MSC implements the Cube_i function described in Section 3.3.1 above. At stage_i, the address lines grouped at a particular switch are those that differ in the i^{th} bit position.

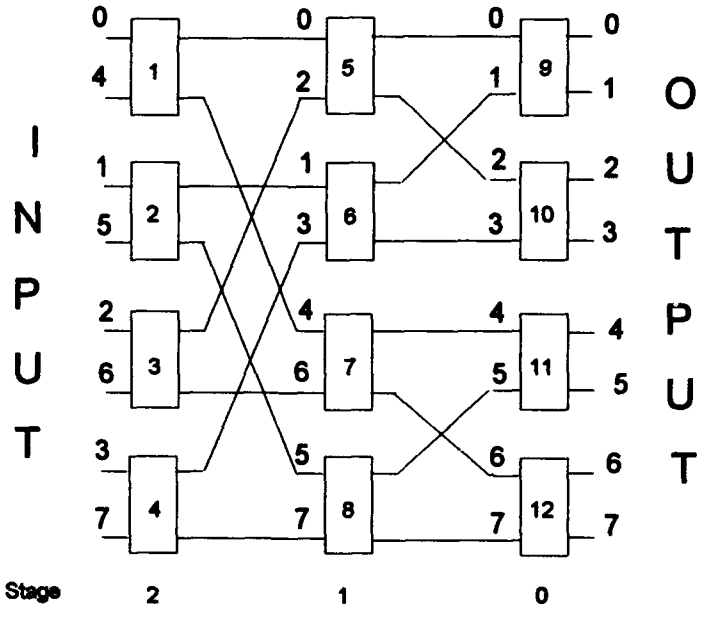


Figure 3.6. The 8-PE multistage cube network.

3.3.2.2 Network Topology The *network topology* of a multistage network can be classified into two categories: one-sided or two-sided networks. One-sided networks are those whose input and output ports are one the same side of the network. Two-sided networks, on the other hand, have inputs on one side of the network and outputs on the other. Two-sided networks contain three distinct classes: *blocking, rearrangeable,* and *nonblocking* [Fen81]. Blocking networks are those in which a connection cannot be established between an input and output pair because an interconnection link is busy servicing another input/output pair. The Data Manipulator [Fen81], Omega [Law75], and Indirect Binary n-cube [Pea77] are examples of blocking networks. Rearrangeable networks [Ben65] perform all possible combinations of input/output connections by rearranging existing connection to allow for new input/output connections. Nonblocking networks are those which contain one-to-one connections between every input and output port. The crossbar is an example of a nonblocking network.

3.3.2.3 Control Structure The *control structure* determines how the network switching elements are to be controlled. Two control strategies exist: *centralized* and *distributed. Centralized* control uses a single control unit to inform individual switching boxes of their routing setting. A *distributed* control strategy relies on individual boxes to determine the routing of the message. Determining which control strategy to use requires a tradeoff analysis. The key advantages of centralized control are constant path set-up and simple logic at each switching box. Disadvantages of centralized control are (1) only one message can be routed at any instance of time, thereby serializing the network access and (2) the central controller becomes a single point of failure for the entire network. The main disadvantage of distributed control is the complexity of the switching logic at individual switch boxes. Though switch boxes in a distributed control system are more complex than in a centralized control system, multiple messages can be routed simultaneously yielding no bottleneck effects.

3.3.3 Fault Tolerant Multistage Interconnection Networks One of the key issues in contemporary computer systems architectural design is the network performance in a faulty environment. As the number of processing elements in parallel processing systems increases, so does the probability of a system failing or having reduced performance due to hardware (switches, interconnection links) failures or faults. To overcome possible system performance degradations, network designers have proposed numerous networks which have the ability to withstand various levels of system failures.

Many multistage interconnection network research efforts have focused on Shuffle-Exchange [Sto71] and topologically equivalent networks [CrG85, GoG83, PfB85, DaS86, RaD88] . Each of the networks studied provide unique paths from the input PE to the output PE. These networks provide throughput necessary to meet real-time application processing requirements, but fail to be robust when network component failures are encountered. As an illustrative example, consider the 8-PE multistage cube network shown in Figure 3.6. Consider a desired communication link between PEs 2 and 6. The path would be through switches 3, 7, and 12. Now consider a failure occurring in either switch number 7 or in the lower output line coming out of the switch. Either one of these failures makes it impossible to route a message from PE 2 to PE 6. Therefore, a failure in this type of network can be catastrophic to interprocessor communications. To overcome these communication obstacles, network designers have proposed systems which maintain robustness in the presence of failures within the network.

A system is considered to be fault tolerant if its primary functions can be maintained in the presence of a hardware and/or software failure (fault). Designers of fault tolerant systems have two approaches for dealing with faults: either eliminate all possible faults during the design and manufacturing phase of the system; or implement redundancy in the system. The former option is seldom achieved. Therefore, the designer must compensate for possible failures by providing auxiliary hardware and/or software to be used in the event of failure by the primary components.

Fault tolerance of interconnection networks, as defined by [AdA87], relates to a chosen fault-tolerance model which consists of two parts: the fault model and the fault-tolerance criterion. The fault model describes the types of faults and how the faults occur. The fault-tolerance criterion states the operating requirements that must be met by the network when a fault is encountered (i.e., can the network still function without extreme degradation given a fault(s) encountered). Networks can have different levels of fault-tolerance. A network that can withstand a single fault and still meet its specified fault-tolerance criterion is considered to be *single-fault tolerant*. If a network can still meet the criterion after encountering m faults, then it is considered to be an *m-fault tolerant* network.

Many different design approaches to fault-tolerant multistage interconnection network architectures have been proposed. One of the more popular fault-tolerant design techniques is to provide extra stages within the network and dynamically reconfigure the network when a fault is encountered [AdA87, KuR87, KuR89]. These extra stages can take the form of either a set of crossbar switches or a set of multiplexers. Networks employing extra stages increase their fault tolerance levels over non-fault tolerant networks from 0 to b-1, where b is the size of the crossbar switch being used to implement the network. The main disadvantage of these type networks is the overall utilization of network hardware. In the works referenced above, hardware added for fault tolerance capabilities remains idle until a fault is detected by the network, leading to poor utilization of the circuitry devoted to fault tolerance. Researchers are examining ways of increasing this hardware utilization in non-faulty network environments. The research to be described in Chapter 4 explores the possibilities of using one such network, the Augmented Shuffle Exchange Network (ASEN) [KuR87, KuR89], in a fault-free network.

3.4 Switching Methods

For communication networks, two types of switching methodologies are used: *circuit switching* and *store-and-forward switching*. Each of these methods is described below.

3.4.1 Circuit Switching The circuit switching method is most commonly associated with communication systems which require dedicated resources for the anticipated duration of usage. The public telephone system is a prime example of the use of circuit switching. In this type of system, an independent virtual circuit is established between the originating source and the receiving destination. The virtual circuit remains established until the communication session is terminated by either the source or the destination. The advantage of using circuit switching lies in the dedicated resources. Communicating users (nodes, PEs) are guaranteed access to these resources for the duration of the session. Two major disadvantages to using circuit switching exist: currently utilized links cannot be accessed by others wishing to use the links simultaneously; and inefficiency of resource utilization. The former disadvantage blocks the establishment of circuits wishing to use only a portion of the circuit presently allocated to another session. The resource utilization inefficiency results from the *bursty* or random interarrival of messages wishing to use the link resources. Consider the following example from [BeG92]. Let λ denote the message arrival rate, $\overline{X}$ be the expected transmission time over the link, and T denote the expected allowable delay from message arrival at the source to delivery at the destination. Obviously, $\overline{X}$ will be less than T when the effects of queuing are included in $\overline{X}$. Thus, $\lambda\overline{X} < \lambda T$. So, if $\lambda T << 1$, the expected link utilization, $\lambda\overline{X}$, will be much less than 1 (100%). To overcome the inefficiencies of circuit switching, designers have looked to store-and-forward methods to improve communication network performances.

3.4.2 Store-and-Forward Switching In a store-and-forward communication system, each communication session is established without prior reservation of resources. As the information entity (message or packet) traverses the communication link from source to destination, it is either stored (buffered) at a switching element or forwarded to the next switching element along the path. Resources along the path are seized by the entity only for the time required to buffer or forward the entity. The process is repeated until the entity reaches its predetermined destination. Many types

of store-and-forward switching methods exist: *message switching*, *packet switching*, *virtual cut-through*, and *wormhole routing*. In a message switching system, the message is sent as one entity. The drawback to a system such as this is as messages become large, associated buffer space must be increased. Packet switching systems decompose the message into packets. The packets then traverse the network and are reassembled at the destination. Virtual cut-through or virtual circuit routing examines the packet header at transmission initiation and sets up a path from source to destination. Virtual cut-through is similar to circuit switching in that the path is maintained for the duration of the communication session, but differs in that it establishes a virtual path which can be shared by other packets wishing to utilize a particular resource. The link is utilized on a demand basis rather than fixed allocation. Wormhole routing is similar to virtual cut-through in the way the packet header is examined to determine the path. The two methods differ in how the packet is handled when a blockage occurs. Instead of buffering the packet, as is done in virtual cut-through, wormhole routing keeps the packet in the network until the blockage has been resolved. The advantage of store-and-forward switching over circuit switching is link utilization. A communication link (path) is fully utilized whenever there exists traffic wishing to use the link. The disadvantage of store-and-forward switching is the control of queuing delays associated with a particular network switch. In a store-and-forward switched implementation, individual switches can receive packets from multiple sources. Unless control mechanisms are in place to notify sending switches of excessive queueing delays, the problem compounds itself. Even if flow control mechanisms are in place, the time delay between a receiving switch notifying senders of congestion allows additional packets to be transmitted.

3.5 Contemporary Parallel Processing Systems

Contemporary parallel processing systems have been implemented to take advantage of the architectural advances made in the design of interconnection networks. These systems have been designed to provide computational speeds necessary to meet the response times of contemporary

applications. This section reviews four major systems which have either been implemented commercially or have been built solely for the purpose of research of present technology. The systems discussed represent diverse approaches to parallel processing systems design.

3.5.1 The Intel iPSC The Intel Personal Super Computer (iPSC) is a research-oriented MIMD machine. The iPSC architecture is more commonly referred to as the *hypercube* or *binary n-cube* based upon its interconnection network, a packet-switched implementation of a single-stage network. The hypercube processing elements are implemented with the i860, a 64-bit, million-transistor Reduced Instruction Set Computer (RISC) microprocessor. The hypercube architecture is scalable, with commercial systems ranging in size from 8 to 128 processing elements. Peak system performance ranges from 480 MFLOPS (8 processing elements) to 7.6 GFLOPS (128 processing elements) [Int92].

3.5.2 The IBM Research Parallel Processor Prototype (RP3) The RP3 is a MIMD machine designed strictly to research the hardware and software aspects of parallel processing [PfB85]. Incorporating much of the NYU Ultracomputer design [GoG83], The RP3 has been designed and constructed at the T.J. Watson Research Center. The RP3 contains 512 32-bit microprocessors. These 512 processors are grouped into eight modules with each module containing 64 processors.

The interconnection network of the RP3 consists of two separate networks; a multistage cube and a combining network which is used for interprocessor coordination functions. The RP3's interchange boxes are implemented using 4-by-4 crossbar switches. The interprocessor communications are a mixture of circuit- and packet-switching.

3.5.3 nCube 3 Supercomputer The nCube 3 is a MIMD supercomputer presently under development [DuB92]. nCube Corporation touts its product as providing programming flexibility in a multi-user, multi-application environment. The nCube 3 is envisioned to provide processor scalability, in 8 processor increments, and support as few as 8 processors to as many as 65,536

processors in a single system architecture. Proposed application performance levels vary from one GigaFLOP to more than 6.5 TeraFLOPS.

The interconnection topology of the nCube 3 is the hypercube structure. When fully equipped, this system forms a 16-dimensional hypercube. Cut-through routing allows the worst case latency between two communicating processors to be on the order of 10 microseconds. Addressing is based on a 48-bit address, thus allowing for over 256 TeraBytes of address space. Interprocessor data communication channels allow for peak data rates of 400 MegaBytes-per-second.

3.5.4 Intel Paragon™-XP/S The Paragon™-XP/S is Intel Corporation's next generation parallel supercomputer [Sco92]. Intel made an engineering decision to move away from the hypercube architecture and its $\log_2 N$ communication times to use a two-dimensional mesh for the Paragon. Though the worst case hop distances are longer in a mesh than in a hypercube, Intel proposes to overcome the increased distances with greater channel speeds. The proposed bi-directional channel communication rates are 200 MegaBytes-per-second.

Each processing node of the Paragon carries two processors: the applications processor, a 50-MHz i860XP; and a message passing processor, also an i860XP. The hardware structure of the Paragon does not limit the number of processors capable of functioning in the system.

3.6 Summary

In this chapter, an overview of parallel processing systems has been presented. Three methods for classifying parallel processing systems were discussed. Flynn's taxonomy which is widely recognized and is probably the most commonly used, provides a simple and compact way of classifying computer systems. Taxonomies of Feng and Händler provide architectural insight but fail to provide enough information to accurately describe the system.

Interconnection networks were defined and discussed from a functional point of view. Insight was provided on the cost/logic complexity tradeoffs associated with connecting multiple

communicating processors. A discussion on the two classes of interconnection networks, single-stage and multistage, was presented. This discussion provided representative examples of each class along with the architectural parameters associated with each. The limitations of non-fault tolerant multistage networks were presented along with a survey of fault tolerant multistage interconnection networks.

Switching methodologies were next presented to provide insight into another architectural issue which requires design tradeoffs. Circuit switched networks provide simple switching logic at each switching element but present possible problems by having a single point of failure in the network central controller. Store-and-forward networks eliminate this problem by having control decisions located at each switching element. Store-and-forward networks also utilize communication links more efficiently than circuit switched implementations.

As a culmination of the previous sections, Section 3.5 briefly examines four parallel processing systems. These systems demonstrate the varied approaches to providing efficient communications in a massively parallel environment.

CHAPTER 4

MODELING AND PERFORMANCE ANALYSIS OF THE AUGMENTED SHUFFLE EXCHANGE NETWORK AND THE MULTISTAGE CUBE NETWORK

4.1 Introduction

This chapter presents the results of independent studies on the performance of interconnection networks for parallel processing. The first study, performed by Shaw, Davis, and this author [ShD93], proposes mathematical models which accurately predict the network delay for three different network topologies. The second study, performed jointly by Ramachandran [Ram92] and this author, develops a simulation model for the Augmented Shuffle Exchange Network (ASEN). This model is used to determine the delay performance characteristics of the network along with the cost associated with implementing the network. Extensions to the metamodeling of interconnection networks are developed and applied to the ASEN for performance comparisons to the multistage cube network described in Chapter 3.

The concept of metamodeling is introduced in Section 4.2 with applications to predictive modeling of network delay. Mathematical models are developed and applied to the MSC, single stage cube, (hypercube), Illiac IV mesh-type network. Section 4.3 provides an overview of the ASEN proposed by Kumar [KuR87, KuR89]. Section 4.4 presents the ASEN and MSC model derivations along with rationale for determining the simulation tool chosen to model the networks. In addition, Section 4.4 discusses the network operating assumptions, the network simulations, and the validations of the network models. A performance analysis of the ASEN and MSC is conducted in Section 4.5. In this section, network performance parameters of packet delays and costs are analyzed and compared. Extensions to the metamodels introduced in Section 4.2 are

applied to the ASEN in Section 4.5. Section 4.6 concludes this chapter with a summary of these research efforts.

4.2 Performance Prediction Modeling of Interconnection Networks

This section presents the results of a study in predictive mathematical modeling of delay performance parameters associated with interconnection networks. This study provides a comparative analysis of three interconnection networks described in Chapter 3. These interconnection networks are the hypercube (single-stage cube), the multistage cube, and the Illiac IV mesh-type network.

4.2.1 Metamodeling of Interconnection Networks Metamodeling provides an avenue for expressing mathematically, a system model in terms of system parameters which effect the model [Agr85]. The value of metamodeling lies in its ability to map complex, time consuming simulation models into models which are simpler to evaluate, but at the same time, retaining the accuracy of the original model. Thus, metamodeling builds a model of the system model. Metamodeling seeks to add relevance to the simulation output by enforcing a structural model which suitably "fits" the data and builds a relationship between the control variables of the simulation (design criteria) and the measured variables determined by the simulation model. Metamodeling provides an experimental mechanism to add meaning to data and allows the researcher to propose a mathematical model to explain natural events.

For interconnection networks, metamodels can be used to estimate the value of some output dependent variable (e.g., network delay or throughput) as a function of a set of input predictor variables (e.g., network size and loading factor). In practice, a metamodel is derived from multivariate analysis of the performance data gathered for a particular network. Two key advantages of using metamodels are their predictive nature and their ability to clearly show the interaction among the input parameters of the system model. When used, metamodeling can reduce the amount of simulation required, and the size of the model simulated [RaR92]. In addition, the

effects of varying one or more system parameters can be clearly seen analytically. This gives to the system designer an improved ability to identify the key factors that influence the overall system performance thereby reducing the number of simulation trials needed to evaluate the impact of parameter variations.

4.2.2 Model Development One of the motivating factors for this research is the belief that differential equation metamodels can be formulated and used for performance appraisals of the interconnection networks studied (hypercube, mesh, MSC, and ASEN) over a practical range of network loading. That is, for network loads which are typical and result in queueing delays that are nearly linear with respect to the load, changes in the network performance can be explained by the design characteristics of the network. The existence of such metamodels enables the rapid prediction of the networks' performance for alternative design choices and allows for the direct comparison of network architectures.

The delay a packet experiences in traversing the network is of great importance to the network designer. This performance parameter, network delay, is the central focus of this metamodeling effort. To construct the metamodel, it is first necessary to consider the factors that affect the network delay. First, the number of nodes (N) in the network directly effect the delay. Second, the size switch (S) used in the network implementation plays a key role in determining delay. The third factor affecting delay is the network architecture (A). And finally, the network loading (L) affects the delay by causing queueing actions within the network. To see how three of the factors (N, S, A) affect delay, consider the minimum delay a packet can experience in traversing a given network. Minimum packet delays can range from $\log_s N$ in the hypercube, MSC and ASEN to $\sqrt{N}$ in the mesh. Addition effects can be due to the interactions among these four factors. These four factors, along with levels of variation are summarized in Table 4.1. The research described herein seeks to characterize the packet delay D_{ijklm}, encountered in traversing

the network as a function of these four factors, their joint interaction with one another, and a lumped error term, e_{ijklm} . Algebraically, this relationship is shown in Equation 4.1 below.

$$D_{ijklm} = A_i + N_j + L_k + S_l + AN_{ij} + AL_{ik} + AS_{il} + NL_{jk} + NS_{jl} + LS_{kl} + ANL_{ijk} + ANS_{ijl} + ALS_{ikl} + NLS_{jkl} + ANLS_{ijkl} + e_{ijklm} \quad (4.1)$$

Equation 4.1 shows the maximum possible combinations of interactions that can affect network delay. A four-factor ANOVA analysis can be used to investigate all interactions and determines their statistical significance to the overall delay model. For this investigation, three replications (simulations per data point) of the experiment are performed to arrive at the final model for network delay.

Table 4.1. Design factors for metamodeling of MSC, mesh, and single stage cube.

FACTOR	LEVELS OF VARIATION	LEVEL VALUES
Architecture (A)	i = 3	MESH, MSC, SINGLE
Nodes (N)	j = 3	1024, 256, 64
Load (L)	k = 5	2N/3, N/2, N/3, N/8, 2
Switch Size (S)	l = 2	16-by-16, 4-by-4

Using simulation data gathered for the MSC, mesh, and single stage cube networks, statistical analysis techniques are used to determine the impact of the four design factors on the network delay. Analysis of Variance (ANOVA) statistics for the experiment are computed using SAS[1]. Table 4.2 shows the results of the ANOVA execution for the experiment at large. This analysis assumes a uniform distribution of packets among sources and destination addresses.

[1]SAS is a product of the SAS Institute, Cary, NC 27512

Table 4.2. Experiment ANOVA Results
(Uniform Distribution)

ANOVA for Packet Time in System			
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
Model	89	16984.41	190.84
Error	180	2.93	0.02
Corrected Total	269	16987.34	
Model F =	11717.59	RF > F= 0.0	
R-Square	C.V.	Root MSE	Mean
0.999827	1.4470	0.1276	8.82

The parameter of particular interest in the ANOVA execution is *R-square*. The R-square parameter measures how much variation in the model's dependent variables can be accounted for by the specific model parameters. The closer the R-square value is to 1.0, the better the model fits the experimental data. From the R-Square value in Table 4.3 (.9998), one can see the highly descriptive power of the four factors in characterizing packet delay. The next step in the metamodeling development is to determine which factors from Equation 4.1 are statistically significant.

Statistical significance is determined by "F-value" and the probability of a factor's variation occurring by chance. The F-values are results of F-tests which gauge the likelihood that a factor's variation does not impact on the performance of the network. The larger the F-value, the less likely that the observed variation in the model occurred by chance.

One of the goals of the metamodeling development is to create a single all-encompassing relationship for packet delay independent of the network architecture. This is not possible due to the different queueing effects caused by the routing of packets through various network architectures. Because of this, separate metamodels are developed for each network architecture.

When the effects of architecture (A) are accounted for by separate models, Equation 4.1 reduces to:

$$D_{ijk} \propto N_i + L_j + S_k + NL_{ij} + NS_{ik} + LS_{jk} + NLS_{ijk} + e_{ijk} \quad (4.2)$$

The general relationship for network delay in Equation 4.2 can be used to formulate regression equations to predict the actual delay values based on the observed simulation data. Again using SAS, the general regression equation for time delay for the networks is:

$$\text{Time Delay, } T = x_0 + x_1 N_i + x_2 L_j + x_3 N_i L_j + x_4 S_k + e \quad (4.3)$$

Equation 4.3 omits factors from Equation 4.2 that have been determined to be statistically insignificant by the ANOVA analysis. F-tests are conducted on the network performance parameters to determine significance to the overall model. Those factors of Equation 4.2 omitted in Equation 4.3 (NS_{ik}, LS_{jk}, and NLS_{ijk}) each have very small F-values (< 1000) relative to the factors remaining in Equation 4.3. The x_i coefficients for the MSC, the Illiac IV mesh, and the hypercube network models, along with their adjusted R-square values are shown in Table 4.3. The high adjusted R-square values[2] indicate that the equations accurately predict the network delay values. The nonapplicability of the switch size as a variable for the mesh and hypercube, shown in Table 4.3, results from the fact that for a given network size, only one switch size can be used to implement the network. Recall from Chapter 3 that the switch size is fixed at 5-by-5 for the mesh and $\log_2 N + 1 - by - \log_2 N + 1$ for the hypercube. In addition, the nodes factor (N) coefficient for the MSC is not significant. Derived directly from Equation 4.3 and Table 4.3 are the explicit delay equations for the MSC, mesh, and hypercube networks under uniform network loading.

---

[2]The adjusted R-square value includes only those parameters which significantly affect the model. Inclusion of additional model parameters effect the R-square value while not necessarily effecting the adjusted R-square value.

Table 4.3. Least square estimates for packet delay models.

Network Architecture	Models' Adjusted R-square	Model Parameters				
		Intercept	Nodes (N)	Load (L)	NxL	Switch (S)
Mesh	0.9695	2.48575	0.01443	0.17222	-.00014	NA
Hypercube	0.9210	4.27617	0.00154	0.01586	-.00001	NA
MSC	0.7103	4.06375	**	0.03282	-.00003	-2.6602

$$\text{Mesh Delay, } T_{mesh} = 2.48575 + 0.01443N + 0.17222L - 0.00014NL \quad (4.4)$$

$$\text{Hypercube Delay, } T_{hypercube} = 4.27617 + 0.00154N + 0.01586L - 0.00001NL \quad (4.5)$$

$$\text{MSC Delay, } T_{msc} = 4.06375 + 0.03282L - 0.00003NL - 2.6602S \quad (4.6)$$

The accuracy of the derived metamodels is seen through the following example. Consider a 256 node hypercube network operating at 50% of the theoretical saturation point (256 packets/cycle). Figure 4.1, taken from [RaD88], shows that the delay for the hypercube is 6 time units. Equations 4.5 calculates the predicted delay to be 6.38 or a 6 percent deviation from the observed value. Similarly, Equation 4.4 provides a predicted value which varies approximately 0.4 percent from the observed value (23.6 predicted to 23.7 observed).

4.3 The Augmented Shuffle Exchange Network (ASEN)

The augmented shuffle-exchange network proposed by Kumar [KuR87, KuR89] is a fault-tolerant variation of the multistage cube network discussed in Chapter 3. The ASEN, a multistage b-by-b. The ASEN uses interconnection network, has $log_b N$ stages, each of which contains N/b crossbar switches of size intrastage loops to provide multiple paths from any source to a desired destination.

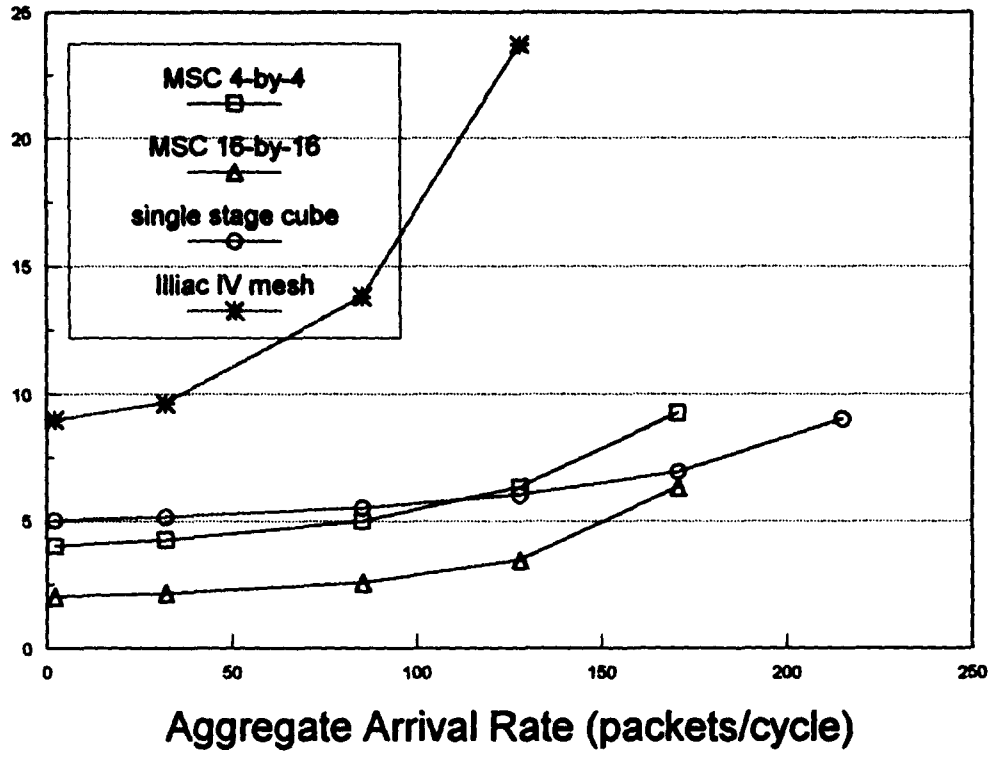


Figure 4.1. Average packet delay for 256 node networks [RaD88].

In [KuR89], the use of *conjugate pairs* of switches are defined in determining which switches within a given stage shall be interconnected to form a closed loop. In general, two switches are defined to be conjugate pairs if they are connected to the same switches in the next stage. For switches to be in the same loop, they cannot be conjugate pairs. One other loop construction constraint must be met. This constraint is that the switches in the loop must belong to the same *conjugate subset*. Kumar defines a conjugate subset as "the switches in a given stage which have output paths leading to the same subset of destinations." Loops can vary in size from stage to stage but must be consistent within a given stage. Figure 4.2 shows a 16-node ASEN with the maximum size loops possible. Also shown in Figure 4.2 are sets of input and output multiplexers used to overcome faults encountered at the input and output stages of the network. The importance

of the intrastage loops is realized when a packet becomes blocked in its normal progression through the network. When a blockage occurs, the packet is routed through the augmenting links in the same stage to another switch which has the capability to continue the packet routing to its destination. Packets routed over the augmented links have priority for transmission at the newly assigned switch. An illustrative example of the augmented routing is given in Section 4.4.3.2.1. Using the fault criteria described in Chapter 3, the ASEN is considered to be single fault tolerant.

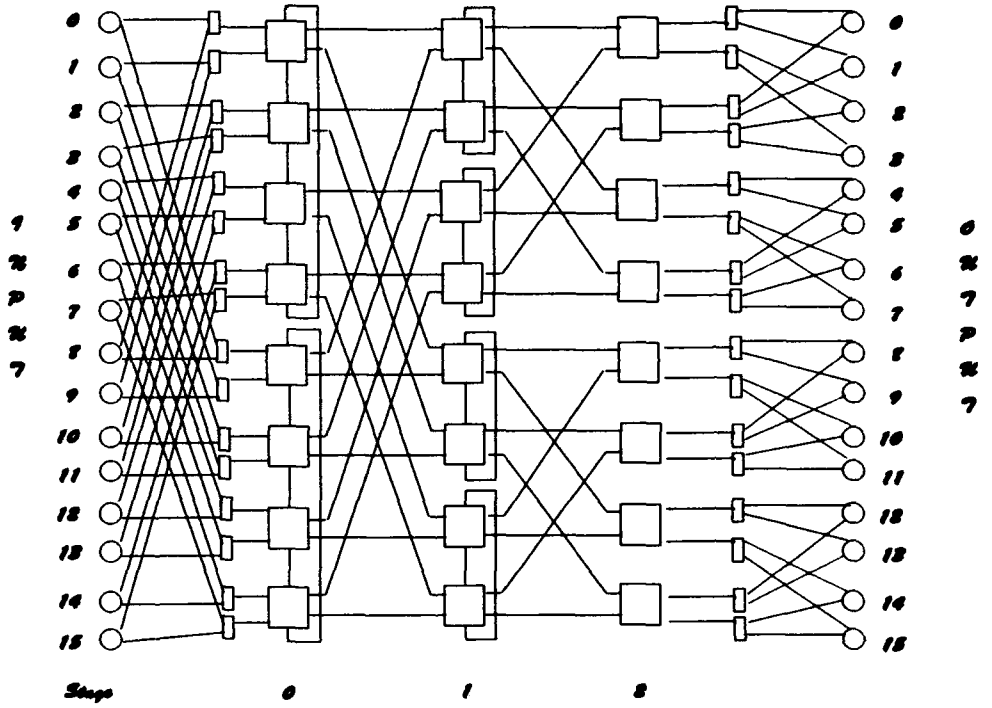


Figure 4.2. 16-node Augmented Shuffle-Exchange Network [KuR89].

Kumar's research investigates the ASEN's performance characteristics in a circuit switched environment. Central to his research are the network delay characteristics of the ASEN relative to the multistage cube (MSC) network and the effects of faults within the network. Kumar concludes that the ASEN, with longer path lengths, outperforms a MSC of similar size and operating

assumptions in terms of network delay. The research presented in this chapter complements Kumar's efforts by examining the performance of the ASEN in a packet switched environment.

4.4 Simulation Modeling of MSC and ASEN

In recent years, there has been an influx of commercial simulation software designed to ease the burden of the system modeler. In very few cases is it practical to model large scale systems using high order languages such as C or FORTRAN. In these few cases, the research is either pioneering to the point where commercial packages have not been developed to support the effort or the problem is not well suited for available simulation packages (i.e., an inordinate amount of tailoring of the package must be performed to meet the needs of the application).

When the application and the chosen simulation tool mesh, extreme benefits can be seen by the system modeler. First, many simulation packages (e.g., MODSIM, BOSS, NETWORK II.5, COMNET II.5) require little to no high order language programming. Built-in libraries are linked within the application via menu-driven displays to form the system being modeled. Second, built-in data gathering and statistical functions reduce the complexity of the analysis and the time required to manually perform the analysis. And thirdly, the burden of validating the simulator is removed from the designer. This is not to say that the simulation model need not be validated. It does mean that language constructs such as queuing, files, and statistical analysis functions have been validated previously by the package distributor.

Choosing which tool is right for the application is another decision that faces the system modeler. Commercial simulation packages can be grouped into two categories: general purpose and application specific. Packages such as SLAM II and Simscript fall into the general purpose category due to their flexibility in performing a wide range of applications. Application specific packages such as Network II.5 and ComNet II.5 perform well in their intended environments but do not provide the flexibility of the general purpose packages. The modeling of interconnection

networks in this research effort requires a tool that provides both flexibility in manipulation of thousands of queuing resource structures and also ease of data collection and system monitoring.

4.4.1 The Simulation Tool The simulation language SLAM II [Pri86] was chosen as the modeling tool for this investigation. The rationale for the choice of SLAM over other simulation languages such as Network II.5 and MODSIM is three-fold. First, the basic constructs of the language directly support queue and resource structures which can be interconnected to form models of a wide range of intercommunication networks. Second, built-in statistic and monitoring functions aid in the model development, data gathering, and validation. Third, system modeler familiarity with the language allowed for immediate development of the model. SLAM also allows for the use of user defined FORTRAN or C routines for tailoring to specific applications. A brief description of the SLAM network language is provided to assist in the understanding of the models to be presented in the latter sections of this chapter.

In modeling a network, the SLAM simulation is considered to consist of two parts: the control statements and the network description. The control statements provide options to the user for determining the initial states, any modifications to the simulation, and when and how to terminate the simulation. The network description is the portion of the SLAM code which represents the modeler's interpretation of the actual operation of the network to be simulated. SLAM provides 23 network statements which allow for in-depth simulations of large-scale computer communication networks [RaD88].

SLAM provides the *entity* which is used to model a message that will flow through the network in a store-and-forward manner. Each entity can have associated with it a set of *attributes* which are used to distinguish one entity from the other. Attributes can be assigned values which represent source addresses, destination addresses, message lengths, the time a message enters the system, and other user defined values. The *file* is used to represent resources such as data channels or memory modules as well as queues which store groups of entities. The basic concept of a

SLAM model is to have the entity(s) generated at some prescribed interarrival rate and then flow through the network, following the routes dictated by the designer. Each entity entering the network is terminated and removed from the network upon completion of routing through the network. As each entity terminates, statistics associated with the entity are available for collection if desired by the designer.

The statistical results of the simulation are provided for in SLAM via a summary report. The summary report includes statistics on the files, activities, and/or variables of the model. The summary report is the primary output of a SLAM simulation. Additional information related to the simulation is obtained from echo reports and trace reports. Echo reports reflect the data input and the initial values which are set prior to execution of the simulation. The trace report is primarily used as a debugging and validation tool for the model. The trace report provides a snap-shot view of the network at each instance of time in which an event is scheduled to occur.

4.4.2 Network Operating Assumptions To facilitate the modeling and the simulation of the multistage cube (MSC) network and the Augmented Shuffle Exchange Network (ASEN), certain assumptions and operating conditions for the networks must be established. The operating assumptions and conditions used herein [Ram92] are based on previous research [DiJ81, KrS83, RaD88] in packet-switched multistage interconnection network modeling and simulation. These assumptions and conditions are described below.

1. Each of the networks to be modeled are assumed to be operating in an MIMD environment.
2. A PE-to-PE architecture is assumed.
3. Packet-switching is used as the method for inter-PE communications. First-end-first-out (FIFO) message buffers are employed at the switches for the storing and forwarding of packets to and from the switches.
4. Message interarrival times are assumed to be Poisson processes.

5. The source and destination addresses are determined from uniform or normal distributions which are specified prior to a simulation trial. The normal distribution is defined to have a standard deviation of $.25N$, where N is the number of PEs in the system. One mean is chosen for the entire network.
6. Messages are assumed to be single packets in length.
7. The unit of measure for determining the average message delay and throughput of the network is the *packet cycle time* [DaS86]. This is the time required by a packet to move from the front of a queue, through its corresponding switch, and arrive at the queue associated with the next point (switch) in its routing scheme. For all simulations, the packet cycle time is normalized to 1 unit.
8. When a link blockage occurs in the ASEN, the individual switch determines if the packet should be buffered or routed through an augmented link within the same stage.
9. Packets sent over an augmented link are assumed to have priority over the other packets vying for a particular output link at a given switch.
10. The network input buffers are assumed to be infinite in length. This assumption allows for accurate assessment of the network performance by not restricting the generation of messages from a given PE.
11. Packet buffers within the network are finite in length. The length is determined experimentally such that the probability of overflow by any internal buffer being less than one percent in heavy, non-saturated network loading.
12. Network loading is based on the mean Poisson packet generation rate. A generation rate of $1/N$ equates to 100 percent loading in that on average, every PE generates a packet in any given cycle.
13. Distributed routing control is assumed. Each switch in the network determines the packet's outgoing channel based on the routing information associated with the packet.

4.4.3 ASEN and MSC Network Modeling. As discussed previously, the choice of simulation package used for network modeling is a critical design issue. Besides flexibility, the package must offer the designer modularity and conciseness in code development. Once the right package has been chosen, the network modeler is tasked with deriving a system model which takes advantage of the inherent capabilities of the simulation package. The work described here does just that.

Central to the development of the MSC and the ASEN models is the crossbar switch. The crossbar switch model is required to be scalable to accommodate switch sizes ranging from 2-by-2 to any reasonably sized switch that is a power of 2. Switches used in system implementations are limited in size to 32-by-32 due to integrated circuit pin-out limitations. The use of simulation though allows the network designer to observe performance characteristics of systems which exceed contemporary implementation capabilites.. The basic crossbar switch model used for this research was developed and validated by the author in [Rai87] with modifications by Ramachandran and Raines made to allow modeling of the ASEN. In [Rai87], switches up to size 64-by-64 were used to show the representative performance of networks implemented with larger sized switches.

4.4.3.1 Multistage Cube Network Model The multistage cube SLAM model is taken from [Rai87]. In that research, networks of size 64 through 1024 nodes were modeled. Switch sizes ranged from 2-by-2 to 32-by-32. Uniform distributions were used to generate source and destination PE addresses. Throughout the modeling of the MSC and the ASEN, mathematical relationships were sought for the appropriate network parameters. Once these relationships were derived, the resulting benefits were in compact and easy to follow code.

Associated with each entity (packet) is a set of attributes which distinguish the network entities. These entities, described in detail in [Rai87], contain the following attributes: source

address, destination address, time packet was created, stage of network where packet is presently located, present queue number, present transmission channel number, and number of bits in the routing tag. Three global variables are also associated with a particular network simulation. These variables remain static throughout the simulation. These variables are for (1) the size of the crossbar switch; (2) the number of routing tag bits that must be examined by a switch; and (3) the number of nodes in the network.

Each simulation begins with the creation of a packet at time zero, and there after, at a rate specified by the Poisson interarrival times. Upon creation, a packet is assigned a source-destination address pair based on either a uniform or normal distribution whose range is restricted by the number of nodes in the network. Next, the packet's appropriate attributes are assigned a value for the present stage and the initial input queue. The packet continues to flow into a sequence of statements which assign and reassign values to the outgoing channel, the present stage, and the input queue to the next stage. This sequence constitutes a loop which steps the packet through the network toward its destination. The packet exits the loop when it reaches its destination. At this point, statistics are gathered on the packet's time in the system.

The SLAM EVENT node serves as the mechanism for controlling the flow of packets through the network. The EVENT node contains the code logic which models the crossbar switch. The EVENT node allows a packet to traverse the network until a blockage is encountered. When a blockage occurs (either the inbound queue is not empty or the outgoing channel is busy), the EVENT node assigns the packet to the appropriate queue and then checks to see if any outgoing channels of the switch are available to route waiting packets.

4.4.3.1.1 Packet Routing in the MSC Model As previously described in Chapter 3, the multistage cube network uses deterministic routing of packets through the network. For each stage in the network, address lines are grouped at switches according to the $Cube_i$ function. The routing

to the next stage is dependent upon the source-destination pair and the routing algorithm used (XOR or destination routing).

Four parameter values are required to determine the outgoing channel of a particular switch. These parameters are: the current stage number; the bits to be examined in the routing tag; the size of the crossbar switch; and the source-destination address pair. The routing algorithm for the MSC is discretely coded in FORTRAN and linked into SLAM by user defined functions labeled USERF. The USERF(1) function computes a value that is added to the present queue number where the packet resides with the resulting value giving the outgoing channel number.

Once the outgoing channel is determined, the calculation of the input queue number at the next stage can be made. An internal numbering scheme for channels and queues was derived so that once the outgoing channel is known, the input queue at the next stage is equal to the present value of the outgoing channel plus the number of nodes in the network. This queue-channel numbering relationship means that only one calculation has to be made for any queue-channel pair. Figure 4.3 shows the internal numbering of a 8-node MSC.

4.4.3.2 Augmented Shuffle Exchange Network (ASEN) Model The ASEN model, developed by Ramachandran and this author, is derived from the MSC model described above. The ASEN model uses the same crossbar switch SLAM code as the MSC. The two models differ as a result of the ASEN augmented intrastage links and associated queues. Packet source and destination addresses are generated in the same manner as for the MSC. Interarrival rates for packets are controlled by a Poisson process. Entity (packet) attributes are similar to those used in the MSC but with additional attributes assigned to indicate assignment to priority queues and pseudo-source/destination addresses. The pseudo-source/destination addresses are internal bookkeeping attributes for routing the packets once an augmented link is taken. Packet flow control decisions are made in the SLAM EVENT node logic.

Modifications to the MSC EVENT node are required due to the additional links and queues present in the ASEN but not the MSC. Associated with each augmented stage are priority queues located at each of the switches within the stage. The EVENT node logic is modified to accommodate routing to and interrogation of the priority queues. As with the MSC model, mathematical relationships between the queues and channels are developed to allow for modularity and scalability of the ASEN model.

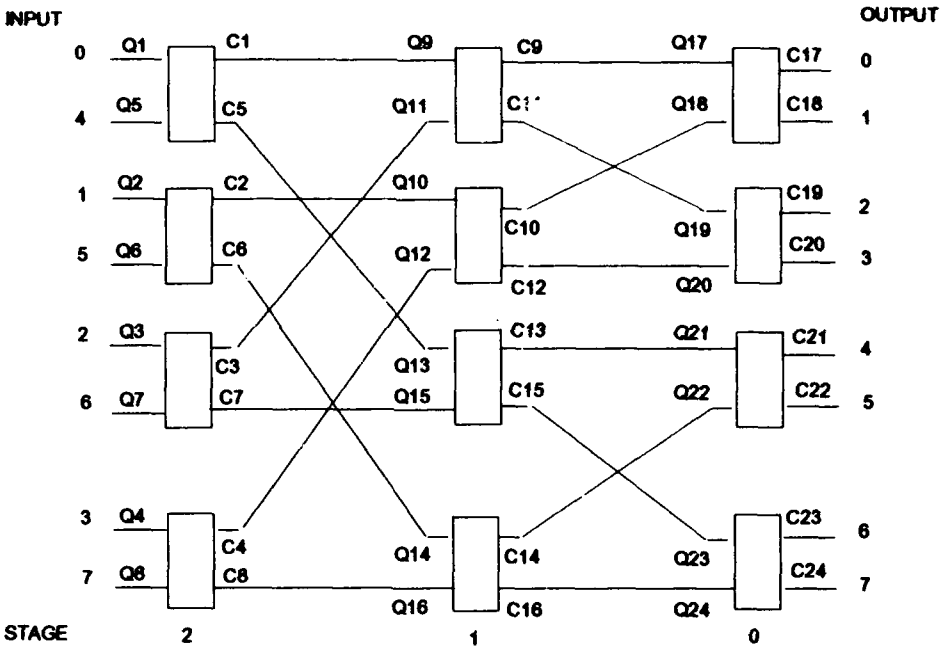


Figure 4.3. Internal channel-queue numbering for SLAM MSC model.

4.4.3.2.1 Packet routing in the ASEN model Packet routing in the ASEN is similar to the method used for the MSC described above. The two methods differ when a packet experiences a blockage due to a busy outgoing channel. When a packet experiences a blockage, one of two possible courses of action takes place. First, the EVENT node logic computes the augmented priority queue and then checks to see if the priority queue is full. If the priority queue is not full,

the packet is allowed to move to the priority queue and check the appropriate outgoing channel for utilization and subsequent movement to the next stage. The second course of action takes place when the priority queue is full. In this case, the packet is filed in its original FIFO queue and then must contend with other non-priority queues of the switch for outgoing channels when they become available.

As a clarifying example, consider the 8-node ASEN shown in Figure 4.4. Suppose PE 1 wishes to send a packet to PE 6. Without blockage, the path would be through switches 1 to 7 to 11. If a blockage occurs at switch 1 (i.e., the bottom outgoing channel is busy), then the control mechanism attempts to route the packet to the augmented priority queue in switch 0. If the priority queue at switch 0 is full, then the packet is filed in the upper FIFO queue of switch 1. If the priority queue at switch 0 is not full, then the packet is routed to the priority queue at switch 0. Once the packet arrives at the priority queue, the appropriate outgoing channel is checked for availability. If the lower outgoing channel of switch 0 is busy, the packet is filed in the priority queue and waits for the channel to become available. Packets in the priority queue will take precedence over non-priority queues when determining which packet can seize the channel next.

4.4.4 Network Simulations The thrust of this research effort is to compare the run-time performance of the ASEN and MSC. Specifically, two performance parameters are of interest: the average message delay and the buffer memory costs associated with implementing either of the networks. The average message delay is defined as the time required by a message to traverse the network from input PE to output PE. The buffer memory cost is the product of the total number of buffers in the network, the maximum buffer length (the length required to ensure that 99% of the time that the network is in operation, this length will not be exceeded), and the unit cost per memory size (assumed to be constant at one unit for this investigation). To obtain this information, simulation trials of the network models described above are required.

Once the network parameters of interest are defined, message delay curves for the ASEN and the MSC must be generated. For each of the curves to be generated, multiple simulation runs are associated with each discrete point on the delay versus loading curve. Determining the number of simulation runs associated with each data point depends upon the desired accuracy and degree of confidence in the mean average delay value obtained from the multiple runs. In addition to determining the number of simulation runs for a given point, the number of points that are required to accurately reflect the average message delay curve characteristics must be determined. These two factors are determined through pilot simulation runs.

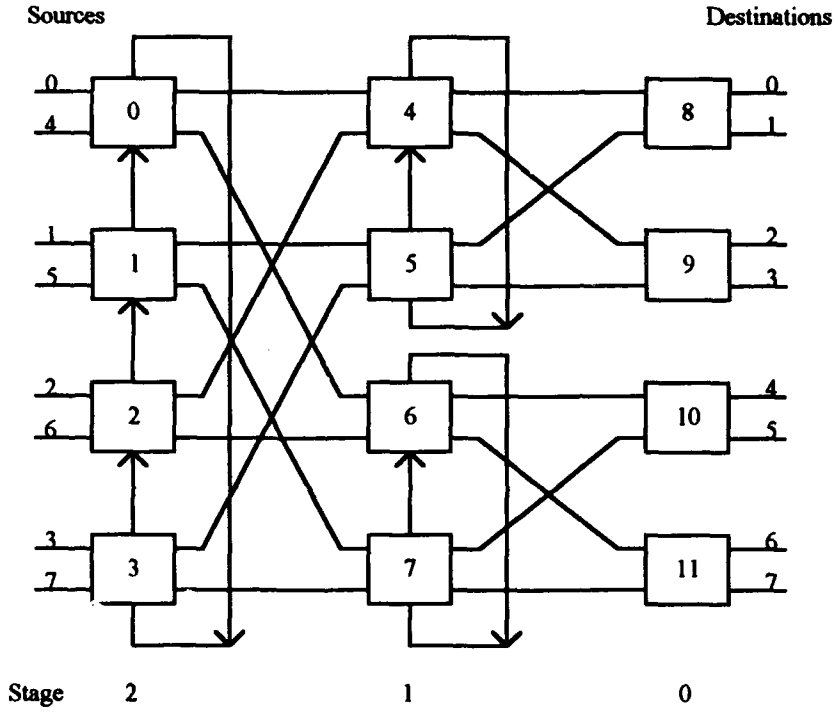


Figure 4.4. An 8-node ASEN network.

Each delay curve generated for the ASEN contains eight data points. Curves for the MSC contain five data points. These numbers are chosen based on the following reason. For light to medium loading, network delays remain approximately constant allowing the emphasis of data

point distribution to be placed in the area of the "knee" of the message delay curve. The "knee" is the area of the curve where queueing delays become more prevalent causing the curve to become nonlinear. The "knee" does not include the portion of the curve where the network is in saturation.

Next, the number of simulation runs per data point must be determined. Two approaches exist in determining the number of simulation trials necessary to ensure the validity of the results obtained. The first approach requires that the desired confidence interval be specified prior to simulation. With this approach, the desired confidence interval determines the number of simulation trials. The second approach is to choose the number of simulation trials to make and allow the confidence interval to directly result from this choice. This approach allows the designer to iteratively "tighten" the confidence interval to a desired level. The latter approach was chosen for this investigation.

Three independent simulation trials per data point were chosen. Three replications provide an acceptable trade-off between the run-time requirements of the experiment and reasonable confidence interval of the performance data. The worst case variation in the average message delay for both the MSC and ASEN is less than 9% and is typically less than 2% (the 9% variation occurs under extremely heavy network loading just prior to the occurrence of saturation). These values indicate, with a high degree of confidence, that the mean message delay values obtained accurately represent the models simulated. These delay values are used for comparison against previously published works to determine the validity of the network models.

4.4.5 Network Model Validation A major concern in any simulation study lies in determining the validity of the model. Possible validation methods are to compare the simulation results against analytic models, compare simulation results against previously published works, or develop testing methodologies for the models. This research effort uses a combination of the latter two methods.

As stated previously, the modeling of the ASEN is an extension to the MSC model developed in [Rai87]. In that effort, the MSC model is validated against previously published works of [DiJ81, KrS83]. The validation of the MSC allows for internal testing methodologies to be used to validate the ASEN.

Validation of the ASEN model requires the validation of the network structure and the routing algorithm used to pass packets through the network. The network structure validation entails determining whether or not queues and link structures are properly interconnected. The routing algorithm validation ensures that packets are routed to the correct queues and transmitted along the correct channels. The validation of both aspects of the ASEN model are accomplished via the generation of packets and the use of SLAM and FORTRAN statements. The initial testing of the network structure and routing algorithm is accomplished by generating packets at a rate slow enough so that the SLAM TRACE statement can be used to track the routing of discrete packets. The TRACE statement allows FORTRAN write statements to be strategically placed throughout the network code to provide a "snap-shot" of the network operation at the desired times. Using these constructs allow for the debugging and ease of locating faults within the system model. Once the initial testing is completed, the packets generation rate is increased to observe the operation of the queueing structures of the ASEN. Once again, the TRACE and other summary constructs inherent to SLAM are used along with pilot simulation trials to validate the correct operation of the network model.

4.5 Performance Analysis of the MSC and ASEN

In modeling the ASEN and the MSC, three network sizes are chosen for implementation: N = 64, N = 256, and N = 1024. These network sizes are representative of systems which are implementable using current microprocessor technology. Another modeling decision lies in the choice of crossbar switch implementations. For a given size network, multiple sizes of crossbar switches can be used for the implementation. In the case of a 64-PE network, switch sizes of 2-by-

2, 4-by-4, and 8-by-8 can be used to model the network (these switch sizes do not count the augmented links). For 256-PE networks, possible switch sizes are: 2-by-2, 4-by-4, and 16-by-16. The ASEN and MSC of network size 1024 can be implemented with switches of size 2-by-2, 4-by-4, and 32-by-32. This investigation examines the effects of various sized switches on network delay.

The ASEN defined in Chapter 3 is a fault-tolerant interconnection network. As with the other fault-tolerant networks referenced in Chapter 3, the ASEN uses redundant hardware to overcome the possible loss of interprocessor communications when a network fault is encountered. This research focuses on using the ASEN's augmented links to reduce congestion in a fault-free environment rather than its performance capabilities in a faulty environment. Analyzing the ASEN in a fault-free environment allows for increased utilization of the additional fault-tolerant hardware which remains idle under the assumptions of Kumar [KuR87, KuR89].

The simulation and analysis of the MSC and ASEN reveals trends that are consistent across various network sizes. Because of this consistency, the results of a representative network size of 256-PE are presented without loss of generality. Network delay performance plots for networks of size 64 and 1024 are presented in [Ram92].

4.5.1 Network Delay Before a comparison of the ASEN and MSC network models' packet delay characteristics can be made, it is necessary to define the minimum obtainable delay a packet can experience in traversing the network. Using a fixed packet cycle time, which includes the processing time internal to the crossbar switch, the minimal number of packet cycles it takes a message to traverse the network is quantifiable. In both of the networks examined, the number of hops from the source PE to the destination PE is dependent upon the number nodes in the network and the size of the crossbar switch used to implement the network. Specifically, the minimum delay a packet can experience in moving from source to destination is $\log_b N$, where b denotes the size of the crossbar switch and N denotes the number of nodes in the network. Consider a network

supporting 256 PEs. For the 4-by-4 crossbar switch implementation, $\log_4 256 = 4$ hops are required to traverse the network with a minimum delay of 4 packet cycles. As the network loading increases, the network delay will increase due to queueing actions within the network.

Shown in Figures 4.5, and 4.6, 4.7 are the average packet times in the system for various loading factors, source-destination distributions, and switching element size for a representative network size of N=256. Note that for light system loading (i.e., the aggregate loading factor is less than 25%), the ASEN and the MSC have approximately the same level of performance. As the loading increases, the delay curves tend to diverge. For the case where 2-by-2 (3-by-3 for the ASEN) switches are used, the divergence of the delay curves is quite dramatic. Figure 4.5 shows that the MSC network, under a uniformly distributed source-destination pairing, begins to saturate at approximately 60 percent of full loading. The ASEN, with the same uniform source-destination distribution and loading (60%), continues to route packets through the network without saturation and experiences average delay times which are approximately 20 percent less than the MSC. The ASEN begins to saturate at 70 percent loading. For the uniform distribution, the non-saturation operating range of the ASEN is approximately 17 percent greater than that of the MSC (70% loading relative to 60%). Comparing the performance of the two networks under a statistically normal source-destination distribution, the ASEN realizes a 25 percent reduction in the average time in system over the MSC for heavy loading. In addition, for the normal distribution, the ASEN does not begin to show signs of saturation until 70 percent loading -- 15 percent higher than the MSC for heavy loading. The major attributing factor to the ASEN's superior performance is that when using 2-by-2 switches, the priority queues provide for a 50 percent increase in the number of queues for the stages which use the ASEN cyclic redundancy. This allows for normally blocked packets to be routed to less congested links for transmission to their respective destinations.

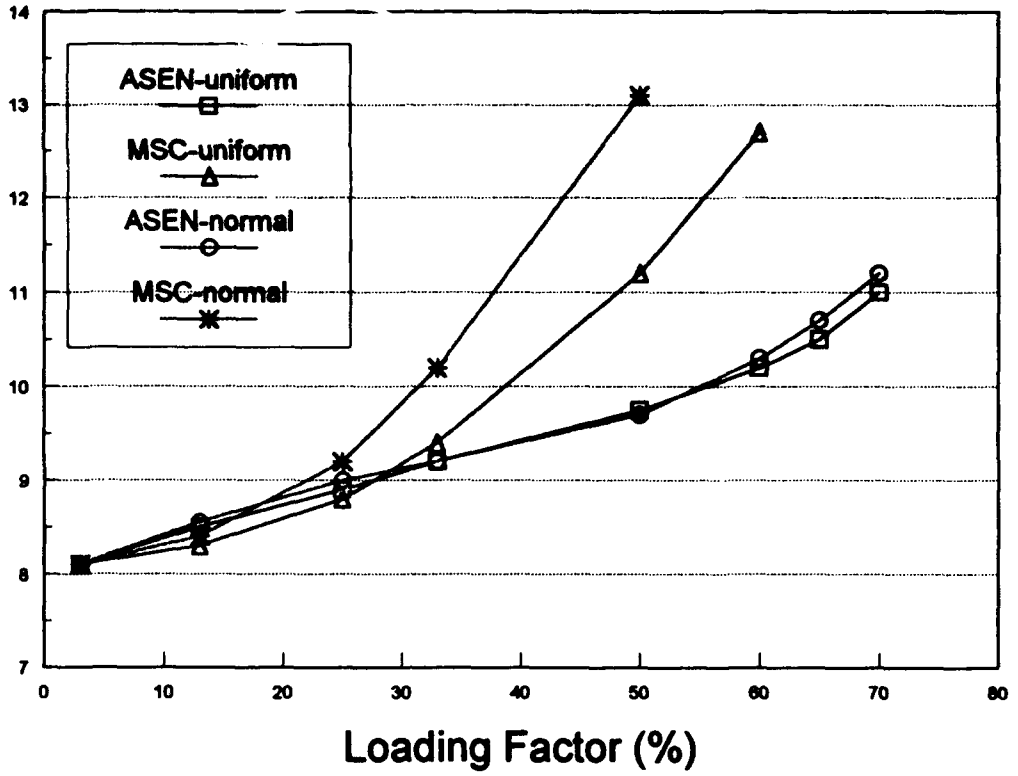


Figure 4.5. Network performance N=256 (2-by-2 switches).

Figures 4.6, and 4.7 show the effects of larger switch sizes on the delay performance of the ASEN and the MSC network. Figure 4.6 shows that the use of 4-by-4 switches allows the ASEN to route packets with an average delay that is 25 percent less than the MSC at a loading factor of 50 percent, using a uniform source-destination distribution. The performance improvement is more dramatic for the same distribution at 60 percent loading, where the ASEN average packet delay is approximately 40 percent less than the MSC. A comparison of the two networks using a normal packet distribution reveals that the ASEN packet delay is approximately 40 percent less than the MSC delay at 50 and 60 percent loading. Figure 4.7, for 16-by-16 switches, reflects the same trends seen in Figure 4.6. At loading factors greater than 50 percent, the packet delay performance of the two networks differs from 25 to 40 percent, depending upon the source-destination distribution chosen. The ability of the ASEN to have lower average packet delay than the MSC is

attributed to its use of augmented links and priority queues. Though the percentage increase in augmented links and priority queues decreases as the ASEN switch size is increased, the ASEN performance relative to the MSC does not degrade.

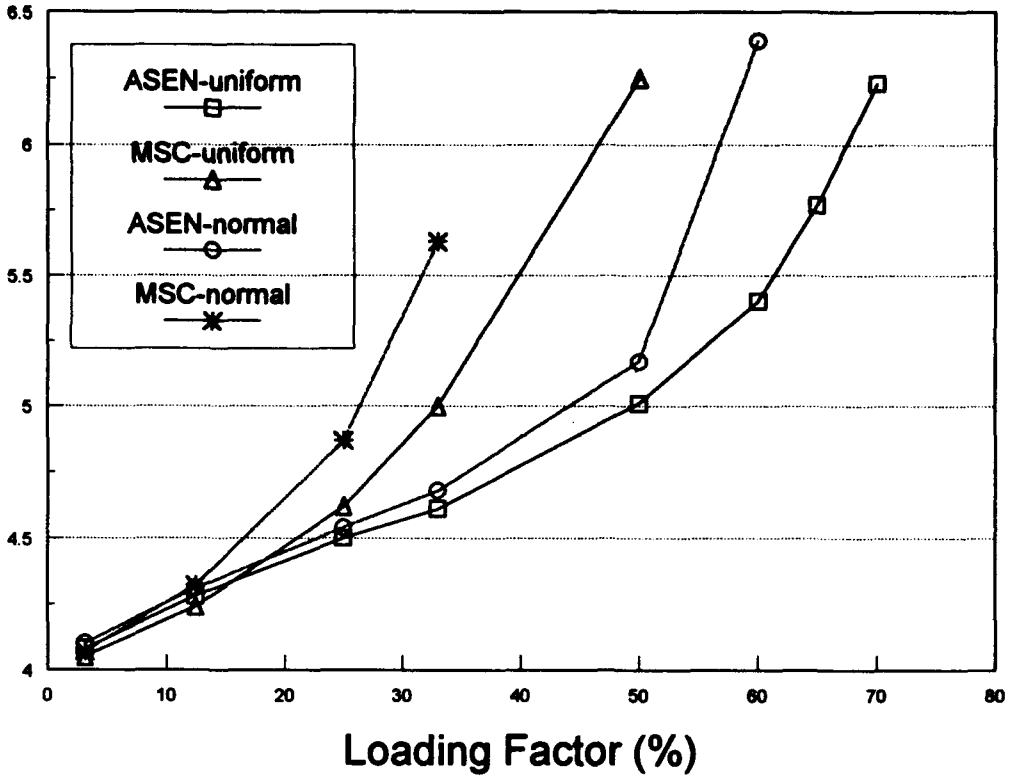


Figure 4.6. Network performance N=256 (4-by-4 switches).

The impact of hot spots resulting from the use of a normal message distribution is also shown in Figures 4.5, 4.6, and 4.7. Hot-spots, as summarized by DeCegama [DeC89], cause a disproportionate number of packets to be routed to a subset of destinations, and can cause the network to saturate. All network messages are ultimately affected. This can be seen clearly by comparing the MSC simulations for uniform and normal message distributions. Under similar operating conditions, the ASEN performance is more robust than that of the MSC. Performance improvement is direct result of the redundant intra-stage links and the alternate routing paths that

they create. By routing packets to other switches in the same conjugate subset (i.e., those switches within a given stage that have the capability to communicate with the same set of destinations), these links ease the load on those switches whose buffers tend to fill up. The degradation in the performance of the ASEN when using a normal message distribution (when compared to its uniform distribution performance) is much less pronounced than the degradation seen by the MSC.

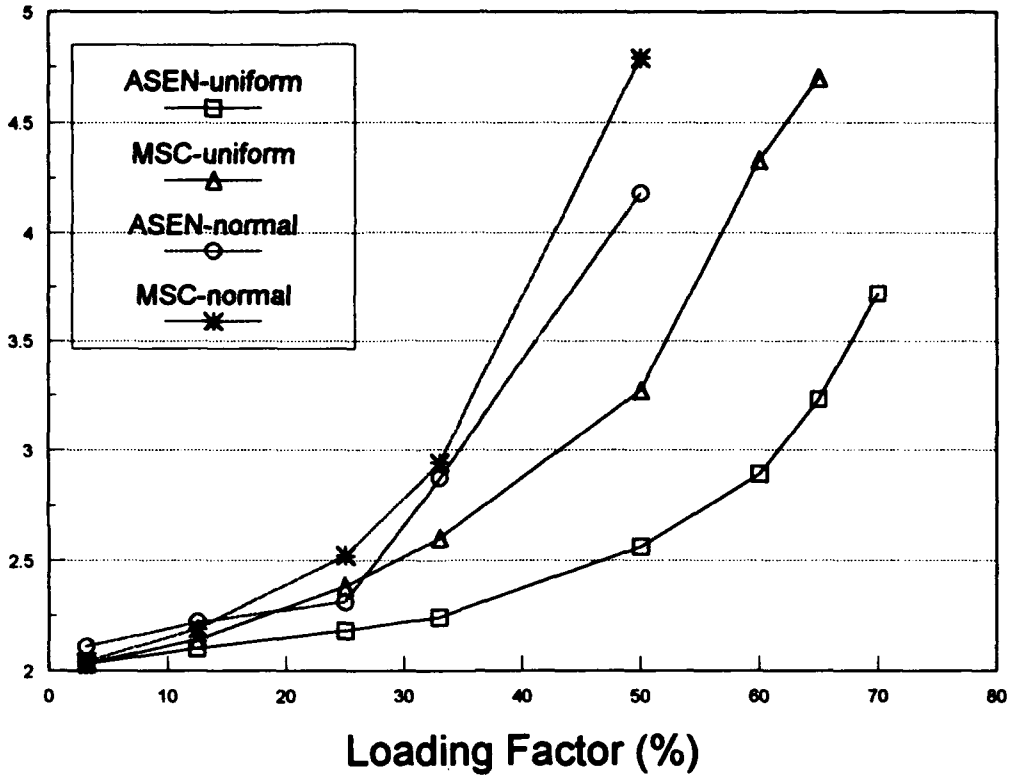


Figure 4.7. Network performance N=256 (16-by-16 switches).

As the switch size and loading factors are increased in the ASEN, the delay curves reveal that a normal distribution of packets causes the ASEN to saturate much faster than the same sized network with packets generated by a uniform distribution. This is because the total number of augmenting links available for use decreases as the number of individual switches in the network is reduced. Also contributing to this divergence of the uniform and normal distribution curves are the "hot-spot" effects of the normal distribution described above. Despite this divergence, Figures 4.5,

4.6, and 4.7 show that the ASEN, under a normal packet generation distribution, outperforms the MSC of the same size, operating under a uniform source-destination distribution. Additionally, the ASEN has a higher saturation point than does the MSC.

The analysis of the delay variance follows directly from the packet delay statistics. The delay variance is found to be approximately the same for both MSC and ASEN networks across various switch sizes. The minimum delay encountered by a blocked packet is one time unit. The worst case delay variance in the ASEN is 1.42 time units (for a 2-by-2 switch implementation). The corresponding variance for the MSC is 1.46 time units.

4.5.2 Network Costs Just as important as the network delay is the cost associated with implementing a chosen network. Network costs are a function of the switches, wiring and queues required to construct the network. For this investigation, network cost is defined as the product of the buffers per queue and the total number of queues in the network. An implementation cost comparison between the MSC network and the ASEN is considered here for the conditions which follow. First, space in an input queue at any switch is defined in terms of unit packet lengths (i.e., a queue capable of storing 5 packets equates to 5 buffers). Second, the network cost is dominated by the total number of buffers used to implement the network queues [Sie90]. Finally, one buffer equates to unit cost. Table 4.4 shows the cost associated with implementing both the ASEN and the MSC network of various network and switch sizes. Implicit to Table 4.4 is the length of each switch queue. Analysis of the network packet delays in steady-state show that to achieve a queue overflow rate of less than one (1) percent, the MSC network requires a capacity of six (6) packets for its $N \log_s N$ network queues while the ASEN requires a capacity of five (5) packets for its $N \log_s N$ non-augmenting queues and a one packet capacity for its $N/s*(\log_s N - 1)$ augmented queues. These queue capacity values are derived from using an infinite queue model and observing, through pilot simulation runs, the overflow rates for specified queue lengths. SLAM statistical gathering summaries provide data on average and maximum lengths for all

network queues. The cost equations for the MSC and ASEN are given in Equations 4.7 and 4.8 below.

$$Buffers_{msc} = 6N \log_s N \tag{4.7}$$

$$Buffers_{ASEN} = 5N \log_s N + (N/s)(\log_s(N) - 1) \tag{4.8}$$

The tabular results shown in Table 4.4 provide insight into the possible implementation cost savings of the ASEN over the MSC. These savings in buffer costs for the ASEN relative to the MSC range from 9.2% (a 1024-PE network implemented with 2-by-2 switches) to 16.4% (a 1024-PE network implemented with 32-by-32 switches). The implementation cost savings of the ASEN as well as the network's superior delay performance characteristics can be attributed to the augmented links and priority queues used to redistribute packets in a congested environment. The combination of reduced costs and increased performance makes the ASEN the preferred network over the MSC.

Table 4.4. Cost of Buffers
(Uniform Distribution)

NUMBER OF BUFFERS IN THE NETWORK*				
Size	Network	64	256	1024
2x2	MSC	2304	12288	61440
	ASEN	2080	11136	55808
4x4	MSC	1152	6144	30720
	ASEN	1024	5504	26624
8x8	MSC	768	Not	Not
	ASEN	648	Used	Used
16x16	MSC	Not	3072	Not
	ASEN	Used	2576	Used
32x32	MSC	Not	Not	12288
	ASEN	Used	Used	10272

*MSC: 6 buffers per queue; ASEN: 5 buffers per queue
1 buffer = unit cost

4.5.3 Interconnection Network Metamodeling Extension The metamodeling presented above is extended to another interconnection network, the ASEN. Following the same model development procedure as described above, a metamodel is derived which accurately predicts network delay. The metamodel design factors, ANOVA analysis results, and least square estimates for the ASEN are shown in Tables 4.5, 4.6, and 4.7 respectively.

Table 4.5. ASEN Experiment Design
(Uniform Distribution)

FACTOR	LEVELS OF VARIATION	LEVEL VALUES
Nodes (N)	i = 3	1024, 256, 64
Load (L)	j = 8	3.125, 12.5, 25, 33, 50, 60, 65, 70%
Switch Size (S)	k = 5	2x2, 4x4, 8x8, 16x16, 32x32

Table 4.6 shows that Equation 4.4 once again has a highly descriptive power when applied to the ASEN (R-square = 0.9995). Table 4.7 summarizes the explicit predictive delay equation for the ASEN.

$$\text{ASEN Delay, } T_{ASEN} = 8.21154 + 0.0027N + 0.0266L - 0.000014NL - 2.2944S \quad (4.9)$$

The usefulness of metamodels is shown through the following examples. Consider a 256 node ASEN with a uniform source-destination distribution implemented using 2-by-2 switches. For a loading factor of 60 percent (which is equivalent to .6 x N packets/unit time), Equation 4.9 estimates that the packet delay will be 10.14 units. Figure 4.5 shows that the average packet delay is approximately 10.2 units for 60 percent loading -- a difference of less than 1%. Similarly,

taking the partial derivative of Equation 4.3 with respect to loading, the incremental change in delay given unit change in loading is:

$$\frac{\partial \text{Delay}}{\partial L} = x_2 + x_3 N \tag{4.10}$$

Table 4.6. Experiment ANOVA Results
(Uniform Distribution)

ANOVA for Packet Time in System			
Augmented Shuffle Exchange Network			
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
Model	53	1836.31	34.64
Error	206	1.009	0.005
Corrected Total	259	1837.31	
Model F =	7073.21	RF > F= 0.0	
R-Square	C.V.	Root MSE	Mean
0.999451	1.344680	0.0700	5.2048

Table 4.7 Least Squares Estimates for ASEN Packet Delay Model
(Uniform Distribution)

Network Architecture	Models' Adjusted R-square	Model Parameters				
		Intercept	Nodes (N)	Load (L)	NxL	Switch (S)
ASEN	0.8469	8.21154	0.0027	0.02660	-.000014	-2.2944

Consider again 256-node ASEN and MSC networks implemented with 4-by-4 crossbar switches. Using Equation 4.10 and the derived coefficients from Tables 4.1 and 4.7, the incremental delay caused by a unit increase in the loading is 9.2% greater for the MSC (.02514) compared to the incremental change in delay of the ASEN (.02302). This example shows the usefulness of the metamodel. Relative sensitivity of a network to change in a given parameter is revealed in a simple, yet accurate quantitative form. Similar regression metamodels can be formulated for other dependent variables of interest such as the delay variance of delay or the maximum delay time.

4.5.3.1 Model Limitations When using metamodels such as these, the range of their accuracy must be kept in mind. For example, from Table 4.7 and Equations 4.3 and 4.10, one could surmise that for a sufficiently large value of N, the incremental change in delay could be negative as the load increases. This is nonsensical due to applying the metamodel outside of the normal operating regions of load and network size. Therefore, the analysis is limited to the non-saturated portion of the delay curve and limited to feasible network designs.

The non-saturated portion of the delay curve is defined as the operating region preceding the "knee" of the classic queueing curve. Once a network is loaded to the point where delay increases rapidly (and nonlinearly) with minor increase in offered packets, explaining the behavior of the network is tenuous. Queueing theory reveals that single server, single queue models saturate at approximately 70-80% of the server capacity. Networks of queues tend to have extremely sharp ramp-ups in delay as the capacity of the network is approached (see Figures 4.5, 4.6, and 4.7). The models described above are intended to be used in the relatively linear, gradually increasing delay region desired in most network designs.

4.6 Summary

This chapter presented the results of two independent research efforts: the modeling and analysis of the Augmented Shuffle Exchange Network, and the development and application of

mathematical metamodels to various types of interconnection networks for parallel processing. Derivation of network models for the ASEN and MSC were presented along with the simulations and validations associated with these networks. Performance comparisons were made between the ASEN and the MSC from network delay and implementation cost standpoints. From these analyzes, it was shown that the ASEN outperforms the MSC in terms of reduced network delay for both uniform and normal source-destination address generation distributions. Additionally, the costs associated with implementing the ASEN were shown to be 9 to 16 percent lower than those associated with a similarly sized MSC.

Mathematical techniques for predicting network delay were presented for the MSC, single stage cube, Illiac IV mesh, and ASEN interconnection networks. The derivation of these models was presented to include ANOVA and regression modeling techniques. These models were shown to have a highly accurate descriptive power when applied to predicting network delay for the interconnection networks investigated. Because of the high accuracy of these models, reductions in simulation times can be realized when deriving delay characteristics for interconnection networks.

CHAPTER 5

PART I CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of Part I Research

Part I of this research effort has focused on the performance analysis and predictive modeling of interconnection networks for parallel processing. Simulation and mathematical models have been developed to aid the network designer in choosing the appropriate interconnection network for implementation.

Chapter 2 provided a background for the research problem at hand: processing needs of society exceed the processing capabilities of uniprocessor systems. Because of the performance limitations of single processor systems, computer architects have proposed and developed multiprocessor systems. These systems allow for the complex computations of present day applications to be performed and results achieved almost instantaneous of the input. This immediate access to computational results has many intrinsic values, the greatest of which results in the saving of human lives in the cases of weather analysis and national defense systems.

Many design factors must be considered in the development of a parallel processing system. Of major importance is the structure used to interconnect multiple processors. Once the interconnection architecture has been determined, factors such as switching methodology, control strategies, and fault tolerant capabilities of the network must be investigated and determined. These factors were discussed in Chapter 2.

An overview of parallel processing systems was presented in Chapter 3. System classification taxonomies were discussed along with inherent problems associated with the use of each taxonomy. An examination of interconnection networks followed the presentation of parallel processing taxonomies. Classification of interconnection network architectures, primary switching

elements used in the network, control structures, and fault tolerant capabilities were topics presented and discussed. Relative merits and demerits of these design factors were also highlighted. A brief overview of two switching methods for interconnection networks, circuit switching and packet switching, lent additional insight into the problems which face the system designer. Chapter 3 concluded with a review of four contemporary parallel processing systems related to the interconnection network scheme used in their respective implementations.

The methodology and analysis used to solve Part I of this dissertation effort was the topic of Chapter 4. In this chapter, the work of Kumar on the Augmented Shuffle Exchange Network (ASEN) was summarized. His work, under a circuit switched and faulty environment showed that the ASEN outperforms the multistage cube (MSC) network in terms of delays encountered by packets as they traverse the network. The choice of simulation tool was discussed to provide insight into application factors a systems designer must consider in determining the simulation tools for a particular application. A set of network operating assumptions was established to form a basis for the comparative performance analysis of the subject networks. Network simulations and validations of network models were addressed for completeness. An analysis of the performance characteristics of the ASEN and the MSC was performed to determine if one network could outperform the other in a common operating environment. Network performance parameters of packet delay and network costs were analyzed and compared for the two networks. The concept of metamodeling was introduced. Predictive delay models were developed for four different interconnection networks: the ASEN, MSC, single stage cube (hypercube), and the Illiac IV mesh-type network. These models were used to provide supplemental information to the network designer relative to incremental changes in network dependent design parameters such as loading, switch size, and number of processors in the network.

5.2 Part I Research Effort Conclusions

This investigation has provided a performance comparison of two multistage interconnection networks under a common set operating assumptions. Direct comparisons of the delay and cost characteristics of the ASEN and MSC have not previously been performed assuming a packet switched environment. It has been shown that the ASEN outperforms the MSC in terms of average delay encountered by packets traversing the network. Average delay times for the 256-PE ASEN were shown to be approximately 20% and 25% less than the MSC for uniform and normal source-destination address generation distributions respectively. Also, the non-saturation operating range of the ASEN is 17% and 15% greater than the MSC for the respective address generation distributions. In terms of network costs, the ASEN once again outperforms the MSC. Network switch queues were examined to determine the length required to ensure that buffer overflow would occur less than one (1) percent of the time in a steady-state, non-saturated operating mode. Results of this examination have shown that the MSC queues require capacities of six (6) packets while the ASEN requires five (5) under uniform loading. This translates to implementation cost savings for the ASEN relative to the MSC ranging from 9.2% (a 1024 PE network implemented with 2-by-2 switches) to 16.4% (a 1024 PE network implemented with 32-by-32 switches).

Metamodels were developed for the MSC, single stage cube, and Illiac IV mesh-type networks. These metamodels were used to show the effects of incremental changes in network parameters on packet delays. A high degree of accuracy in predicting packet delays was shown by an ASEN example in which the predicted values and the simulated values differed by less than one (1) percent. It was also shown through the metamodels that the delay characteristics of the MSC were more susceptible to change than the ASEN. These findings also correspond to the results obtained through simulation. Similar results were noted for the Illiac IV network.

5.3 Recommendations for Future Research

This investigation has expanded the knowledge base in performance modeling and analysis of interconnection networks for parallel processing. New and innovative models for accurately predicting and comparing the performance of dissimilar networks provide insights not previously seen in published works. While the work described above is original and noteworthy, extensions to the research are envisioned. The foremost extension to this research is to model and analyze the ASEN in a faulty, packet switched network environment. This extension would require the development of dynamic routing algorithms and fault detection methods in existing system models.

The overall results of the efforts described in Part I have been presented in three technical papers. The first [RaR92], describes the performance comparisons of the ASEN and the MSC. The second technical paper [ShD94] derives and applies metamodeling techniques to the single stage cube, MSC, and Illiac IV networks. Metamodeling of the ASEN and comparative analysis of the ASEN and the MSC are the subjects presented in the third technical paper [RaD94].

PART II

MODELING AND PERFORMANCE ANALYSIS OF A LOW EARTH ORBIT SATELLITE COMMUNICATIONS NETWORK

CHAPTER 6

INTRODUCTION

6.1 Background

In Part I, an analysis was conducted on communications systems whose delay and throughput performance characteristics are predominantly affected by delays incurred by messages waiting in network switch queues. The research to be presented in Part II of this effort analyzes communication systems which are affected by both queuing and propagation delays.

Providing global communications has been a design goal of communication engineers for years. Satellite systems such as the geosynchronous INTELSAT series have provided transcontinental communications for over 25 years. These communications have been in the form of voice circuits and television signals. As discussed in Chapter 2, communication trends have been to provide machines and services which function efficiently, at high rates of speed, and at the same time, at minimal cost. These trends hold true for satellite communications.

For years the satellite communications industry has been dominated by geostationary systems. The reasons behind this have been the large investments required to place a satellite into orbit and the ability of the geostationary satellite to provide reliable communications. But probably the most important reason for the emergence of geostationary satellite systems as the predominant satellite communication system lies in the ease of its tracking by a ground station. As its name indicates, a geostationary satellite appears as a fixed point in the sky when viewed from

the surface of the earth. This "fixture in space" characteristic allows earth station antennas to remain approximately fixed (slight adjustments must be made periodically due to satellite drifts) with respect to elevation and azimuth pointing angles. This advantageous characteristic of the geostationary satellite is not provided without cost. Because of the orbital altitude (approximately 35,870 kilometers) required by the geostationary satellite, transmitter power requirements must compensate for the energy losses encountered when propagating a signal to earth. In addition, placing a geostationary satellite into orbit requires that the satellite pass through the Van Allen radiation belts. This requires the electronic equipment on board the satellite to be hardened to avoid radiation damage. The hardening of the electronic equipment adds both weight and cost to the orbital vehicle. Added to the cost of the geostationary satellite is the cost required to launch the vehicle. These satellites require large booster rockets such as the Delta, Atlas, Ariane, or Titan [PrB86] to place the satellite into orbit. Another more recent launch approach has been to use the Space Shuttle to place the satellite in the initial orbit and then transfer the vehicle into the geostationary orbit by means of payload assist modules.

Keeping in mind the goals of providing efficient, low cost global communications, researchers have begun to reexamine the possibilities of placing multiple cooperating satellites into orbital planes much lower than those of the geostationary orbit. This concept is not a new one. Research performed over twenty years ago [Wal70] proposed low earth orbiting satellite constellations for global communications. Only recently have advances in microelectronic circuit design and mobile telecommunication technology made it possible to place multiple satellite constellations into orbit. These advances allow for satellite components to be orders of magnitude smaller and more efficient than similar components developed ten to twenty years ago.

The development and deployment of multiple satellite constellations has been intriguing to satellite communications system researchers. Specifically, the deployment of low earth orbit (LEO) satellite systems is of interest for many reasons. First, propagation delays associated with

signal transmission are greatly reduced in comparison to geostationary systems (1 to 10 milliseconds versus 120 milliseconds for one-way propagation). Reduced orbital altitudes permit reduced transmitter power levels at both the satellite and earth station. This, in turn, translates into smaller antennas and associated electronic equipment. Second, low earth orbital planes have lower altitudes than the Van Allen radiation belts. This means that satellites placed at these orbital altitudes are not required to have equipment which is radiation hardened. The results of this factor are lower component weights and fabrication costs. The effects of lower orbital altitudes and smaller system components are also beneficial when it comes to placing the vehicles into orbit. Small orbital vehicles can now be placed into orbit planes by the use of smaller launch vehicles, such as Pegasus, which can be airborne launched.

It should not be misconstrued that low earth orbit satellite constellations are the panacea of satellite communications. Along with the potential benefits of using such a system go distinct drawbacks that must be considered in the engineering analysis. First to be considered is global coverage. The number of satellites required to provide this type of coverage must be determined. Many factors, addressed in Chapter 7, affect this determination. Second, if global coverage is required, decisions must be made on how the communication of end-users is to take place. It must be decided if communications are to be performed via intersatellite links (crosslinks) or to rely on terrestrial systems for the majority of the transmission using the satellite only for "bent-pipe" communications as performed by geostationary systems. If the communications are performed via intersatellite links, analyses must determine the processing capabilities and memory space that the satellite will have. With intersatellite links, individual satellites must have the ability to determine which route a message should take in the transmission of a message from generating source to final destination. A third system consideration is that of tracking. Since low earth orbit satellites do not have the "fixed point in the sky" characteristic of geostationary satellites, earth stations must have tracking equipment capable of maintaining communications with the satellite passing overhead.

As was the case in the examination of interconnection networks for parallel processing systems, performance modeling of satellite systems relies heavily on simulation. Modeling via simulation is required for two reasons. First, and most obvious, is the proof of concept of system design. Where actual system implementations are costly to field and restricted by the economics associated with the system, simulation provides a means for unrestricted "what if" analyses to be made. The second reason for simulation modeling of satellite communication systems results from the dynamic nature of the system. Because of differing transmission rates and time-space dependence of communication links, LEO systems are not well suited for representation by large scale queuing networks.

These factors which affect low earth orbit satellite communication systems provide the impetus for the research to be described in Part II of this dissertation. Specific research goals are provided in the next section.

6.2 Research Goals

The research presented in Part II has goals similar to those presented in Chapter 2 of Part I in that the research described herein is focused on two goals. The first goal is to extend the knowledge base in the area of low earth orbit satellite communications systems performance analysis. A summary of the existing knowledge base is presented in Chapters 7 and 8. The second goal of Part II's examination is to develop and apply mathematical models which can be used to predict the overall packet delay performance of the satellite network.

To extend the knowledge base of performance evaluations of low earth orbit satellite communication systems, it is necessary to define the overall system modeling goals and the figures of merit important to this investigation. The analysis of low earth orbit satellite communication systems is composed of two parts.

The first part examines the performance of this system in a packet-switched data environment. By this, individual users transmit and receive data packets at various points in time.

Because known workloads do not exist for the systems being modeled, possible system workloads are derived for characterization of the system delay performance. The packet-switched data environment offers two important figures of merit for study: network delay and satellite resource utilization. The network delay is the time required by a packet to traverse the system from generating source to final destination. Satellite resource utilization is the percentage of total resources (slots) in use averaged over the period of the investigation. Each of these figures of merit is network dependent parameters.

The second part of the LEO analysis is to develop mathematical models to quantify the delay performance of the system as it is affected by changes in the network dependent parameters. A similar approach to that taken in Part I of this effort is used to develop metamodels for the satellite communication network.

6.3 Summary

This chapter has presented a background discussion on the design factors that the satellite communications engineer must consider when developing a system. Extensive trade-off studies must be accomplished to determine the appropriate system configuration for the application. With a goal of low cost global communications, commercial corporations are examining the potential benefits low earth orbit satellite systems have to offer. These systems also provide for interesting research in the area of system performance. Part II of this investigation analyzes the performance characteristics of low earth orbit satellite communication systems.

Chapter 7 provides an overview of the parameters which affect satellite communication systems. Coverage area geometry is introduced to show the potential surface area covered by a single satellite. Types of satellite orbits are discussed to lend insight to the reader about the numerous orbital geometries used in present systems. Included in the discussion of orbital types are the relative advantages and disadvantages of systems implemented by a particular geometry. Continuous whole-earth coverage by multiple satellites is also discussed along with the factors

affecting this type of coverage. Chapter 7 concludes with an overview of multiple access techniques used in satellite communication systems.

A review of relevant research with low earth orbit satellite communication system applications is the topic of Chapter 8. This chapter touches many different research efforts. Summaries of these works include the modeling and analysis of satellite channels from a multiple access technique point of view. Summaries of studies conducted to determine relative channel capacities are provided. Satellite constellation modeling and analysis examinations are reviewed with limitations to the efforts discussed. Satellite crosslink studies are reviewed to lend insight into the problems associated with particular architectures and scheduling of resources. Problems associated with the routing of messages through multiple satellite systems are addressed and reviewed in the latter portion of Chapter 8.

Chapter 9 discusses the design and modeling of low earth orbit satellite communication systems. In this chapter, the research problem is defined along with the system services, performance metrics, operating assumptions, system design parameters and model factors. A discussion of the simulation model is included in Chapter 9. This discussion includes the model development, testing, verification and validation. Also discussed is the time required to perform the simulations.

Chapter 10 presents the analysis of low earth orbit (LEO) satellite communication systems. Communication link, packet delay, satellite resource utilization, and metamodeling analyses are presented in Chapter 10.

Chapter 11 summarizes the research performed in Part II. This summary includes research effort conclusions and recommendations for future research.

CHAPTER 7

AN OVERVIEW OF SATELLITE COMMUNICATION PARAMETERS

7.1 Introduction

The establishment of a satellite communications system results from numerous analytical and trade-off studies which examine system design parameters. These parameters include orbital geometry, orbital altitude, coverage patterns, multiple access techniques, antenna types, modulation schemes, and error correcting techniques to mention a few. All of these parameters interact in the determination of the satellite-earth link budget and the overall satellite communications system. This chapter presents a discussion of most of the parameters mentioned above. Antenna types, modulation schemes, and error correcting codes are not addressed in this chapter due to modeling and analysis assumptions detailed in the chapters to follow.

Section 7.2 presents the analytic approach to determining the circular surface coverage area that a single satellite can provide. Types of satellite orbits are discussed in Section 7.3. These orbits comprise three basic categories: geostationary, highly elliptical, and low earth. Advantages and disadvantages of using a particular orbit relative to the others are presented. The concept of whole-earth coverage by multiple satellites is the topic of Section 7.4. A general overview of the factors that affect whole-earth coverage is discussed. Section 7.5 covers multiple access techniques for satellite communications. Techniques presented include frequency division multiple access (FDMA), time division multiple access (TDMA), demand assigned multiple access (DAMA), and code division multiple access (CDMA) also referred to as spread spectrum multiple access (SSMA).

7.2 Single Satellite Coverage Area

The coverage area of a single satellite is depends on two factors: the orbital altitude and the elevation look angle (or elevation angle) of the earth station. The orbital altitude is the straight-line distance from the orbiting satellite to the center of the earth minus the radius of the earth (R_E). The elevation look angle is defined as the angle between the local horizontal (referenced at the earth station) and the orbiting satellite. This angle, whose minimum is typically 10°, is the angle required for the earth station to communicate with the satellite [PrB86]. Figure 7.1 shows the instantaneous coverage area of a single satellite.

Mathematically, the coverage area or footprint of a single satellite is given by [INT72]

$$A_s = 2\pi R_E^2 (1 - \cos\theta) \tag{7.1}$$

where

$$\theta = \left[\cos^{-1}\left(\frac{R_E \cos E}{R_E + h} \right) \right] - E \tag{7.2}$$

where R_E is the radius of the earth (approximately 6,370 kilometers), θ is the central angle, E is the minimum elevation angle of the radio path from the earth station to the satellite, and h is the orbital altitude of the satellite.

The spherical coverage area A_S defined by Equation 7.1 increases with the orbital altitude of the satellite, given a fixed elevation look angle E. Consider the following examples. With h equal to 36,000 kilometers (approximately geostationary altitude), and E equal to 0°, A_S is approximately 2.16 x 10^8 square kilometers. This corresponds to forty-one percent of the surface area of the earth. Using a more typical value for E, 10° (elevation angles of less than 10° are subject to blockage due to terrain and foliage), the value obtained for A_S is approximately equal to one-third of the surface area of the earth. Next consider an orbiting satellite with h equal to 700

kilometers. With E equal to 0°, A_S is approximately 25.25 x 10^6 square kilometers or five percent of the earth's surface. If E is increased to 10°, A_S decreases to approximately 2.3 percent of the earth's surface or a circular footprint with radius of 1,934 kilometers. A satellite orbiting at 700 kilometers would have footprint which spans the continental United States and most of Canada.

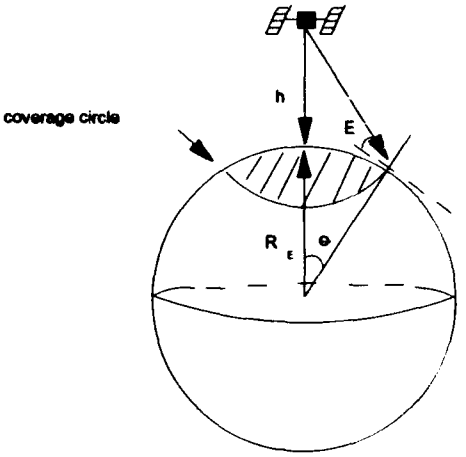


Figure 7.1. Spherical coverage area of a single satellite.

The discussion presented thus far is related to single antenna beam coverage. To increase the capacity of satellite communication systems, researchers have developed multiple beam antennas which are used to cover the same area as single beam antenna systems. The impetus behind this development is two-fold: reduced power levels and frequency reuse. The reduced transmitter power levels result from high gain antennas which have smaller beamwidths. Frequency reuse within a given satellite coverage area is necessitated by the scarcity of available frequency spectrum for satellite communications. System capacity increases in multiple beamed systems result from the reuse of portions of the overall satellite frequency bandwidth. This frequency reuse translates into increases in the number of communication channels available. With

researchers exploring the possibility of using low earth orbit satellite constellations for real-time communications, the reuse of assigned frequencies becomes a more important design issue [Rus92]. Figure 7.2 shows a representative frequency reuse pattern for a multiple beam antenna proposed by Motorola in their Iridium system [Mot90]. The overall beam shown in Figure 7.2 covers approximately 161,000 square kilometers. The total area is divided into 37 cells, each of which covers approximately 580 square kilometers. The frequency reuse factor for this example is seven since each overall satellite coverage area reuses seven different frequencies distributed over the 37 cells.

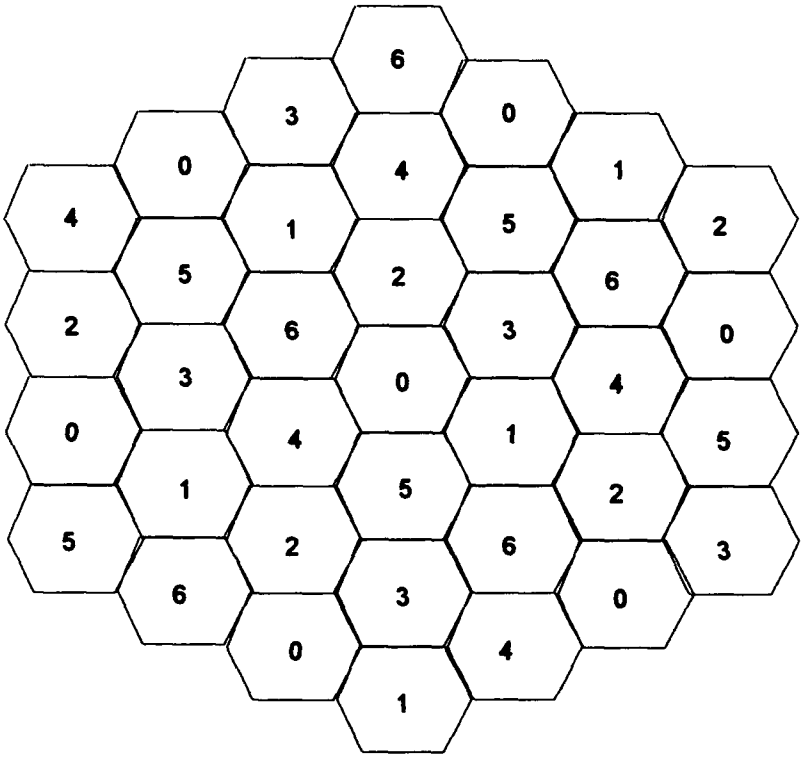


Figure 7.2. 37-cell frequency reuse pattern for proposed Iridium system [Mot90].

The drawbacks associated with multiple beam antennas lie in the complexity of hardware used by the satellite to control the different spot beams. Additionally, earth stations must change frequencies as particular beams of the frequency reuse pattern pass overhead.

7.3 Satellite Orbits

Different strategies exist for classifying satellite orbits. Polaha [Pol89] categorizes satellite orbits into six distinct classifications based on orbital altitudes and geometry. A later presentation by Maral et al. [Mar91] states that satellite orbits can be grouped into three categories. These three categories encompass the six presented by Polaha. The three basic concepts presented in [Mar91] are the geostationary satellite system (GEO), the highly elliptical orbit system (HEO), and the low earth orbit satellite system (LEO). Each system has unique properties, which will be discussed below.

7.3.1 The Geostationary Satellite System A geostationary satellite is placed at an altitude of 35,797 kilometers (km) so that its period of revolution about the earth is exactly one sidereal day (23 hours, 56 minutes, and 4.09 seconds). The satellite orbit, which lies in the equatorial plane, is circular in nature so that from the surface of the earth, the satellite would appear to be at a fixed location in the sky. Because of the fixed location of a GEO system, global coverage can, for all practical purposes, be achieved by using three satellites spaced 120 degrees apart. GEO systems generally do not provide coverage above latitudes of 70 degrees.

The benefits of the GEO system are two-fold. First, world-wide communications can be achieved with a small number of satellites. This translates to minimal number of launches relative to other systems which require many satellites to provide the same coverage. The second benefit of a GEO system is simple tracking or no tracking required by the earth station to locate and communicate with the satellite. The earth station maintains set elevation and azimuth coordinates necessary to communicate with the satellite. Minor adjustment to the coordinates may be required

periodically due to drifts in the orbital location. The level of adjustments depend on the diameter of the earthstation antenna beamwidth and the angular drift of the satellite.

The major disadvantages of a GEO system relative to other systems in lower orbits are transmitter power requirements, propagation delays, radiation hardened equipment (for passing through the Van Allen belts), and cost per launch. With any satellite or terrestrial communications system, the greater the distance between communicating nodes, the greater are the expected transmitted energy losses and propagation delays. To compensate for these expected energy losses, design choices are to increase the transmitter power and/or use high gain antennas. Propagation delays can only be overcome by reducing the orbital altitude of the satellite. With the proposed LEO systems, propagation delays of GEO systems are now more of an issue than in the past. The one-way propagation delay for a GEO system is approximately 120 milliseconds compared to three to four milliseconds for a LEO system. The third drawback to the use of a GEO system is that its electronic equipment should be radiation hardened to avoid damage while passing through the Van Allen radiation belts in transit to the geostationary orbit. This increases both the weight and the cost of the orbital vehicle. The added weight, and orbital altitude require that the launch vehicle be substantially larger than those used to launch satellites of lesser weight and lower orbital altitude. As a result, the cost of launching a GEO satellite is much greater than the cost associated with launching a LEO satellite.

7.3.2 Highly Elliptical Orbit Systems Satellites in elliptical orbits vary in altitudes from approximately 1,000 kilometers at perigee, the closest point to the earth, to approximately 40,000 kilometers at apogee or the farthest point from the earth. HEO system orbits can be inclined relative to the equator to provide coverage for a chosen area. Orbital periods of HEO satellites vary from system to system. Molniya-type systems [Joh88], which have inclination angles of 63° relative to the equator, have orbital periods of 12 hours. Tundra type orbits [Ash88] have 24 hour

orbital periods. HEO systems are designed to communicate best when the satellite is at apogee [Mar91].

Disadvantages to using HEO systems are similar to those of GEO systems. First, the satellite electronic equipment must be hardened to radiation. Second, because the satellite is moving relative to an observer on the earth's surface, a Doppler shift between the transmitted and received signal must be corrected to maintain proper ranging and communication data. A third relative disadvantage of using a HEO system versus a GEO system is antenna pointing. Because the satellite is not stationary relative to the surface of the earth, steerable antennas must be used to maintain coverage over a desired area. In addition, a greater number of HEO satellites must be used to maintain the same continuous area coverage as a GEO system.

7.3.3 Low Earth Orbit Satellite Systems Low earth orbit satellite systems, or LEO systems, maintain orbits in the range of 500 to 1500 kilometers. LEO system orbits are circular and can be either polar or inclined or a combination of the two. In a polar orbit, all satellites within a constellation have orbital crossing points at both the poles of the earth. Pole crossings can be staggered in time to avoid satellite collisions. Inclined orbits are those whose orbital plane is inclined relative to a fixed latitude. Inclination is normally measured relative to the equatorial plane [PrB86].

Advantages of using LEO systems over GEO and HEO systems are reduced propagation delays, increased fault tolerance in terms of satellite failure, frequency reuse, lower transmitter power requirements, reduced weight, and lower launch costs per satellite. As mentioned above, propagation times for LEO systems are on the order of thirty to forty times less than those of GEO systems. Constellation fault tolerance results from the use of multiple satellites for a given coverage area. Since coverage overlaps exist, failure of a single satellite does not dictate the loss of communication coverage. Frequency reuse by a satellite system is almost required due to the limited amount of spectrum available [Rus92]. Because of the lower altitude of LEO systems,

transmitter power levels can be reduced while still meeting link budget requirements. The reduced weight of a LEO satellite (relative to a GEO or HEO satellite) results from reduced power requirements and lighter electronic components since LEO satellite orbits are not in the Van Allen radiation belt regions. All of these factors are additive to result in a satellite which is smaller than GEO or HEO satellites. This in turn allows for multiple launch vehicles such as the Space Shuttle or Pegasus to be used in placing LEO satellites into orbit, thereby reducing the cost per launch.

While advantages of using LEO systems over GEO or HEO systems are numerous, many disadvantages exist. First, as with a HEO system, a LEO system requires a greater number of satellites to provide the same coverage area as a GEO system. Also, LEO system receivers must compensate for Doppler shifts in frequencies due to the satellite's motion relative to the receiver. A third point which makes the use of LEO constellations challenging is the intersatellite links (crosslinks) required to provide communications between geographically separated end users. Knowledge of neighboring satellites must be maintained by communicating satellites along with routing tables or sophisticated methods for message delivery.

7.4. Whole-earth Coverage

It is widely known that three satellites placed in geostationary orbits and spaced 120° apart can essentially provide coverage of the entire earth (with exception of latitudes above 70°). Because of the costs and propagation delays associated with GEO satellites, researchers [Wal70, Wal77, Bes78, Rid85, Rid86, AdR87] have performed studies which examine placing multiple satellites into a single constellation to provide continuous coverage of the entire earth. These research efforts have played an important role in establishing the feasibility of global communication systems using non-GEO satellites. This section presents a summary of such research.

Many design factors are taken into consideration in formulating a system to provide whole-earth coverage. First to be considered is the orbital altitude of the satellites. From Equation 7.1

above, the orbital altitude plays a key role in the coverage area of a single satellite. The greater the orbital altitude, the greater is the potential coverage area of a single satellite, which thereby reduces the number of satellites required to provide whole-earth coverage. Next, the type of orbital geometry of the constellation must be considered. Determination must be made if circular or elliptical orbits and polar or inclined orbits are to be used. Once the orbital altitude of the satellites and constellation geometry has been established, the number of satellites per plane and the number of orbital planes can be determined. Intertwined in this analysis is the relative phasing (angular displacement) of the satellites between adjacent orbits. Last, but certainly not of least importance are the economics associated with providing whole-earth coverage. The cost of satellites and launch vehicles play an important role in determining the final constellation configuration.

Early works of [Wal70, Wal77] form the "classical" foundations of circular orbit analysis for whole-earth coverage. Walker's research , based on spherical analysis of coverage projections onto the earth's surface, examines circular orbital patterns which ensure that every point on the surface of the earth can always be seen by at least one satellite. Central to the research described in [Wal70, Wal77] is the description and use of two orbital patterns: the *star pattern* and the *delta pattern*. The star pattern shown in Figure 7.3, has common nodal crossings (as referred to by Walker) for all satellites in the constellation. Each orbit is co-rotating with adjacent planes except for the first and last orbital planes, denoted A and D respectively. Co-rotating planes A through D allow for phasing between satellites in these orbits to remain relatively constant. The early analysis of Walker concludes that for polar orbits, the star pattern provides optimal single or double global coverage. The second orbital configuration examined by Walker is the delta pattern. The delta pattern is defined as having T satellites in equal-period circular orbits, evenly-spaced and all of which have the same inclination relative to the reference plane. In addition, the satellites of a delta pattern are uniformly distributed among and within the orbital planes [Wal77]. Figure 7.4 shows a four-orbit delta pattern. Walker's research deduces that the delta pattern is preferred over

the star pattern when whole-earth coverage zones require coverage by more than one satellite (i.e., multiple satellites cover the same surface area at a given instant of time). Walker's work provides a coverage analysis for satellite constellations consisting of 24 or fewer satellites positioned in 24 or fewer orbital planes with varied interorbital phasing.

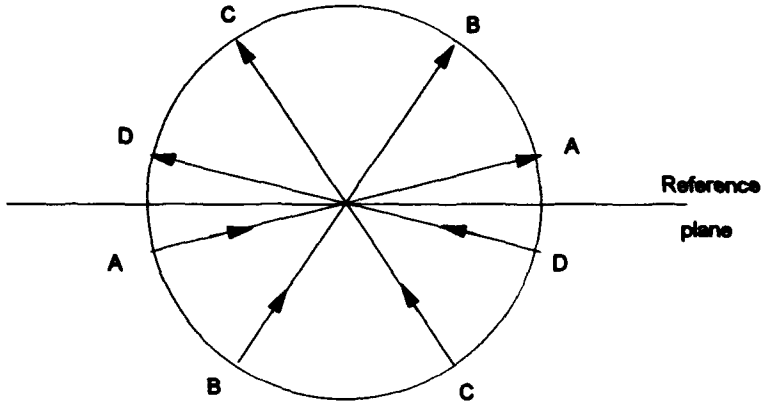


Figure 7.3. Star pattern for 4 orbital planes [Wal70].

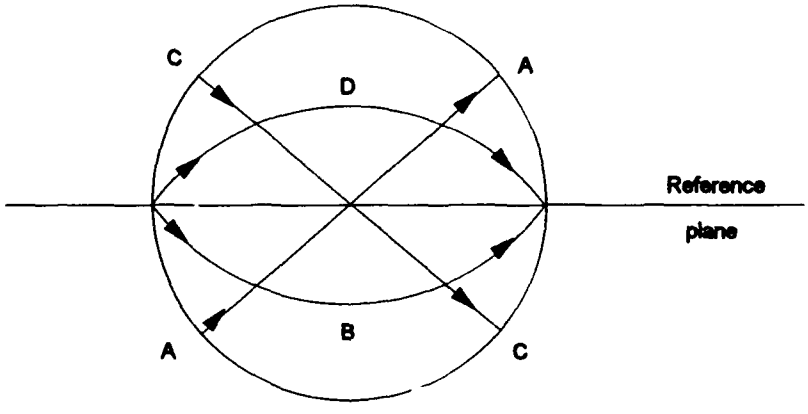


Figure 7.4. Delta pattern for 4 orbital planes [Wal70].

To facilitate the understanding of his work, Walker developed an ordered triple notation to represent a given configuration. This triple, denoted by T/P/F, defines the total number of satellites in the constellation (T), the number of orbital planes (P) in which the T satellites are uniformly distributed, and the relative phasing of satellites in different orbital planes (F). Later researchers [Rid85, Rid86, AdR87] refer to the Walker terminology in presenting their respective works.

Adams and Rider [AdR87] expand previous whole-earth coverage analysis by using the *streets-of-coverage* technique first developed by [Lüd61] and later generalized by [Rid85] to characterize a complete family of minimal polar orbit constellations providing single or multiple coverage above a specified latitude using optimal or arbitrary inter-orbital plane phasing. This technique defines a *street-of-coverage* as the intersection of coverage circles in a constellation which has symmetrically distributed satellites in each orbital plane. Figure 7.5, taken from [AdR87], illustrates this technique where levels of instantaneous coverage can vary from single to 5-fold for the seven satellites shown. Note that the intra-plane satellite spacing is $2\pi/s$ radians, where s is the number of satellites residing in a given plane.

Adams and Rider also address inter-orbital satellite phasing for collision avoidance. To avoid satellite collisions in large polar constellations, deviations from optimal phasing must occur. This is to ensure that no two satellites are at their ascending nodes simultaneously. The use of the streets-of-coverage techniques along with phasing analysis, provides the methodology for Adams and Rider to tabulate the first fifty families of minimal constellations for coverage ranging from single satellite to 4-fold coverage with various sized constellations. These constellations range in size from six to 220 satellites uniformly distributed in orbital planes ranging from two to ten where applicable. This analysis gives the system designer a concise database allowing for various trade-off studies to be performed on the number of satellites, orbital altitudes, levels of coverage, and

phasing of satellites of a particular conceptual system. One such commercial designer, Motorola, has used the results of this analysis for their proposed Iridium system [Mot90].

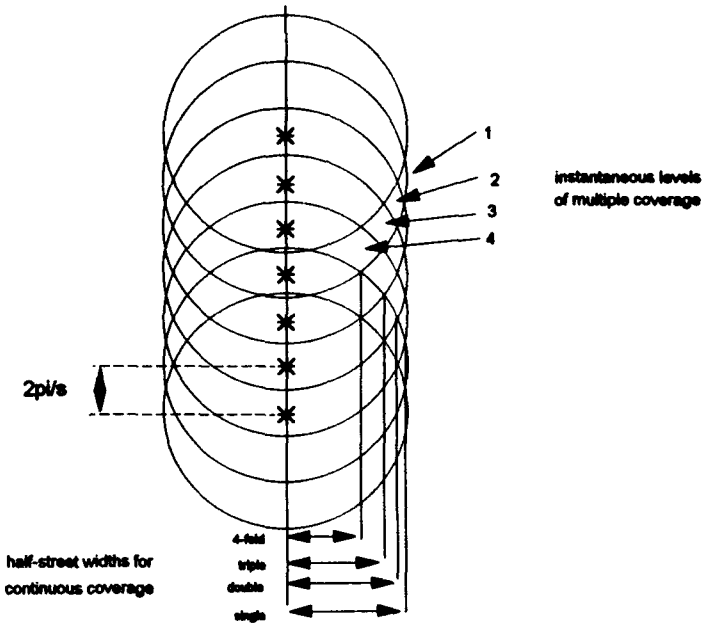


Figure 7.5. Streets of coverage [AdR87].

7.5 Multiple Access Techniques

Multiple access techniques for satellite communications can be categorized into three classes: fixed allocation, demand assigned allocation, and random access [Abr92]. Fixed allocation multiple access can occur either in assignment in time or frequency. Two fixed allocation multiple access techniques are predominantly used in satellite communication systems. They are frequency division multiple access (FDMA) and time division multiple access (TDMA). Demand assigned multiple access (DAMA) is used to overcome inherent efficiency limitations of utilizing transponder capacities for the fixed allocation techniques. Random access techniques

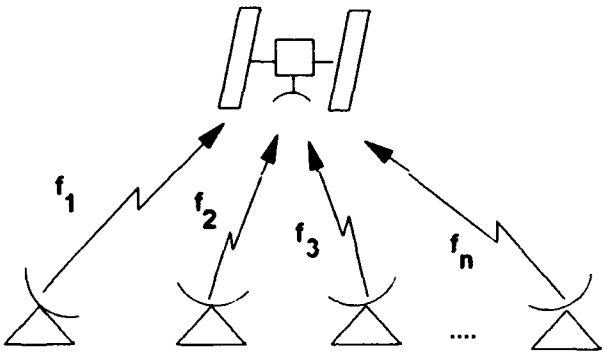
such as ALOHA and code division multiple access (CDMA) are beneficial in systems whose transmission characteristics are fundamentally random and bursty in nature. Each of these techniques is briefly discussed in the subsections which follow.

7.5.1 Frequency Division Multiple Access (FDMA) One of the earliest multiple access techniques used in satellite communications is Frequency Division Multiple Access (FDMA). With this technique, each earth station transmits one or more carriers at different center frequencies to the satellite transponder. Each carrier is assigned a frequency band and an associated guard band. The guard band, a portion of the transponder bandwidth, is used between carrier bands to avoid frequency overlap (interference) between adjacent carriers. On earth, the receivers are tuned to listen for their particular carrier band and then selectively receive the messages intended for them.

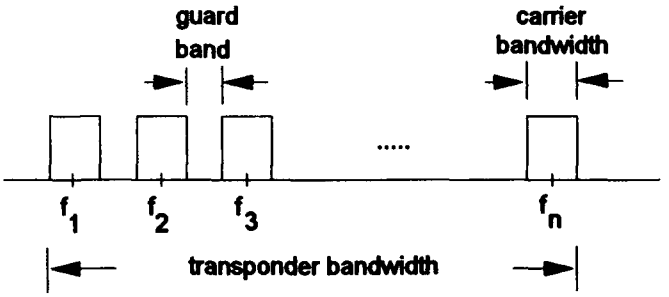
Two main types of FDMA techniques exist. The first type, multichannel-per-carrier (MCPC) transmission, is used for large capacity transmission links. MCPC use single-sideband suppressed carrier channels which are frequency division multiplexed into one carrier baseband assembly. The second predominant FDMA technique used is the single-channel-per-carrier (SCPC) technique. This technique, used when few channels per link are required, is characterized by channels which are independently modulated with separate radio frequency (RF) carriers. SCPC can employ either analog (FM) or digital (Phase Shift Keying (PSK)) modulation schemes [Ha90]. Figure 7.6 shows a typical FDMA system.

7.5.2 Time Division Multiple Access (TDMA) In a TDMA system, the earth stations use a single carrier for transmission to a satellite's transponder on a time division basis. Each earth station has a time burst allocated to it for transmission. For each earth station's burst period, the entire bandwidth of the transponder is available to the earth station. Because the available transmission bursts are multiplexed in time, all earth stations must be synchronized so that transmissions arriving at the satellite are not overlapped in time. Typical satellite transponders

receive one burst at a time, amplify the signal, and then retransmit it back to earth. Earth stations receive the entire satellite burst and extract portions intended for them. Figure 7.7 shows a typical TDMA system.



a.) FDMA uplink concept



b.) FDMA downlink concept

Figure 7.6. FDMA system.

The synchronization of earth station transmissions requires TDMA message overhead (in terms of bits) to be greater than in systems which use FDMA or other types of multiaccess techniques. A TDMA frame consists of reference bursts, traffic bursts, and guard times between the two type bursts. Reference bursts are used to provide timing synchronization for earth stations accessing a particular satellite transponder. They contain information such as the *unique word*

which is decoded by the earth station receiver and indicates the beginning of a TDMA frame. Traffic bursts consist of the information to be transmitted by the earth station. Traffic bursts can be located anywhere within the TDMA frame following the reference bursts. Exact locations of earth station bursts are predetermined by a central burst time plan provided by the coordinating system station. Guard times, as their name indicates, are used between bursts (both reference and traffic) to ensure that earth station bursts do not overlap at the satellite transponder.

To increase the capacity of a TDMA system, designers [BeB87, YaS90, IsM91, MiI91] have extended the TDMA principle to include the use of multiple satellite spot beams. This principle, called satellite-switched TDMA (SS-TDMA), uses the multiple spot beams to provide spatial reuse of a given frequency band. The multiple spot beams allow for the use of parallel uplink and downlink TDMA channels. Additionally, the use of multiple narrow beams provides high gain for a coverage area, thus allowing reduced power levels for both the uplink and downlink channels. The main drawback of SS-TDMA is the increased complexity of coordinating the uplink and downlink beams.

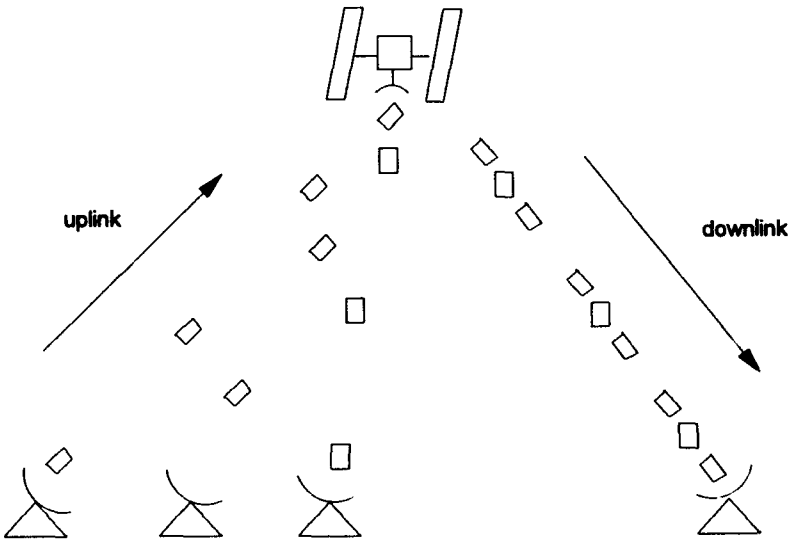


Figure 7.7. TDMA system concept.

7.5.3 Demand Assigned Multiple Access (DAMA) DAMA can be based on either a fixed allocation or a random access. In either case, system bandwidth is only assigned upon request and availability. This means that idle stations will not occupy portions of the system channel capacity as in a fixed allocation system. Assignment on demand results in greater efficiency in use of available channel capacities. When a DAMA system is based on FDMA or TDMA, a portion of the channel capacity is required for allocation requests [Abr92]. These requests must be sent to the system control mechanism which in turn grants or denies the request. The amount of overhead associated with the allocation request is dependent on the number of nodes (earth stations) in the network. Therefore, for a given request word structure, the number of nodes that can exist in a system is limited. When the number of nodes in the network is large (greater than 150), random access allocations appear to be preferable [Abr77].

7.5.4 Code Division Multiple Access (CDMA) CDMA systems, also known as spread spectrum multiple access (SSMA), takes a given signal and spreads it over a much larger frequency band. Each user of the system is allocated a specific code which defines how the signal is to spread across the frequency spectrum in time. Correspondingly, the receiver must also know the spreading code so that the incoming signal can be despread. Two types of CDMA most commonly used are: frequency hopping (FH) and direct sequence (DS) [CaQ93].

The spreading of the signal is performed by a pseudo-noise (PN) sequence which converts the narrow-band signal into a wide-band noise-like spectrum. This PN sequence is a binary sequence with almost random properties. The "almost" disclaimer on the PN sequence's randomness is because the PN sequence is periodic. The narrow-to-wide band conversion is accomplished by modulating the outgoing data signal (rate R_b) with the PN sequence (rate NR_b, where N is the number of elements in the PN sequence). This type of modulating is referred to as *direct sequence* (DS) CDMA. Bits of the outgoing data are decomposed into pieces called *chips*.

Each chip corresponds to an element of the PN sequence. The amplitude (positive or negative) of the transmitted chip is dependent upon the amplitude of its corresponding PN sequence element. In a frequency hopping CDMA system, the PN code controls the frequency of the local oscillator used to mix the input signal and the PN sequence.

The benefits of a CDMA system are three-fold. First, the transmitted signal has a low probability of intercept (LPI) by unauthorized receivers. This results from the PN sequence spreading which requires receivers to know the sequence so that the received signal can be despread. The second benefit lies in power spectral density distribution. By spreading the original signal, the total spectral power is also spread over a much larger frequency range. This causes a CDMA signal to appear as noise to unauthorized receivers, hence hiding the transmitted signal. And third, a CDMA signal is more resistant to jamming signals than the aforesaid FDMA, TDMA, and DAMA techniques. Jammer rejection (JR) is achieved by the despreading receiver. The authorized receiver only despreads the desired signal while spreading unwanted (jamming) signals. As with the spread transmitted signal, the power spectral density of the jamming signal is distributed over a wide band reducing its overall effect on the desired signal being received.

Disadvantages of CDMA are the increased complexity of the transmitter and receiver hardware. In addition, as the number of users in a CDMA system is increased, the noise floor (minimum noise power level of the channel) of the system is increased. This increase in noise floor causes the number of system users to be more limited than in a system implemented with TDMA or FDMA [CaQ93].

7.6 Summary

This chapter has presented an overview of the parameters that affect satellite communications. The information presented above does not constitute a complete discussion of every affecting parameter, but highlights aspects pertinent to previously performed research and the efforts of this investigation which are described in later chapters. Described above are the

basics associated with satellite communications: the orbits, types of coverage, and multiple access techniques. Not covered in this chapter are parameters such as antenna types, forward error correcting codes, modulation techniques, and earth station design.

Section 7.2 presented a discussion on single satellite surface coverage. Maximum coverage areas are shown to be dependent upon the orbital altitude of the satellite and the elevation look angle of the earth station. In addition, Section 7.2 briefly discussed multiple beam antennas and their advantages and disadvantages relative to single beam antennas. Section 7.3 described the different satellite orbits configurations used in satellite communication systems. Design tradeoffs are associated with each of the choices of orbital configurations resulting from performance merits and demerits attributed to each configuration. An overview of whole-earth satellite coverage was the topic of Section 7.4. Classical analysis by Walker and later expansions by Adams and Rider provide insight to the satellite constellation designer. Section 7.5 discussed multiple access techniques for satellite communications. This discussion noted the major characteristics of a particular technique along with potential pros and cons of using a chosen technique. The following chapter provides a review of research performed in the areas of satellite communication channels, satellite constellations, and satellite-to-satellite crosslinks. As will be shown, each of the efforts is affected by the parameters presented in this chapter.

CHAPTER 8

PERFORMANCE MODELING AND ANALYSIS OF SATELLITE COMMUNICATION SYSTEMS WITH LOW EARTH ORBIT APPLICATIONS

8.1 Introduction

Over the past 25 years, there has been an increased interest in the performance analysis of satellite communication systems. With the predominance of operational geostationary satellite systems, researchers have focused on ways of improving the system channel capacities while at the same time reducing the power required to establish reliable communications. These studies have focused on various ways of accomplishing the above stated goals.

The literature review contained herein presents the results of numerous studies in satellite communications. This review is not intended to be a complete report covering every aspect of satellite communications. Synopses of significant research which can be applied to low earth orbit satellite systems for real-time communications are presented.

The review of previously published works starts in Section 8.2 with a look at the satellite channel modeling and analysis performed from a multiple access point of view. Three basic types of multiple access studies exist: time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA). Derivatives of the three also exist. The efforts discussed by this review deal mainly with channel capacities and benefits of their implementations relative to the other access methodologies. Section 8.3 reviews the satellite constellation modeling efforts. In these efforts, performance parameters such as reliability, delay, throughput, and costs (in terms of queuing capacities) are analyzed. Satellite crosslink analysis is the topic of Section 8.4. The concepts of survivability and crosslink architectures are introduced and discussed. In addition, Section 8.4 reviews problems encountered with the scheduling of crosslink access. Section 8.5 presents representative works in the area of message routing in

satellite constellations. Different approaches to efficient message routing are addressed with potential advantages and disadvantages of their use discussed. Section 8.6 briefly discusses proposed LEO systems and pertinent design characteristics of the satellite constellation architecture. The concepts of "little LEOs" and "big LEOs" are also introduced in this section. Concluding remarks are provided in Section 8.7.

8.2 Satellite Channel Modeling and Analysis

In Chapter 7, multiple access techniques for satellite communications were introduced along with relative advantages and disadvantages of their implementations in communication systems. This section presents a review of the research performed in satellite channel analyses and the pertinent aspects which can be applied to low earth orbit satellite systems.

Of primary interest to the satellite channel researcher are link margins and channel capacities. Link margins are the excess power (transmitted power minus the received power) designed into the system to overcome loss effects due to atmospheric conditions, thermal noise, and shading due to foliage, to mention a few. Channel capacities are of importance in providing the bandwidth necessary to meet end-user requirements.

Numerous research efforts have focused on the determination of channel capacities under various operating conditions including [DiB74, GaK74, Abr77, Vit85, JoR88, RaJ88, GiJ90, GiJ91, Abr92].

8.2.1 TDMA channels Time Division Multiple Access channel studies have been ongoing since the late 1960s [SeP68, ScG69, GaK74, Mur74, AnT90, FoO90, GaK91, IsM91, Mil91]. The early studies of Sekimoto and Puente [SeP68], and Schmidt et al. [ScG69] dealt with satellite experiments which touted the efficiencies of TDMA channels to be greater than those of similar FDMA channels. These two experiments were shown to have operating TDMA channels with bit rates ranging from 6 to 60 megabits per sec (Mbps). Gabbard and Kaul [GaK74] later provided a tutorial paper on the advantages and disadvantages of using TDMA versus FDMA while using the

INTELSAT IV system as a representative example. Some of the advantages (single frequency plan) and disadvantages (synchronization of all earth stations) to using TDMA presented by Gabbard and Kaul still exist today.

Researchers have extended the use of TDMA channels by proposing and developing multiple beamed antennas which provide switching mechanisms for the TDMA channels [Mur74, BeB87, YaS90, IsM91, Mil91]. This concept, satellite-switched TDMA (SS-TDMA), presented in Chapter 7, is used to increase the satellite channel capacity by the parallel use of the TDMA channels. The early work of Muratani [Mur74] compares the channel performance of a SS-TDMA system to that of a FDMA system. Muratani shows that for a system with 400 MHz available bandwidth, five frequency channels, and a TDMA frame length of 750 microseconds, a SS-TDMA system has a total channel capacity of 37,900 voice channels (64 kbps/ch) compared to 29,870 voice channels for TDMA/FDMA. This results in an increase in system capacity of over 30 percent.

Yabusaki and Suzuki [YaS90] extend the application of SS-TDMA by analyzing the traffic performance of a variable-channel-per-burst (VCPB) system which dynamically reconfigures traffic bursts. Traffic bursts are assigned to each earth station one-by-one. The authors state that in a VCPB system, idle channels can be shared among all earth through reconfiguration. The reconfiguration of bursts is executed without interruption to calls in progress. The reconfiguration process is performed as follows. Calls from each earth station are assigned the required number of idle channels in its burst based on demand. If the number of idle channels in a burst is less than the number required for the call, idle channels are transferred from other bursts.

The benefits of the reconfiguration results in channel capacity increases over single-channel-per-burst (SCPB) systems. The authors show that for systems with less than 300 earth stations, the channel capacity of the VCPB system is greater than a similar SCPB system.

Two studies performed by Rubin and Zhang [RuZ91, RuZ92], analyze a TDMA channel in circuit switched and packet switched environments. Both studies examine the delay characteristics and queue size requirements of the system. The authors consider a TDMA system consisting of $M = N + L$ slots per frame, where N is the number of slots allocated to the tagged (selected) station, and L is the number of contiguous slots per frame required to reserve a circuit. In [RuZ91], a circuit switched TDMA channel is analyzed. A $Geom^{[x]}/Geom/N$ type queuing system (arrivals and services are forms of geometric distributions) is first analyzed with the results applied to the circuit switched TDMA channel. Upper and lower bounds are derived for the mean message delay and system queue sizes. From these upper and lower bounds, estimates for the mean reservation delay (time required to reserve a circuit) and the mean service request wait time (time from reservation request to obtaining the circuit) can be expressed.

Similar bounds are derived for the packet switched system [RuZ92]. The expected system queue size can be expressed as

$$Q_L \le E(X) \le Q_L + ((N-1)/2)(Lq/(N+L)) \tag{8.1}$$

where q is the geometric distribution parameter, $0 < q < 1$, used to generate the number of packets in a message (q^{-1}), and N, L, are the parameters defined above. Q_L , the system queue size lower bound, is a function of N, L, q, and α (the average number of message arrivals per slot).

Three different estimates for the expected system delay are derived. These derivations are based on the type of distribution used for message arrivals. Distributions used to model arriving messages are the geometric, geometric with geometric branch, and Poisson. Equations for these estimates are provided in [RuZ92].

Jabbari and McDysan [JaM92] examine the capacity requirements of demand assignment TDMA systems under various operating conditions. The first condition, fixed assignment (FA),

preallocates capacity to earth stations according to offered traffic. The second condition, variable destination (VD), has static assignment with alterable destinations. Demand assignment (DA) constitutes the third condition where both the destination and the assignment are alterable. Each of these conditions is investigated with and without using the traffic activity compression technique of digital speech interpolation (DSI). The authors conclude from the comparative analysis of these three conditions that the required capacity of the DA condition is significantly lower than those of the other two conditions for any given offered traffic loading. In addition, it is shown that DA has an implementation advantage over FA and VD in that it can better adapt to changing traffic levels from the earth stations.

8.2.2 FDMA channels The oldest and most commonly used multiple access technique is Frequency Division Multiple Access (FDMA) [PrB86]. Systems such as INTELSAT, TELESAT, DCSC, and Molniya/Orbita in the late 1960s and early 1970s used this technique because of its simplicity, low cost, and ease of adaptation to varying international networks [DiB74]. Current systems such as the INTELSAT VI and VII/VII-A employ FDMA techniques [BrD90, ThS92]. Improved FDMA techniques have resulted in increases in frequency reuse which translates into increases in the system channel capacity.

Shinonaga and Ito perform two investigations [ShI86, ShI92] on the applicability of a subchannel switched frequency division multiple access (SS/FDMA) system for satellite communications. The impetus for their first effort was to enhance the interbeam connectivity of transponder banks of multibeamed satellites operating in an FDMA environment. Interbeam connectivity is sought to improve flexibility in the use of allocated frequencies. The results of a more flexible network are shown in [ShI92]. The authors state that with SS/FDMA, it is possible to separate different types of communication networks (e.g., VSAT (Very Small Aperture Terminal) or business) by allocating different transponder banks for different type carriers. This approach prevents cochannel interference in a low density carrier channel. A second benefit of

SS/FDMA is its ability to limit power densities of carriers within a transponder. This results in lower intermodulation interference. By using SS/FDMA, the authors show that separations on transponder banks can be made between high and low density carriers. The authors develop a proof-of-concept SS/FDMA router used to provide interbeam connectivity flexibility. The mass, volume, and dc power properties of the router, along with a reliability rating of 15 years make the router technically feasible for implementation.

A performance comparison of CDMA, FDMA, and TDMA is made by Wachira et al. [WaB89] in a mobile satellite (MSAT) system environment. The authors find that the system implemented using FDMA techniques exhibits better performance characteristics than those of the CDMA and TDMA techniques. The system characteristics which make FDMA more advantageous are larger system capacities (channels) in a fading environment, flexibility in spectrum coordination and incremental growth, and compatibility with existing systems (e.g., INMARSAT).

Petr et al. [PeM92] analyze an FDMA system from a modeling and simulation approach. The system modeled is a bandwidth-on-demand FDMA channel. Examined in their research are the effects on resource requirements (i.e., number of transponders, channels) caused by different levels of request blocking. Analyses are performed on three levels of blocking, 2, 5, and 10 percent with resulting resource requirements obtained.

8.2.3 CDMA channels Recent studies [JaG88, GiJ90, GiJ91, Abr92] have proposed the use of CDMA or spread spectrum techniques and derivatives in satellite channels. In [JaG88, GiJ90, GiJ91] comparisons are made between CDMA and FDMA. These comparisons examine the spectral efficiencies of spread spectrum CDMA (SS-CDMA) and single channel per carrier (SCPC) FDMA.

Jacobs et al. [JaG88] perform a specific comparison of SS-CDMA to FDMA for the proposed MobileStar system. Their analysis compares the two schemes under two different

scenarios given the following operating assumptions. For the FDMA channel, Quadrature Phase Shift Keying (QPSK) modulation with a 7/8 coding rate (the ratio of information bits to codeword bits) is used. The CDMA coding is 1/3. The authors state that these two types of modulation schemes are thought to be appropriate for use and comparison in the MobileStar channel analysis. The first comparison scenario performs a straight-forward analysis of the spectral efficiency (throughput) based on the modulation schemes. The results show that for large signal to noise ratios, FDMA has approximately four (4) times the channel capacity of the CDMA channel. This is because as the signal to noise ratio becomes large, the CDMA capacity is limited by self noise of transmitted into the channel by end users. The capacity results differ when additional factors are included into the analysis. In the second comparative scenario, voice activation (the percentage of time an end user is speaking (transmitting)), and satellite antenna discrimination (elimination of unwanted signals) factors are included in the analysis. Both the FDMA and CDMA models are affected by the inclusion of these factors. The effects are more pronounced though on the CDMA channel. The use of these factors cause the CDMA channel to have a throughput capacity which is approximately 2.8 times greater than the FDMA channel.

Two later studies [GiJ90, GiJ91] expand the work of [JaG88] by including additional factors: cross-polarization frequency reuse and discrimination between multiple satellites providing coverage of the same area. These two studies apply the use of CDMA techniques to mobile satellite and cellular systems. For comparisons of FDMA and CDMA, [GiJ90] uses the FDMA channel characteristics proposed by the American Mobile Satellite Consortium (AMSC) [AgB88]. As a result of the use of these additional factors, it is shown that the CDMA channel has a throughput capacity seven (7) times greater than the FDMA system proposed by the AMSC.

Abramson [Abr92] also investigates spread spectrum performance in satellite channels by applying the technique to the random access ALOHA method. Part of Abramson's work focuses on showing how the application of spread ALOHA can reduce the complexity of the packet

receiver at the hub station and also simplify synchronization problems by transmitting a single signal from the hub station. The primary difference between the spread ALOHA channel and the CDMA channels described in Chapter 7 lies in how the spreading of the information signal is performed. Where CDMA multiplies a low bandwidth packet by a high bandwidth carrier to spread the signal in frequency, spread ALOHA generates the packet directly at a high bandwidth and then spreads it in time. In a spread ALOHA system, all users employ the same spreading sequence which means that all receivers can use the same matched filters, thus simplifying the receiver hardware. Also, since only one spreading sequence is used, synchronization of transmissions from users is simplified by feedback from the hub station using a pilot of master signal [Abr92].

8.2.4 Combined FDMA-TDMA channels An investigation into the feasibility and applicability of using both FDMA and TDMA in a satellite channel is summarized by Jabbari in [Jab82]. In this type of system, the subchannels are established by partitioning the transponder bandwidth into a number of bands and then using the time sharing properties of TDMA in each of the bands. In this manner, a transponder of bandwidth b can be channeled into k subchannels, where k is the number of distinct carrier frequencies. Figure 8.1 shows the combined structure of the system. The combined structure is still required to have the same frame synchronization as that of a TDMA system.

Jabbari conducts a performance-cost analysis on the combined system. From this analysis, the author concludes that the performance (in terms of channel utilization) remains relatively constant for a low number of system carriers (50 in a 1000 earth station system) and then drops dramatically as the number of carriers is increased. Corresponding increases of the bit rate for the satellites and decreases for the earth stations are seen as the number of carriers is increased. By examining these trends, the author is able to determine the optimal number of carriers for a given system with a predetermined number of earth stations.

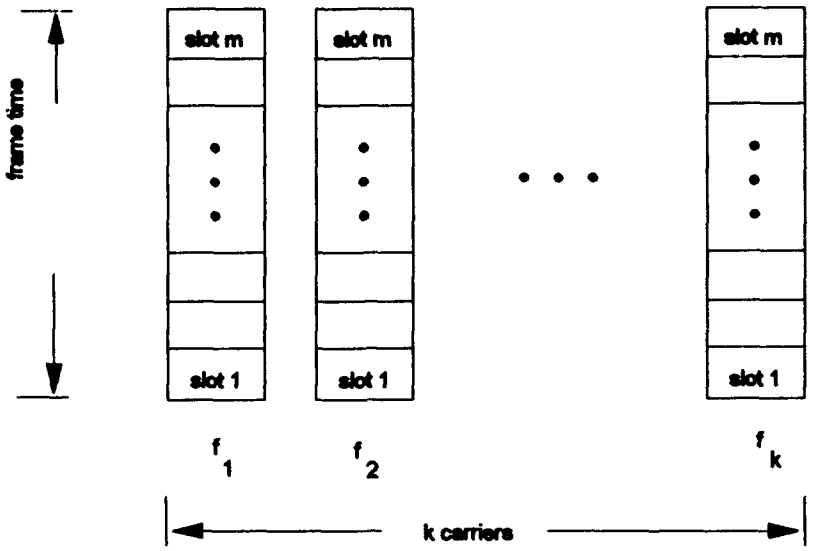


Figure 8.1. Combined FDMA-TDMA structure [Jab82].

8.3 Satellite Constellation Modeling

Satellite constellations can be composed of any number of satellites. The particular number of satellites in a constellation is dependent upon many factors which were addressed in Chapter 7 and also upon the application of the system. Pertinent to this investigation are the previous research efforts performed in modeling and analysis of constellations capable of providing real-time communications. Many performance metrics exist to describe the potential benefits of a particular satellite communication system. First is the overall delay associated with messages as they traverse the network. A second metric is the system throughput which is closely associated with the system delay. Cost is a third metric that designers must examine when analyzing the performance of a particular configuration. Cost can have many different meanings. For the purpose of this investigation, system cost is associated with the number of buffers required to handle incoming and outgoing messages at the satellite and also, the number of satellites used in the

constellation. A fourth performance parameter is associated with the time between a call initiation and the establishment of the circuit. This metric is referred to as the *call set-up time*.

8.3.1 Propagation and Reliability Optimization In Chapter 7 the potential benefits to be gained in the use of low earth orbit satellite constellations were addressed. Foremost of these benefits is the reduced altitude of the orbital vehicle which thereby reduces the propagation delays incurred. Also of importance is the reliability of the network, since in a multiple low earth constellation, the network topology relative to the surface of the earth changes with time. Minimizing the propagation delays and maximizing the system reliability provides the impetus for research such as that performed by McLochlin et al. [McW87]. In [McW87], a topological satellite constellation model is developed to provide single-satellite whole-earth coverage. Satellite constellations within the orbital altitudes of 759 to 11662 kilometers are evaluated with the number of satellite per plane ranging from 3 to 9 and the number of orbital planes varied from 2 to 5. Analytic equations are derived to determine the number of satellites per orbital plane (N_S) and the number of orbital planes (N_P) which optimizes either propagation delay or reliability (in terms of connectivity of the network). The authors conclude that the development of one all encompassing expression which optimizes both propagation delay and reliability is not possible as user defined priorities must be considered in choosing a topology for application.

8.3.2 Throughput Analyses In the modeling and analysis of satellite constellations, a wide range of approaches, based on the multiple access technique used, are possible. This statement holds true for throughput and delay analyses presented in this section. The four studies presented below [ClW87, GaK91, GaL92, KaT92], each use the ALOHA random access technique for their analyses.

Clare et al. [ClW87] describe a multiple satellite LEO communications system study which examines the delay-throughput characteristics using the ALOHA protocol. This effort analyzes two different constellation patterns which provide uniform earth surface coverage. The first

constellation consists of 12 satellites configured in a cuboctahedron (cubo) while the second constellation is composed of 30 satellites in an icosidodehedron (icos) pattern. The 12 satellites in the cubo pattern are uniformly placed three per orbital plane with the four planes having equal altitude and circular orbits. The icos case places five satellites in each of six orbital planes. The orbital altitudes are 1126.5 kilometers for the icos constellation and 2607 for the cubo. These altitudes are chosen to ensure that no intersatellite link is closer than 483 kilometers.

For their delay-throughput analysis, Clare et al. do not model earth station transmissions. It is assumed tha. all traffic originates and terminates at the satellites. In addition, message generations at the satellites are assumed to be Poisson uniformly distributed among the constellation satellites. The network throughput versus end-to-end delay for each constellation is compared for two cases: using random routing and deterministic routing (in this case, the shortest path is always taken). It is shown that with shortest path routing, the network throughput (packets delivered to a destination per second) is greater than that achieved using random routing. It is also shown that the iso constellation has a higher throughput capability than the cubo case. These results are as expected due to the assumptions made by the authors relative to the method in which the messages are generated (uniformly distributed) and the effectiveness of shortest path routing over random routing.

Ganz and Karmi [GaK91] analyze a satellite "cluster" from a performance standpoint to include the metrics of mean delay, throughput, and buffer overflow probability. Their model is a two satellite system with multiple transponders available for uplink and downlink transmissions. Slotted ALOHA is used as the multiple access technique for orbital-to-terrestrial duplex communications. Intersatellite links are modeled using a time division multiplexing (TDM) scheme. Variables in the system model are message loading, traffic distribution, number of uplinks, number of buffers, number of earth zones covered by a given satellite, number of transponders at each satellite, and the intersatellite link capacity. System operating assumptions

are geostationary orbital altitudes (42,000 km), Poisson packet arrival rate, uniform distribution of message destinations within a particular zone, and that packet lengths are exponentially distributed. Analytical approximations are developed for throughput, buffer overflow probability, and mean packet delay. The approximations to these performance parameters are shown in Equations 8.2 through 8.5. Equation 8.2 shows the average delay a message incurs if the source and destination addresses lie within the same earth zone. Equation 8.3 shows the average delay for a message requiring crosslink transmissions.

$$Delay_{avg} = 1 + R + W + E(1 + A + (K-1)/2) \tag{8.2}$$

where R is the round trip propagation delay time (in slots), W is the mean processing delay of the satellite for a packet (in slots), E is mean number of retransmission, A is the acknowledgment delay (in slots), and K is the number of satellites in the system.

$$D_{21} = 1 + .5R_2 + W_{21}^{(2)} + J + W_{21}^{(1)} + .5R_1 + E_2(1 + A_2 + (K-1)/2) \tag{8.3}$$

In Equation 8.3, W_{jl}^{k} represents the mean delay on board satellite k for traffic originating in domain j with destination in domain 1. The approximation for throughput is given as:

$$S_k = \sum_{l=1}^{B} l\theta_k(l) \tag{8.4}$$

where B is the number of buffers at satellite k and $\theta_k(l)$ is the probability of transmitting l packets per slot. The probability of buffer overflow is given as the ratio of H, the mean number of packets rejected due to buffer overflow, to C, the mean number of packets arriving at a given satellite on the uplink. This ratio is shown in Equation 8.5.

$$V = H/C \tag{8.5}$$

By varying the system loading, Ganz and Karmi are able to determine that the system implemented with six on-board buffers is sufficient to obtain maximal throughput. The maximum throughput is determined by the slotted ALOHA access technique to be 0.184 and 0.552 when one and three uplinks are used respectively. Reducing the number of buffers from six to four decreases the maximum achievable throughput from 100 percent to 98 percent. Simulation models and analytical models for throughput, delay, and overflow probabilities are compared for confidence interval correlation. The two models provide a 99% confidence interval of correlation.

Limitations to the work of Ganz and Karmi are in the size of the constellation, the distribution of packet destination addresses (only use uniform distribution), and the traffic distribution matrix (used uniform traffic in five of six examples, last example varied only one entry in distribution matrix).

A later investigation by Ganz and Li [GaL92], extends the earlier works of Ganz and Karmi by using the ALOHA policy to analyze satellite clusters consisting of m satellites. Operating assumptions for [Gal92] are the same as those addressed in [GaK91]. As a result of using an arbitrary number of intersatellite links for their analysis, Equation 8.4 is modified and shown in Equations 8.6 below. Equation 8.6, the throughput from a particular satellite to another, is modified by taking the minimum of either the number of buffers (B_j) or the number of transponders (T_j) in the summation. The intersatellite delay and overflow probability equations are not affected by the increase in crosslinks from 2 to m.

$$S_k = \sum_{l=1}^{\min(T_j, B_j)} l\theta_k(l) \tag{8.6}$$

As was shown in [GaK91], Ganz and Li show that the maximum throughput that can be achieved in the system is 0.55. Also, buffers of size 4 allow for the throughput to reach 98 percent of the maximum. An extension to [GaK91] lies in the investigation of the effects on delay caused by varying the number of transponders. The authors show that for a system with one transponder, severe traffic congestion is encountered. Increasing the number of transponders to three (3) allows for throughput to reach 85 percent of the maximum value.

Another constellation modeling effort uses the ALOHA scheme for the analysis of a low density traffic, low earth orbit system [KaT92]. Kaniyil et al. propose a global message passing network consisting of satellites deployed uniformly at 5000 kilometers in either six or ten nonpolar orbital planes. Elevation angles of 45° are used for ground user access to the satellite. Neither whole-earth coverage nor real-time communications is an issue in their research. Coverage analysis is performed by modeling the system with 30, 60, and 120 satellites. The earth coverage varies in the ten orbit analysis from 72% (30 satellites) to approximately 100% (120 satellites). The six orbit analysis yields coverage percentages ranging from 75% (30 satellites) to approximately 98% (120 satellites). It is shown that the coverage of 60 satellites achieve roughly the same percentage of coverage regardless of the number of orbital planes.

Kaniyil et al. also examine the message routing and throughput estimates. For their analysis, the satellite system is designed for low density traffic such as electronic mail and database accesses. A message can be routed either from initiating user to a single satellite and then to end user or through multiple satellites before reaching the destination. In the case of multiple satellite hops, two different crosslink routing schemes are considered. These schemes are orbit-independent routing and orbit-dependent routing. Routing tables are provided to each satellite for interrogation when messages arrive. Messages routed by the orbital independent method are assured of taking the shortest path (i.e., the minimum number of satellite hops). This is not the case when using the orbital dependent routing method. In the orbital dependent routing mode, messages are routed via

satellites in two steps. First, the message is routed between satellites in the same orbital plane progressing toward the destination satellite. The message then changes orbits when the first orbital plane intersects (or has connectivity) with the orbital plane of the destination satellite. Once reaching the destination orbital planes, the message is routed within the plane until it reaches the destination satellite and is then downlinked to its terrestrial destination.

To estimate the throughput of system, Kaniyil et al. make the following operating assumptions. First, uniform loading across satellite uplinks is assumed. Second, buffers at the satellites are sufficiently large (50) so that uplink and crosslink messages are not rejected due to insufficient storage space. Third, the length of messages equate to one ALOHA slot. And fourth, crosslinks channels are slotted ALOHA in nature. Throughput estimates are broken down into the cross-link throughput (S_C) and the downlink throughput (S_g) expressed in Equations 8.7 and 8.8 respectively.

$$Throughput_{crosslink} = G_c(1-G_c)^{N_c} \tag{8.7}$$

where G_C is the channel traffic of any satellite and N_C is the number of neighboring satellites. The throughput of the downlink is given by:

$$Throughput_{downlink} = (1/\overline{n})G_c(1-G_c)^{N_c} + (1/N)G_g \tag{8.8}$$

where $\overline{n}$ is the average number of satellite hops and N is the number of satellites in the constellation.

Limitations of [KaT92] lie in the distribution of messages throughout the network. While uniform distributions are the easiest to model and understand, real-world representations are not captured by this type of modeling. In addition, as stated by the authors, real-time communications

is not addressed in the paper. This omission is of great importance since the majority of proposed uses of low earth orbit satellite constellations are in real-time communications [Mot90, LoC91].

8.4 Satellite Crosslink Analysis

With the emergence of multiple satellite constellations, there has also been a need to design satellite systems capable of performing intersatellite communications. To facilitate the intersatellite communications, network architectures must provide connectivity in the event of failures by satellites within the network. Because of connectivity issues associated with satellite constellations, increased emphasis has been placed on studies which analytically model the motion of satellite constellations so that crosslink pointing accuracy and tracking can be analyzed [GrK88]. The concept of *survivable constellations* has long been a design requirement of military space systems [Cha89]. Even though Cold War adversarial roles no longer exist among nations, the need for survivable constellations still persists in the commercial communications arena. Satellite crosslinks can be based on either radio frequency (RF) or optical (laser) techniques. In either of these cases, crosslink architectures and transmission scheduling for intersatellite links must be considered.

8.4.1 Crosslink Architectures Binder et al. [BiH87] investigate three candidate crosslink communication architectures for packet-switched, low-altitude, multiple satellite systems (MSS). The first architecture is a frequency-division approach utilizing a cellular concept. The second architecture uses an unsynchronized space-time division approach with directional antennas and random accessing. The third architecture implements an unsynchronized random access scheme and toroidal antennas.

The overall system architecture consists of 240 satellites placed in semi-random orbits relative to each other. The "semi-random" orbits result from the design choice that satellites will not have on-board station-keeping capabilities and have varied launch inclinations. The orbital altitude of the system satellites ranges from 350 to 400 nautical miles (nmi). Each satellite has, on

average, six neighboring satellites for intersatellite communications. The maximum crosslink distance is approximately 1200 nmi. Each satellite has an orbital velocity of 15000 miles per hour and a corresponding orbital period of 92 to 98 minutes. This allows a maximum in-view time to be nine minutes.

The key goal of [BiH87] is to develop a cost-effective MSS crosslink architecture: an architecture which minimizes the crosslink propagation times while at the same time considering the economical effects of large satellite constellations. Binder et al. propose an architecture called *PRS* (Pseudorandom Scheduling) synthesis. This architecture uses directional antennas, time sharing of a single frequency throughout the system, pseudorandom scheduling, and half-duplex operation. The key operation of the PRS is the pseudorandom scheduling. The pseudorandom nature of the scheduling results from the fact that each pair of satellites in a given crosslink uses the same random number seeds and sequences to generate timing for transmission and reception of signals. While one end of the crosslink pair is transmitting, the other end is listening. This technique performs a pair-wise scheduling for each crosslink. It is proposed that each satellite use this pair-wise scheduling to anticipate packet arrivals from neighboring satellites. The anticipation allows for the directional antennas to be pointed to the correct locations. Scheduling for each crosslink is performed independently which allows for potential conflicts among the scheduled times of each satellite's set of crosslinks.

The performance of the PRS is found to be comparable to an ALOHA channel in terms of throughput. PRS suffers greater delays than ALOHA due to the waiting times caused by the scheduling technique chosen. A comparison of transmitter power and bandwidth requirements is made for the same throughput-delay performance. It is found that the PRS approach requires far less average transmitter power (approximately 500 times less) and bandwidth (700 times less) than ALOHA type channels compared under the same performance criteria.

8.4.2 SS-TDMA Crosslinks The use of SS-TDMA for satellite crosslinks has been the topic of several research efforts [Inu81, RaD84, BeB87, Tak87, GaG92]. The interest for this research has been fueled by the desire to provide global communications through a set of small and relatively inexpensive satellites. The key problem addressed by the above referenced works deals with how to properly assign traffic to time slots to avoid conflicts or interference within separate transmissions. The objective is then to schedule all traffic in time slots such that transponder utilization is maximized, which in turn can be achieved by minimizing the overall duration of the schedule.

Inukai [Inu81], examines the intersatellite link (ISL) time slot assignment problem for clusters of two SS-TDMA satellites. These satellites are provided with on-board buffering capabilities for incoming messages. Inukai's analysis shows that the scheduling problem for this type of system can be equivalently reduced to single satellite scheduling problems studied by [Itu77, BoC81, GoB85]. In Inukai's approach, the traffic transmitted over the ISL is stored at the receiver in an ISL buffer and then sent to its destination zone when no conflicts may arise. He concludes that the delays encountered by messages are slightly increased, but the benefits of no retransmissions due to conflicts outweighs the added delays.

Bertossi et al. [BeB87] study the time slot assignment problem for unbuffered satellites. In their operating environment, the authors state that all traffic received by a satellite must be immediately sent to earth zones and over the ISL according to the internal settings of the RF switches at the satellites. As in the works of Inukai [Inu81], Bertossi et al. perform their initial analysis on clusters of two satellites. The authors later show that the results obtained from the two satellite clusters can be extended to a cluster with k intersatellite links. Another important operating assumption made by the authors is to neglect the intersatellite propagation time as a factor in the time slot approximation development. This omission is possible when comparing the propagation time to the time slot duration. The authors develop two heuristic algorithms which

generate "near optimal" scheduling across the ISL. The "near optimal" scheduling results from the fact that the scheduling problem is shown to be computationally intractable and only lower bounds on the duration of a schedule can be achieved. The authors show that as the size of the ISL matrix increases (the possible intersatellite links), the complexity of the problem increases exponentially.

Takahata [Tak87] provides an ISL study based on the INTELSAT traffic database. His work shows that the benefits of using ISL is strongly dependent on the traffic distribution patterns of the network. Further, he shows that transcontinental distances provide significant advantages in the effective utilization of satellite resources at high elevation angles.

Ganz and Gao [GaG92] examine the SS-TDMA slot assignment problem for a satellite cluster with an arbitrary number of satellites connected via ISL. The ISL have arbitrarily assigned bandwidth with each satellite having the capability of covering an arbitrary number of earth zones.

Because Bertossi et al. [BeB87] proved that the optimal time slot assignment problem is NP-complete, Ganz and Gao [GaG92] investigate the possibilities of analyzing this problem by analogy with a classical scheduling problem, the *open-shop scheduling problem* [CoW67]. Briefly, the open-shop scheduling problem attempts to schedule jobs optimally with multiple tasks per job to multiple machines, each of which perform different tasks. In this problem, machines can only have one task assigned. Ganz and Gao modify this problem to fit the satellite cluster SS-TDMA slot scheduling problem. The modifications to the open-shop problem are needed because a satellite may be required to perform more than one task simultaneously (e.g., crosslink and up/downlink transmissions). Heuristic scheduling algorithms are developed and applied to the satellite clusters to determine accuracy of the algorithms. The authors show that the results obtained from the use of these scheduling algorithms are within 10 percent of optimum.

8.5 Message Routing

Efficient message routing methods are important in satellite network applications. With one-way propagation times ranging from a few milliseconds in LEO applications to approximately

120 milliseconds in GEO applications, it is critical that the underlying network efficiently transmit messages to intermediate destinations when applicable. In a dynamic system such as a low earth orbit constellation, routing algorithms and tables must maintain current information pertaining to nearest neighbor and availability of the neighbor. This section reviews some of the methods purposed to accomplish the above.

Brayer [Bra84] introduces a routing concept which can be summarized in two parts. In the first part, the routing algorithm calculates a shortest path (minimum data links that must be traversed) between a given source and destination address pair. The second portion of Brayer's concept allows the nodes (satellites) to learn the connectivity of the network. In this connectivity "learning" approach, the author assumes that the satellite network has no prior knowledge of the connectivity of adjacent satellites. At system start-up, satellites transmit a start-up message indicating its network identification. After transmitting the start-up message, a given satellite "listens" for start-up messages transmitted by other satellites in its line-of-sight area. After a given amount of time, a satellite will know the identification of all the other satellites within its communications area. With this approach, satellites will only know their immediate neighbors and have no information about the connectivity of the satellites beyond its line-of-sight view. This approach is similar to that used by present-day workstations to determine their network connectivity [Com91]. The drawback to this approach is that until all possible paths have been established and updated in a satellite's routing table, the satellite has two inefficient methods for transmitting the message to the next satellite: randomly select a satellite to transmit to or broadcast the message to all satellites within its communications area. Both approaches have inherent problems. With the random selection method, the wrong satellite may be chosen and longer paths to the destination address could result. Flooding or broadcasting the message to all satellites has the problem that it consumes bandwidth. Even with the drawbacks associated with the message

routing and path determinations, the author states that, because each satellite has its own routing table, the loss of a single satellite will not cripple the network.

The message routing approach used in the system study of Clare et al. [ClW87], described above in Section 8.3.2, takes a different approach when considering message routing. In that study, it is assumed that every node (satellite) knows the connectivity of the network. This assumption requires that the routing tables be complex in nature due to the dynamic structure of the network. For comparative purposes, Clare et al. also study the effects of random routing versus the deterministic routing which results from all nodes knowing the network connectivity. The authors show, as would be expected, that deterministic routing outperforms random routing in terms of average hop distances as the end-to-end distance between source and destination nodes is increased. The authors do not address the cost tradeoff associated with maintaining an up-to-date routing table at each satellite.

Chakraborty [Cha89] investigates the message routing problem by examining the potential benefits and drawbacks of using dynamic routing within a satellite network. His approach assumes that each node (satellite) individually calculates the cost of every possible path from source to destination and then a least cost route is chosen. The factors which affect cost are the internodal distance and the perceived congestion of the links. This routing approach decentralizes the control for routing but increases the computational workload and memory requirements at each satellite. Chakraborty assumes that all paths are not of the same length and that satellite utilization varies based on whether or not the satellite is part of an optimal (in terms of cost) path. A potential cost benefit approximation is derived and shown in Equation 8.9 below.

$$P \simeq (1+z)\rho[1-\{(1+z)\rho\}^{(n-1)}] \tag{8.9}$$

where z is an improvement factor over assuming that all paths have the same distance and all satellites have equal utilization, ρ is the satellite utilization factor, and n is the number of satellites in the network. In a representative example, Chakraborty shows that for low (10-20 percent) or high (70 percent or greater) satellite capacity utilization, dynamic routing does not necessarily provide the most cost effective routing method.

Though packet switched networks are generally not well suited for voice communications due to variable delays in receiving voice packets [BeG92], Chakraborty introduces a Fixed Path Protocol (FPP) for handling voice transmissions in a packet switched environment. This protocol, which is very similar to virtual circuit switching, uses a signaling message to setup the path for a call. Pointers are set at the satellites to indicate the appropriate path for each voice packet. The FPP protocol is used in conjunction with the dynamic routing approach described above to ensure that the optimal path is chosen. This is proposed to eliminate the delay variability of packets arriving at the destination address. Further refinements are made on the packet switched network for voice by making it a priority-oriented demand assignment (PODA) network. The author addresses two distinct reservation mechanisms in the PODA scheme: one for data and one for voice. The author concludes that by using PODA under the assumptions made above, network throughput can be increased with data and voice packets utilizing the network.

8.6 Proposed LEO Systems

Within the past four years, there have been numerous applications filed with the Federal Communications Commission (FCC) by private organizations to provide LEO mobile satellite communication systems [Rei92]. These systems propose to provide a range of services from radio location to global cellular telephony.

LEO systems are generally categorized by their transmission frequency. Those systems whose transmission frequencies are less than one gigahertz are considered to be "little LEOs." These little LEOs are characterized by low-cost and low-data rates (less than 10 kb/s), and provide

two-way digital communications and position locations services. Systems which operate above one gigahertz are commonly known as "big LEOs." Big LEOs propose to provide worldwide services in personal communications similar to services provided by cellular terrestrial systems. The sections which follow briefly discuss nine proposed LEO systems whose applications have been filed with the FCC. Design specifications for these systems have been taken from [Rei92] unless otherwise noted.

8.6.1 Leosat The Leosat system is a little LEO proposed by Leosat Incorporated of Ouray, Colorado. The system architecture consists of 18 satellites placed in three orbital planes. The orbital altitude of the Leosat system is approximately 1000 kilometers. The proposed operating frequencies are 148-149 MHz for the uplink and 137-138 for the downlink. Primary services to be provided by Leosat are two-way data communications and radio locations of equipped vehicles.

8.6.2 Orbcomm The Orbcomm system is proposed by Orbital Communications Corporation of Fairfax, Virginia. The fully operational system will contain 20 satellites located in five circular orbiting planes. Two of the orbital planes will be polar in nature. One satellite will reside in each of the polar orbits. The remaining 18 satellites will be placed in three equally spaced, equally distributed, inclined orbital planes. Inclination angles for these planes range from 40 to 60 degrees [Har91]. The orbital altitude of the 20 satellites is approximately 970 kilometers. Continuous coverage is proposed for latitudes less than 65 degrees. As with the proposed Leosat system, the operating uplink and downlink frequencies are 148-149 MHz and 137-138 MHz, respectively. Basic systems services to be provided by Orbcomm include two-way data communications and position determinations. The Orbcomm system is classified as a little LEO.

8.6.3 Starnet The Starnet system, a little LEO, is proposed by Starsys Incorporated of Washington, D.C. The system constellation will consist of 24 satellites placed in 24 random orbits. The orbital altitude of the constellation is approximately 1300 kilometers. The uplink channel is designed to operate in the 148-149 MHz range, while the downlink operates in the 137-

138 MHz range. Proposed system services include global two-way data communications and radio location.

8.6.4 Vitasat Vitasat is proposed by Volunteers in Technical Assistance (VITA) of Arlington, Virginia. This little LEO system will contain two satellites orbiting in a single circular plane. The orbital altitude of this plane is 800 kilometers. Operating frequencies for Vitasat have not been firmly established. Proposed combinations for frequencies are either 400.2 MHz uplink and 137.7 MHz downlink; or 149.8 MHz up and 400.2 MHz down. Services proposed by Vitasat include data and file transfers geared primarily for use by developing third world countries.

8.6.5 Constellation Constellation, originally named Aries, is a big LEO system proposed by Constellation Communications Incorporated (CCI), of Herndon, Virginia. This system will contain 48 satellites operating in four inclined planes (the original FCC filing called for four polar planes). The satellite orbital altitude was originally proposed to be 1,018 kilometers but design changes to the system now place the altitude at 2,000 kilometers. This orbital altitude is proposed to eliminate mobile to satellite communications with viewing angles less than 15 degrees [Kla94]. Mobile communication frequencies are 1610-1625.5 MHz uplink and 2483.5-2500 MHz downlink. Gateway communications are at 5150-5216 MHz for the downlink and 6525-6541 MHz for the uplink. Services proposed by Aries include position determination, two-way telephony, facsimile, and data collection, distribution, and control services. The proposed multiple access technique is spread spectrum/code division multiple access (SS/CDMA). The satellites which comprise Constellation function as simple "bent pipe" relays with no inter-satellite communication links. Since its original filing with the FCC, design changes by CCI have increased the voice channel capacity of Constellation by 10-fold, to over 1,000 channels.

8.6.6 Ellipso Ellipso, another big LEO system, is proposed by Ellipsat of Washington, D.C. The original constellation design consisted of 24 satellites operating in three highly elliptical orbits. Each orbit was characterized by an apogee at 2903 kilometers and a perigee of 426

kilometers. Subsequent design changes to the system have changed both the number of satellite and the operating orbital planes of Ellipso. The present configuration calls for a 16-satellite system operating in two different types of planes. The first plane type, called Borealis, has an apogee of 7,800 kilometers and a perigee of 520 kilometers. These figures yield an orbital period of approximately three hours. The second orbital type to be used by Ellipso is an equatorial elliptical orbit, referred to as Concordia. This orbit has an apogee of 7,800 kilometers and a perigee of 4,000 kilometers which provides an orbital period slightly greater than three hours [Kla94]. The operating frequencies of Ellipso are 1610-1626.5 MHz uplink and 2483.5-2500 MHz downlink. Proposed services are to connect to a cellular phone to convert 800 MHz cellular to the 2.5/1.6 GHz RDSS bands. A form of spread spectrum which also uses FDMA is proposed as the multiple access technique for this system. Twenty-four hour coverage for mid-latitude regions north and south of the Equator is proposed.

8.6.7 Globalstar The Globalstar system is a big LEO to be developed by Loral Cellular Systems Corporation, of New York, New York. A fully operational Globalstar system will contain 48 satellites. These 48 satellites will be uniformly distributed among eight circular orbital planes. Each plane has an inclination angle of approximately 52 degrees with inter-planar phasing of 7.5 degrees [Rou93]. The orbital altitude of the constellation is 1389 kilometers. The Globalstar system operational concept is to provide voice and data communications in a cellular or mobile environment. Communications between mobile users occurs through uplinks with in-view satellites which in turn downlinks the information to the nearest system gateway. The system gateway then determines the location of the gateway nearest the final destination user. Connection is made through the local Public Switched Telephone Network (PSTN) with the information being routed to the next system gateway. From this gateway, the information is uplinked to the satellite whose footprint covers the destination user. The covering satellite then downlinks the information to the destination user. The operating frequencies for the Globalstar system are 1610-1626.5 MHz

uplink and 2483.5-2500 MHz downlink for the mobile links while 5199-5216 MHz uplink and 6525-6541 MHz downlink are used for the gateway links. A SS/CDMA multiple access technique is used for all communication links.

8.6.8 Iridium The Iridium system, proposed by Motorola Incorporated of Chandler, Arizona, is being designed to perform communication tasks unparalleled by competing manufacturers. The system, whose application was originally filed with the FCC in December 1990, is continually undergoing design changes in attempts to optimize the system architecture. The original system design was envisioned to incorporate 77 satellites, uniformly distributed over 11 orbital planes. Baseline changes reflected in [LeM93] presently have the constellation consisting of 66 satellites still operating in 11 orbital planes. Orbits for the Iridium system are circular in nature with inclination angles approximately 86 degrees (this is a change from the December 1990 filing which maintained that polar orbits would be used). The operating altitude of the constellation has been set at 765 kilometers. L-band communications is proposed for the mobile-to-satellite links (1.6 GHz). Gateway-to-satellite links are being designed to operate in the K-band region (30 GHz uplink, 20 GHz downlink). The unique communications aspect associated with Iridium is its proposed use of inter-satellite links to provide global data communications and telephony. Inter-satellite links are being designed to operate at 23 GHz. The system is claimed to provide single satellite, global coverage for latitudes below 70 degrees. Cellular beams on-board the satellites will allow for critical spectrum to be reused approximately 5-6 times. The multiple access technique to be employed by Iridium is a combination of FDMA and TDMA.

8.6.9 Odyssey The Odyssey system is proposed by TRW Incorporated of Redondo Beach, California. The satellite constellation will consist of 12 satellites, uniformly distributed among three inclined orbital planes. The orbital altitude for this system is quoted at 10,370 kilometers [Rei92]. The Odyssey system is proposed to use the same mobile frequency bands as Aries and Globalstar described above. The gateway links are being designed to operate at 19,700-20,000

MHz (downlink) and 29,500-30,000 MHz (uplink). The primary services to be provided by Odyssey include voice, radio location, messaging, and data communications. SS/CDMA is the proposed multiple access technique for the system.

8.7 Summary

In this chapter, a review of the previously published research efforts in the performance modeling and analysis of satellite communication systems has been presented. The focus of this presentation was to summarize the efforts that have low earth orbit applications. Very few of the above referenced works were performed in a low earth orbit application environment.

Section 8.2 discussed the satellite channel modeling and analysis. The research conducted on four types of multiple access channels was addressed. First, TDMA modeling and analysis efforts which examined capacity requirements were presented. This presentation was followed by satellite-switched TDMA analyses which looked at capacity improvements over TDMA and demand assignment methods which better utilize available resources. The TDMA discussion was followed with similar presentations but from a FDMA viewpoint. CDMA channel analyses were then presented with the main points being increased channel capacities over FDMA channels. Section 8.2 concluded with brief summary of an effort which combined FDMA and TDMA techniques for satellite channel implementations.

Constellation modeling and analysis was the topic of Section 8.3. Various studies were performed on propagation and reliability optimizations and throughput analyzes. Satellite crosslink architectures and scheduling problems associated with SS-TDMA crosslinks were examined in Section 8.4. Proposed methods for intersatellite message routing were detailed in Section 8.5. Potential advantages and limitations of these methods were addressed.

Proposed LEO systems were the topic of Section 8.6. Four "little LEOs" and five "big LEOs" were briefly discussed. The discussion presented design features associated with orbiting

constellations, proposed operating frequencies, and primary services to be provided. Each of these systems propose to provide services unparalleled by competitors.

Many limitations to the works presented above were observed and noted. These limitations, coupled with the scarcity of open literature which addresses low earth orbit communication systems, provide the motivation for the research to be performed in Part II of this dissertation effort.

CHAPTER 9

LEO SATELLITE NETWORK DESIGN AND MODELING

9.1 Introduction

Part II of this research effort consists of the modeling and analysis of low earth orbit satellite communications systems. Section 9.2 describes the research effort along with the analysis scope and limitations. System services of the networks being analyzed are discussed in Section 9.3. This discussion is followed by performance metric definitions presented in Section 9.4. Network operating assumptions are detailed in Section 9.5. System design parameters are defined in Section 9.6. The system design parameter definitions are followed by a presentation of design factors. These factors are presented in Section 9.7. Section 9.8 presents a discussion on the simulation model itself. In this section, the operations and functions of the main components of the simulation model are addressed. The verification and validation of the simulation model are detailed in Sections 9.9 and 9.10. In these sections, the approaches taken to ascertain the correctness of the model are discussed. Section 9.11 concludes this chapter with a summary of the information contained within.

9.2. Problem Definition

This investigation performs a utilization and delay/throughput analysis for low-earth orbit satellite communications networks. The utilization performance metric is defined as the amount of time a given satellite is busy performing message processing tasks out of the total in-view time the satellite is over a specific geographical location. The delay/throughput metrics measure the number of packets/messages transmitted and received per unit time along with the associated delay caused by the network.

Mathematical metamodels are sought which concisely capture the performance of the network. The purpose of the metamodel development is to provide detailed information to the satellite system designer for insight into the effects on system performance caused by factor variations.

Systems such as the ones to be evaluated present new challenges in satellite communication design and development. These challenges include the design of multiple cooperating satellite constellations providing inter-satellite communications, advanced multi-beam antenna design and control, more complex earth station tracking requirements, and sophisticated beam switching for cellular-type coverage and frequency reuse, to mention a few. All of these factors add to the complexity and cost of the system design.

The following subsections define the scope and the limitations associated with this research effort. Subsequent sections provide a detailed description of the system parameters and factors.

9.2.1 Geographic Area of Investigation This investigation focuses on the network communications performance over the geographic locations encompassed by the latitude and longitude coordinates 20°-50°N and 65°-130°W. These coordinates bound and enclose the continental United States, its coastal waters, Cuba, a small portion of Canada, and a large portion of Mexico. The bounds are chosen to investigate the performance in a well-populated, industrialized region of the world. Locations such as Europe could have as easily been chosen as a representative area of study. Additionally, these coordinate bounds limit the number of inter-satellite hops required by a packet/message in traversing the network from generating source to final destination.

9.2.2 Communication Systems Architecture The communications network consists of three parts: the satellite constellation, earth station gateways, and individual user units (IUU). The satellite constellation comprises multiple communications satellites operating in low-earth orbit environments (500-2000 kilometers in altitude above the earth's surface). These satellites form the

backbone for inter-user communications. The communication networks contain two gateway stations, one located on the east coast of the United States at approximately 77° W longitude and 39° N latitude (a close proximity to Washington DC) and the other on the west coast at coordinates 118°W, 34°N (the approximate location of Los Angeles). The decision to model two gateways rather than some other number of gateways is to allow for the performance analysis to examine "worst case" distances between gateways. The models to be described later can easily be modified to include additional gateways. The gateway stations function to connect Individual User Units (IUU) with the Public Switched Telephone Network (PSTN). The coverage area is graphically depicted in Figure 9.1. Design parameters associated with the gateway stations are summarized in Section 9.6.1.1. The IUUs are similar in nature to cellular telephone units with the capability of providing both voice and data communication services. IUU design parameters are listed in Section 9.6.1.2. Satellite design parameters are addressed in Section 9.6.2.

9.2.3 Research Scope In the systems being evaluated, two types of user information switching (the transferal of information from one system resource to another) are being performed. The first type involves switching within a given satellite. This type of switching relies on interchangeability of antenna beams to maintain user communications while the satellite is within the viewing range of the user. The second type of switching being performed is intersatellite switching. Intersatellite switching is required for two reasons: the distance between two users, and the motion of the satellites relative to the user. The analysis performed in this investigation considers beam switching as an internal operation of the satellite. This investigation does not evaluate beam utilization or the possible interference caused by adjacent beams (other than to add a carrier-to-interference ratio (C/I) factor into the link analyses). These exclusions are not meant to downplay the importance of beam switching and its affects on system performance. Rather, the intersatellite switching is considered as having a much greater affect on the overall performance of the network.

Figure 9.1. Geographic area of investigation.

Additional research scope limitations are made with regards to multiple access techniques, digital modulation schemes, and error correcting code implementations. Many multiple access techniques (e.g., CDMA, FDMA, TDMA, Random Access, etc.) have been either implemented or proposed for satellite communication systems. Each of these techniques has relative advantages and disadvantages when compared to other approaches within the class. This study assumes that TDMA is used as the multiple access technique. This decision is based on a goal of efficient utilization of channel resources. Similar choices are made for QPSK modulation and convolutional coding of the information stream.

9.3. System Services

The system services modeled and analyzed operate in a packet-switched, real-time environment. Real-time communications in this system is defined as information receipt (whether

it be data or voice) no later than 400 milliseconds after transmittal. Data communications are supported within the scope of this investigation.

9.4. Performance Metrics

Numerous performance metrics can be analyzed in a study such as this one. Two important metrics are analyzed without limiting the importance of the research. These metrics are network delay and satellite utilization rates. These two metrics are chosen because of their importance relative to system performance and system costs. Network delay is defined as the time required by a packet in its traversal of the network from generating source to final destination. Satellite utilization is important from an economic as well as a performance standpoint. The satellite utilization metric is defined as the percentage of time that the satellite is busy processing packets. Both of these metrics are affected by the system factors defined below.

9.5. Operating Assumptions

The modeling and analysis of satellite communications networks requires that numerous system operating assumptions be made. These operating assumptions are representative of similar decisions made by commercial corporations in designing their proposed systems [Mot90, LeM93].

9.5.1 Satellite Coverage The network is designed to provide whole-earth coverage in a low earth orbit environment. The system is designed such that overlapping satellite coverage occurs only at latitudes above 70°. For this investigation, the geographic area of interest lies within the North American bounds specified above. The entire bounded area may require more than one satellite for full coverage, but no area within these bounds shall be simultaneously covered by more than one satellite.

9.5.2 Period of Evaluation of the Network The network is designed for continuous operation, 24 hours per day, seven days per week. The period of evaluation for the network shall be a ten to twelve minute interval in which a representative satellite is located over the bounded

geographic area. This period of time is chosen to correspond with the approximate in-view time of the satellite. Approximate in-view times for satellites in each of the constellations evaluated are summarized in Section 9.7.

9.5.3 Simulation Epoch One epoch is used across all the simulations performed. This epoch is randomly chosen to be 10:00 am on July 1, 1993. The epoch is used for initial satellite and mobile transmitter location determinations.

9.5.4 Traffic Distribution Distinct data traffic generation distributions are used for gateway and IUU transmissions. Presently, LEO systems, such as those being evaluated, do not exist. Because of this lack of application data, the traffic distributions described below are based on a "best-guess" scenario along with cellular radio issues addressed in [Lee89]. The packet-switched operating environment assumed for this investigation ideally translates into a data communications system environment. Four types of communication links exist in this system. The first is a gateway to gateway link. This type of link has as its source, a gateway station with another gateway station serving as the destination. The second type of link is a gateway to IUU link. Once again the source is a gateway but the destination is an IUU. The two other links are IUU to gateway and IUU to IUU. The four links are modeled rather than two due to the differences in transmission rates.

9.5.4.1 Source Generation Rates Two Poisson processes are used to control the generation of packets by the system. The choice of Poisson processes for traffic distributions results mainly from the fact that data workloads for the systems being examined are unknown. The Poisson processes are used to provide a possible workload characterization of the system. For comparative purposes, a constant traffic distribution is also used for the 36-satellite system.

A separate process is used for both gateway sources and IUU sources. The generation rates are based on the number of TDMA channels and slot capacities of a single beam in one system satellite. A detailed description of these rates and capacities are provided in Sections 9.6 and 9.7.

9.5.4.2 Source Address Distribution Source addresses are uniformly distributed among the earth-based transmitters. The distribution of IUU and gateway addresses per unit time is controlled by the TDMA channel structure, as well as the number of channels assigned to each type of transmitter and associated data rates.

9.5.4.3 Destination Address Distribution The following distribution is used for destination address generations. Of the total traffic generated during the examination period, approximately 80 percent of the generated traffic is destined for termination at the gateway stations (40 percent at each station) with the remaining 20 percent destined for particular IUUs. Equivalent source and destination addresses are not permitted for transmission. The large percentage of traffic associated with the gateways is due to the intended purpose of the network, to supplement the PSTN and not replace it. To measure the performance of the network, statistics for message times associated with the gateways begin and end with the generation and termination of the message at the gateway, respectively. In an implemented system, statistics would include the additional time of the PSTN in processing messages.

9.5.5 Satellite Link Availability The satellite link is designed to provide an availability of 99.5 percent. The link analyses presented later in the dissertation are performed in a clear sky environment.

9.5.6 Message Routing Message routing through the network is based on shortest path principles. The shortest path between any communicating source-destination pair is defined as the route taken to minimize the number of inter-satellite hops, thereby minimizing the total physical length a message must propagate. Routing tables are predetermined based on the configuration of the system.

9.5.7 Multiple Access Technique Time division multiple access (TDMA) is chosen as the satellite access technique for this investigation. This choice is based on similar decisions made by commercial corporations [Mot90, LeM93].

9.5.8 IUU Locations At system startup, the locations and number of the IUUs are known. These locations shall be randomly distributed over the coverage area. The random distribution of IUU locations is chosen for the purpose of modeling a proposed system as accurately as possible. Directional movements of a specific IUU is considered to be random and independent of other IUUs. IUU movements are periodically updated during the network evaluation period.

9.5.9 Minimum Look Angles The minimum look angles for the gateways and the IUUs are 10° relative to the local horizon. This figure is chosen to minimize the effects of vegetation and terrain on the propagation path.

9.5.10 Satellite Crosslink Communications Each satellite has the ability to communicate with the nearest neighboring satellites forward and aft in its orbital plane. In addition, crosslink communications are possible with the nearest neighbor satellite in each of the adjacent orbital planes.

9.5.11 Bit Error Rate (BER) The minimum BER for data communications modeled in this investigation is 10^{-6}. This rate represents a commonly referenced uncoded value for data communications [PrB86].

9.5.12 Control of Satellite Capacity Capacity control is distributed. Each satellite within the network possesses the logic and memory required to maintain an idle-busy table of its particular satellite channel pool.

9.5.13 TDMA Frame Length The TDMA frame length is chosen to be 60 milliseconds. This value is derived from a proposed system [Mot90, LeM93] and is representative of a LEO class system.

9.5.14 Packet Lengths Packet lengths are fixed for this evaluation. The size of the packet is 1024 bits. This figure is based on the product of the TDMA frame time (60 milliseconds) and the burst bit rate (180 kilobits per second) divided by the number of IUU users per frame (11).

The computed value for the packet length is 981 bits but is rounded up to 1024 for ease of processing and decoding by the receiver. These values are discussed further in Section 9.6.2.10.

9.5.15 System Queues All system queues are first-in-first-out (FIFO). Satellite input queues have finite capacity with the size set at 100 packets in length. The overall memory space for an individual queue is 1 Mb. This value is representative of proposed systems [Mot90].

9.5.16 Satellite System Capacity The overall capacity of a single satellite, in terms of number of channels, is fixed across the constellations evaluated. Derived values are presented in Section 9.6.2.10.

9.5.17 Packet Retransmission Packets which are blocked due to insufficient storage space at the satellites shall be dropped with no retransmission. The number of blocked packets shall be tabulated with the resulting value used to determine blocking probabilities. In implemented systems, blocked packets must be resubmitted for transmission. The approach of dropping blocked packets is chosen so that the measured performance of the network is independent of the retransmission protocol being used.

9.5.18 Satellite Beam Independence The example system used to derive the design parameters associated with each satellite consists of 37 antenna beams per satellite. Each beam is assumed to operate independent of other beams residing on the same satellite. This assumption corresponds to the actual operation of a proposed system. Though the beams perform independently of each other, there still exists a co-channel interference factor that must be compensated for in the link analysis. This factor is assumed to be approximately 18 dB. Statistical independence of the beams allow for the modeling and analysis of a single beam with the results obtained being scaleable to encompass the performance of a fully implemented system. A percentage of the total resources available in a single beam will be modeled and analyzed without the loss of generality or accuracy of the model. A detailed discussion of this analysis is presented in the chapter on system modeling.

9.5.19 Regenerative Links The systems being analyzed are assumed to have regenerative capabilities at the satellites. This means that the incoming signal is demodulated, the data detected and processed, and then remodulated prior to transmission. This assumption is consistent with the digital nature of the satellite TDMA mode of operation.

9.5.20 Satellite Processing Time Satellite processing time is defined as the time required to decode an incoming packet and determine the next location in its traversal of the network. It is assumed that this time can be represented by a normal distribution with mean equal to 100 microseconds and variance of 5 microseconds. These values are derived from [ClJ89] and applied to the satellite communications environment.

9.6. System Design Parameters

The analysis of a low earth orbit satellite communications system requires that various system parameters be defined. The system comprises three parts: terrestrial, orbital, and links. The terrestrial portion of the system consists of the earth based transmitters and receivers. The orbital portion is composed of the orbiting satellites. Uplink, downlink, and crosslink channels make up the link portion of the system. The discussion which follows defines the system parameters associated with each of the three system parts. The parameter values are chosen so that the simulation model depicts, as accurately as possible, a proposed commercial system. The parameter characteristics chosen for this study closely resemble those proposed by Motorola [Mot90] for their Iridium system satellites. The Iridium system characteristics are chosen for two reasons: it is a representative example of a LEO system, and the availability of proposed design data for this system. Motorola has been continually refining and making changes to the system design, evident from the constellation changes from 77 to 66 satellites and the antenna changes of 37 to 48 spot beams, since their December 1990 FCC application. To facilitate this investigation, the December 1990 design baseline is used. These parameter values have been obtained as a result of review of pertinent literature related to the Iridium system [Mot90, LeM93, Com93].

9.6.1 Terrestrial Parameters Terrestrial parameters encompass the system design parameters associated with fixed (gateway) and mobile (TUU) transmitting and receiving stations.

9.6.1.1 Gateway Parameters The following design parameters are associated with the two gateway stations located at the above defined coordinates. These parameters represent both physical attributes of the gateways and characteristics associated with the gateway links.

9.6.1.1.1 Antenna Type The antenna type is a passive array.

9.6.1.1.2 Antenna Size The antenna array has a diameter of 3.5 meters.

9.6.1.1.3 Antenna Polarization Antenna polarization is right hand circular.

9.6.1.1.4 Antenna Gain The antenna gain is 54.0 dBi at the frequency of 20 GHz (downlink) and 57.5 dBi at 30 GHz (uplink).

9.6.1.1.5 Transmitter Power The transmitted power is variable dependent upon the constellation being evaluated. Maximum values are shown in Table 10.2.

9.6.1.1.6 Antenna Noise Temperature The antenna noise temperature for the uplink is 290°K and 30°K for the downlink.

9.6.1.1.7 System Noise Temperature The system noise temperature, T_S, is 1450 °K for the uplink and 1200°K for the downlink.

9.6.1.1.8 Center Frequencies The uplink center frequency is 30 GHz and the downlink center frequency is 20 GHz.

9.6.1.1.9 Modulation The modulation scheme chosen for the uplink and the downlink is QPSK.

9.6.1.1.10 Error Correction Coding Coding for the uplink and the downlink is 1/2 rate convolutional with constraint level, K=7.

9.6.1.1.11 Bandwidth The channel bandwidth of the gateway uplink and downlink are 15 MHz.

9.6.1.1.12 Data Rate The coded data rate for the gateway uplink and downlink is 12.5 Mbps.

9.6.1.1.13 3-dB Beamwidth The 3-dB beamwidth for the gateway antenna at 20 GHz (downlink) is 0.36° and 0.24° at 30 GHz (uplink).

9.6.1.2 IUU Parameters The following design parameters apply to the hand-held IUUs. These parameters represent physical attributes of the IUUs as well as characteristics associated with the IUU communication link.

9.6.1.2.1 Antenna Type The antenna type for the IUU is a Quadrifliar Helix.

9.6.1.2.2 Antenna Polarization The polarization of the Helix is right hand circular.

9.6.1.2.3 Antenna Gain The gain of the Helix is variable from 1.0 to 3.0 dBi. To allow for worst case conditions, the gain is assumed to be 1.0 dBi.

9.6.1.2.4 Transmitter Frequency The center frequency of the IUU is 1.61825 GHz. A 16.5 MHz frequency bandwidth is used for the IUU operation.

9.6.1.2.5 Transmitter Power The transmitter power is variable. Variability is required due to slant range differences caused by variations in constellation architecture. Maximum values are shown in Table 10.2.

9.6.1.2.6 Antenna Noise Temperature The antenna noise temperature is 150°K.

9.6.1.2.7 System Noise Temperature The system noise temperature looking towards the IUU from the satellite is 300°K.

9.6.1.2.8 Modulation QPSK is chosen as the uplink and downlink modulation scheme.

9.6.1.2.9 Error Correction Coding A 3/4 rate convolutional code with constraint level K=7 is used for both uplink and downlink.

9.6.1.2.10 Bandwidth The uplink receiver noise bandwidth is 90 kHz. The downlink receiver noise bandwidth is 200 kHz.

9.6.1.2.11 Data Rate The coded data rate for the uplink channel is 180 Kbps and 400 Kbps for the downlink channel. Channel data rates are 135 Kbps (uplink) and 300 Kbps (downlink).

9.6.2 Satellite Parameters The satellite design baseline consists of the following design parameters.

9.6.2.1 Antenna Types The uplink/downlink and crosslink antennas are linear phased array antennas.

9.6.2.2 Antenna Size Each array panel is rectangular in shape with dimensions 2 meters by 1 meter.

9.6.2.3 Antenna Polarization The antenna polarization is right hand circular for the uplink/downlink antennas. Vertical polarization is used for the crosslink antennas.

9.6.2.4 Antenna Gain The net transmitter antenna gain for the IUU link is 22.3 dBi. This figure takes into account 2.8 dB of losses associated with tapering, scanning, and edging. The gateway link antenna gain is 18.0 dBi. Crosslink antenna gain is 36.0 dBi.

9.6.2.5 Transmitter Power The transmitter power shall be fixed for a given constellation. The transmitted power level shall vary from constellation to constellation. Maximum values are shown in Table 10.2.

9.6.2.6 Antenna Noise Temperature The satellite antenna noise temperature is 290°K.

9.6.2.7 System Noise Temperature The receiver system noise temperature at the satellite is 552.6°K.

9.6.2.8 Receiver Noise Bandwidth The uplink receiver noise bandwidth is 90 kHz. The downlink receiver noise bandwidth is 200 kHz. The satellite to gateway link receiver noise bandwidth is 6.25 MHz. Crosslink receiver noise bandwidth is 12.5 MHz.

9.6.2.9 Antenna Beams The uplink/downlink antenna system consists of 37 cellular beams. The 37 cellular beam configuration is composed of seven different type antennas. Antenna types 1-6 are phased array while type 7 is a cupped dipole located on the nadir face of the satellite. To maintain whole-earth coverage, 3-dB circular beamwidth of a particular spot beam must be variable from 8.79° (77 satellite constellation) to 10.0° (36 satellite constellation). These variations in beamwidth appear to be possible since Motorola [Mot90] lists aperture gains of 20.0 dBi for antennas having 3-dB beamwidth of 26.0° azimuth and 11.5° elevation.

9.6.2.10 Satellite Capacity The uplink capacity of each cell is 102 frequency channels, spaced every 160 kHz. Each uplink channel occupies 126 kHz bandwidth. The total IUU uplink capacity of the satellite is approximately 8800 user slots. This figure is based on a 60 millisecond satellite TDMA frame which contains multiple transmit times (satellite) of approximately 1.3 milliseconds, receive times of approximately 3.0 milliseconds, and guard times of approximately 43 microseconds [Mot90]. Within a single TDMA frame, approximately 14 mobile users or 366 gateway slots can be accommodated. The value of 8800 mobile slots is derived from the product of the number of users per frame, the number of channels per beam, and the number of beams per satellite divided by the frequency reuse factor. The frequency reuse factor for this system is between 5 and 6 and denotes the number of times that a discrete frequency can be simultaneously reused within the 37 beam cellular configuration.

9.6.3 Orbital parameters Table 9.1 summarizes the orbital parameters necessary to define a given constellation and the location of individual satellites. Configurations for the specific constellations to be evaluated are summarized in Table 9.2 of Section 9.7.1.

9.7. Factors

In a study such as this, many engineering tradeoffs relative to system design are possible. The goal in this investigation is to limit the number of varying factors pertinent to the design of the communications network. Two factors have been identified for variation in this communications network study: the orbiting satellite constellation, and the loading applied to the network. Each of these factors are discussed in detail below.

Table 9.1. Orbital System Parameters

Parameter	Possible Variation Levels/Choices	Chosen Value
Coverage Type	limited area, whole-earth	whole-earth
Levels of coverage	1 to N	single
Number of Satellites (N)	N	variable
Orbital planes (s)	1 to N	variable
Satellites per plane (p)	N/s	variable
semi major axis (a)	dependent upon coverage requirement, look angle	
orbit inclination (i)	0° to 90°	90°
mean anomaly (M)	dependent upon coverage requirement	see Note 1 below
eccentricity (e)	0 for circular orbits	0
argument of perigee (ω)	N/A for circular orbits	0°
argument of right ascending node (Ω)	dependent on coverage requirement	see Note 2 below

Note 1: $M_t = t(360^\circ/p) + x(360/2p)$ where 0 < t < p-1, x=1 (even numbered plane), x=0 (odd numbered plane).

Note 2: $\Omega_t = (t-1)\alpha$ where 1 < t < s, and α is the interplane spacing defined by [AdR87] and shown in Table 9.2.

9.7.1 Constellations Evaluated Many design factors must be considered when choosing a constellation to implement. From a commercial standpoint, the economic returns on the investment are the foremost consideration. Trade-off studies weigh system performance and engineering design choices against the potential profit margins of the respective designs. In this investigation, a similar consideration is made in determining the number of satellites contained in the constellation. Of the proposed LEO systems described in Chapter 8, none have initial designs which contain more than 77 satellites within the constellation. For this reason, the number of satellites contained in the constellations to be evaluated will not exceed the 77 proposed by Motorola in their 1990 Iridium FCC filing [Mot90].

It is also important to note that the number of possible constellation configurations is restricted when continuous whole-earth coverage requirements are assumed. Adams and Rider [AdR87] show that not all numeric combinations of orbital planes and satellites per orbit are possible in a polar circular orbit environment.

The constellations chosen for evaluation of satellite utilization rates and delay/throughput characteristics are based on orbital altitudes and single-satellite, continuous whole-earth coverage requirements of commercially proposed LEO constellations. On the low altitude end (413 nautical miles) is the Iridium system proposed by Motorola. Constellation Communications, which proposes the Aries system, plan for a 48-satellite system operating at approximately 550 nautical miles above the earth's surface. Loral Cellular, who proposes the Globalstar system, also plans to design and implement a 48-satellite global constellation with an orbital altitude of 750 nautical miles. It is noted that both the Aries and Globalstar systems have circular orbits which are not polar in nature. Both have inclination angles relative to the Equator much less than Iridium's, which is approximately 90°.

The orbital altitude range for this study is 400 to 800 nautical miles. Only circular polar orbiting constellations will be evaluated. Within this specified range, Adams and Rider [AdR87]

conclude that twelve (12) unique constellations exist for the coverage requirements described above. In these twelve constellations, the number of satellites range from 36 to 77. Six of the twelve constellations will be evaluated without loss of generality. The characteristics of these constellations are summarized in Table 9.2.

Table 9.2. Constellation Configurations.

SYSTEM	PLANES	SATELLITES PER PLANE	TOTAL	θ RADIUS OF COVERAGE	α PLANE SPACING	ω INTER-PLN PHASING
A	4	9	36	27.607	47.043	20.000
B	5	9	45	24.153	37.983	20.000
C	6	9	54	22.206	32.059	20.000
D	5	11	55	22.171	37.341	16.364
E	6	11	66	19.907	31.402	16.364
F	7	11	77	18.450	27.091	16.364

The constellations listed above are denoted as Constellation A, B, C, D, E, and F respectively in Table 9.3. This table summarizes the orbital altitudes and maximum slant ranges used in the link analyses. The numeric range values calculated in Table 9.3 use an elevation look angle of 10° and an average radius of the earth value (R_E) of 6371 kilometers. The orbital altitude is denoted by *H*. The terrestrial to satellite slant range is determined by the following equation:

$$d_{earth-satellite} = \sqrt{(R_E + H)^2 + R_E^2 - 2R_E(R_E + H)\cos\theta} \tag{9.1}$$

where θ is the satellite radius of coverage needed for whole-earth coverage by the constellation. The determination of the maximum slant range for satellite crosslinks is obtained from analysis of two cases: intraplanar distance and interplanar distance. For the intraplanar distance, the *n*

satellites per orbital plane are equally spaced over 360°. So the angular separation, β, of the satellites within a plane is given by:

$$\beta = 360^{\circ}/n. \tag{9.2}$$

From Equation 9.2, the linear distance between adjacent satellites in a given plane is found from:

$$d_{\text{int raplane}} = 2(R_E + H)\sin(\beta/2). \tag{9.3}$$

Table 9.3. Constellation Orbital Distances (maximum distance in kilometers)

SYSTEM	ALTITUDE H (KM)	ORBITAL PERIOD (MINUTES)	APPROX. IN-VIEW (MINUTES)	SLANT RANGE CROSSLINK		
				EARTH	INTER-PLANE	INTRA-PLANE
A (36)	1549	117.1	17.9	3726.8	6717.0	5417.6
B (45)	1211	109.7	14.7	3150.3	5460.5	5186.4
C (54)	1044	106.1	13.1	2845.6	4731.8	5072.2
D (55)	1041	106.0	13.0	2840.2	5104.0	4176.4
E (66)	867	102.3	11.3	2502.5	4355.3	4078.4
F (77)	765	100.1	10.3	2293.2	3853.2	4020.9

For the interplanar distance, the maximum distance between two communicating satellites in adjacent planes occurs when one of the satellites is located at the Prime Meridian and the Equator while the other is located at latitude α and longitude ω, as given by Adams and Rider [AdR87]. In spherical latitude and longitude coordinates, the two satellites are located at (0°, 0°) and (α°, ω°), respectively. The conversion to Euclidean coordinates is given by:

$$(x,y,z) = (r\cos v\cos u, r\cos v\sin u, r\sin v) \tag{9.4}$$

where $r = (R_E + H)$, $u = \alpha = latitude$, $v = \omega = longitude$.

Once coordinate conversions have been made, the maximum interplanar slant range is obtained from:

$$d_{int\ orplane} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}\,. \tag{9.5}$$

The maximum slant range values for the six constellations to be evaluated are determined from Equations 9.1, 9.3, and 9.5. The values are used as one-way propagation distances for earth-satellite and satellite-satellite links analyses.

9.7.2 Network Loading Network loading consists of packets generated by individual users and gateway stations. The evaluation of the network's performance is accomplished by varying the number of simultaneous user requests within the specified geographic area. Variations in the number of simultaneous requests are controlled such that the analysis examines five loading cases (10, 30, 50, 70, and 90 percent). The network loading cases range from light (10 percent) to heavy (90 percent) with incremental values taken every 20 percent. One hundred percent loading occurs when the maximum number of system users transmit within the chosen unit time. For this investigation, the unit of time is the 60 millisecond TDMA frame time. Capacity or number of simultaneous users in a TDMA system is determined by the frame length in time, the channel bit rate, and the packet length in bits.

As stated in Section 9.5, Poisson processes are used to control the traffic distributions of the gateways and the IUUs. Other controlling processes such as uniform or arbitrary could have easily

been used as traffic distributions. Because the workload environment of the LEO systems investigated is unknown, Poisson processes are used to provide a possible characterization of the system workload. In Chapter 10, Section 3.1, a comparison of average delay times resulting from the use of Poisson processes and deterministic (constant) processes is discussed.

9.8 LEO System Simulation

This section discusses the simulation of six low earth orbit satellite constellations. The simulations are performed using the Designer (formerly known as BONeS) and SatLab commercial simulation packages [Com93] supplemented with C language primitive subroutines. The simulation of the LEO constellations requires that both Designer and SatLab be used for the modeling effort. Designer is used to model the communications portion of the system while SatLab performs the positioning functions for the earth stations and satellites. The subsections which follow describe the design hierarchy of the simulation model. Figure 9.2 shows the top level simulation flowchart for the systems being examined.

9.8.1 The SatLab Model As stated above, positioning information associated with terrestrial and orbital transmitters is performed by SatLab. The system modeler is required to provide three types of information for the system: orbital parameters associated with the constellation, positioning information of earth transmitters, and the epoch for the simulation.

The orbital parameters consist of the satellite identification number (referenced externally by a number-letter pair which denotes the orbital plane and the satellite within the plane), semi-major axis (the radius of the earth plus the altitude above the earth's surface of the satellite), the orbital inclination measured with respect to the equator, the mean anomaly of a satellite, the eccentricity of the orbit, the argument of perigee, and the argument of right ascending node. The combination of these parameters represents one entry in the constellation definition.

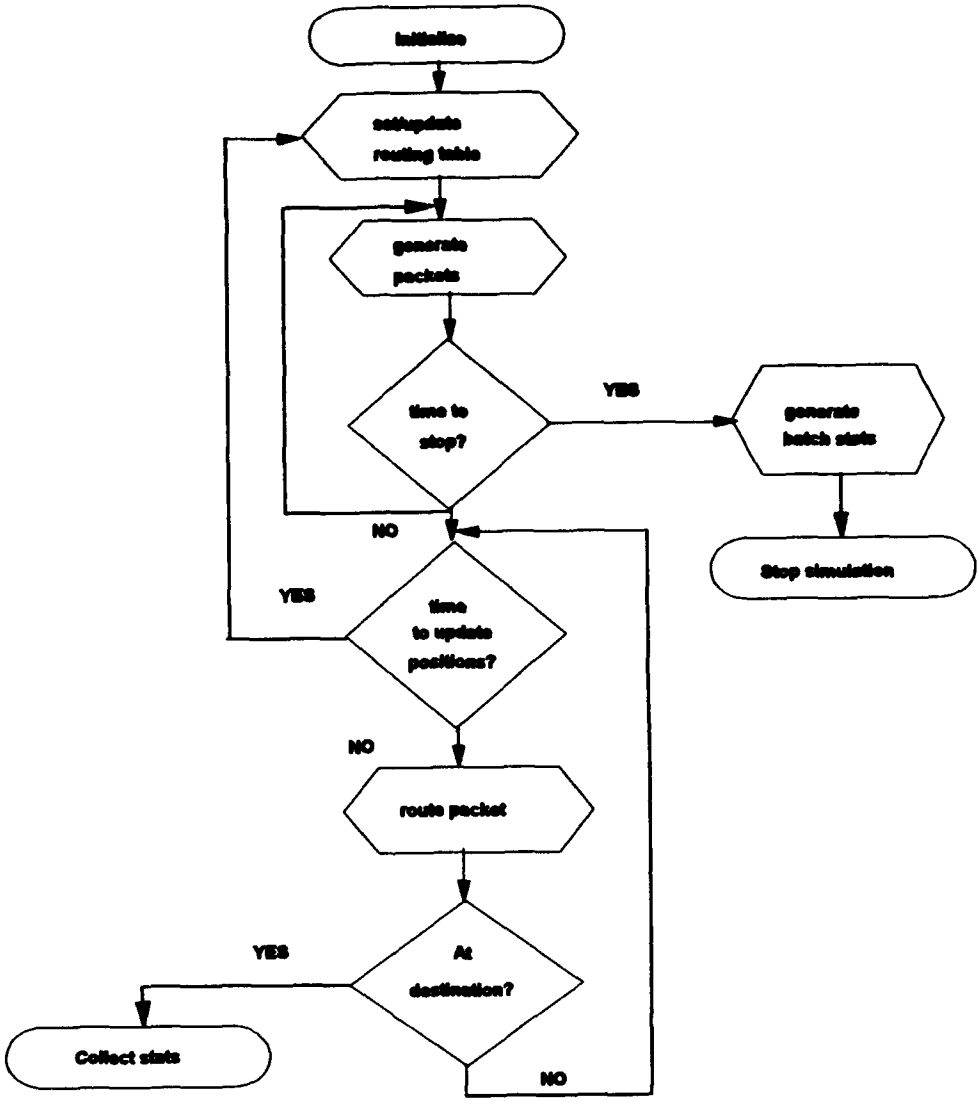


Figure 9.2. LEO system flowchart.

The earth station data defines either fixed location or mobile transmitters. Fixed location transmitters are defined by their latitude, longitude, and altitude positions relative to the Prime Meridian. Western hemisphere longitudes are denoted by negative values rather than standard western reference. The same holds true for southern hemisphere latitudes. Altitudes are referenced

to mean sea level. Mobile transmitters are defined by their locations relative to the system epoch. The SatLab User's Guide [Com93] states that a separate file is used to define each mobile earthstation transmitter. A mobile earthstation transmitter file consists of time in seconds (after epoch), longitude, latitude, altitude, velocity, and direction of movement. During the testing of the positioning portion of the simulation model, it was discovered that the mobile earthstation positioning capability of SatLab did not work. A work-around solution was to define the mobile earthstations as fixed location transmitters. The stationary location assumption for the mobile transmitters was valid because these transmitters could physically move at most 5 to 6 miles during the simulation period.

The epoch denotes the starting date and time of the simulation. This information is required so that initial locations for all communicating nodes (transmitters) can be ascertained. This information is passed to the communications portion of the simulation. Additional information such as number of nodes, line-of-sight distances between pairs of nodes, and relative velocities are also available.

Once these pieces of information are provided, positioning and coverage simulations can be performed. Initial positions of satellites and earth station transmitters are established from the epoch (the calendar date and time of the simulation start) and the positioning information supplied by the user. Coverage simulations reveal the percentage of time that a given area is covered by at least one satellite. Color coding of coverage areas is used to denote the number of satellites covering a given location.

9.8.2 The Satellite Communications Model The satellite communications model is developed and simulated using the Designer modeling tool supplemented with C code subroutines where necessary. The Designer tool requires that the top level module be at the system level. This means that the module cannot have external input or output ports for sending or receiving data. For this investigation, the top level system module is named *Leo Satellite Communications*. This

system module, shown in Figure 9.3, comprises a positioning portion and a communications portion. Each of these portions are briefly described below.

9.8.2.1 Positioning The positioning portion of the *LEO Satellite Communications* system module consists of two sublevel modules: *initialize* and *update_node_position*. At the beginning of a simulation, these two modules are given the higher priority for execution over the other modules of the same hierarchical level. At the start of simulation execution, the *initialize* module executes before the *update_node_position* module. The ordering of execution is controlled by the *Init* and the *EIO* modules. The *Init* module contains a simulation flag which tells the simulator that the blocks connected to this module have execution precedence over other system modules. The *EIO* module is an execute in order module. Any blocks connected to the 1 output port execute before those connected to the 2 output port. The *initialize* module calls the Comdisco provided *BSIM* module to obtain information from SatLab related to the number of nodes (satellites and earthstations) and the locations of fixed earthstations. Once this information is received, global memories containing the number of nodes, number of earthstations, earthstation latitudes, and altitudes are initialized. Upon completion of the *initialize* module execution, control of the simulation execution is next given to the *update_node_position module*. This module's function is two-fold. First, an update delay time (user specified) controls how often positional information associated with satellites and mobile earthstation transmitters are updated. Once again, the *BSIM* module is called to receive positioning information from SatLab. The second function of the *update_node_position* module is to create or update a global memory used to store routing information and relative positions between communicating nodes (earthstations and satellites). The routing table contents are based on visibility and nearest neighbor constraints. Global memories containing relative distances and elevation look angles between node pairs are also updated within this module. The *update_node_position* module execution is repeated at user-specified intervals throughout the simulation. This is accomplished by a parameter-based delay module and a

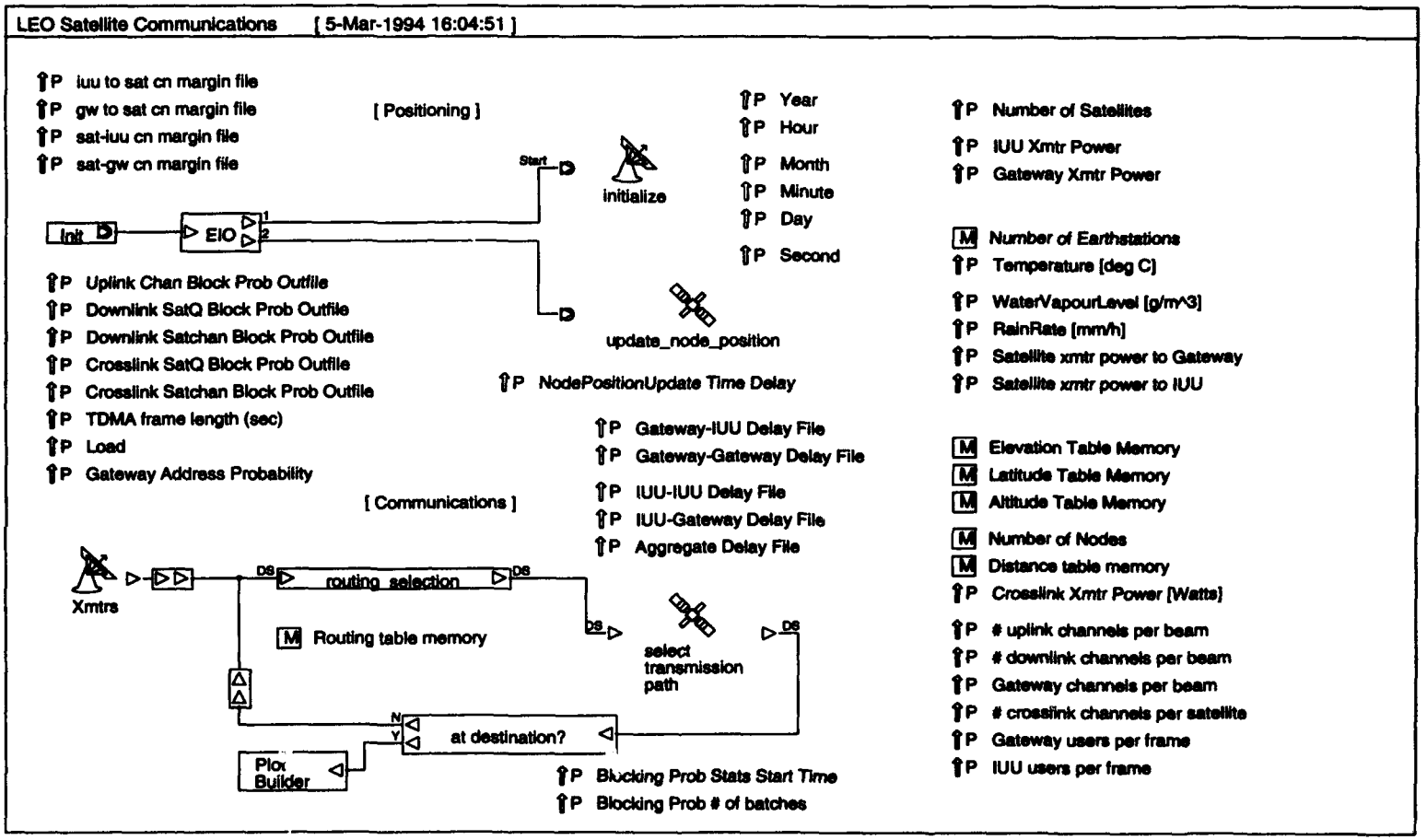


Figure 9.3. LEO Satellite Communications system module.

feedback loop to the input. This feedback loop is internal to the *update_node_position* module and is not shown in Figure 9.3.

9.8.2.2 Communications The communications portion of the *LEO Satellite Communications* module performs four basic functions: transmitters (either earth or satellite based), routing selection, transmission path, and arrival at final destination determination. These four functions form a closed loop path for packets to traverse the system. Each of these functions is briefly described below.

The communications simulation begins with the generation and transmission of packets by source nodes. For this investigation, only earth-based transmitters generate packets. Simple modifications to source address generation distributions can be made to allow for satellite generated packets. The creation and transmission of packets is controlled by the *Xmtrs* module. This module generates packets based on two Poisson processes. One Poisson process is used by the gateways while the other is used by the mobile units. The use of two separate processes results from the differing transmission rates of the two types of transmitters. The *Xmtrs* module is also responsible for the creation of two data structures used throughout the simulation: the Sat_DS and the Sat_Route_DS structures. These structures are defined in Tables 9.4 and 9.5. The Sat_DS data structure contains packet information relative to the source and destination types, the sequence number of the packet, and the time of creation. This data structure is encapsulated within the Sat_Route_DS data structure in its Data field. The Sat_Route_DS represents the packet being sent through the network. Additional fields of the Sat_Route_DS data structure are the address of the current location, the address of the next location in the path, the bit error rate, and the E_b/N_o absolute ratio. At various points within the simulation model, field information is extracted from the Sat_Route_DS data structure to determine routing and packet integrity.

Once a packet is generated, it flows into the *routing selection* module where the next node along the packet's path from source to destination is determined. This module examines the

current location and the final destination to calculate an index value into the routing table global memory (discussed in detail in Section 9.8.2.7). The output of the memory read is the next node to where the packet shall be routed. The Next field of the Sat_Route_DS is updated and the packet flows into the *select transmission path* module.

Table 9.4. The Sat_DS Data Structure.

FIELD	DATA TYPE	VALUE
source	Integer	0 to number of earthstation - 1
xmtr_type	Integer	0 = gateway, 1 = mobile
destination	Integer	0 to number of earthstation - 1
rcvr_type	Integer	0 = gateway, 1 = mobile
packet length	Integer	1024 bits
sequence number	Integer	0 to infinity
time stamp	Real	time of creation

Table 9.5. The Sat_Route_DS Data Structure.

FIELD	DATA TYPE	VALUE
Current	Integer	Current node location
Next	Integer	Next node in path
Data	Sat_DS	encapsulated packet
BitErrorRate	Real	BER of link
EbN0	Real	EbN0 of link

The *select transmission path* module selects one of three paths for the packet to take: uplink, crosslink, or downlink. The determination of which path to take is made by analyzing the current location and the next node along the path. Each of these three paths is represented by a block module which encapsulates two lower levels of logic. These lower levels perform the channel allocation, deallocation and packet rejection tasks as well as atmospheric delay, multiple

access delays, and attenuation calculations. While the channel allocation, deallocation, and rejection tasks are performed by modules designed and implemented using Designer library blocks, the atmospheric and multiple access calculations must be performed by user written C language routines referred to as *primitive modules*. Because of their importance to the accuracy of the model implementation, each primitive is individually discussed below.

The *at destination?* module is used to determine if a packet has reached its final destination. This module compares the address contained in the Current field of the Sat_Route_DS data structure with the address contained in the Destination field of the Sat_DS data structure. If the two addresses are the same, the packet is routed to the *Plot Builder* module where packet delay statistics are gathered. If the two addresses differ, the packet is sent back to the *routing_selection* module.

9.8.2.3 Multiple Access Primitive With version 2.0 of Designer, Comdisco provided a multiple access model primitive with one of the library example simulations. This primitive module computes the delay incurred by packet transmissions for the type of multiple access technique implemented. The original module allowed for three types of access: frequency division multiple access (FDMA), time division multiple access (TDMA), and Demand Assign-TDMA (DA-TDMA). The delay equations implemented by the Comdisco model were taken from [Ha90]. Examination of this primitive module revealed logic errors in the method used for calculating the delays. Corrections were made to this primitive via code changes and recompilation. Verification of the correctness of this module will be discussed in subsequent sections.

9.8.2.4 Uplink and Downlink Channel Model Primitives The uplink and downlink channel model primitives perform four basic functions. These functions are packet delay, carrier to noise (C/N) level calculations, energy per bit to noise power density (E_b/N_0) calculations, and bit error rate computations for a given link.

Associated with each of the primitives are five inputs. The uplink and downlink primitive models have four common inputs: altitude, latitude, elevation, and distance. The fifth input is the transmitter type for the uplink module and the receiver type for the downlink module. The difference in the fifth input is required due to the differences in the types of earthstations (fixed and mobile) used for this investigation.

Two sources contribute to packet delay: propagation distance and, in certain cases, delays associated with rain. The propagation delay is the quotient of the range between the two communicating nodes divided by the speed of light (3×10^8 meters/second). Delays associated with rain can occur if the attenuation due to rain is greater than three decibels. For these occurrences, the rain delay is approximately twenty (20) microseconds.

Signal attenuation is due to propagation distances, rain losses, gaseous losses, fade losses, and scintillation losses. Each of these possible attenuation factors are accounted for in the uplink and downlink primitives. Rain, gas, and scintillation losses are modeled using the CCIR approved models taken from [StD82, Ipp89, Had93]. Modifications and additions to originally supplied Comdisco primitive codes were necessary to correct logic errors and to improve the accuracy of the model.

9.8.2.5 Crosslink Channel Model Primitive The crosslink channel model primitive performs functions similar to those of the uplink and downlink channel model primitives. Functions not necessary in the crosslink primitive are those associated with atmospheric attenuation.

9.8.2.6 BONeS SatLab Interface Module (BSIM) The BSIM module, provided by Comdisco, provides the interface between the two Comdisco products used for this investigation. This interface is accomplished via a Unix socket established at the time of simulation execution. All positioning data and information relative to the satellite constellation or earthstation

transmitters are obtained through this module. A detailed description of the functions and capabilities of the BSIM module are presented in the SatLab User's Guide [Com93].

9.8.2.7 Satcom_router Module The Satcom_router module is also a Comdisco provided library module. Its function is to determine the shortest path (distance-wise) between two communicating nodes given the constraint that the two nodes must be visible to one another. The visibility constraints are determined within SatLab from user-specified elevation look angles. The input to this module is a vector containing distances between all possible pairs of communicating nodes. The module takes the input and computes a routing table based on the constraints above noted. The output of the module is an integer vector which is stored in a global routing table memory. This memory is later accessed by modules requiring positioning information for routing, delay, and atmospheric attenuation calculations.

9.8.3 Simulation Platform and Execution Times The simulations performed for this investigation were accomplished via the use of a SUN SPARC II workstation. For the duration of this investigation, this workstation operated in a project dedicated mode (i.e., 100 percent of the non-operating system CPU cycles are dedicated toward this effort). To appreciate the complexity and duration of this simulation investigation, Table 9.6 is provided to detail the average simulation execution times per data point. For value shown in Table 9.6, three independent replications of the simulation trial were performed. In addition, each data point resulted from a 1/10 scale simulation version of the satellite TDMA frame capacities defined above. This means that the number of gateway users per frame and IUU users per frame were reduced from 366 and 14 to 36 and 1, respectively. In addition, the number of mobile earthstations were reduced from 150 to 15. This scaling was necessary to achieve results within a reasonable period of time, while at the same time achieving the same accuracy of results. Pilot simulations are performed on the 36-satellite system using the full-scale capacities defined above. For 10 percent system loading, the simulation execution time was approximately 48 calendar hours. Thirty percent loading required 6.5 days for

the simulation execution. These enormous amounts of simulation execution times, in conjunction with the fact that 90 simulation runs needed to be performed across the chosen constellations, led to the scaling down of the simulation model. Additional simulations are performed on the 36-satellite system using the scaled-down capacities and number of users. Comparisons of delay results between both the full-scale and the scaled-down models were made for each of the loading values defined above. Differences in delay values were less than one (1) percent.

Table 9.6. Average Simulation Execution Time Per Data Point (in calendar hours)

Loading %	36-satellite (System A)	45-satellite (System B)	54-satellite (System C)	55-satellite (System D)	66-satellite (System E)	77-satellite (System F)	GEO
10	3.2	4.7	6.6	6.8	8.0	9.6	1.8
30	9.6	14.7	19.9	23.7	23.9	29.5	5.5
50	16.2	24.2	33.1	38.4	40.7	54.9	9.3
70	23.5	34.3	46.3	53.8	56.6	76.3	13.2
90	30.5	45.0	63.2	70.7	73.7	102.2	17.7

The cumulative simulation time required for this investigation was approximately 3274 calendar hours which translates to roughly 136 days of constant CPU execution. These values do not include simulation model testing and pilot executions. Table 9.6 shows that the simulation execution times were highly dependent upon the number of communicating nodes (fixed/mobile earthstations and satellites). An entry in the overall routing table and relative distance table was required for each possible pairing of transmitters and receivers. Because the locations of most of the transmitters (excluding two gateways and mobile users) changed with time, periodic updates to the routing table and distance table were made. The time required to update these tables was the predominant factor in the total simulation execution time. As an example of the size of these tables, consider a 77-satellite system with 15 earthstations. This translates into a 1155 entry table.

Additionally, SatLab determined the visibility between any two possible pairs. This was accomplished by line-of-sight calculations for the satellites and coverage area calculations for the satellite-earthstation pairings.

9.9 Model Verification

As was the case with the model development, the model verification was composed of two portions: the positioning of the satellites and earthstations, and the inter-transmitter communications. Each of these are discussed below.

9.9.1 Positioning Verification The verification of the positioning portion of the simulation model entailed determining the correct location of the orbiting satellites and mobile terrestrial transmitters. Correctness of satellite location was determined via the use of SatLab in the stand-alone mode. Data files containing earthstation and satellite parameters were loaded and subsequent positioning simulations performed. Each positioning simulation began at a user specified epoch. At the epoch, the simulation animation was viewed to reveal the locations of each orbiting satellite. These simulation positions were compared against positioning information manually calculated from Keplerian orbital mechanics parameters. These comparisons revealed an exact correlation between the simulation positioning and the analytic positioning. This was expected since the SatLab modeling tool uses identical techniques in the positioning simulation. Once the correctness of the initial positioning was determined, the simulation was allowed to run so that orbital paths, polar crossings, and periods of revolutions can be examined. Orbital path, polar crossings, and satellite period verifications were similar to initial positioning analysis in that expected locations and periods determined from Keplerian orbital mechanics.

9.9.2 Communications Verification The verification of the system model followed a bottom-up testing approach. The lowest level module, primitives in some cases, were tested prior to incorporation into higher level modules. Two major types of testing were required for the

system components: the user-designed C code primitives, and the Designer-built hierarchical modules.

9.9.2.1 Primitive Verification As described above, the simulation model used four C code primitives for the link models. In the uplink module, two primitives were used: the multiple access primitive and the uplink channels model primitive. The crosslink module used one primitive, the crosslink channel model primitive. The downlink module used the downlink channel model primitive. Verification of each of these modules entailed the exercising of the C code modules prior to their incorporation into the Designer system model. The exercising of the modules consisted of compilation of module code as independent programs and then performing test cases to verify the correctness of the execution. These test cases exercised atmospheric models for attenuation due to propagation distances and rain fading as well as models for scintillation effects and TDMA packet delay. Results from these executions were compared to analytic results obtained from algorithms used to construct the C modules. Algorithms used to implement the C modules were taken from previously published independent research [Ha90, Had93] and CCIR approved techniques [StD82, Feh83, Ipp89].

Verification of operational correctness was also necessary for two additional primitive blocks, *BSIM* and *Satcom_router*. These two primitives were supplied by Comdisco for one of the library example models. The need to verify these modules rather than assume their correctness was necessitated by the errors found in other library functions previously mentioned. The verification of each of these modules consisted of initiating single requests to/from SatLab and observing the resultant actions. Following this procedure with *BSIM* requests revealed that Version 1.0.1 of SatLab functioned incorrectly in the recognition of mobile users and their associated positioning data. A work-around solution was arrived at by considering mobile users to be fixed for the duration of the simulation. The mobile positioning problem was subsequently corrected in Version 1.0.3. The verification of the correct operation of the *Satcom_router* module was accomplished

using single packets with known source and destination addresses and transmittal path through the network. Comparison of the path taken was made with the animated view provided by SatLab.

9.9.2.2 Designer Module Block Verification The verification of the Designer blocks which compose the communications portion of the model was accomplished by testing modules at the lowest possible level and building upward. Encapsulation of verified lower level modules allowed for the testing of modules at the next highest level. This process was continued until the four basic functional modules described above were integrated to form the communication system.

The testing of block modules was accomplished using the data structures defined above and probe modules supplied within Designer. Verification testing using the data structures was most commonly performed for the cases when routing functions were invoked. The routing functions and general packet flow through the network were accomplished by the use of a single packet with user-specified source and destination address. Textual probes were placed throughout the network to monitor the packet's progression through the network. The data collected by these probes were then analyzed to ensure that the packet was routed correctly. These probes also allowed for the verification of the delay incurred by the packet as it moved from source to destination. Delay values collected by the probes were compared with the theoretical values for propagation delays.

9.10 Model Validation

The validation of a simulation model consisted of validating the operating assumptions, input parameter values and distributions, and the output values and conclusions associated with the model [Jai91]. Validity tests on these three model aspects can be accomplished by a combination of expert intuition, measurement from real systems, and/or comparison with theoretical results. For certain applications, all three comparative processes can apply. In other cases, only one may apply. The validation of the LEO satellite communications model fell into the category where only one comparative type applied: expert or engineering intuition. Comparison with measurements from real systems does not apply since no LEO system of the type being analyzed exists. Also, the

systems being modeled do not fit classic queuing models. This is because of the differing transmission rates and the time-space dependent relationship of communication transmitters/receivers. Therefore, the determination of model validity followed a step-wise approach for the operating assumptions, input parameters, and output results.

9.10.1 Validation of Operating Assumptions The overall operating environment of the systems being modeled closely match those of a proposed LEO system [Mot90, LeM93]. These operating assumptions, presented in Section 9.5, included the global satellite coverage, system availability, beam independence, and gateway/mobile communications to mention a few. In certain instances, engineering intuition was used to create operating assumptions not addressed in previously published related works. Such was the case for the traffic distributions associated with source and destination packet addresses.

9.10.2 Validation of Input Parameters To facilitate this investigation, 74 communication input design parameters, defined in Section 9.6, were set for each simulation. Of these 74, 66 remained constant across the constellations being evaluated. In addition, with each new constellation analyzed, a new set of design parameters was loaded from SatLab. The number of constellation parameters varied with the size of the constellation and was a function of the number of satellites and orbital configuration. Most of the input parameters presented earlier in this chapter represent a set of conditions proposed by Motorola for their Iridium system [Mot90, LeM93]. To extend the analysis to encompass the constellations being examined, additional assumptions, such as polar orbits and coverage considerations, were made based on the works of Adams and Rider [AdR87]. Additional operating assumptions related to the delays associated with packet processing by the satellite were made using the results presented by Clark et. al. [ClJ89]. Validation of atmospheric attenuation design parameters was possible from the previously published works noted in Section 9.9.2.1.

Within the scope of this investigation, there existed conditions where design parameter definitions were required but no previously published data was available. One such case was for the traffic distributions assumed herein. Extensive literature searches and conversations with Electrical Engineering faculty members revealed that very little is published relative to traffic distributions in a mobile cellular environment. Because of this lack of information, engineering intuition and consultations with faculty members were used to derive the source and destination address distributions previously discussed in this chapter.

9.10.3 Validation of Output Results The validation of the output results followed a similar approach used in the verification of the model. A bottom-up approach to validation was used for the system model. The output results of interest were the delay encountered by a packet as it traverses the network and the satellite resource utilization rates. Each of these outputs are discussed below.

9.10.3.1 Packet Delay Validation Packet delay is defined as the difference in time between the transmittal of the first bit of a given packet at the originating transmitter and the receipt of the last bit of the packet by the destination receiver. The delay encountered by a packet as it moves through the network is a function of many factors. These factors include: the transmitted data rate, multiple access technique employed, propagation delay, satellite processing delay, and queuing delays.

To validate the packet delay portion of the model, it was necessary to control the environment of the system. This included choosing a representative satellite constellation to serve as the testbed case. The 36-satellite constellation, referred to previously as Constellation A, was used as the test case constellation. The two gateway locations previously defined served as source and destination for the packet transmittal. Testing of both gateway and mobile user transmitters was accomplished with the chosen locations by changing specific parameters between simulations. A single packet was chosen to transmit from point A (Washington, D.C. gateway location) to point

B (Los Angeles gateway location). By transmitting only one packet, the queuing delays at the satellites were eliminated. Also eliminated, was the variability of the propagation delay due to the movement of the satellites relative to the fixed locations of the transmitters. Therefore, the packet delay was a function of the transmission time, the TDMA access time, the round-trip propagation time, and the satellite processing time. Each delay factor was implemented via C code and verified for correctness of operation as previously discussed. The validation of these factors come from known physical and engineering laws (transmission and propagation delays) as well as from previously published works (TDMA delay [Ha90]). The satellite processing delay was derived from a terrestrial analysis which examined Transmission Control Protocol (TCP) processing overhead [ClJ89].

Because the specific effect on delay caused by queuing at the satellite was not available from previously published works, a type of sensitivity analysis was performed to validate the satellite queuing effects. This was accomplished by incrementally increasing the arrival rate of packets at given satellite, setting probes on inputs and outputs of the satellite processing module, and then analyzing the resultant delays. The effects of queuing at the satellite follow a classic response relative to delay versus loading characteristics. As the load was increased, the delays associated with queuing also increased. A detailed analysis of these delays is presented in the following chapter.

9.10.3.2 Satellite Resource Utilization Verification and Validation The verification and validation of the usage of satellite resources followed an approach similar to that described above. Three cases were verified and validated: uplink resource usage; downlink resource usage; and crosslink resource usage. The verification and validation process looked at the overall utilization of resources rather than utilization of a given satellite. The reason for this approach was that prior to performing the simulation, it was not exactly known how the usage of resources were distributed among the satellites. The overall percentage of utilization was based on the capacity (in

terms of number of TDMA slots) of one satellite. Therefore, if a single satellite had a capacity of N slots, ten percent network loading should result in an aggregate usage of $0.1N$ slots at any given time in a packet switched environment.

To verify the uplink and downlink utilization, Constellation A was used once again as the test environment. Recall that Constellation A was a 36-satellite configuration. With this configuration and the required orbiting altitude, a minimal number of requests for crosslinks (0.16 percent of total requests) was used for transmission within the geographic bounds of this investigation. Verification of the uplinks and downlink resource usage was accomplished by placing probes on the internal resource modules to collect pertinent information concerning usage requests. A test case was setup for ten percent network loading. The outputs resulting from these probe placements were analyzed to verify that the aggregate usage of resources was at ten percent.

The verification of the crosslink resource usage was slightly more difficult since not all requests were routed from one satellite to another. In the case of Constellation A above, there was minimal utilization of crosslink resources. To observe the utilization rates for the crosslink resources, Constellation B, a 45-satellite system was used as the test case. Probes were placed on the inputs to the crosslink channel module to count the number of packets requiring crosslinks, along with probes on the internal resources to observe the distribution of crosslink requests. It was observed that this configuration required approximately 27 percent of the total requests for satellite resources require a crosslink as part of the transmission path. Verification of correct operation followed the approach described above for the uplink and downlink cases.

The validation of the resource utilization distributions and rates, though by no means considered to be rigorous, was assumed to be accurate based on the verification and validation of the operation of the portions of the simulation model used to derive these utilization results. By this, the channel models and routing algorithms were verified to operate correctly, while the delay algorithms and modules were validated against previously published works. In addition, extensive

testing was performed on the resource modules themselves to verify correct operation. Detailed analysis of the utilization rates is presented in the following chapter.

9.11 Summary

This chapter has presented the design and development of the LEO satellite communication network model. Section 9.2 provided a discussion on the problem definition associated with the LEO simulation. Research scope, systems architecture, and geographic area of evaluation were presented in the section. A brief overview of the system services followed Section 9.2 in Section 9.3. In this section, a real-time packet switching environment was defined. A presentation of the performance metrics being investigated was presented in Section 9.4. The two metrics chosen for analysis were network packet delay and satellite resource utilization. Following the discussion of performance metrics, Section 9.5 defined the network operating assumptions. These assumptions were based on similar assumptions made by commercial ventures in the low earth orbit satellite communications environment. In cases where assumptions needed to be made and previously published works did not exist for comparison, engineering intuition and judgment were used as a basis for the assumption. The operating assumptions discussed in Section 9.5 were followed in Section 9.6 by a presentation of the design parameters which compose the system. The design parameters were categorized into three areas: terrestrial parameters (to include gateway and IUU designs), satellite parameters; and orbital parameters. Section 9.7 defined the two simulation factors analyzed in this investigation: 6 constellation configurations and network loading. The design and modeling of the LEO system simulation model were the topics of Section 9.8. Salient features of the model were explained in this section. Model verification and validation followed Section 9.8 in Sections 9.9 and 9.10, respectively.

CHAPTER 10

LEO SATELLITE SYSTEM ANALYSIS

10.1 Introduction

This chapter addresses important portions of the design issues which face the satellite system designer. Analyses are conducted for six low-earth orbit (LEO) satellite systems in a packet-switched operating environment. Section 10.2 describes the communications link analyses for the systems. In that section, five links are defined for the LEO systems. Power requirements for each of the links are defined for a clear-sky operating environment. Link budget analyses are performed for each of the six configurations. Also included in Section 10.2 is a brief discussion of the Motorola Iridium link analyses and the inconsistencies noted in their filing with the Federal Communications Commission (FCC). Section 10.2 also includes a rain fade analysis for the portions of the system most affected. This analysis is used to determine the decibel level margins that must be added to meet specified link availability rates. Site diversity and diversity gains are also addressed in Section 10.2. The rain fade and site diversity analyses are presented for completeness of the link budget analyses. Average packet delay analyses are the topic of Section 10.3. In this section, the parameters which affect the overall packet delay are defined. The delay performance of five differing system communication links are examined and analyzed. A comparison of the delay performance between different constellations is performed in Section 10.3 as well as a comparison of the performance of LEOs to a GEO of similar capacity. This discussion is followed in Section 10.4 by a utilization analysis. This analysis is performed for each constellation and examines the number of satellites being utilized by the simulations in addition to the time-average loading on satellites within the constellation. Section 10.5 extends the modeling

and analysis of LEO systems by the application of metamodeling to the delay analysis. Metamodels are derived which provide concise representations of the system performance.

10.2 Link Analyses

The strategy used in this study was to define and hold constant as many design parameters as possible and thereby limit the number of design factors. Based on the discussion in the previous chapter, a decision was made to hold constant the physical characteristics of a satellite while evaluating the six different constellation configurations described above. The sole exception to this decision was the variation of the transmitter power of the satellite for a given constellation. The variation of the transmitted power was required for three reasons: (1) fixed antenna gains across the evaluated constellations, (2) constant flux density (ideally) across constellations, and (3) Friis' transmission equation which shows that the received power decreases with the square of the path length.

Because one of the design goals was to maintain a constant flux density across the constellations evaluated, it is necessary to establish this value for the particular links to be examined. The flux density values used for reference are derived from [Mot90] and are the energy per distance squared within the area illuminated by the satellite. Analytically, the flux density is defined by

$$F = \frac{P_t G_t}{4\pi R^2} \text{ W/m}^2 \tag{10.1}$$

where P_t is the transmitted power, G_t is the gain of the transmitter antenna, and R is the slant range distance between the transmitting source and the receiving destination. Flux density values determined by Equation 10.1 yield an overall value without regard to the occupied bandwidth of the channel. Flux density values are more commonly referenced with regard to 4 kHz of channel

bandwidth [PrB86]. To arrive at the 4 kHz reference flux density, Equation 10.1 must be modified to allow for the contribution associated with 4 kHz of the total occupied channel bandwidth. Equation 10.2 shows the resulting flux density per 4 kHz of bandwidth:

$$F = \frac{P_t G_t (4/BW)}{4\pi R^2} \tag{10.2}$$

where BW is the occupied channel bandwidth (kHz).

Table 10.1 shows the referenced flux density values. From Table 10.1 and the defined gains of the transmitter antennas, the transmitter power required for each link and each constellation can be determined. Rearranging Equation 10.2 into decibel notation and solving for Pt yields Equation 10.3. Equation 10.3, used in conjunction with the slant range values shown in Table 9.3, provide the worst-case power requirements, summarized in Table 10.2. Note that for crosslinks, the slant range value used in the power calculation is the maximum of the interplanar and intraplanar distances.

Table 10.1. Reference Flux Density Values per 4 kHz BW

LINK	FLUX DENSITY (dBW/m^2)
(1) IUU to satellite	-147.6
(2) satellite to IUU	-132.1
(3) Gateway to satellite	-118.8
(4) satellite to Gateway	-155.7
(5) satellite to satellite	-140.2

$$P_t(dBW) = F(dBW/m^2) - G_t(dBi) + 20\log R(dBm^2) - 10\log(4/BW) + 11dB\,. \tag{10.3}$$

From the parameters defined in Table 10.2, it is possible to analyze the satellite links for the various constellations. In a digital satellite communications system, the performance of the link is specified by its bit error rate (BER). Bit error rate is defined as the probability that a received bit is in error. Link designs for data communications normally require that the BER be small (10^{-6}) for quality service. Bit error rates in a communications link can be determined by using the ratio of the energy per bit (E_b) to the noise density (N_o) of the communications link measured at the receiver and the complementary error function (erfc). Link analyses are affected by many factors, to include the type of modulation scheme used. For this analysis, QPSK is chosen as the digital modulation scheme This relationship is shown in Equation 10.4, where *PB* represents the probability of bit error [PrB86]

$$PB = \tfrac{1}{2} erfc(\sqrt{E_b / N_0}) . \tag{10.4}$$

The choice of QPSK and a BER of 10^{-6} dictate the E_b/N_o required to maintain the BER. It is shown in [PrB86] that E_b/N_o must be 10.6 dB or greater for uncoded QPSK and 10^{-6} BER.

Table 10.2. Transmitter Power Requirements for Specific Links.

Constellation	Transmitted Power in dBW (W) for Communications Links				
	IUU to satellite	satellite to IUU	gateway to satellite	satellite to satellite	satellite to satellite
A (36 satellites)	8.82 (7.6W)	6.53 (4.5W)	1.87 (1.5W)	4.47 (2.8W)	9.30 (8.5W)
B (45 satellites)	7.37 (5.5W)	5.07 (3.2W)	0.41 (1.1W)	3.01 (2.0W)	7.50 (5.6W)
C (54 satellites)	6.48 (4.5W)	4.18 (2.6W)	-0.48 (0.9W)	2.12 (1.6W)	6.86 (4.9W)
D (55 satellites)	6.47 (4.5W)	4.17 (2.6W)	-0.49 (0.9W)	2.11 (1.6W)	6.92 (4.9W)
E (66 satellites)	5.37 (3.4W)	3.07 (2.0W)	-1.59 (0.7W)	1.01 (1.3W)	5.54 (3.6W)
F (77 satellites)	4.61 (2.9W)	2.31 (1.7W)	-2.35 (0.6W)	0.25 (1.1W)	4.85 (3.1W)

A standard practice in link budget designs is to reference the carrier to noise ratio (C/N) instead of E_b/N_0 [PrB86]. This C/N usage results from the early dominance of analog satellite systems. These type of systems used this ratio as a measure of a received signal's fidelity. The design of the satellite link is such that a minimum overall C/N is maintained in the worst case atmospheric conditions. Normally, an additional margin is provided in the clear sky design to allow for attenuation due to inclement weather or foliage attenuation. All link analyses presented in this chapter are performed in a clear sky environment at maximum slant ranges. In addition, these analyses are performed for a single beam in the multiple beam satellite antenna configuration. The single beam analysis does not limit the validity or scope of the research. The effects of co-channel interference caused by multiple beams has been compensated for within the link budget requirements by adding an 18.0 decibel interference factor. The effects of this addition is shown in Table 10.3 and also in Appendix A.

Given that the E_b/N_0 value must be 10.6 dB for the design conditions assumed above, a conversion to C/N must be made to accomplish the link analysis. The conversion equation is

$$E_b/N_0 = (B_N/R)(C/N), \tag{10.5}$$

where B_N is the receiver noise bandwidth and R is the symbol transmission rate. When ideal Nyquist filter approximations are used, along with BPSK modulation, $B_N = R$. For the case of QPSK, $R = 2B_N$. For this case, Equation 10.5 reduces to $E_b/N_0 = 1/2(C/N)$. It should be noted that the reduced E_b/N_0 value is given as a ratio and not in decibels. Once the conversion to decibels has been made, the required link C/N is 13.6 decibels prior to coding being applied.

In reality, ideal Nyquist filters do not exist. Link analyses allow for the ideal filter assumptions in making the Eb/No to C/N conversion. This assumption is compensated for by an implementation margin (approximately 0.5 dB for low rate systems with rates on the order of

hundreds of kilobits per seconds, and 2.0 dB for higher rate systems) added to the overall link C/N requirements.

For the link analysis, three cases must be analyzed: the uplink, possible crosslinks, and the downlink. The following link analyses use Constellation F, the 1990 Iridium 77-satellite system, as a representative system configuration for the presentation. Equation 10.6 is used throughout the following analyses to determine the received C/N value at the receiver:

$$C/N(dB) = P_t + G_t + G_r - B_N - k - T_s + L_p \,. \tag{10.6}$$

Parameter L_p in Equation 10.6 includes the path loss and two additional losses: polarization (0.5 dB) and gaseous (0.3 dB) for the L-Band calculations. For the gateway link calculations which follow, the gaseous losses are more predominant (2.5 dB). This higher gaseous loss results from the use of higher transmission frequencies for the gateways than the IUUs. Polarization losses for the gateway and IUU links are identical.

Tables 10.3, 10.4, and 10.5 summarize the link analyses for Constellation F (the December 1990 Iridium baseline). Appendix A summarizes the link analyses for constellations A through E, respectively. Before presenting the link analyses for Constellation F, a review of the analyses provided by Motorola [Mot90] is provided.

10.2.1 The Iridium Link Analyses The Iridium link analyses are of interest due to the inconsistencies and errors presented in their analyses. The first inconsistency noted in the review of the FCC filing is that of noise bandwidth (B_N) used in the link analyses for the L-band links (ISU-to-satellite and satellite-to-ISU). ISU is the abbreviation for Iridium Subscriber Unit. Motorola quotes throughout the document burst symbol rates of 180 kbps (uplink) and 400 kbps (downlink). For these values, using QPSK modulation, the ideal noise bandwidths of the raised cosine filters are 90 and 200 kHz for the uplink and downlink, respectively. In [Mot90], it appears that the

channel bandwidth was used instead of the ideal noise bandwidth. This adds 1.76 dB to the uplink and 1.46 dB to the downlink C/N calculations. These errors propagate to the overall margin of the link. The second inconsistency in the link analyses is in the form of the required E_b/N_o for the link. Recall that the required E_b/N_o for a 10^{-6} BER in a QPSK link is 10.6 dB. This is the required value prior to coding. Coding for the L-band links is quoted to be 3/4 rate, K=7 convolutional coding. For this type of coding, the achievable gain is approximately 4.8 dB. This implies that the required "after coding" E_b/N_o is 5.8 dB. The required values quoted by Motorola in Appendix A [Mot90], Tables A-2 and A-4 are 3.1 dB (satellite to mobile link) and 3.6 dB (mobile to satellite), respectively. From the information provided in the FCC filing, there is no explanation for this discrepancy in required E_b/N_o values. In using these E_b/N_0 values, Motorola has erroneously added 2.7 dB (downlink) and 2.2 dB (uplink) to the overall margin calculations of the particular link. The overall impact of these noted errors is to allow Motorola to claim lower transmitter power levels than those actually required to maintain stated (10^{-6}) bit error rates. The lower output power levels in-turn mean lower input power levels needed to operate the IUUs and the satellites.

10.2.2 L-Band IUU-Satellite Link Analyses The L-band (1.6 GHz) analyses consist of the examination of the communication links between the mobile user (IUU) and the communicating satellite operating in a clear sky environment. Table 10.3 provides the information necessary to perform these analyses. Design parameter values summarized in Table 10.3 are those previously defined in Section 6 of Chapter 9. The slant range value used represents the maximum distance between the IUU and the satellite, given a 10° minimum elevation look angle constraint. Also addressed in Table 10.3 is the carrier to interference ratio (C/I) of 18.0 decibels which results from the physical implementation of multiple beams and adjacent channels. Though this investigation assumes independent operation of the satellite beams, the effects of adjacent beams must be considered to accurately model the system operation. With the 18.0 decibels C/I factor included,

the received C/N is reduced by 3.0 decibels. From Table 10.3, the clear sky margins are 5.6 and 1.6 decibels for the uplink and downlink, respectively.

Table 10.3. Constellation F (77-satellite) IUU-Satellite Clear Sky Link Analyses.

	UPLINK	DOWNLINK
Slant Range (km)	2293.2	2293.2
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Antenna Transmitter Power (dBW)	4.61	2.31
Antenna Transmitter Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	163.7	163.7
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	164.5	164.5
G_r - Antenna Receiver Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R - Coded Data Rate (Mbps)	0.180	0.400
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dBW/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.9	11.9
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	10.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

10.2.3 Gateway-Satellite Link Analyses Table 10.4 summarizes the uplink and downlink design for the gateway-satellite links. The values shown in Table 10.4 represent clear sky conditions with no rain in the path. As was the case in Table 10.3, maximum slant range distances are used in the gateway-satellite link analyses. The required C/N value is derived from a BER of

10^{-6} (13.6 dB) minus the coding gain (5.8 dB) for a 1/2 rate convolutional code with K=7. Channel data rates are 1/2 the coded data rates defined in Chapter 9. With the increase in transmission frequencies, the atmospheric gas loss effects are increased over those presented in the L-band analysis. Clear sky margins for the gateway-satellite links are 4.5 and 3.6 decibels for the respective uplinks and downlinks. These values assume a 10^{-6} BER.

Table 10.4. Constellation F (77-satellite) Gateway-Satellite Clear Sky Link Analyses.

	UPLINK	DOWNLINK
Slant Range (km)	2293.2	2293.2
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Antenna Transmitter Power (dBW)	-2.35	0.25
Antenna Transmitter Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	189.2	185.7
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	192.2	188.7
G_r - Antenna Receiver Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.96	67.96
k - Boltzmann's constant (dBW/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.3	13.4
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.5	3.6

10.2.4 Satellite-Satellite Link Analyses Table 10.5 summarizes the intersatellite crosslink design. The required E_b/N_0 value is derived from a BER of 10^{-6} (10.5 dB) minus the coding gain (5.8 dB) for a 1/2 rate convolutional code with K=7. Channel data rates are 1/2 the coded data rates defined above. Note that no atmospheric losses are encountered in these links and the elevation angle parameter is not applicable. Recall from Chapter 9 that four possible crosslinks are available at any satellite. These crosslinks are to the satellites forward and aft of a chosen satellite in addition to the nearest neighboring satellite in each of the two adjacent orbital planes. The slant range value used in Table 10.5 represents the maximum value obtained from either the interplanar or intraplanar separation of the satellites.

10.2.5 Rain Analysis All of the link analyses are performed in a clear sky environment. The simulation models developed in support of this investigation have built-in routines to allow for rain analyses. A design decision was made to perform the simulations in a clear sky environment rather than one containing a rainy environment because of the infinite number of transmitter power levels that can result from the variation of rain rates. The clear sky environment assumption does not limit the scope nor the importance of the research. The effects of rain on the communications model has the impact of attenuation which can cause higher bit error rates. These effects, though, do not considerably affect the delay performance parameter. This section is provided for completeness and also to explain the design issues that must be considered in a rainy environment.

While the clear sky design is valid for more than 90 percent of a calendar year, additional analysis and design of operational systems should include the periods of time when rain is present. For radio transmissions, signal attenuation due to rain is significant for radio transmissions operating at frequencies at or above 10 GHz [PrB86]. For this investigation, the gateway-to-satellite and satellite-to-gateway links are affected by rain due to the frequencies used (30 GHz up and 20 GHz down).

Table 10.5. Constellation F (77-satellite) Satellite-Satellite Link Analysis.

	CROSSLINK
Slant Range (km)	4020.9
Elevation Angle (deg)	---
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Antenna Transmitter Power (dBW)	4.85
Antenna Transmitter Gain (dBi)	36.0
L_p - Path Loss (dB)	191.8
L_a - Atmospheric Loss (dB)	---
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	192.3
G_r - Antenna Receiver Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R - Coded Data Rate (Mbps)	25.0
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dBW/k/Hz)	-228.6
Received C/N (dB)	13.2
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.4

Two well-accepted models for determining the affects due to rain are the Crane Global Rain Attenuation Model [Cra80] and the Simple Attenuation Model (SAM) [StD82]. The SAM model is used in this investigation. The decision to use the SAM model rather than the Crane model results from its simplicity of implementation and accuracy of results relative to the Crane model. In each of these models, estimations of attenuation due to rain can be made from a set of specified parameters. These models are used to determine the link margin, in decibels, necessary to withstand specific rain rates. The rain rates, in turn, are used to specify link availability.

The traditional approach to link design and margins is to specify a link availability percentage. Link availability is normally desired for 99.99 percent of the total annual operation time [PrB86]. This availability is desired when a large number of voice or data channels are transmitted on a single carrier. This investigation assumes a link availability of 99.5 percent. This assumption is based on a similar analysis performed by Motorola [Mot90] for their Iridium system. From the link availability rate, references to [Cra80, StD82] can be made to determine the rainfall rate which causes a link outage of 0.5 percent. From this rain rate, the attenuation due to rain can be calculated and the system designed to withstand this attenuation. The analysis presented herein assumes a specified margin (obtained from the clear sky analysis) and works backwards to determine the link availability.

10.2.5.1 SAM Calculations From Table 10.5, the uplink and downlink margins are 4.6 and 3.7 decibels, respectively. The SAM model uses Equation 10.7 to determine the attenuation caused by a specific rainfall rate when the rate is less than 10 millimeters per hour.

$$A = aR^b L \text{ (dB)} \tag{10.7}$$

In Equation 10.7, the variables a and b are transmitter frequency (f) dependent and defined for this investigation by Equations 10.8 and 10.9 [PrB86]. The R value in Equation 10.7 denotes the rainfall rate, while L denotes the path length in rain. The path length L is defined by Equation 10.9 where H_0 denotes the earthstation kilometer height above sea level, and H_e denotes the effective storm height in kilometers. The effective storm height, H_e, is related to the zero degree isotherm by Equation 10.11, shown below. In Equation 10.11, Λ_e represents the earthstation latitude. Values for f in the following equations are expressed in GHz.

$$a = 4.21 \times 10^{-5} f^{2.42} \qquad 2.9 \le f \le 54 \text{ GHz} \tag{10.8}$$

$$b = \begin{cases} 1.41f^{0.0779} & 8.5 \le f < 25 \text{ GHz} \\ 2.63f^{0.272} & 25 \le f < 164 \text{ GHz} \end{cases} \tag{10.9}$$

$$L = \frac{H_e - H_0}{\sin(El)} \text{ km} \tag{10.10}$$

$$H_e = \begin{cases} 7.8 - 0.1|\Lambda_e| & |\Lambda_e| > 30°, R \le 10 \text{ mm/hr} \\ 7.8 - 0.1|\Lambda_e| + \log_{10}(R/10) & R > 10 \text{ mm/hr} \end{cases} \tag{10.11}$$

Using Equations 10.7 to 10.11 along with the uplink and downlink transmission frequencies and calculated link margins, the outage rain rate can be determined. From this obtained rain rate, the link availability can be calculated. The earthstation of interest for rain attenuation is located near the Washington, D.C. area as defined in Chapter 9. The parameters associated with this earthstation are latitude of 39° and H_0 of approximately 0.5 kilometers. Using these parameters and the above equations, the outage rain rates are approximately 1.46 mm/hr and 2.85 mm/hr for the uplink and downlink, respectively. These outage rain rates translate to a link availability of approximately 98 percent.

To achieve the desired link availability of 99.5 percent defined in Chapter 9, it is necessary to determine the rainfall rate which is exceeded 0.5 percent of the calendar year. This is obtained from the CCIR rainfall intensity exceeded table listed in [Ipp89]. From this table, the approximate rainfall rate for Region D_2 in which the rate is exceeded 0.5 percent of the time is 5.2 mm/hr. Using the approach defined above and assuming a worst case elevation angle of 10°, which yields a worst case rain path length, the uplink and downlink attenuations are 17.3 and 7.3 decibels, respectively. The difference between the rain attenuation and the clear sky link margins (12.7 dB up, 3.6 dB down) represent additional requirements that must satisfied to maintain the link availability rates.

10.2.5.2 Rain Fade Mitigation Many possible approaches exist to combat the effects of rain fading. Obvious approaches are to increase the Effective Isotropic Radiated Power (EIRP (P_tG_t)) and/or G_r/T_s, and reduce the transmission frequency below 10 GHz. These approaches have limitations due to technological, regulatory, and radio interference constraints [Ipp89]. Because of these constraints and limitations, system designs are examining the benefits of diversity either through orbital or site techniques. Motorola proposes site diversity for their Iridium system [LeM93].

In the Iridium system, site diversity is accomplished through the separation of gateway receivers at a distance of approximately twenty (20) miles (32 kilometers). This separation of receivers allow for a signal gain advantage or diversity gain, G_D, not obtainable with a single-site receiver. To calculate the approximate gain achieved through site diversity, the model developed by Hodge [Hod82] is used. Hodge's model takes into account factors such as site separation distance, single-site attenuation, link frequency, elevation angle, and an intersite baseline to link path angle. Diversity gain, as defined by Hodge's model is highly dependent upon the amount of single-site attenuation experienced by the earthstation. As an example of the use of this model, uplink and downlink single-site attenuation values of 12.7 and 3.6 decibels, respectively, are considered. These values represent the difference between the clear sky margins assumed above for the link analyses and the margins required to provide 99.5 percent link availability. Best and worst case diversity gains are calculated for the gateway-satellite uplinks and downlinks. The best case gains for these links are 8.55 (uplink) and 1.73 (downlink) decibels. Worst case gains are 5.08 (uplink) and 1.03 (downlink) decibels. Using the worst case gains, the approximate values of 7.6 (uplink) and 2.6 (downlink) decibels of signal margin must be provided to maintain the desired link availability.

10.3 Delay Analyses

The delay analyses consist of the simulations of six constellations, defined previously, and the observation of the average delay, in milliseconds, incurred by a packet as it traverses the network from source to destination. Because two types of transmitters and receivers exist (the gateways and the mobiles) with differing transmission rates, the average delay observed by the system simulations are broken down into five separate categories. These categories represent the following communication combinations: gateway-to-gateway, gateway-to-IUU, IUU-to-gateway, IUU-to-IUU, and an aggregate delay which measures the delay regardless of the type of transmitter/receiver. The average delay, T_{packet}, is represented symbolically by Equation 10.12:

$$T_{packet} = T_{TDMA} + T_{transmission} + T_{uplink} + N\left(T_{satproc} + T_{satq}\right) + \left(N-1\right)T_{cross} + T_{downlink}, \quad (10.12)$$

where T_{TDMA} is the average delay incurred in using the TDMA scheme, $T_{transmission}$ is the transmission time of the packet, T_{uplink}, $T_{downlink}$, and T_{cross} are the propagation times for the respective links, $T_{satproc}$ and T_{satq} are the satellite processing time (decoding and route selection) and time a packet spends in a satellite queue, respectively and N is the number of satellites a packet encounters as it traverses the network from source to destination.

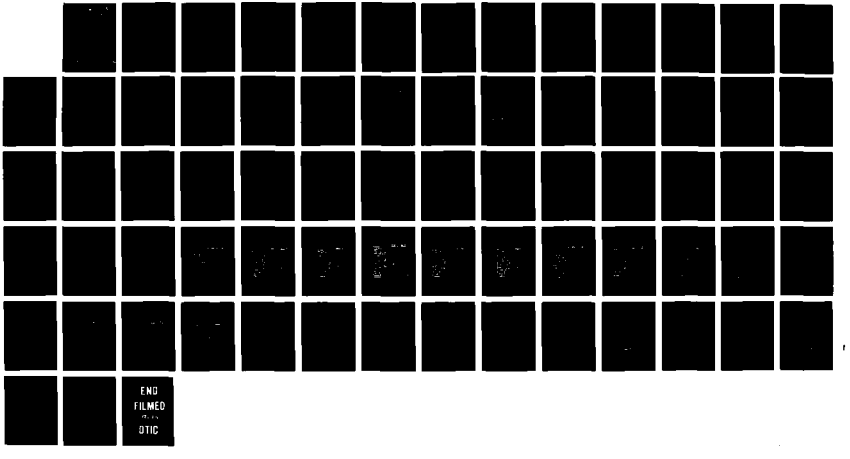
10.3.1 Individual Network Delay Performance To gauge the delay performance of each of the networks modeled, a set of five common system loading values is defined. As defined in Chapter 9, the system loading is the percentage of transmitters active per TDMA frame times the number of channels allocated to a particular type of transmitter. For this investigation, 2 gateway channels and 15 IUU channels are available per satellite beam. These channel numbers represent a 1/10 scale version of the actual beam capacity, as defined in Chapter 9. Within each of the channels are the TDMA frames which support slots assigned to individual users or gateways. In Chapter 9, it was shown that a gateway frame can support approximately 366 users while an IUU

AD-A283 243 PERFORMANCE MODELING AND ANALYSIS OF PARALLEL PROCESSING AND LOW EARTH ORBIT SATELLITE COMMUNICATION SYSTEMS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH R A RAINES 1994 AFIT-CI-94-027D

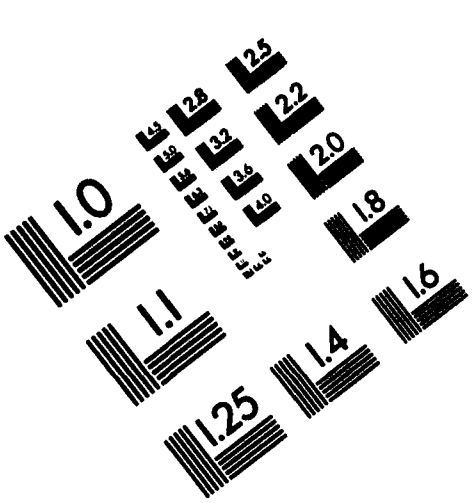
3/3

UNCLASSIFIED

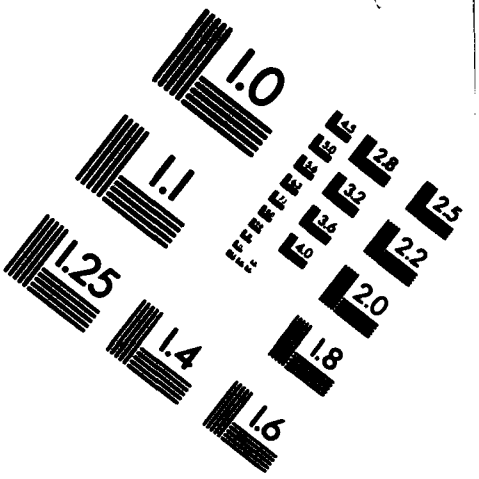
NL



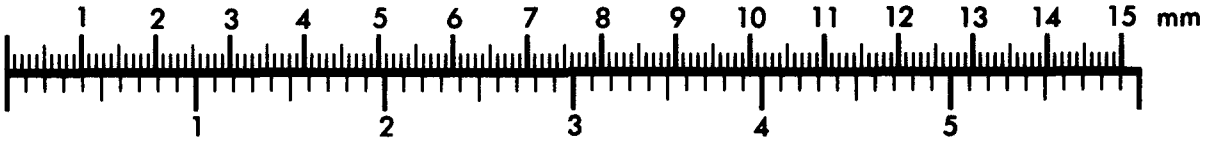
END
FILMED
DTIC



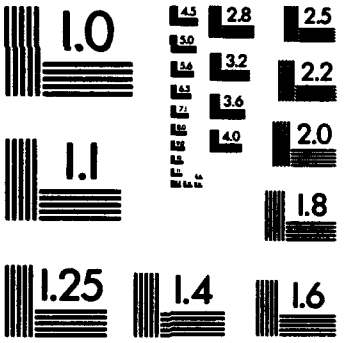
Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

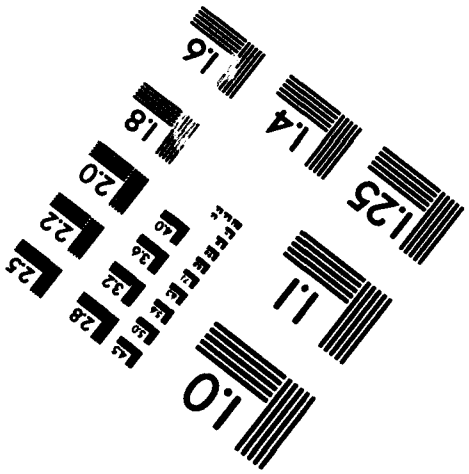


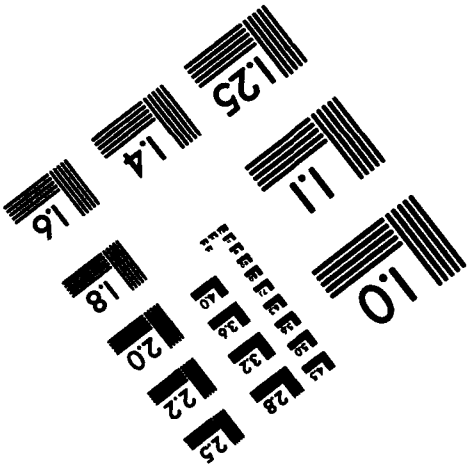
Centimeter



Inches







MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

frame can support 14 users, based on a 60 millisecond frame. In the scaled model, these numbers are reduced to 37 and 1 for the gateway and IUU frames, respectively. As an example of the loading percentage, a loading factor of 100 percent would physically translate to every transmitter transmitting at the maximum rate possible under the TDMA timing constraints and thereby filling each of the available slots. The five common loading factors used to evaluate the networks are 10, 30, 50, 70, and 90 percent of the maximum user capacity of a single beam.

Figure 10.1 shows the delay curves for the 77-satellite (Constellation F) network. The delay curves associated with this constellation are representative of the trends found in the modeling and analysis of the six LEO networks. This figure shows the average packet delay as a function of the system loading. Each data point represents the average packet delay from three independent replications of the simulation trial. A unique seed value is used with each replication for number and distribution generators. For each replication, a mean delay value is obtained. The degree of variance from the replication mean is highly dependent upon the type of source-destination transmission path as well as the fact that the satellites are moving relative to the earth-based transmitters. The standard deviation from the mean of a particular replication ranges from 1 percent for a gateway-gateway path to 13.5 percent for a gateway-IUU link. The differences in standard deviation are to be expected for the following reasons. First, the low percentage deviation for the gateway-gateway link results from the fact that the mean communication distance between the two gateways is only affected by location of the communications satellite. The higher variability for the gateway-IUU link results from multiple IUUs located at differing geographic locations, non-uniformly distributed, along with non-equidistant spacing from the two gateways. As with all the links, the replication mean variance is affected by the location of the mobile satellite used for communications. With regard to the variance of the replicate means, the standard deviation does not exceed 1 percent of the overall mean value for any given link.

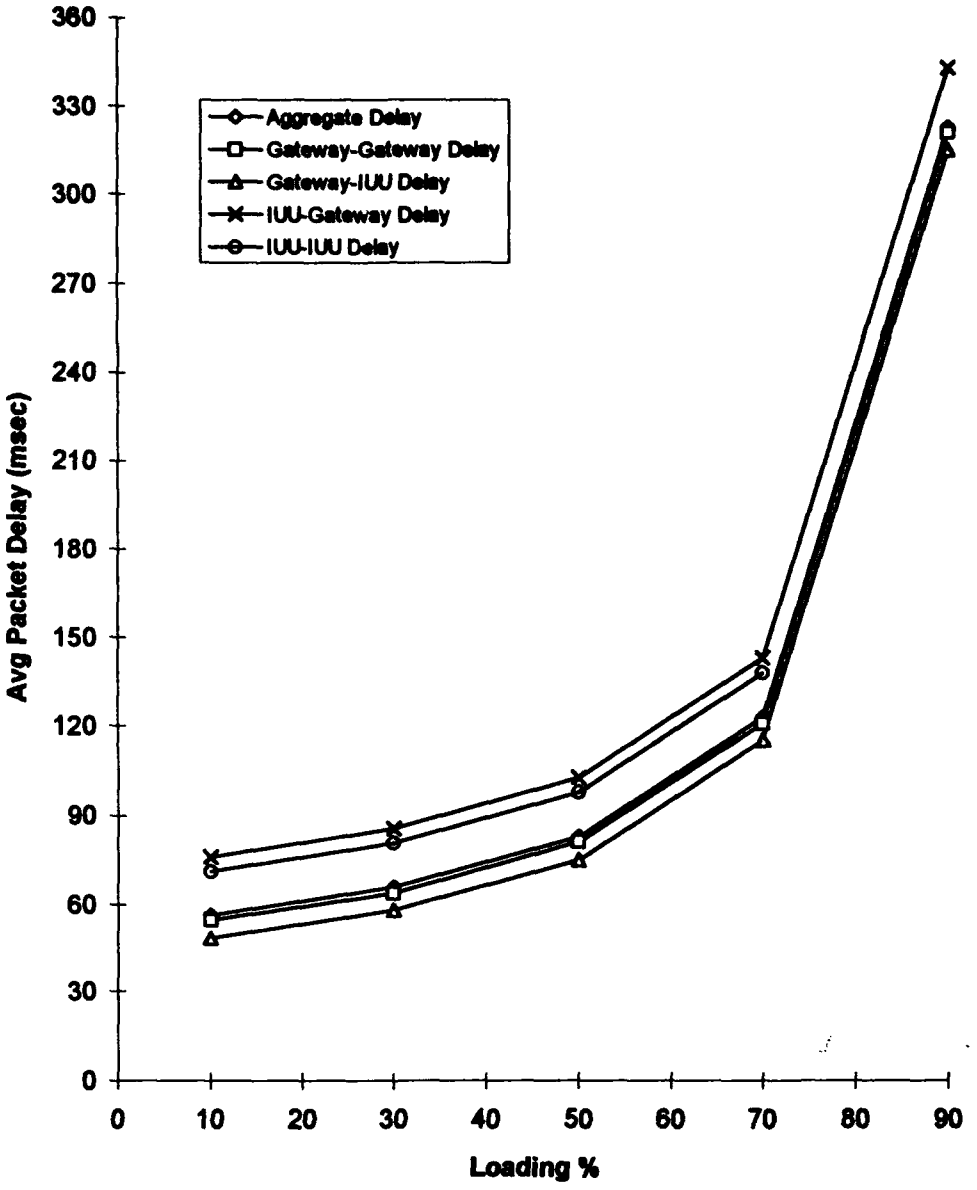


Figure 10.1. Communication link delay curves for a 77-satellite constellation.

Figure 10.1 shows the IUU to gateway delay to be the greatest of the four communication links modeled. This is to be expected due to the following reasons. First, the uplink transmission burst bit rate is slower (180 kbps) for the IUU as compared to that of the gateway (12.5 Mbps).

This factor causes greater transmission delays for lower rates. Second, the TDMA delay associated with the overall delay is also affected by the burst bit rate [Ha90]. For a given loading value and fixed frame length, higher burst rates result in lower TDMA delays. A third reason for the IUU to gateway link to have the greatest delay stems from the position locations of the IUU mentioned above. Because the IUUs are not uniformly spaced nor equidistant from the gateways, transmissions to the communicating satellites are more greatly affected when the lower burst rate is used. The IUU to IUU delay is slightly less (approximately five milliseconds) than the IUU to gateway. This is due in large part to the fact that the IUUs are located within the footprint of a single beam from the satellite. While the round trip propagation distances (for this investigation) are the smallest for the IUU to IUU path, this path has the two slowest transmission burst rates (180 kbps for the uplink and 400 kbps for the downlink). As in all cases, the TDMA delay is also affected by the transmission rate. In all the constellations examined, the gateway to IUU links experienced the least average packet delay of the four links modeled. This is because of the greater transmission burst rates (12.5 Mbps up and 400 kbps down) than any other link pair, except the gateway to gateway link, and also the locations of the destinations relative to the source. As defined in Chapter 9, 80 percent of the total packet traffic is assumed to be generated by the two gateways and, that by random placement, the IUU grouping is located closer to the Los Angeles gateway than the D.C. gateway. Also recall that destination addresses are uniformly distributed among the earthstations. This means that 50 percent of the total gateway traffic destined for IUUs has shorter propagation distances than the other 50 percent. It is believed that this is the reason why the gateway to IUU link experiences less average delay than the gateway to gateway link which has a constant coast-to-coast distance to traverse. It is also noted that all the delay times shown in Figure 10.1 are dependent on the locations of the communications satellites which are varying with time.

Appendix B contains the delay curves for Constellations A through E as well as individual link comparisons to similar links of a geostationary satellite. The figures shown in Appendix B show trends similar to those observed in Figure 10.1. In each set of delay curves, the IUU to gateway delay is the greatest among the four links. Slightly lower delays are experienced for the IUU to IUU link. A comparison of IUU to gateway link delays with IUU to IUU link delays reveals a variance of 1 to 7 milliseconds depending upon the system configuration. The 36-satellite system (Constellation A) experiences the minimum difference (one millisecond), while the 45-satellite system (Constellation B) experiences the greatest delay difference (seven milliseconds). The minimum delay differences of these two links in the 36-satellite system can be attributed to the fact that satellites in this system have the longest in-view times (17 minutes) of the constellations modeled. This factor favorably impacts the delay performance by minimizing the number of crosslinks required by a packet transmission. A similar relationship for the delay differences can be seen for the 36-satellite gateway to gateway and gateway to IUU links.

As stated previously, the expected workload for a LEO system operating in a packet-switched data environment is unknown. Poisson processes used for traffic generation provided one possible, among many, characterization of the system workload. To examine the sensitivity of the delay model to changes in the traffic generation distribution, a constant traffic distribution was also examined. Because of the long periods of time required to collect the simulation data (summarized in Table 9.6), representative sample simulations were performed to observe the effects on packet delay caused by the different distributions. The 36-satellite constellation was chosen as the representative constellation. Simulations were performed for 10 and 90 percent system loading. These loading values were chosen to capture the traffic distribution type effects at each end to the loading spectrum. The results of these simulations are summarized in Table 10.6. The delay values shown in Table 10.6 indicate that the obtained delay values are sensitive to the input traffic distribution. In all cases, the constant traffic distribution rate yields lower average delay times than

the Poisson distributions. All delay values obtained by using of the constant distribution are within 2 percent of the values obtained from the Poisson distributions.

Table 10.6. Average Packet Delay by Traffic Distribution

		Average Packet Delay (msec)				
Loading %	Traffic Distribution	Gateway-to-Gateway	Gateway-to-IUU	IUU-to-Gateway	IUU-to-IUU	Aggregate
10	Poisson	52.8	51.9	79.6	78.7	57.1
10	Constant	52.0	51.0	78.7	77.7	56.2
90	Poisson	319.5	318.5	346.3	345.3	323.7
90	Constant	318.7	317.6	345.7	344.4	322.9

10.3.2 Network Delay Performance Comparison To get a "feel" for how the six different LEO constellations' delay performances compare, it is necessary to display each of the five possible combinations for the communication links. This type of comparison is believed to be the first of its kind since no such comparison has been found in the open literature. This comparison is important for the following reason. Until this study, there has been no clear and accurate presentation of the relative delays experienced by differing LEO systems. This comparison makes it easy to see how one system performs relative to the other. Subtle and contrary behaviors, such as those described below, can be readily observed from this comparison.

Figures 10.2 through 10.6 graphically display the link delay comparisons. For each of the figures, the trends are consistent; the packet delays tend to increase rapidly beginning at the 70 percent loading factor. It is therefore in the best interest of the system performance to operate at less than 70 percent loading. In doing so, all of the delay values shown in Figures 10.2 through 10.6 meet the CCIR real-time communications constraint of 400 milliseconds. If the average packet delay is used as the sole system design consideration, one could conclude from Figures 10.2

through 10.6 that the delay differences between a 36-satellite system and a 77-satellite system do not warrant using the 77-satellite system over the 36-satellite system.

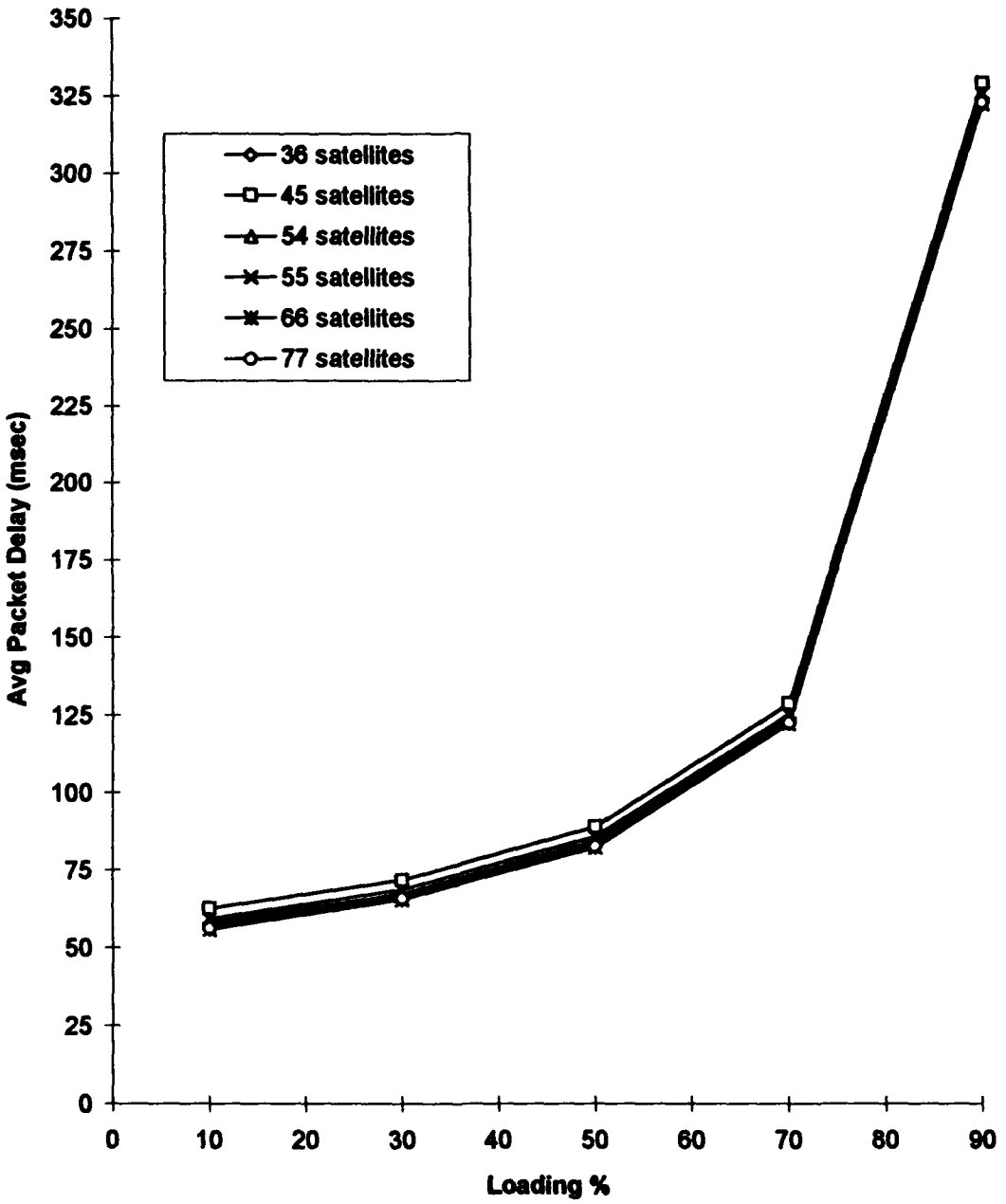


Figure 10.2. Aggregate delay curves for LEO constellations.

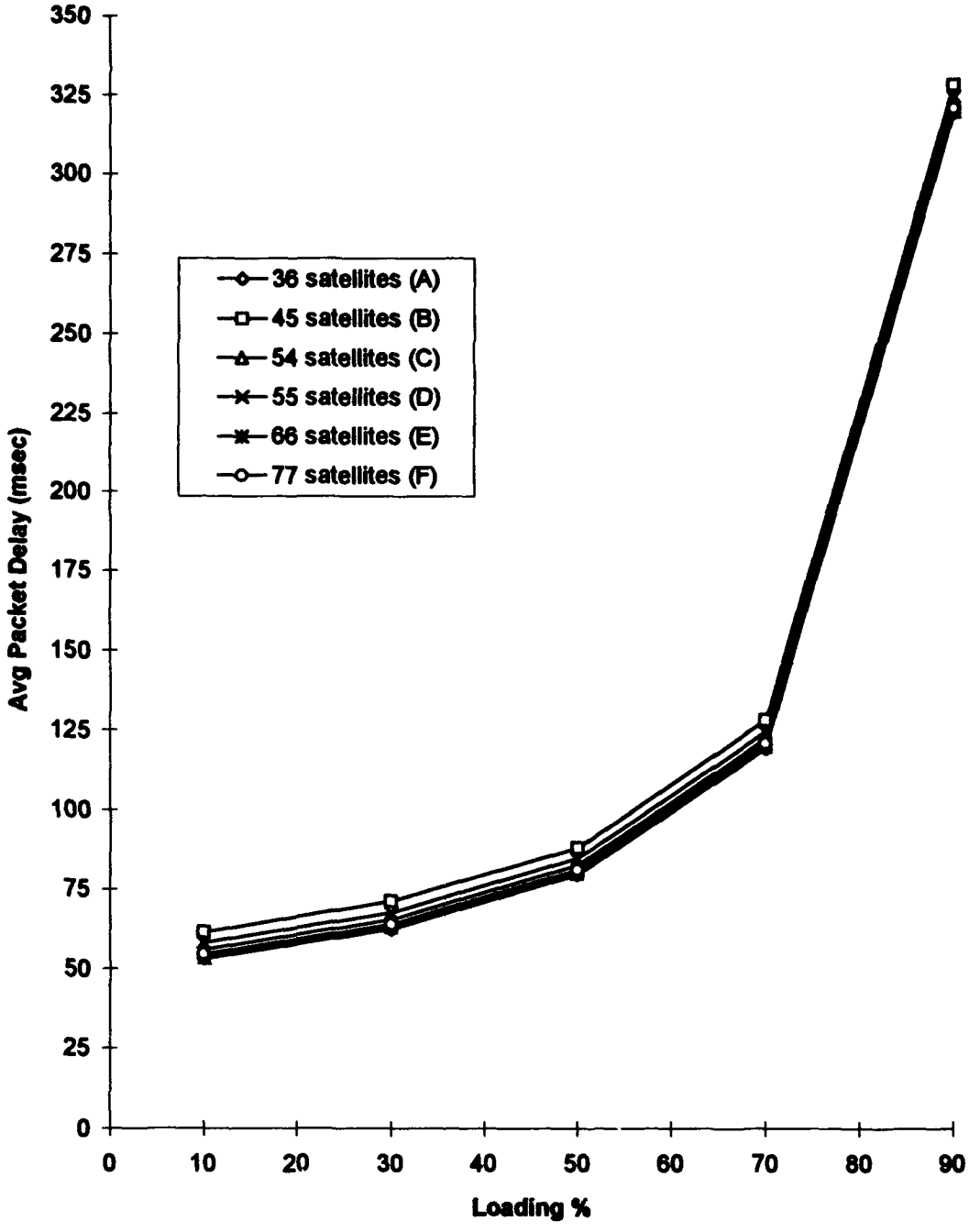


Figure 10.3. Gateway to Gateway link delay curves for LEO constellations.

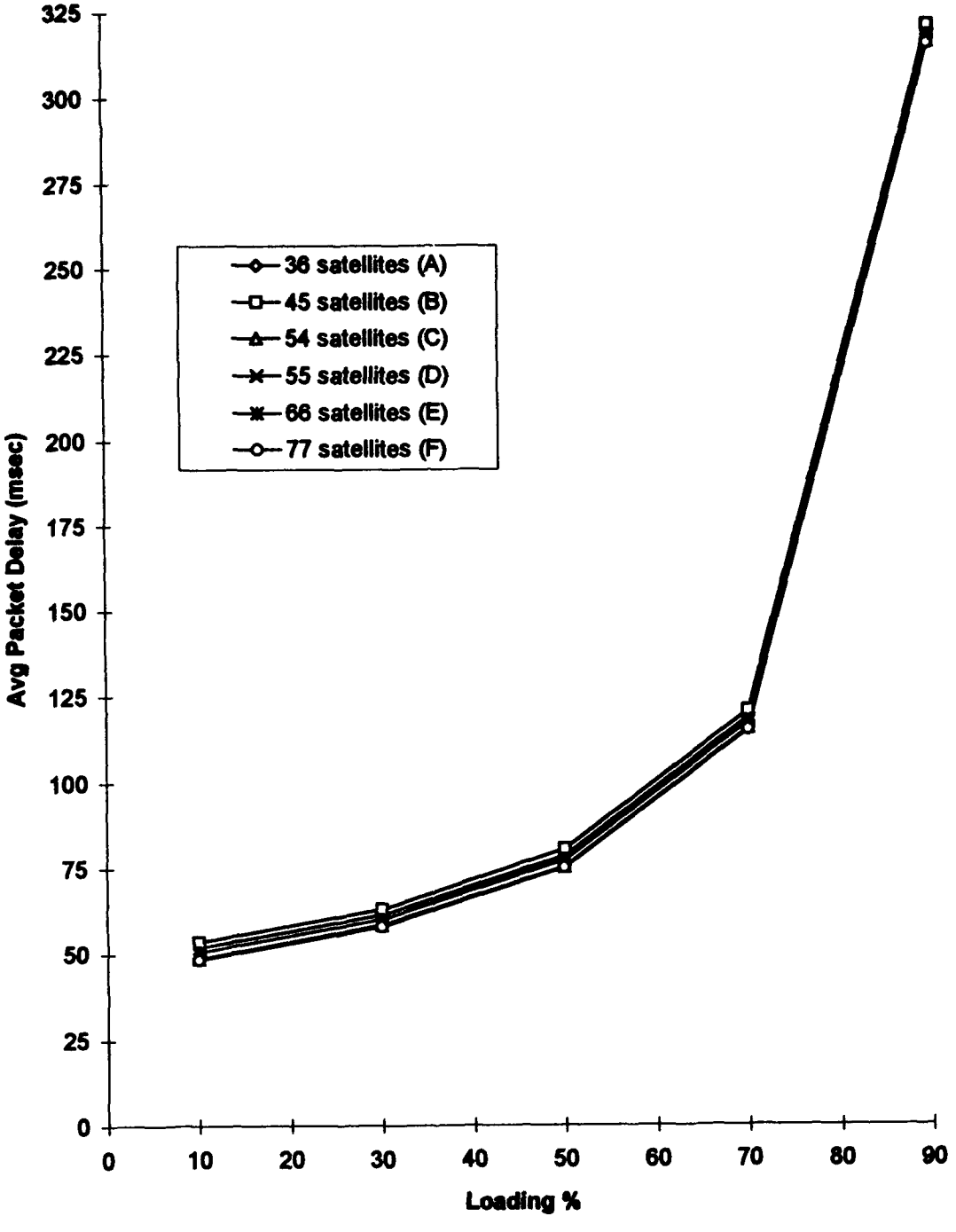


Figure 10.4. Gateway to IUU link delay curves for LEO constellations.

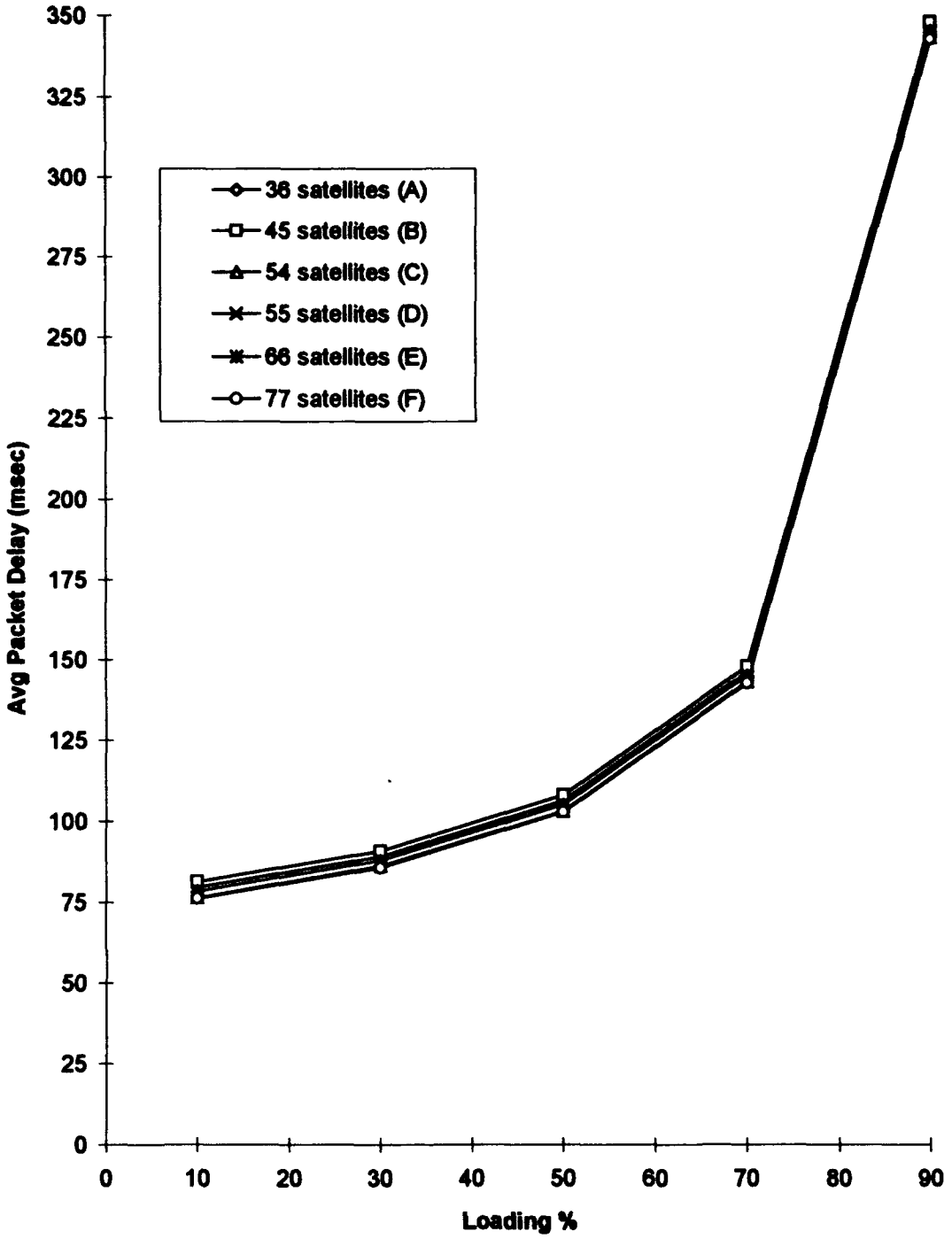


Figure 10.5. IUU to gateway link delay curves for LEO constellations.

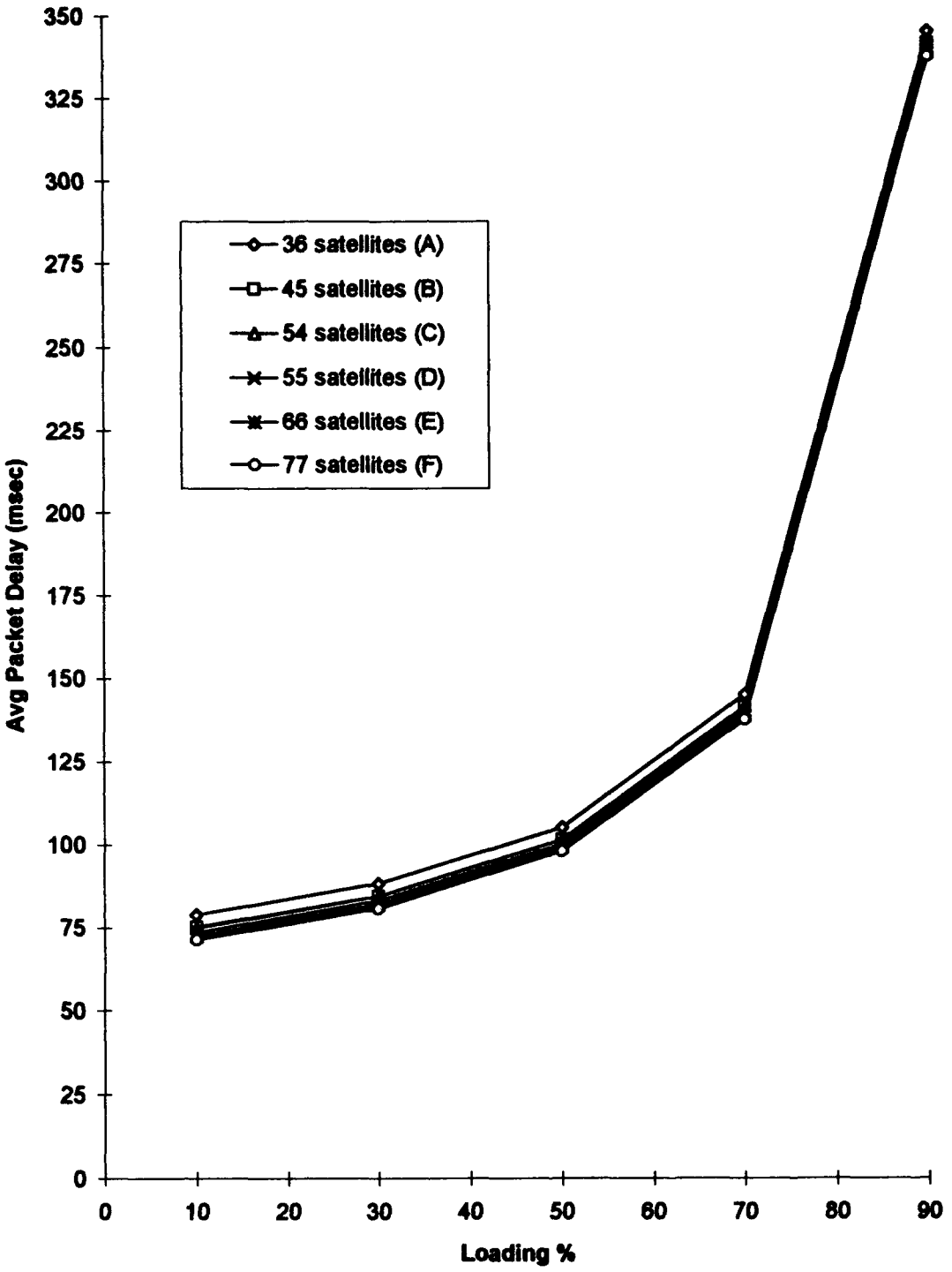


Figure 10.6. IUU to IUU delay curves for LEO constellations.

A second interesting observation is the delay associated with the 45-satellite system (Constellation B). In each of the comparative delay curves except one, the IUU to IUU link, this system experiences the greatest link delays. This tends to be contrary to first thoughts that the 36-satellite system should experience the greatest delays since the earth-space links potentially have the greatest propagation distances. As discussed earlier, the 36-satellite system experiences a minimal number of required crosslinks. This appears to be the likely reason for the delays of the 36-satellite system to be less than those of the 45-satellite system. For the IUU to IUU link delay curve shown in Figure 10.6, the relative delays are as to be expected. The system with the highest orbital altitude (the 36-satellite system) has the greatest delay. The IUU to IUU link delay decreases as the orbital altitudes of the particular system's satellites decrease. The reason for this expected trend results from the fact that all the IUUs are located within the footprint of a single beam of a single satellite. No crosslinks are required for IUU to IUU transmissions.

Figure 10.7 provides a comparison of the aggregate delay times of the LEO system versus a geostationary (GEO) satellite. Comparisons for the four other communication links are displayed in Appendix B. The GEO to LEO link comparisons shown in Appendix B show delay trends similar to those described below. For these comparisons, the same single-satellite capacity is assumed for both LEO and GEO satellites. Figure 10.7 shows the dominant effect of propagation distance on the overall delay. Recall that for a GEO satellite, the round-trip propagation delay is approximately 250 milliseconds. Only when the LEO systems experience 90 percent system loading does the aggregate delay come close to the baseline propagation delay of the GEO. But, at this loading factor, the GEO aggregate delay exceeds the CCIR real-time communications limit of 400 milliseconds. Once again, if average time delay was the only factor to consider in the design of a satellite system, the LEO system clearly outperforms the GEO system.

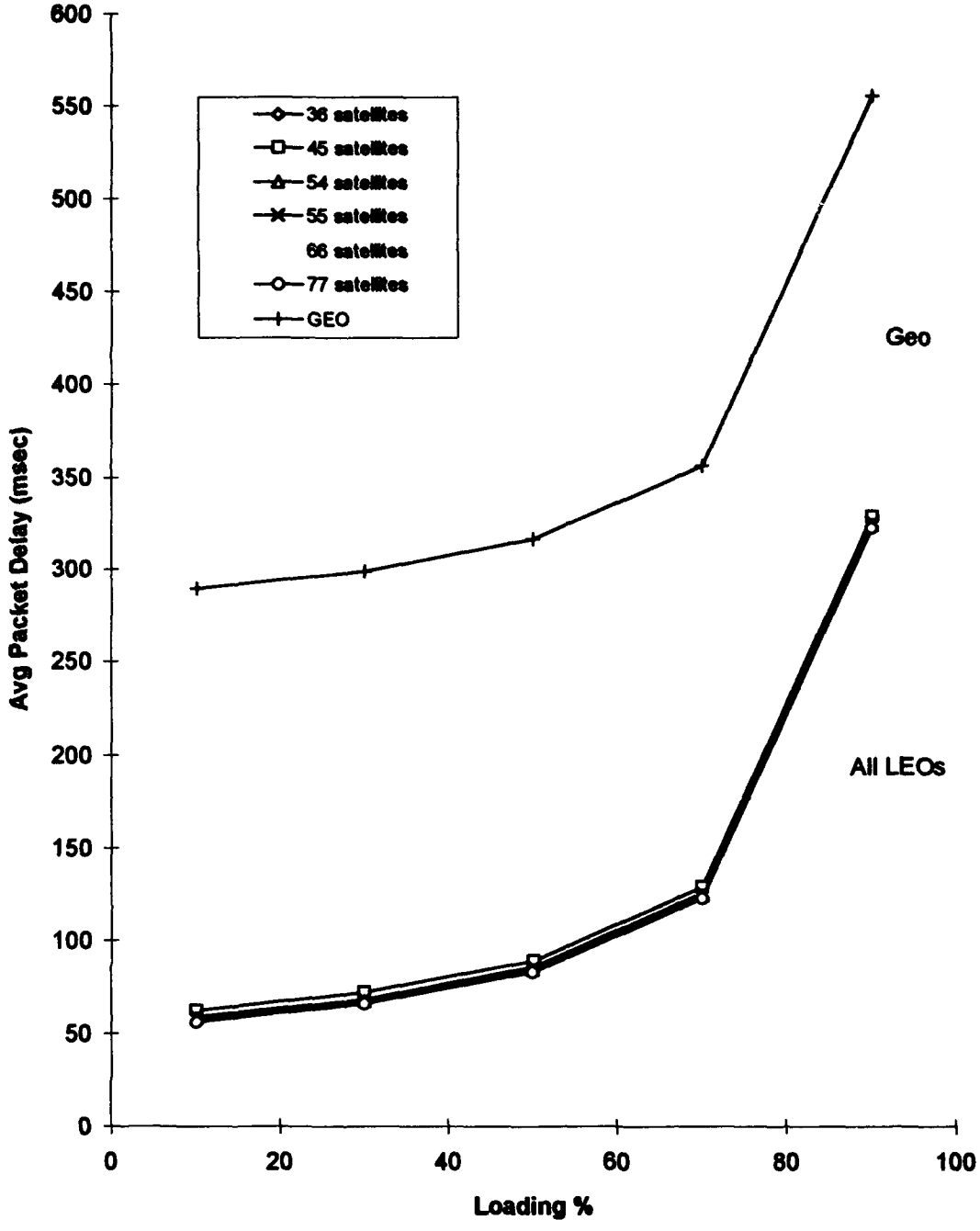


Figure 10.7. Aggregate delay comparison for LEO versus GEO.

Unfortunately, the design of a satellite system is not based on one design consideration. Many factors and analyses must be considered before a final design configuration is made. What the delay curves show is one important performance criteria that must be factored into the total design equation. These delay curves represent an addition to the LEO satellite system analysis process not provided in previously published works. The following section adds yet another factor, satellite resource utilization rates, to the design process.

10.4 Utilization Analysis

The second important area of interest in this research is the resource utilization of the satellites. The determination of resource utilization rates is important because it provides an idea of how the system capacity is being used at any instant of time. One of the design goals of any communications system is to have, as much as possible, a balance of utilization among independent system resources. By this, the system design should try to avoid a configuration which can cause overloading or "bottle-necks" of an individual communications device. This area of the investigation adds another piece to the trade-off analyses needed to determine the system design and constellation configuration.

Each of the satellites modeled in this investigation has a fixed capacity (in terms of channels). This fixed allocation applies to the uplink, crosslink, and downlink capabilities of a satellite. The resources allocated for crosslink usage are much less (approximately 50 percent) than those allocated to the satellite uplinks and downlinks. For the LEO systems examined, requests for uplink and downlink channels are predominant. Examination of the resource requests by link category revealed that crosslink requests never exceeded 27 percent of the total link requests. This occurred for the 45-, 55-, and 77-satellite configurations. The lowest percentage of crosslink requests occurred for the 36-satellite configuration, where the request rate was approximately 0.16 percent of the total requests. It should be noted that these values for crosslink requests are highly dependent upon the physical locations of the satellites. Because of the dynamic

nature of the LEO systems being analyzed, an optimal configuration of satellites over a geographic area is only obtainable for a small period of time. Hence, satellite resource requests will vary with time and across constellations.

To understand how the satellites of a chosen constellation are being utilized, the resource requests by satellite and type of request (uplink, crosslink, or downlink) are captured by the simulations. The results of this data capture are shown in Table 10.7. This table provides insight into the number of satellites being utilized in the ten minute simulation as well as how these satellites are being utilized. Table 10.6 shows that the number of satellites used during the simulation increases, with one exception, as the number of satellites in the constellation increases. This is to be expected since a lower orbital altitude is used in the constellations with a greater numbers of satellites. This lower altitude also means that the satellites have lower in-view times that those in higher orbits. The fact that the 77-satellite configuration uses fewer satellites during the simulation than the 66-satellite configuration can possibly be attributed to the constellation configuration over the geographic area at the simulation start time. Because the two constellations differ in the number of orbital planes, the total number of satellites located over the area of investigation at any instant of time will also differ. This constellation configuration difference can allow a satellite located on the fringe of the investigation area at the simulation start time to have a very low percentage of utilization over the period of the simulation. The very low time-average utilization in turn causes a satellite to not appear to be accessed when compared to other satellites.

Table 10.7 also shows how the uplink and downlink requests are distributed among the satellites used by the simulation model. Note that, the loading distributions shown in this table are time averages. The implication of this time average is the fact that a satellite can have a low overall utilization but can be highly utilized in the short term. Such is believed to be the case for the 36-satellite constellation. From Table 9.3, the approximate in-view time for a satellite in this configuration is 17 minutes. Table 10.7 shows that 92.2 percent of the total resource requests are

routed to a single satellite while two other satellites share the remaining 7.8 percent of the requests. The seventeen minute in-view time along with the coverage area encompassed by a single satellite in the 36-satellite configuration, translates into the high utilization for the single satellite. The two satellites with lower utilization are believed to be fringe satellites. By this, a satellite is only in the in-view period for a short length of time due to its positioning at the start of the simulation time. Further examination of Table 10.7 shows that when load balancing among the satellites is a desired system function, the 54-satellite configuration performs the best in this investigation. The 66-satellite configuration proposed by Motorola in its Iridium system has the highest number of satellites being utilized during the simulation. The main drawback to this configuration is the disproportionately high utilization rate (66 percent) of a single satellite as compared to the other five satellites utilized by the simulation.

Table 10.7. Average Loading Percentage on Satellite

		Distribution Percentage of Total Uplink and Downlink Resource Requests by Constellation					
Satellites in system	Satellites used in simulation	Satellite 1	Satellite 2	Satellite 3	Satellite 4	Satellite 5	Satellite 6
36	3	3.3	92.2	4.5	N/A	N/A	N/A
45	4	20.9	39.9	6.6	32.6	N/A	N/A
54	4	31.2	25.6	32.7	10.5	N/A	N/A
55	4	42.8	17.8	13.2	26.2	N/A	N/A
66	6	13.3	6.8	0.3	66.0	0.01	13.6
77	5	0.7	35.8	22.1	9.1	32.3	N/A

Table 10.7 supports the IUU to gateway and IUU to IUU link delay discussions previously presented for the 36-satellite system. This is seen through the fact that 92.2 percent of the resource

requests in the 36-satellite simulation are routed to a single satellite. This implies, as was observed, that a minimal number of crosslinks (0.16 percent) are requested. Also implied from Table 10.6 and observed through probe data collection and analysis is the correlation between system load balancing and number of crosslinks requested. As the system becomes more balanced in terms of the distribution of requests among satellites, the number of crosslink requests increases.

10.5 Performance Prediction Modeling of LEO Systems

This section presents the results of a predictive mathematical study of delay performance parameters associated with LEO systems. The approach taken here to derive mathematical metamodels parallels the approach used in Chapter 4 to develop metamodels for interconnection networks for parallel processors.

From Equation 10.12, the average packet delay is shown to be a function of many factors. These factors are interrelated by design as well as by simulation parameters. In Equation 10.12, the delay factors T_{TDMA} and T_{satq} are affected by the system loading; T_{uplink} , $T_{downlink}$, T_{cross} , and N are affected by the constellation configuration; and T_{TDMA} and $T_{transmission}$ are affected by the transmission rates of the communication links. The goal of applying metamodeling to LEO systems is to simplify the delay model presented in Equation 10.12, while at the same time developing a model of the system model.

The simplification of Equation 10.12 requires that a system level view be taken of the delay model. Seven of the eight delay factors listed in Equation 10.12 are affected by some or all of the following parameters: the constellation architecture (A), the system loading (L), and the type of communications link (C) as well as the interaction between these three factors. Recognizing this fact forms the basis for, and is the key to, the metamodel development. Algebraically, the average packet delay in a LEO system, D_{ijkl} , can be represented by Equation 10.13, where e_{ijkl} is a lumped error term:

$$D_{ijkl} = A_i + L_j + C_k + AL_{ij} + AC_{ik} + LC_{jk} + ALC_{ijk} + e_{ijkl}. \qquad (10.13)$$

Equation 10.13 shows the maximum possible linear combinations that can affect delay. A three-factor Analysis of Variance (ANOVA) procedure is used to investigate the interactions of these factors and determine the statistical significance of these factors. As noted above, three independent replications for each data point are used in the delay and metamodeling analyses. Table 10.8 shows the design factors for the metamodeling of the packet delay experienced by the six LEO networks. The ANOVA statistics for the experiment are computed using the MINITAB Statistical Software package [Min91]. Table 10.9 shows the results of the ANOVA execution for the experiment.

As discussed in Chapter 4, the parameter of particular interest in the ANOVA analysis is *R-square*. From Table 10.9, it is observed that the R-square is close to 1.0, thereby verifying the descriptive power of the three factors in characterizing the packet delay. The next step in the metamodeling process is to determine the statistical significance of the factors and their interactions.

Table 10.8. Design Factors for Metamodeling of LEO Systems

FACTOR	LEVELS OF VARIATION	LEVEL VALUES
Constellation Architecture (A)	i = 6	36, 45, 54, 55, 66, 77
System Loading Percentage (L)	j = 5	10, 30, 50, 70, 90
Communications Link Type (C)	k = 4	1 - gateway to gateway 2 - gateway to IUU 3 - IUU to gateway 4 - IUU to IUU
m = 3 replications translates to 360 observations		

Table 10.9. ANOVA for Average Packet Delay for LEO Experiment Results

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
Model	4	3380291	845073
Error	355	202518	570
Corrected Total	359	3582809	
Model F = 1481.35	R-Square = 0.943		

The statistical significance of a factor is determined by its "F-value" and the probability of the factor's variation occurring by chance. The F-values result from F-tests which are used to determine the effect of a factor's variation on the performance of the system. In addition, the F-values are products of the ANOVA analyses. For this investigation, the delay data used in the metamodeling meets the normality assumption. The normality of the input data is verified through graphical analysis. For each of the F-tests performed, an α value of 0.05 is used. Large F-values indicate small probabilities that variations in the model occurred by chance. Table 10.10 shows the metamodeling factors, their interaction terms, and corresponding F-values. From Table 10.10, it is concluded that the factors *A, L, C,* and *AC* affect the system delay performance.

One of the goals of the metamodeling effort is to create an all-encompassing relationship for the packet delay, independent of the constellation configuration. This is accomplished via the use of the results of the ANOVA analysis and regression modeling. Using these techniques, a single relationship is formed for the packet delay. This relationship is shown by the general regression equation shown in Equation 10.14.

$$\text{Average Packet Delay, } T = x_0 + x_1 A_i + x_2 L + x_3 C + x_4 L^2 \qquad (10.14)$$

Table 10.10. ANOVA Results for Each Experiment Factor

SOURCE	DF	ANOVA SS	F VALUE	PR > F
NETSIZE (A)	5	1104	110000	0.000
LOAD (L)	4	3531113	440000000	0.000
LINK TYPE (C)	3	49823	8200000	0.000
NETSIZE*LOAD (AL)	20	0	1.44	0.106
NETSIZE* LINK TYPE (AC)	15	768	25000	0.000
LOAD* LINK TYPE (LC)	12	0	3.16	0.000
NETSIZE*LOAD* LINK TYPE (ALC)	60	0	2.17	0.000

Equation 10.14 omits factors from Equation 10.13 that were determined by F-tests to be statistically insignificant. In addition, the interaction term AC was omitted and the term L^2 was added. The addition of the L^2 term resulted from observations of the delay curves which appeared to be quadratic in nature. After this term was added, the AC term no longer had a significant impact on the delay model. This insignificance was observed through separate trials of the regression modeling where the adjusted R-square value changed only slightly with and without the inclusion of the AC term. The adjusted R-square value for this model is 0.943. Equation 10.15 shows explicitly the x_i coefficients derived from the regression model. The high R-square value for this model indicates that Equation 10.15 accurately predicts the system packet delay values.

$$T = 99.1 - 0.11A - 4.26L + 8.28C + 0.0721L^2 \text{ (msec)} \qquad (10.15)$$

The significance of Equation 10.15 lies in its simplistic, yet accurate, representation of the system model. For a particular system being examined, one only needs to provide the number of satellites being used (A), the system loading percentage (L), and the type of communication link

being used (C) to derive the expected packet delay. As an example, consider a 45-satellite system operating at 90 percent loading, using an IUU to gateway link. Equation 10.15 yields a predicted value for the delay of 319.6 milliseconds. The observed simulation value is 348.1 milliseconds. This 28.5 millisecond deviation from the mean represents an 8.4 percent deviation from the observed value. The deviation of the predicted value from the observed simulation value varies across the constellations modeled. Typically, the predicted value is within 20 percent of the observed value and never exceeds 30 percent in the worst case. The deviations between predicted and observed values are attributed to the high variability of packet delays across the systems being modeled.

Additional metamodels are derived for each individual link in the communications system model. The least square estimates for the packet delays of these links are shown in Table 10.11. The values shown in Tables 10.11 are derived from Equation 10.16, a general regression equation for the individual communication links.

$$T = x_0 + x_1 A + x_2 L + x_3 L^2 \tag{10.16}$$

Derived directly from Equation 10.16 and Table 10.11 are the explicit delay equations for the four system communication links. The communications link packet delays are shown in Equations 10.17 through 10.20.

$$T_{gateway-gateway} = 108.0 - 0.056A - 4.26L + 0.0721L^2 \tag{10.17}$$

$$T_{gateway-IUU} = 106 - 0.114A - 4.26L + 0.721L^2 \tag{10.18}$$

$$T_{IUU-gateway} = 134.0 - 0.115A - 4.26L + 0.721L^2 \tag{10.19}$$

$$T_{IUU-IUU} = 132 - 0.155A - 4.26L + 0.721L^2 \qquad (10.20)$$

Table 10.11 lists a high R-square value for each of the derived link models. As was the case for the system model, this R-square value indicates a high degree of accuracy in the predictive nature of the model. The predicted versus observed packet delay value accuracy varies from approximately 4 percent for a gateway-to-gateway link operating at 10 percent loading in a 36-satellite system, to approximately 25 percent for an IUU-to-IUU link operating at 30 percent in a 45-satellite system. The lack of precision in all of the metamodels can be attributed to the high degree of variability in delays observed in each simulation run caused by the dynamic nature of the system being investigated, as discussed in Section 10.3.

Table 10.11. Least Square Estimates for LEO Packet Delay Models

Communications Link	Models' Adjusted R-square	Model Parameters			
		Intercept	Network Size (A)	Load (L)	L^2
gateway-gateway	0.946	108.0	-0.056	-4.26	0.0721
gateway-IUU	0.946	106.0	-0.114	-4.26	0.0721
IUU-gateway	0.946	134.0	-0.115	-4.26	0.0721
IUU-IUU	0.946	132.0	-0.155	-4.26	0.0721

10.6 Summary

This chapter presented the analysis of six low-earth orbit satellite communication networks. The analysis was performed in a packet-switched operating environment. Section 10.2 provided an analysis of the communication links associated with each of the systems. Also addressed in

Section 10.2 were the effects of rain on the gateway links and methods for circumventing these effects. A brief discussion was included on the inconsistencies and errors noted in the Motorola FCC filing for their Iridium system. This section was followed by Section 10.3 in which an analysis of the average delay a packet can expect as it traverses the network was performed. It was shown that if the only design consideration was packet delay, the difference in delay between a 36-satellite system and a 77-satellite system did not warrant using the 77-satellite system over the 36-satellite system. Additionally, a delay performance comparison between LEOs and a GEO of the same capacity was made. Once again, if delay were the only consideration, the LEO systems clearly performed better than the GEO satellite. Satellite utilization was the topic of Section 10.4. In this section, an analysis was conducted to determine both the number of satellites utilized in the investigation area and how the resources among the utilized satellites were being requested. It was shown that from a load balancing criteria, the 54-satellite system provided a better balance of requests than the other configurations. It was also noted that the loading percentages were time averages and that because of the mobile nature of the systems being examined, better utilization and distribution of requests were possible at different times of the orbital configuration. Section 10.5 presented new innovations to LEO satellite system analysis by the application of metamodeling to LEO system packet delays. Compact and concise models were derived to lend insight to the system designer into how incremental changes in the system operating parameters affect the overall performance. These models encapsulate many factors that physically affect packet delay in the LEO system. The beauty of these models lie in their simplistic, yet accurate representation of the system delay behavior. With only the knowledge of the number of satellite in the system, the system loading percentage, and the type of communications link, system packet delay can be predicted. These predicted values have been shown to typically be within twenty percent of the observed value. Predicted accuracies of less than ten percent are achievable. The information contained in this chapter provides only a portion of the information required to fully

analyze and design a LEO system. This information is considered to be of value since there are no previously published works which address these issues.

CHAPTER 11

PART II CONCLUSIONS AND RECOMMENDATIONS

11.1 Summary of Part II Research

Part II of this research effort focused on the performance analysis and predictive modeling of low earth orbit (LEO) satellite communication systems. These systems were modeled to operate in a packet-switched environment. Simulation and mathematical models have been developed to assist the system designer with some of the trade-off analyses necessary before the final design and implementation of a system. The analyses conducted in Part II of this dissertation comprise only a portion of the total analyses that must be performed in the design of a LEO satellite system.

Chapter 6 provided an introduction to the problem to be investigated: the delay and utilization performance of LEO systems. The importance of this research is summarized by the need to have efficient low-cost global communications. This type of communications is necessary to link mobile users in industrialized nations as well as portions of the globe where terrestrial communication is either not feasible or cannot be accomplished due to hostile terrain. Developments in microelectronics and mobile communications have allowed for renewed interest in LEO communications on a global scale.

Satellite communication parameters were the topic of Chapter 7. In Chapter 7, an overview of the basic satellite communication principles were discussed. This discussion included single-satellite coverage area, multiple beam systems, frequency reuse, satellite orbits, global coverage techniques, and multiple access techniques. While not all encompassing, the principles discussed were chosen in direct support of the research to be performed in this effort.

Chapter 8 addressed recent literature with respect to LEO communications. Since satellite channel characteristics are basically common across systems, an in-depth review of channel modeling and analysis was presented. Most of the studies reviewed dealt with geostationary (GEO) systems. Analyses of the "classic" FDMA, TDMA, and CDMA channels were presented along with their perceived relative advantages and disadvantages in the implementation of a given type. The search for literature pertinent to LEO constellation modeling and analysis revealed a scarcity of open literature in the area. Those which did apply were limited in scope and size of the constellations examined. The literature review contained in Chapter 8 also addressed crosslink analyses and routing studies for mobile satellite communications.

The problem definition, system design, and modeling of LEO systems were the topics of Chapter 9. In this chapter, the research scope, operating assumptions, design parameters, and factors were defined. Two design factors, constellation configuration and network loading, were defined in detail for the systems to be evaluated. Chapter 9 also presented a discussion of the simulation model developed to perform this investigation. Run-time performance of the simulations was discussed and summarized. Model verification and validation were also presented in this chapter.

The analyses conducted for the LEO systems were presented in Chapter 10. Four analyses were performed. First, communication link analyses were addressed. These analyses included power level determinations and link budget calculations for each of the four system link types. The link budget analyses were presented for a representative 77-satellite system with the remaining five constellation configuration budget analyses presented in Appendix A. Also included in Chapter 10 was a brief discussion of the inconsistencies noted in the 1990 Motorola Iridium filing to the FCC. Rain analyses for the affected links were also presented. The second analysis in Chapter 10 dealt with average packet delay in a LEO system. Delay curves and analysis of these curves were presented. Third, satellite resource utilization was examined. This analysis consisted of

determining how total resource requests were distributed among the satellites in-view during the simulation period. The fourth analysis developed mathematical metamodels of the system delay characteristics. These models were used to predict the delay performance of the LEO systems.

11.2 Part II Research Effort Conclusions

This portion of the research effort examined areas in LEO satellite communications modeling and analysis not previously found in the open literature. Direct comparisons of the delay and utilization performance of systems ranging in size from 36 to 77 satellites were made.

The overall average packet delay in a LEO system was determined to be a function of eight factors. These factors included propagation times, network size (in terms of the number of satellites), satellite processing times, access method, data rates, and satellite queuing times. The complex interrelationship between these factors made it necessary to perform system simulations in order to observe the delay characteristics of a given system.

It has been shown that the delay performance of a LEO system is superior (lower average delays) to that of a GEO satellite with similar capacity. This was to be expected, due in large part to the greater propagation delays experienced in GEO communications. It was also shown that all the LEO systems investigated meet CCIR real-time requirements for the system loading factors used. In addition, if the only or primary design criteria for implementation of a system was delay performance, the LEO analysis revealed that differences in delay times between a 36-satellite system and a 77-satellite system would not warrant the implementation of the 77-satellite system over the 36-satellite system.

Observed from the satellite resource utilization analysis is that requests for particular satellite resources are highly dependent upon the constellation configuration. Utilization of individual satellite resources were seen to vary from 92 percent down to 0.01 percent. It was also

observed that when balancing of resource requests is a design goal, the 54 satellite system yields the best observed performance.

Mathematical metamodels were derived to provide a predictive tool in determining the expected packet delay for a given LEO system. These metamodels encapsulated the complex eight-factor delay equation to form a concise representation of the behavior of the system in response to varying network size, load, and communication link type. Through the use of statistical analysis techniques, these metamodels were shown to have a high degree of accuracy in representing the system model. Further, it was shown that with only the size of the network, the system loading, and the communication link type given, packet delay could be predicted within twenty percent of the observed values. Best case predictions were within 8 percent of the observed values at the system level and 4 percent within a given link. The implication of the metamodeling capability is reduced simulation requirements and the possibility of accelerated decision making.

11.3 Recommendations for Future Research

This investigation has expanded the knowledge base in the performance modeling and analysis of low earth orbit satellite communication systems. New and innovative models have been developed to predict the delay performance of dissimilar LEO system architectures. While this work has been noteworthy, it has only scratched the surface in terms of possible extensions to the research. The following recommendations are made for future research in the area of LEO satellite communications.

1. Develop and analyze the LEO system models to operate in a circuit-switched environment. This effort would provide insight into how dynamic networks handle various call loading requests, the average call setup times, and the blocking probabilities associated with fixed capacity systems.

2. Extend the packet-switched models to include protocols for retransmissions of lost or corrupted packets along with integration of ATM-based systems. A trade-off analysis performed on different flow-control and retransmission schemes would be of value in determining which protocol best suits LEO applications.

3. Develop a spread-spectrum model for the multiple access portion of the system model. Use this model for comparison of delays and interference with TDMA and FDMA techniques.

4. Extend the geographic area of interest to analyze the performance of the crosslinks in an environment with heavily utilized crosslinks.

5. Examine the possibility of dynamic routing and load balancing among the satellites. Compare the delay and resource utilization performances of load balancing systems to those modeled within this investigation.

6. Model and analyze two LEO systems operating simultaneously over the same geographic area. Examine the effects of interference caused by the operation of these systems.

7. Perform fault-tolerance investigations on LEO systems. By this, allow a percentage of the satellites to become non-operational. Examine the effects of satellite outages on the overall capacity and delay performance metrics of the system. These studies could be performed for both packet and circuit-switched environments.

CHAPTER 12

CONCLUSIONS

This research examined the performance characteristics of two types of communication systems. The first system, a set of cooperative processors, operates in an environment where propagation delays are on the order of a few nanoseconds and queuing delays dominate system performance. The second system, earth- and space-based, operates in an environment where the system is substantially affected by both propagation and queuing delays. Analyses were conducted on both systems to determine the delay performance characteristics of each. For the parallel processing systems examined, system cost metrics were evaluated. For the satellite communication systems examined, resource utilization studies were performed. Mathematical metamodels, which characterize the system delay performance, were developed for both systems. These metamodels were developed to aid the network analyst in determining optimal designs for a chosen application.

The work presented in this dissertation was discussed in two parts. Part I examined the communications performance of parallel processing systems from an interconnection networks point of view. These interconnection networks function to support the inter-processor communications of system containing processors ranging in number from tens to thousands. In these type systems, queuing delays dominate the system performance while propagation delays are considered to be relatively negligible. Four different types of interconnection networks were evaluated in Part I: the multistage cube (MSC), single stage cube (SSC), mesh, and augmented shuffle exchange network (ASEN). From a delay and cost perspective, the MSC was compared to the ASEN. The ASEN was shown to outperform the MSC by 20 to 25 percent in average delay times while having a non-saturation operating region of 15 to 17 percent greater than the MSC. In a system cost comparison, which addressed the cost of the system in terms of buffers, the cost savings of the ASEN relative to the MSC was shown to be between 9.2 percent for a 1024

processor system implemented with 2-by-2 switches to 16.4 percent for a 1024 processor system implemented with 32-by-32 switches.

The metamodels developed for the MSC, SSC, mesh, and ASEN were used to show the effects of incremental changes in network parameters on packet delay. The predictive nature of these models was shown to be highly accurate. An ASEN example of the application of metamodeling revealed less than a 1 percent difference between the predicted delay value and the observed delay value.

Part II of this research investigated the delay and utilization performance of six low earth orbit (LEO) satellite communication systems. These systems were analyzed under various system loading levels in a packet-switched environment. In this type of operating environment, propagation delays have greater effect on system delay than those observed for the interconnection networks evaluated in Part I. The analyses conducted in Part II of this effort provide important contributions to the overall analysis required to design and develop a LEO communications system. Many trade-offs, such as capacity, launch vehicles, satellite costs, and overall system costs were not addressed in this effort.

The delay analysis portion of Part II provided a comparative view of the different communication links within a system as well as how the different configurations performed relative to each other. It was shown that the delay experienced by packets in these systems tend to increase rapidly beginning at the 70 percent loading factor. Also revealed by this portion of the research was the minimal difference in delay between constellation configurations. An implication of the minimal difference in delay values lies in the design choice for the number of satellites to be used in the constellation configuration. If delay were the only design issue, the data obtained within this research supports the use of a 36-satellite system rather than a 77-satellite system. Even though the 36-satellite packet delay was greater than those experienced in the 77-satellite system, the overall system costs would be much lower for the 36-satellite system.

A similar analysis was performed for the satellite resource utilization. Data was collected to investigate how resource requests were being distributed among the satellites. Observed values for

single-satellite resource requests ranged from 0.01 percent to 92 percent. In addition, when the balancing of resource requests was a design goal, it was shown that a 54-satellite system provided the best performance within the constraints of the operating environment.

To aid the system designer in the trade-off analyses process, mathematical metamodels were developed for the purpose of predicting packet delays. These models encapsulated a complex, 8-factor equation into a simple, but concise representation of the system model. The benefit of these models was seen through the fact that only the network size (in terms of the number of satellites), system loading value, and the type of communication link being used were needed to predict, with reasonable accuracy, the packet delay. The predicted to observed value accuracy was shown to range from approximately 4 to 25 percent. This variability in accuracy was attributed to the dynamic nature of the systems being observed.

In conclusion, this research has extended the knowledge base in two areas of communications research: parallel processing and LEO satellite communications. The results of this dissertation effort have been presented in three technical papers and an abstract submission. The first paper [RaR92], presented the performance of the ASEN and the MSC. The second [ShD94], derived and applied metamodeling techniques to three differing interconnection network architectures. The third paper [RaD94], presented a comparative analysis of the MSC and ASEN as well as the application of metamodeling to the ASEN. Technical paper submissions for the LEO satellite systems analyses are planned.

Bibliography

Abr77. N. Abramson, "The throughput of packet broadcast channels," *IEEE Transactions on Communications*, Vol. COM-25, No. 1, January 1977, pp. 117-128.

Abr92. N. Abramson, "Fundamentals of packet multiple access for satellite networks," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 309-316.

AdA87. G. B. Adams III, D. P. Agrawal and H. J. Siegel, "A survey and comparison of fault-tolerant multistage interconnection networks," *Computer*, Vol. 20, No. 6 June 1987, pp. 14-27.

AdR87. W. S. Adams and L. Rider, "Circular polar constellations providing continuous single or multiple coverage above a specified latitude," *The Journal of the Astronautical Sciences*, Vol. 35, No. 2, April-June 1987, pp. 155-192.

AdS82. G. B. Adams III and H. J. Siegel, "The Extra Stage Cube: a fault tolerant interconnection network for supersystems," *IEEE Transactions on Computers*, Vol. C31, No. 5, May 1982, pp. 443-454.

AdS84. G. B. Adams III and H. J. Siegel, "Modifications to improve the fault tolerance of the extra stage cube interconnection network," *1984 International Conference on Parallel Processing*, 1984, pp. 169-173.

AgB88. C. E. Agnew, J. Bhagat, E. A. Hopper, J. D. Kiesling, M. L. Exner, L. Melillo, G. K. Noreen, B. J. Parrott, "The AMSC mobile satellite system," *Proceedings of the Mobile Satellite Conference*, May 1988, pp. 3-9.

Agr85. S. C. Agrawal, *Metamodeling*, The MIT Press, Cambridge, MA, 1985.

AnT90. S. Annelli and V. Trecordi, "Performance evaluation of a distributed reservation TDMA to access ATM user-oriented satellite system," *1990 SBT/IEEE International Telecommunications Symposium - ITS 90*, September 1990, pp. 82-86.

Ash88. C. J. Ashton, "Archimedes: land mobile communications from highly inclined satellite orbits," *IEE 4th International Conference on Satellite Systems for Mobile Communications and Navigation*, October 1988.

BaB68. G. H. Barnes, R. Brown, M. Kato, D. J. Kuck, D. L. Slotnik and R. A. Stokes, "The Illiac IV computer," *IEEE Transaction on Computers*, Vol. C-17, No. 8, August 1968, pp. 746-757.

BeB87. A. A. Bertossi, G. Bongiovanni, and A. Maurizio, "Time slot assignment in SS/TDMA systems with intersatellite links," *IEEE Transactions on Communications*, Vol. COM-35, No.6, June 1987, pp. 602-608.

BeG92. D. Bertsekas and R Gallager, *Data Networks*, Second edition, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

Ben65. V. Benes, *Mathematical Theory of Connecting Networks*, Academic Press, N. Y. 1965.

Bes78. D. C. Beste, "Design of satellite constellations for optimal continuous coverage," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-14, No. 3, May 1978, pp. 466-473.

BiH87. R. Binder, S. D. Huffman, I. Gurantz, and P. A. Vena, "Crosslink architectures for a multiple satellite system," *Proceedings of the IEEE*, Vol. 75, No. 1, January 1987, pp. 74-81.

BoC81. G. Bongiovanni, D. Coppersmith, and C. K. Wong, "An optimal time slot assignment for an SS/TDMA system with variable number of transponders," *IEEE Transactions on Communications*, Vol. COM-29, No. 5, May 1981, pp. 721-726.

Bra84. K. Brayer, "Packet switching for mobile stations via low-orbit satellite network," *Proceedings of the IEEE*, Vol 72, No. 11, November 1984, pp. 1627-1636.

BrD90. M. P. Brown, Jr., R. W. Duesing, L. M. Nguyen, W. A. Sanders, and J. A. Tehrani, "INTELSAT VI transmission design and computer system models for FDMA services," *COMSAT Technical Review*, Vol. 20, No. 2, Fall 1990, pp. 373-399.

CaQ93. D. Campana and P. Quinn, "Spread spectrum communications: reasons to choose this path just keep on widening," *IEEE Potentials*, April 1993, pp. 13-16.

Cha89. D. Chakraborty, "Survivable communication concept via multiple low earth-orbiting satellites," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 25, No. 6, November 1989, pp. 879-889.

CiS86. L. Ciminiera and A. Serra, "A connecting network with fault tolerance capabilities," *IEEE Transactions on Computers*, Vol. C-35, No. 6, June 1986, pp. 578-580.

ClJ89. D. D. Clark, V. Jacobson, J. Romkey, and H. Salwen, "An analysis of TCP processing overhead," *IEEE Communications Magazine*, Vol. 27, No. 6, June 1989, pp. 23-29.

ClW87. L. P. Clare, C-Y Wang, and M. W. Atkinson, "Multiple satellite networks: performance evaluation via simulation," *MILCOM 87: 1987 IEEE Military Communications Conference*, October 1987, pp. 404-411.

Com91. D. E. Comer, *Internetworking with TCP/IP, Volume I, Principles, Protocols, and Architecture, Second Edition*, Prentice Hall, Englewood Cliffs, N. J., 1991.

Com93. BONeS SatLab User's Guide, Comdisco Systems, Incorporated, May 1993.

CoW67. R. W. Conway, W. L. Maxwell, and L. W. Miller, *Theory of Scheduling*, Addison-Wesley, Reading, MA, 1967.

Cra80. R. K. Crane, "Prediction of attenuation by rain," *IEEE Transactions on Communications*, Vol. COM-28, No. 9, September 1980, pp. 1717-1735.

CrG85. W. Crother, J. Goodhue, E. Starr, R. Thomas, W. Milliken and T. Blackadar, "Performance measurements on a 128-node Butterfly parallel processor," *1985 Internation Conference on Parallel Processing*, August 1985, pp. 531-540.

DaS86. N. J. Davis IV and H. J. Siegel, "Performance studies of multiple-packet multistage cube networks and comparison to circuit switching," *1986 International Conference on Parallel Processing*, August 1986, pp. 108-114.

DeC89. A. L. DeCegama, *The Technology of Parallel Processing, Volume 1*, Prentice Hall, Englewood Cliffs, NJ, 1989.

DiB74. J. L. Dicks and M. P. Brown, Jr., "Frequency division multiple access (FDMA) for satellite communication systems," *Proceedings of the IEEE Electronic Aerospace Systems Convention (EASCON)*, October 1974, pp. 167-178.

DiJ81. D. M. Dias and J. R. Jump, "Analysis and simulation of buffered delta networks," *IEEE Transactions on Computers*, Vol. C-30, No. 4, April 1981, pp. 273-282.

DiJ82. D. M. Dias and J. R. Jump, "Augmented and pruned NlogN multistage networks: topology and performance," *1982 International Conference on Parallel Processing*, 1982, pp. 10-11.

DuB92. B. Duzett and R. Buck, "An overview of the nCube 3 supercomputer," *The Fourth Symposium on the Frontiers of Massively Parallel Computation*, October 1992, pp. 458-464.

Feh83. K. Feher, *Digital Communications, Satellite/Earth Station Engineering*, Prentice Hall Inc., 1983.

Fen72. T. Y. Feng, "Some characteristics of associative/parallel processing," *1972 Sagamore Computer Conference on Parallel Processing*, 1972, pp. 5-16.

Fen81. T. Y. Feng, "A survey of interconnection networks," *IEEE Transactions on Computers*, Vol. C-30, No. 12, December 1981, pp. 12-27.

Fly 66. M. J. Flynn, "Very high-speed computing systems," *Proceedings of the IEEE*, Vol. 54, No. 12, December 1966, pp. 1901-1909.

FoO90. G. Forcini and T. Oishi, "An efficient satellite transmission scheme for node interconnection in wideband packet networks," *IEEE International Conference on Communications - ICC '90*, April 1990, pp. 1069-1074.

GaG92. A. Ganz and Y. Gao, "SS/TDMA scheduling for satellite clusters," *IEEE Tranasctions on Communications*, Vol. 40, No. 3, March 1992, pp. 597-603.

GaK74. O. G. Gabbard and P. Kaul, "Time-division multiple access," *Proceeding of the IEEE Electronic Aerospace Systems Convention (EASCON)*, October 1974, pp. 179-184.

GaK91. A. Ganz and G. Karmi, "Satellite clusters: a performance study," *IEEE Transactions on Communications*, Vol. 39, No. 5, May 1991, pp.747-757.

GaL92. A. Ganz and B. Li, "Performance of packet network in satellite clusters," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 6, August 1992, pp. 1012-1019.

GiJ90. K. S. Gilhousen, I. M. Jacobs, R. Padovani, and L. A. Weaver, "Increased capacity using CDMA for mobile satellite communications," *IEEE Transactions on Selected Areas in Communications*, Vol. 8, No. 5, May 1990, pp. 503-514.

GiJ91. K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, May 1991, pp. 472-480.

GoG83. A. Gottlieb, R. Grishman, C. P. Kruskal, K. P. McAuliffe, L. Rudolph and M. Snir, "The NYU Ultracomputer - designing an MIMD shared memory parallel computer," *IEEE Transactions on Computers*, Vol. C-32, Vol. 2, February 1983, pp. 276-290.

GoW85. I. S. Gobal and C. K. Wong, "Minimizing the number of switchings in an SS/TDMA system," *IEEE Transactions on Communications*, Vol. COM-33, No. 6, June 1985, pp. 497-501.

GrK88. M. Greene and D. T. Konkle, "Simulation of satellite motion for intra-satellite pointing," *Proceedings of the Annual Southeastern Symposium on System Theory*, March 1988, pp. 486-490.

Ha90. T. T. Ha, *Digital Satellite Communications*, Second Edition, McGraw-Hill Inc., New York, 1990.

Had93. F. M. Haidara, "Characterization of tropospheric scintillations on earth-space paths in the Ku and Ka frequency bands using the results from the Virginia Tech OLYMPUS experiment," Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg Virginia, May 1993.

Har91. G. E. Hardman, "Engineering Orbcomm: a digital satellite communications system exploiting a range of modern technologies," *Third IEE Conference on Telecommunications*, 1991, pp. 251-256.

Hay88. J. P. Hayes, *Computer Architecture and Organization, Second Edition*, McGraw-Hill Inc., New York, 1988.

Hän77. W. Händler, "The impact of classification schemes on computer architecture," *1977 International Conference on Parallel Processing*, August 1977, pp. 7-15.

Hod82. D. B. Hodge, "An improved model for diversity gain on earth-space propagation paths," *Radio Science*, Vol. 17, No. 6, November-December 1982, pp. 1393-1399.

HsY87. W. T. Hsu, P. C. Yew, and C. Q. Zhu, "An enhancement scheme for hypercube interconnection networks," *1987 International Conference on Parallel Processing*, August 1987, pp. 820-823.

HwB84. K. Hwang and F. A. Briggs, *Computer Architecture and Parallel Processing*, McGraw-Hill, N. Y. , 1984.

Int86. iPSC User's Guide, Intel Corporation, April 1986, pp. 6.14-6.22.

Inu81. T. Inukai, "SS/TDMA networking via ISL," *Proceedings of the International Conference on Communications*, 1981, pp. 70.6.1-70.6.5.

Ipp89. L. J. Ippolito, *Propagation Effects Handbook for Satellite System Design - A Summary of Propagation Impairments on 10 to 100 Ghz Satellite Links with Techniques for System Design*, Fourth Edition, NASA Reference Publication 1082(04), N89-17060, February 1989.

IsM91. Q. Islam, M. S. Mahmoud, and B. G. Evans, "Satellite link performance for an SS-TDMA experiment via Olympus," *International Journal of Satellite Communications*, Vol 9, No. 6, November-December 1991, pp. 415-424.

Itu77. Y. Ito, Y. Urano, T. Muratani, and M. Yamaguchi, "Analysis of a switch matrix for an SS/TDMA system," *Proceedings of the IEEE*, Vol. 65, No. 3, March 1977, pp. 411-419.

Jab82. B. Jabbari, "Combined FDMA-TDMA: a cost effective technique for digital satellite communication networks," *IEEE 1982 International Communications Conference*, Vol. 3, June 1982, pp. 7F.4.1-7F.4.5.

JaG88. I. M. Jacobs, K. S. Gilhousen, L. A. Weaver, K. Renshaw, and T. Murphy, "Comparison of CDMA and FDMA for the MobileStar(sm) system," *Proceedings of the Mobile Satellite Conference*, May 1988, pp. 303-308.

Jai91. R. Jain, *The Art of Computer System Performance Analysis*, John Wley and Sons, Inc., New York, 1991.

JaM92. B. Jabbari and D. McDysan, "Performance of demand assignment TDMA and multicarrier TDMA satellite networks," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 478-486.

JeS86. M. Jeng and H. J. Siegel, "A fault-tolerant multistage interconnection network for multiprocessor systems using dynamic redundancy," *6th International Conference on Distributed Computing Systems*, 1986, pp. 70-77.

Joh88. N. L. Johnson, "Satcom in the Soviet Union," *Satellite Communications*, June 1988, pp. 21-24.

JoR88. K. Joseph and D. Raychaudhuri, "Simulation models for performance evaluation of satellite multiple access protocols," *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 1, January 1988, pp. 210-222.

KaT92. J. Kaniyil, J. Takei, S. Shimamoto, Y. Onozato, I. Oka, and T. Kawabata, "A global message network employing low earth-orbiting satellites," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 418-426.

Kla94. P. J. Klass, "New investors, contenders vie for low-earth orbit service," *Space Technology*, Vol. 14, No. 4, April 4, 1994, pp. 57-58.

KrS83. C. P. Kruskal and M. Snir, "The performance of multistage interconnection networks for multiprocessors," *IEEE Transactions on Computers*, Vol. C-32, No. 12, December 1983, pp. 1091-1098.

KuR87. V. P. Kumar and S. M. Reddy, "A fault-tolerant technique for shuffle-exchange multistage networks," *Computer*, Vol. 20, No. 6, June 1987.

KuR89. V. P. Kumar and A. L. Reibman, "Failure dependent performance analysis of a fault-tolerant multistage interconnection network," *IEEE Transactions on Computers*, Vol 38, No. 12, December 1989, pp. 1703-1713.

Law75. D. H. Lawrie, "Access and alignment of data in an array processor," *IEEE Transactions on Computers*, Vol. C-24, No. 12, December 1975, pp. 1145-1155.

LeM93. R. J. Leopold and A. Miller, "The Iridium™ communications system: person to person on a truly global scale," *IEEE Potentials*, April 1993, pp. 6-9.

Lee89. W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill Book Company, New York, 1989.

LoC91. Loral Cellular Systems, Corporation, FCC filing for GLOBALSTAR system application, Washington D. C., June 1991.

Lüd61. R. D. Lüders, "Satellite networks for continuous zonal coverage," *American Rocket Society Journal*, Vol. 31, No. 2, February 1961, pp. 179-184.

MaR91. G. Maral and J. D. Ridder, "Low earth orbit satellite systems for communications, "*International Journal of Satellite Communications*, Vol 9, No. 4, April 1991, pp. 209-225.

McW87. C. McLochlin, C. Ward, Y. C. Chow, R. Newman-Wolfe, J. N. Wilson, and T. B. Hughes, "Optimizing the delay and reliability of low altitude satellite network topologies," *MILCOM 87: 1987 IEEE Military Communications Conference*, October 1987, pp. 417-422.

Mil91. T. Mizuike, Y. Ito, L. N. Nguyen, E. Maedo, "Computer-aided planning of SS/TDMA network operation," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 1, January 1991, pp. 37-47.

MIN91. MINITAB Reference Manual, Minitab Incorporated, 3081 Enterprise Drive, State College, PA.

Mot90. Motorola Satellite Communications, Inc., FCC filing for application of a low earth orbit mobile satellite system (IRIDIUM), Washington, D. C. , December 1990.

Mur74. T. Muratani, "Satellite-switched time-division multiple access," *Proceeding of the IEEE Electronic Aerospace Systems Convention (EASCON)*, October 1974, pp. 189-196.

NTI72. National Technical Information Service - U. S. Department of Commerce, *Satellite Communications Reference Data Handbook, Defense Communications Agency*, Virginia, July 1972.

Pea77. M. C. Pease III, "The indirect binary n-cube microprocessor array," *IEEE Transactions on Computers*, Vol. C-26, No. 5, May 1977, pp. 458- 473.

PeM92. D. W. Petr, K. M. S. Murthy, V. S. Frost, and L. A. Neir, "Modeling and simulation of the resource allocation process in a bandwidth-on-demand satellite communications network," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 465-477.

PfB85. G. F. Pfister, W. C. Brantley, D. A. George, S. L. Harvey, W. J. Klienfelder, K. P. McAuliffe, E. Melton, V. A. Norton and J. Weiss, "The IBM research parallel prototype (RP3): introduction and architecture," *1985 International Conference on Parallel Processing*, August 1985, pp. 764-771.

Pol89. J. H. Polaha, *An Analysis of Low-Earth Orbit Satellite Communication Systems*, Master's Thesis, The Bradley Department of Electrical Engineering, Virginia Polytechnic Institute & State University, May, 1989.

PrB86. T. Pratt and C. W. Bostian, *Satellite Communications*, John Wiley & Sons, New York, 1986.

Pri86. A. A. B. Pritsker, *Introduction to Simulation and SLAM II*, Systems Publishing Corporation, West Lafayette, IN, 1986.

RaD84. M. Ramaswasmy and P. Dhar, "Comments on 'An efficient SS/TDMA time slot assignment algorithm'," *IEEE Transactions on Communications*, Vol. COM-32, No. 9, September 1984, pp. 1061-1065.

RaD88. R. A. Raines, N. J. Davis IV, and W. H. Shaw Jr., "Modeling, simulation, and comparison of interconnection networks for parallel processing," *Proceedings of the 1988 Summer Computer Simulation Conference*, July 1988, pp. 87-92.

RaD94. R. A. Raines, N. J. Davis IV, V. Ramachandran, and J. S. Park, "Performance modeling and analysis of packet switched augmented shuffle exchange networks," *IEEE Transactions on Computers*, May 1994, (submitted for publication).

Rai87. R. A. Raines, *The Modeling, Simulation, and Comparison of Interconnection Networks for Parallel Processing*, Master's Thesis, School of Engineering, Air Force Institute of Technology, December 1987.

RaJ88. D. Raychaudhuri and K. Joseph, "Channel access protocols for Ku-band VSAT networks: a comparative evaluation," *IEEE Communications Magazine*, Vol. 26, No. 5, May 1988, pp. 34-44.

Ram92. V. Ramachandran, *Performance Analysis of Augmented Shuffle Exchange Networks*, Master's Thesis, The Bradley Department of Electrical Engineering, Virginia Polytechnic Institute & State University, June, 1992.

RaR92. V. Ramachandran, R. A. Raines, J. S. Park, and N. J. Davis IV, "Performance studies of packet switches augmented shuffle exchange networks," *The Fourth Symposium on the Frontiers of Massively Parallel Computation*, October 1992, pp. 566-568.

RaV84. C.S. Raghavendra and A. Varma, "INDRA: a class of interconnection networks with redundant paths," *1984 Real-time Systems Symposium*, 1984, pp. 153-164.

Rei92. E. E. Reinhart, "Mobile communications," *IEEE Spectrum*, Vol. 29, No. 2, February 1992, pp. 27-29.

Rid85. L. Rider, "Optimized polar orbit constellations for redundant earth coverage," *Journal of the Astronautical Sciences*, Vol. 33, No. 2, April-June 1985, pp. 147-161.

Rid86. L. Rider, "Analytic design of satellite constellations for zonal earth coverage using inclined circular orbits," *Journal of the Astronautical Sciences*, Vol. 34, No. 1, January-March 1986, pp. 31-64.

Rou93. D. Rouffet, "GLOBALSTAR: a transparent system," *Electrical Communication*, First Quarter 1993, pp. 84-90.

Rus92. C. M. Rush, "How WARC '92 will affect mobile services," *IEEE Communications Magazine*, Vol. 30, No. 10, October 1992, pp. 90-96.

RuZ91. I. Rubin and Z. Zhang, "Message delay and queue-size for circuit switched TDMA system," *IEEE Transactions on Communications*, Vol. 39, No. 6, June 1991, pp. 905-914.

RuZ92. I. Rubin and Z. Zhang, "Message delay analysis for TDMA schemes using contiguous-slot assignments," *IEEE Transactions on Communications*, Vol. 40, No. 4, April 1992, pp. 730-737.

Sat93. SatLab User's Guide, Version 1, Comdisco Systems, Inc., 919 East Hillsdale Blvd., Foster City, CA, April 1993.

Sco92. D. S. Scott, "Out of Core Dense Solvers on Intel Parallel Supercomputers," *The Fourth Symposium on the Frontiers of Massively Parallel Computation*, October 1992, pp. 484-487.

ScG69. W. G. Schmidt, O. G. Gabbard, E. R. Cacciamani, W. G. Maillet, and W. W. Wu, "MAT-1: INTELSAT's experimental 700-channel TDMA/DA system," *Proceedings of the 1969 INTELSAT/IEE Conference on Digital Satellite Communications*, London, England, pp. 428-440.

SeP68. T. Sekimoto and J. G. Puente, "A satellite time-division multiple access experiment," *IEEE Transactions on Communications Technology*, Vol. COM-16, No. 4, August 1968, pp. 581-588.

ShD94. W. H. Shaw Jr., N. J. Davis IV, and R. A. Raines, "The application of metamodeling to interconnection networks," *ORSA Journal on Computing*, June 1994, (to appear Summer 1994).

ShI86. H. Shinonaga and Y. Ito, "SS/FDMA system for digital transmission," *Proceedings of the Seventh International Conference on Digital Satellite Communications*, May 1986, pp. 163-170.

ShI92. H. Shinonaga and Y. Ito, "SS/FDMA router for flexible satellite communications networks," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 391-400.

Sie77. H. J. Siegel, "Analysis techniques for SIMD machine interconnection networks and the effects of processor address masks," *IEEE Transactions on Computers*, Vol. C-26, No. 2, February 1977, pp. 153-161.

Sie90. H. J. Siegel, *Interconnection networks for large-scale parallel processing: Theory and case studies*, Second Edition, McGraw Book Co., NY, 1990.

SiM81. H. J. Siegel and R. J. McMillen, "The multistage cube: a versatile interconnection network," *IEEE Computer*, Vol. 14, No. 12, December 1981, pp. 65-76.

StD82. W. L. Stutzman and W. K. Dishman, "A simple model for the estimation of rain-induced attenuation along earth-space paths at millimeter wavelengths," *Radio Science*, Vol. 17, No. 6, November-December 1982, pp. 1465-1476.

Sto71. H. S. Stone, "Parallel processing with a perfect shuffle," *IEEE Transactions on Computers*, Vol C-20, No. 2, February 1971, pp. 153-161.

Tak87. F. Takahata, "An optimal traffic loading to intersatellite links," *IEEE Journal on Selected Areas in Communications*, Vol. SAC-5, No. 4, May 1987, pp. 662-673.

YaS90. M. Yabusaki, and S. Suzuki, "Approximate performance analysis and simulation study for variable-channel-per-burst SS-TDMA," *IEEE Transactions on Communications*, Vol. 38, No. 3, March 1990, pp. 318-326.

ThS92. P. T. Thompson, R. Silk, and A. Herridge, "The INTELSAT VII/VII-A genration of global communications satellites," *International Journal of Satellite Communications*, Vol. 10, No. 4, July-August 1992, pp. 183-198.

Vit85. A. J. Viterbi, "When not to spread spectrum - A sequel," *IEEE Communications Magazine*, Vol. 23, No. 4, April 1985, pp. 12-17.

WaB89. M. Wachira, D. Bossler, and B. Skerry, "FDMA implementation for domestic mobile satellite systems," *1989 IEEE Globecom*, Vol. 2, pp. 21.4.1-21.4.6.

Wal70. J. G. Walker, "Circular orbit patterns providing continuous whole earth coverage," *Royal Aircraft Establishment*, Technical Report 70211, November 1970.

Wal77. J. G. Walker, "Continuous whole-earth coverage by circular-orbit satellite patterns," *Royal Aircraft Establishment*, Technical Report 77044, September 1977.

APPENDIX A

LEO SATELLITE CONSTELLATION LINK ANALYSES

Table A.1. 36-Satellite IUU-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	3726.8	3726.8
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Transmitter Antenna Power (dBW)	8.82	6.53
Transmitter Antenna Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	168.0	168.0
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	168.8	168.8
G_r - Receiver Antenna Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R_c - Coded Data Rate (Mbps)	0.180	0.4
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.8	11.8
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	10.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

Table A.2. 36-Satellite Gateway-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	3726.8	3726.8
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Transmitter Antenna Power (dBW)	1.87	4.47
Transmitter Antenna Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	193.4	189.9
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	196.4	192.9
G_r - Receiver Antenna Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R_s - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.9	67.9
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.4	13.5
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.6	3.7

Table A.3. 36-Satellite Crosslink Analysis

	CROSSLINK
Slant Range (km)	6717.0
Elevation Angle (deg)	--
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Transmitter Antenna Power (dBW)	9.3
Transmitter Antenna Gain (dBi)	36.0
L_p - Path Loss (dB)	196.2
L_a - Atmospheric Loss (dB)	--
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	196.7
G_r - Receiver Antenna Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R_s - Coded Data Rate (Mbps)	25
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dB/k/Hz)	-228.6
Received C/N (dB)	13.2
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.4

Table A.4. 45-Satellite IUU-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	3150.3	3150.3
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Transmitter Antenna Power (dBW)	7.37	5.07
Transmitter Antenna Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	166.5	166.5
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	167.3	167.3
G_r - Receiver Antenna Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R_c - Coded Data Rate (Mbps)	0.180	0.400
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.9	11.9
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	10.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Required C/N (dB)	8.8	8.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

Table A.5. 45-Satellite Gateway-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	3150.3	3150.3
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Transmitter Antenna Power (dBW)	0.41	3.01
Transmitter Antenna Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	191.9	188.4
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.9	191.4
G_r - Receiver Antenna Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R_c - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.9	67.9
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.4	13.5
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.6	3.7

Table A.6. 45-Satellite Crosslink Analysis

	CROSSLINK
Slant Range (km)	5460.5
Elevation Angle (deg)	--
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Transmitter Antenna Power (dBW)	7.5
Transmitter Antenna Gain (dBi)	36.0
L_p - Path Loss (dB)	194.4
L_a - Atmospheric Loss (dB)	--
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.9
G_r - Receiver Antenna Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R_s - Coded Data Rate (Mbps)	25.0
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dB/k/Hz)	-228.6
Received C/N (dB)	13.2
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.4

Table A.7. 54-Satellite IUU-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2845.6	2845.6
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Transmitter Antenna Power (dBW)	6.48	4.18
Transmitter Antenna Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	165.6	165.6
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	166.4	166.4
G_r - Receiver Antenna Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R_s - Coded Data Rate (Mbps)	0.180	0.400
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.9	11.9
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	10.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

Table A.8. 54-Satellite Gateway-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2845.6	2845.6
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Transmitter Antenna Power (dBW)	-0.48	2.12
Transmitter Antenna Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	191.1	187.5
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.1	190.5
G_r - Receiver Antenna Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R_s - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.96	67.96
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.3	13.5
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.5	3.7

Table A.9. 54-Satellite Crosslink Analysis

	CROSSLINK
Slant Range (km)	5072.2
Elevation Angle (deg)	—
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Transmitter Antenna Power (dBW)	6.86
Transmitter Antenna Gain (dBi)	36.0
L_p - Path Loss (dB)	193.8
L_a - Atmospheric Loss (dB)	—
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.3
G_r - Receiver Antenna Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R_c - Coded Data Rate (Mbps)	25.0
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dB/k/Hz)	-228.6
Received C/N (dB)	13.2
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.4

Table A.10. 55-Satellite IUU-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2840.2	2840.2
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Transmitter Antenna Power (dBW)	6.47	4.17
Transmitter Antenna Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	165.6	165.6
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	166.4	166.4
G_r - Receiver Antenna Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R_s - Coded Data Rate (Mbps)	0.180	0.400
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.9	11.9
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	11.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

Table A.11. 55-Satellite Gateway-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2840.2	2840.2
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Transmitter Antenna Power (dBW)	-0.49	2.11
Transmitter Antenna Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	191.0	187.5
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.0	190.5
G_r - Receiver Antenna Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R_s - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.96	67.96
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.4	13.5
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.6	3.7

Table A.12. 55-Satellite Crosslink Analysis

	CROSSLINK
Slant Range (km)	5104.0
Elevation Angle (deg)	—
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Transmitter Antenna Power (dBW)	6.92
Transmitter Antenna Gain (dBi)	36.0
L_p - Path Loss (dB)	193.9
L_a - Atmospheric Loss (dB)	—
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	194.4
G_r - Receiver Antenna Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R_s - Coded Data Rate (Mbps)	25.0
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dB/k/Hz)	-228.6
Received C/N (dB)	13.1
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.3

Table A.13. 66-Satellite IUU-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2502.5	2502.5
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	1.6	1.6
λ (meters)	0.1875	0.1875
Transmitter Antenna Power (dBW)	5.37	3.07
Transmitter Antenna Gain (dBi)	1.0	22.3
L_p - Path Loss (dB)	164.5	164.5
L_a - Atmospheric Loss (dB)	0.3	0.3
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	165.3	165.3
G_r - Receiver Antenna Gain (dBi)	22.3	1.0
T_s - System Noise Temp. (dBK)	24.6	24.8
R_s - Coded Data Rate (Mbps)	0.180	0.400
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	49.5	53.0
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	17.9	11.9
C/I (dB)	18.0	18.0
Received C/(N+I) (dB)	14.9	10.9
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	4.8	4.8
Implementation Margin (dB)	0.5	0.5
Link Margin (dB)	5.6	1.6

Table A.14. 66-Satellite Gateway-Satellite Link Analyses

	UPLINK	DOWNLINK
Slant Range (km)	2502.5	2502.5
Elevation Angle (deg)	10.0	10.0
Transmitter Frequency (GHz)	30.0	20.0
λ (meters)	0.01	0.015
Transmitter Antenna Power (dBW)	-1.59	1.01
Transmitter Antenna Gain (dBi)	57.5	18.0
L_p - Path Loss (dB)	189.9	186.4
L_a - Atmospheric Loss (dB)	2.5	2.5
L_{po} - Polarization Loss (dB)	0.5	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	192.9	189.4
G_r - Receiver Antenna Gain (dBi)	22.3	54.0
T_s - System Noise Temp. (dBK)	31.6	30.8
R_c - Coded Data Rate (Mbps)	12.5	12.5
Modulation	QPSK	QPSK
B_N - Receiver Noise BW (dBHz)	67.96	67.96
k - Boltzmann's constant (dB/k/Hz)	-228.6	-228.6
Received C/N (dB)	14.4	13.5
Required C/N (before coding)	13.6	13.6
Coding Gain (dB)	5.8	5.8
Implementation Margin (dB)	2.0	2.0
Link Margin (dB)	4.6	3.7

Table A.15. 66-Satellite Crosslink Analysis

	CROSSLINK
Slant Range (km)	4355.3
Elevation Angle (deg)	—
Transmitter Frequency (GHz)	23.0
λ (meters)	0.013
Transmitter Antenna Power (dBW)	5.54
Transmitter Antenna Gain (dBi)	36.0
L_p - Path Loss (dB)	192.5
L_a - Atmospheric Loss (dB)	—
L_{po} - Polarization Loss (dB)	0.5
Total Losses ($L_p + L_a + L_{po}$) (dB)	193.0
G_r - Receiver Antenna Gain (dBi)	36.0
T_s - System Noise Temp. (dBK)	29.0
R_s - Coded Data Rate (Mbps)	25.0
Modulation	QPSK
B_N - Receiver Noise BW (dBHz)	70.97
k - Boltzmann's constant (dB/k/Hz)	-228.6
Received C/N (dB)	13.2
Required C/N (before coding)	13.6
Coding Gain (dB)	5.8
Implementation Margin (dB)	2.0
Link Margin (dB)	3.4

APPENDIX B

LEO SATELLITE SYSTEM PACKET DELAY CURVES

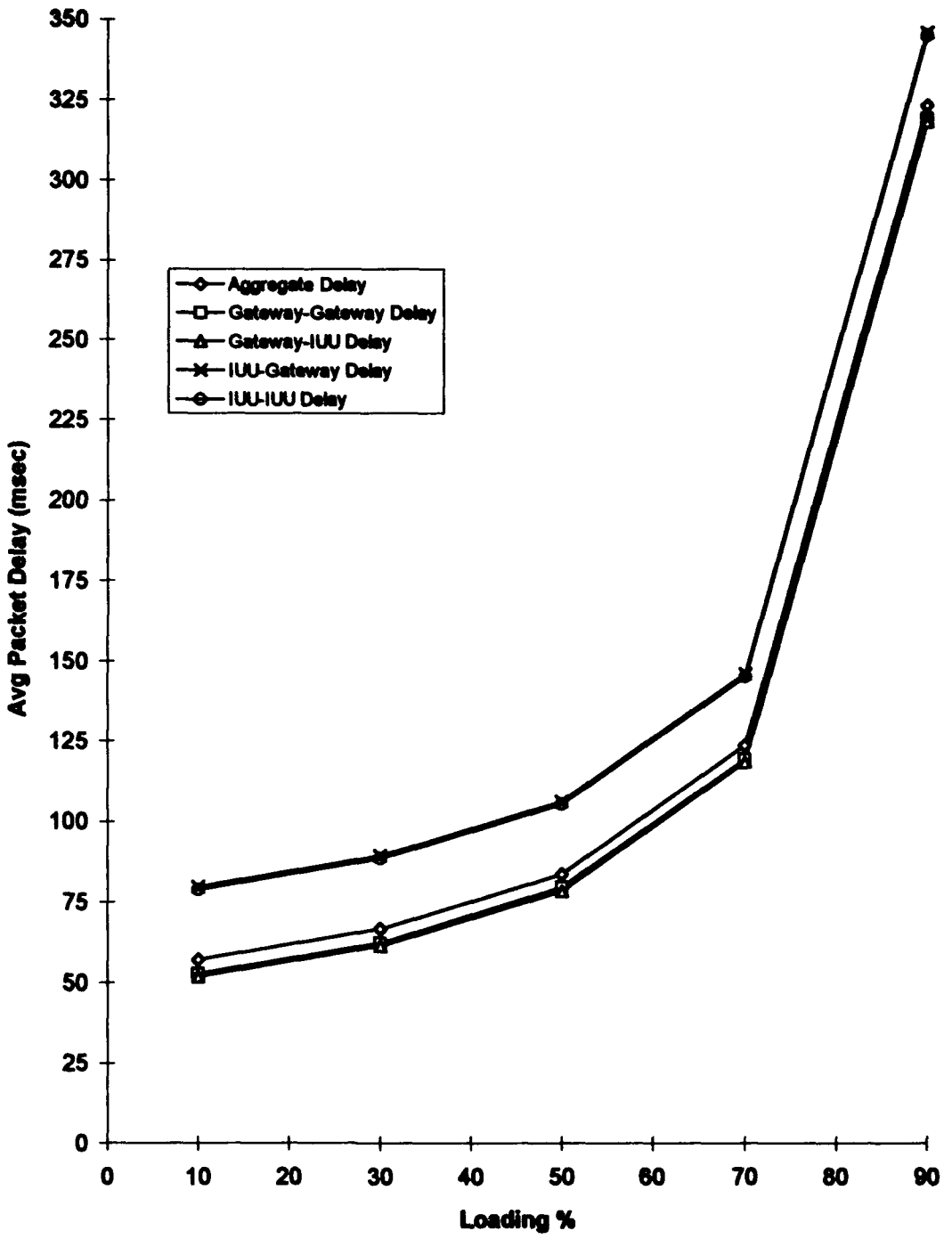


Figure B.1. Packet delay curves for a 36-satellite constellation.

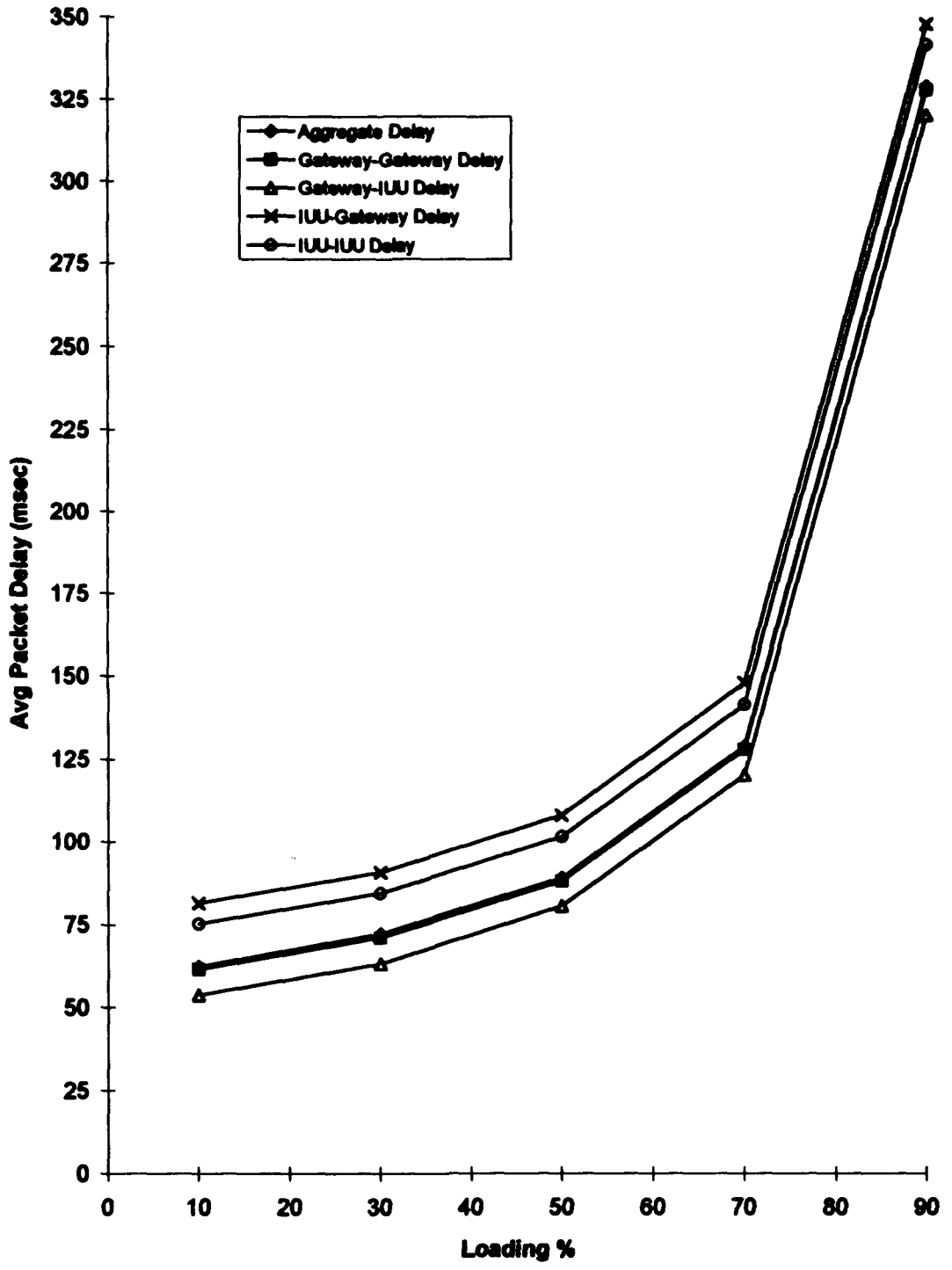


Figure B.2. Packet delay curves for a 45-satellite constellation.

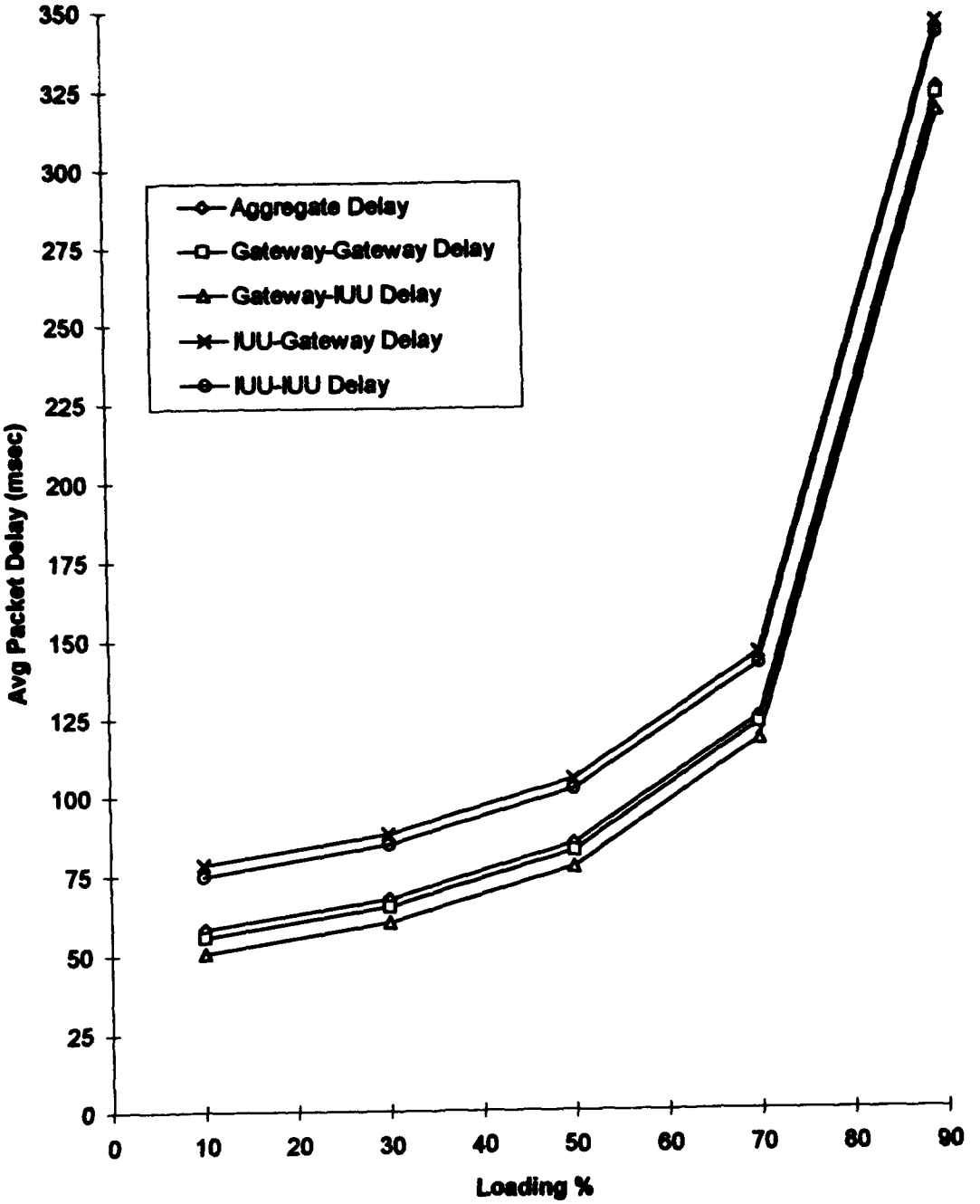


Figure B.3. Packet delay curves for a 54-satellite constellation.

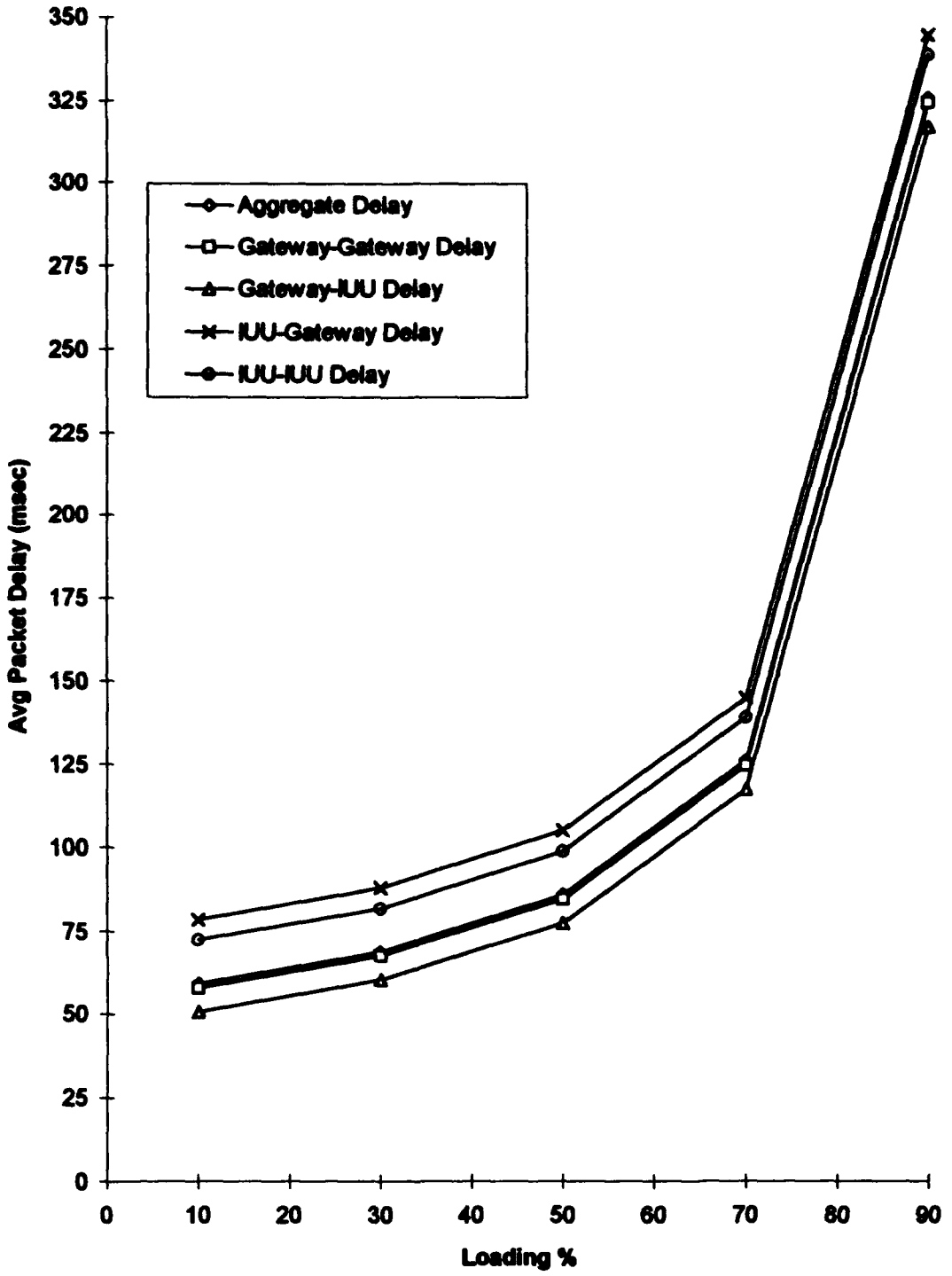


Figure B.4. Packet delay curves for a 55-satellite constellation.

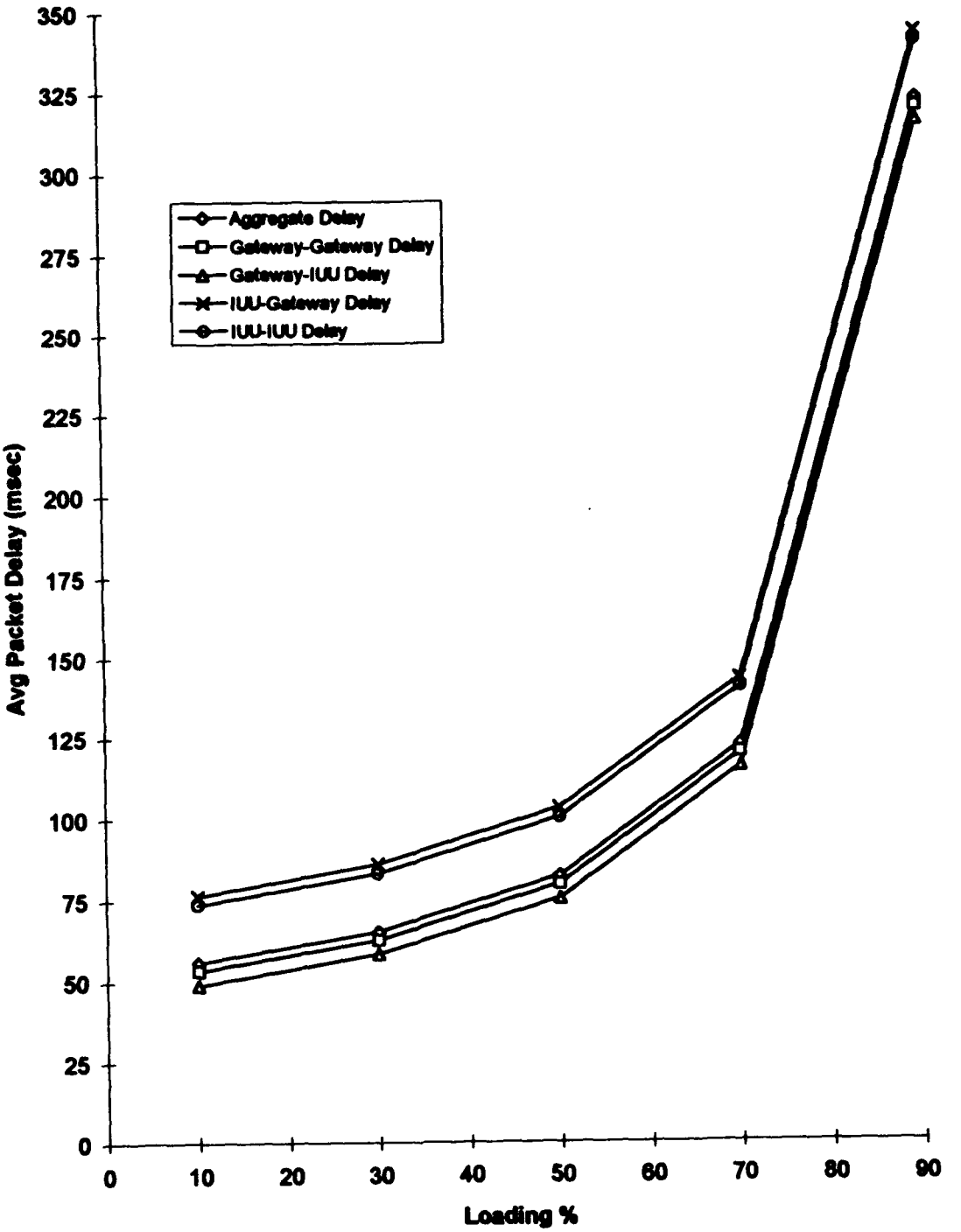


Figure B.5. Packet delay curves for a 66-satellite constellation.

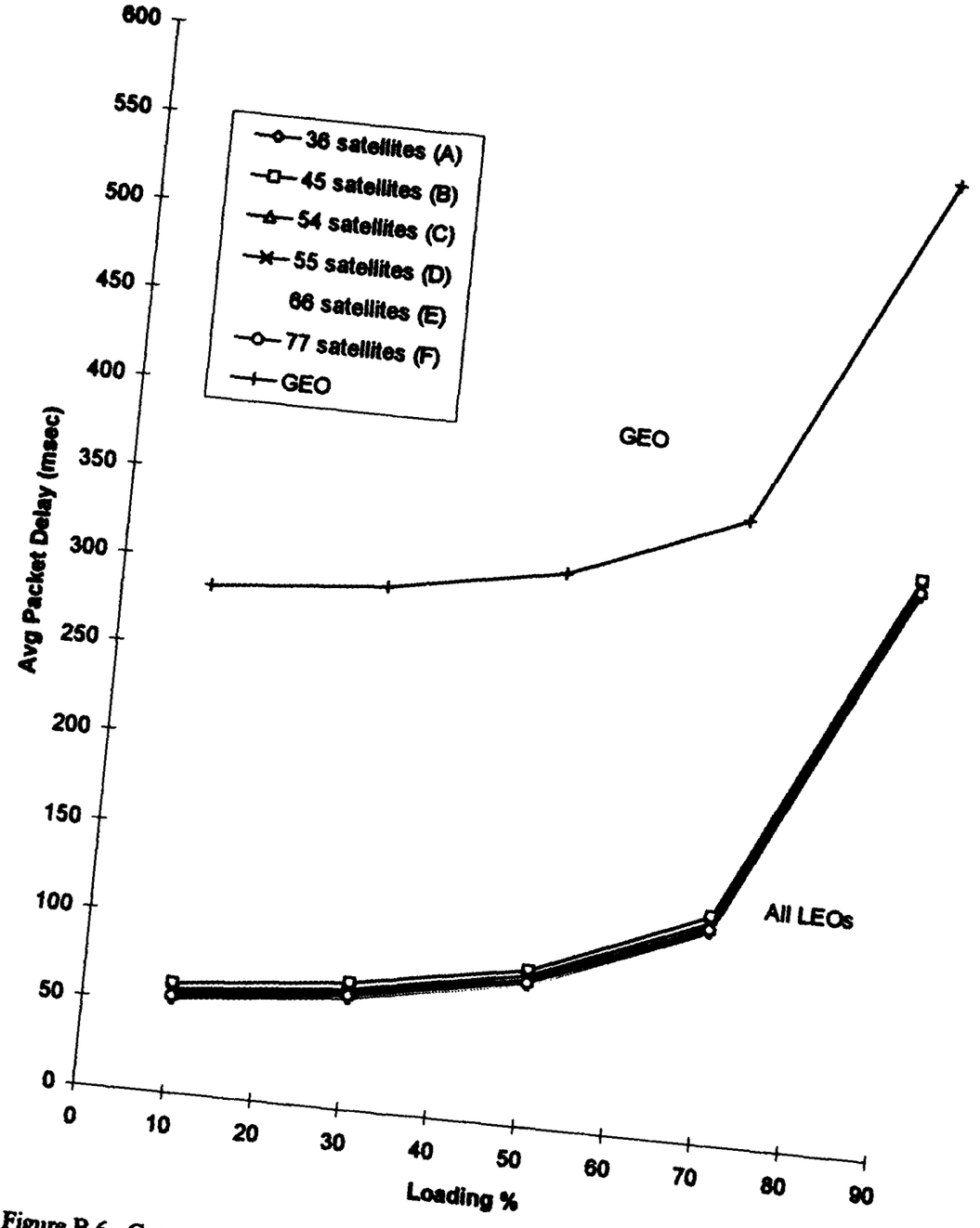


Figure B.6. Gateway to gateway packet delay for GEO versus LEO constellations.

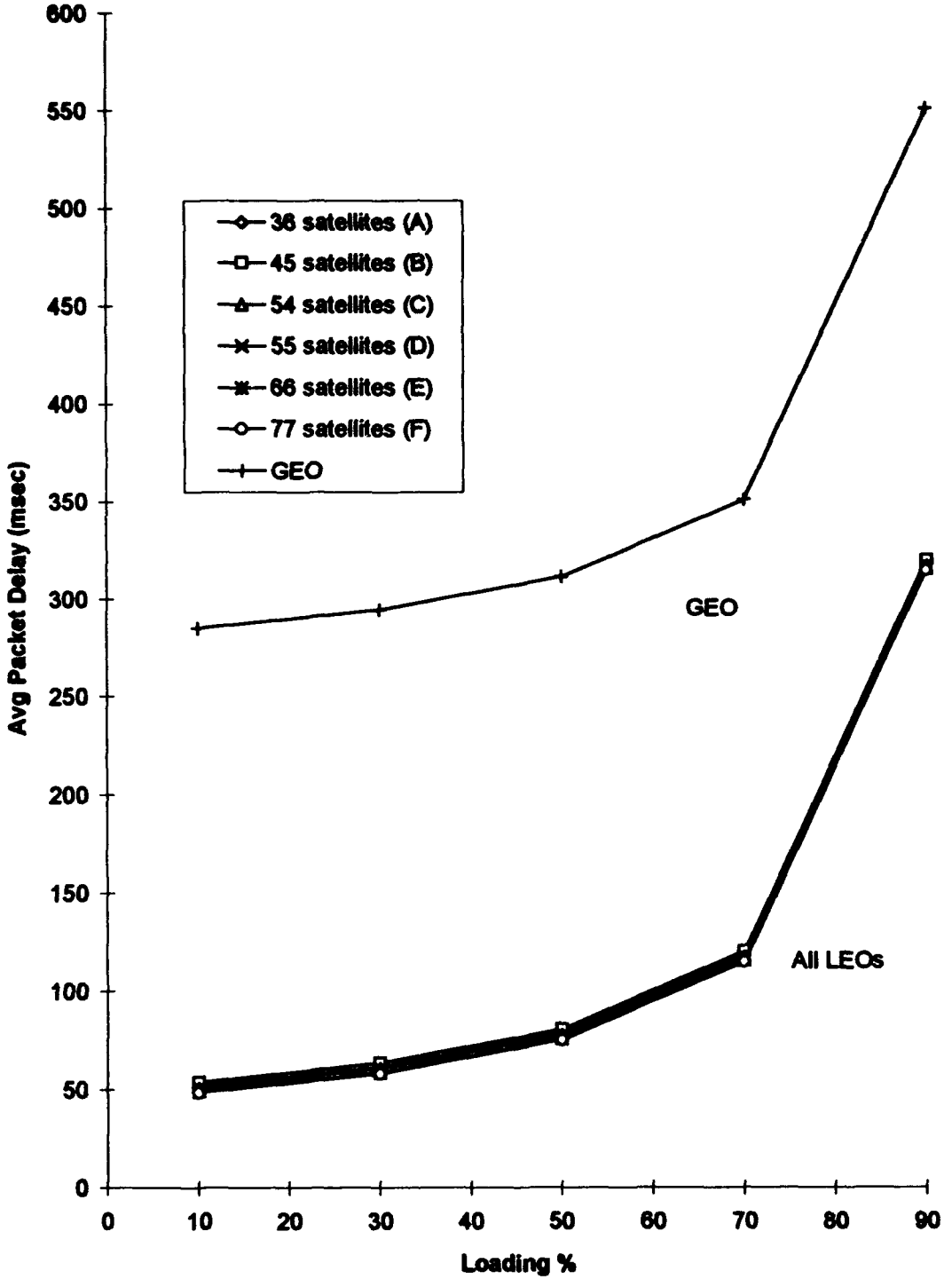


Figure B.7. Gateway to IUU delays for GEO versus LEO constellations.

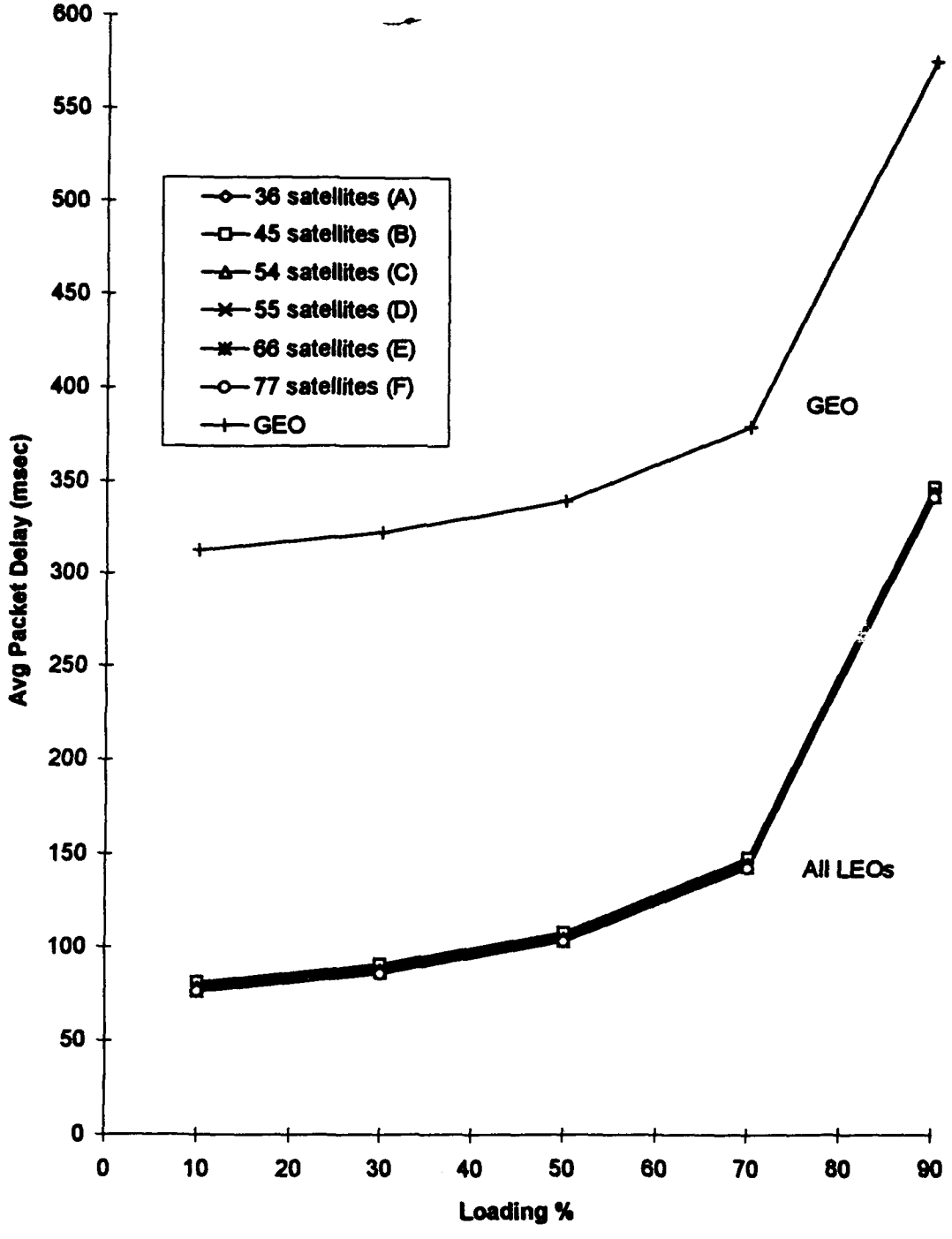


Figure B.8. IUU to gateway delay for GEO versus LEO constellations.

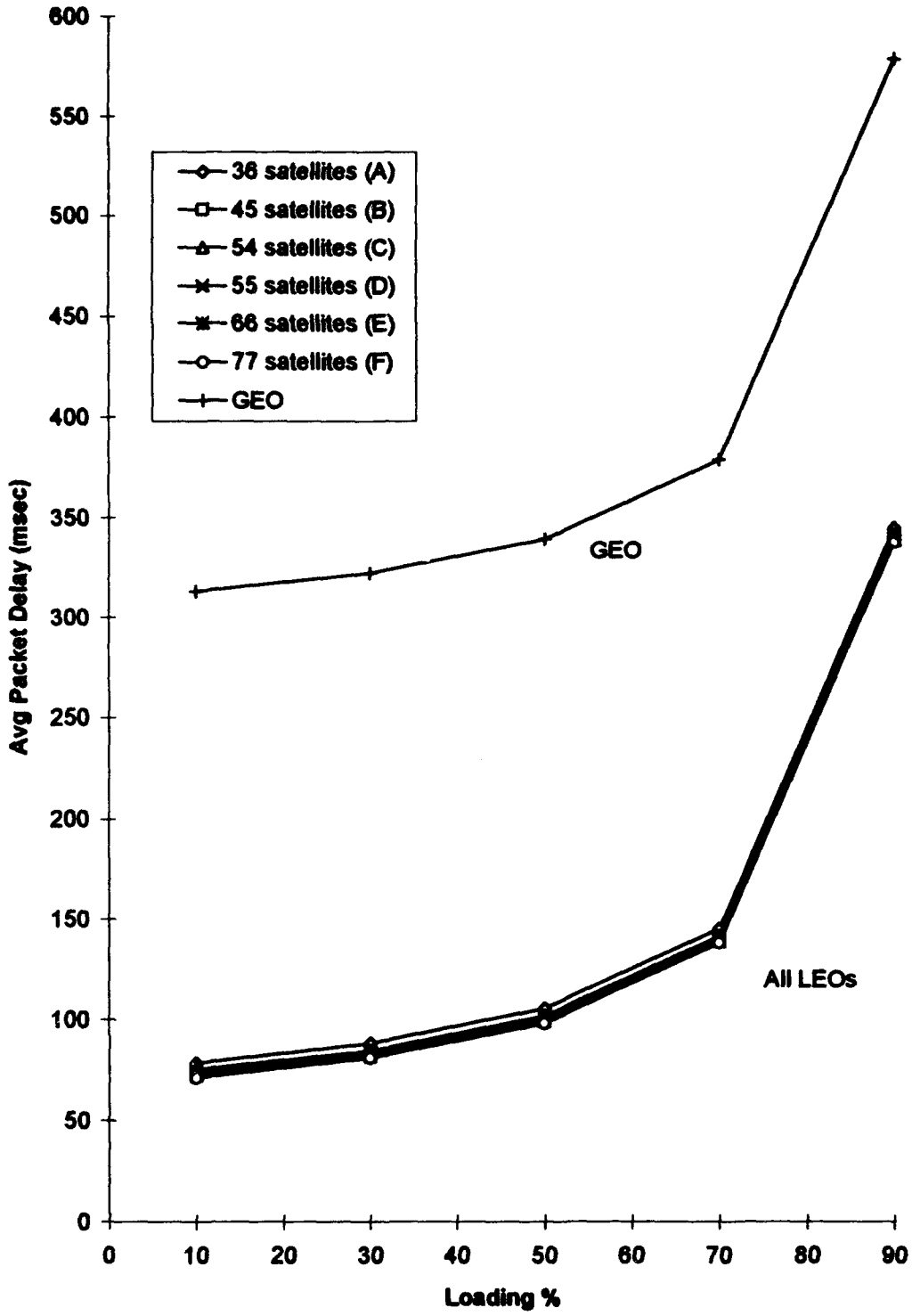


Figure B.9. IUU to IUU delays for GEO versus LEO constellations.

VITA

Richard A. Raines

June 1994

Education

August 1991 - June 1994	*Virginia Polytechnic Institute and State University*, Blacksburg, VA. Ph.D., Electrical Engineering.
June 1986 - December 1987	*Air Force Institute of Technology*, Wright-Patterson AFB, OH. M.S., Computer Engineering. Thesis: "The Modeling, Simulation and Comparison of Interconnection Networks for Parallel Processing."
March 1981 - December 1985	*The Florida State University*, Tallahassee, FL. B.S. Electrical Engineering, Cum Laude.

Work Experience

January 1988 - June 1991	Lead Systems Engineer for the National Training Center/ Air Warrior Air Combat Maneuvering Instrumentation program, USAF, Eglin AFB, FL.
January 1984 - January 1986	Electronics Technician, High Energy Physics Group, The Florida State University, Tallahassee, FL.
October 1976 - December 1980	Nike Hercules radar and computer technician, U.S. Army, Stuttgart, Germany.

Honors/Awards

February 1991	Top Performer/Outstanding Contributor Squadron Officer School, Maxwell AFB, AL
October 1990 - December 1990	ASD/YI Company Grade Officer of the Quarter.
July 1990 - September 1990	ASD/YI Military Employee of the Quarter.
July 1989 - September 1989	MSD/YI, MSD/CZ Company Grade Officer of the Quarter.

November 1977	Distinguished Graduate Advance Individual Training, Redstone Arsenal, AL.
May 1977	Soldier of the Month, 2nd Student Battalion, Redstone Arsenal, AL.

Publications

R. A. Raines, N. J. Davis IV, and W. H. Shaw, "The modeling, simulation, and comparison of interconnection networks for parallel processing," *1988 Summer Computer Simulation Conference*, July 1988, pp. 87-92.

V. Ramachandran, R. A. Raines, J. S. Park, and N. J. Davis IV, Performance studies of packet switches augmented shuffle exchange networks, *The Fourth Symposium on the Frontiers of Massively Parallel Computation*, October 1992, pp. 566-568.

W. H. Shaw Jr., N. J. Davis IV, and R. A. Raines, The application of metamodeling to interconnection networks, *ORSA Journal on Computing*, June 1994, (to appear Summer 1994).