TOPOLOGIES AND LINK ASSIGNMENT PROBLEMS IN LOW ALTITUDE SATELLITE NETWORKS

By

CHRISTOPHER WARD

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1987

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deep appreciation to Professor Randy Chow, chairman of my dissertation committee, for his guidance, assistance, and support. During my graduate level academic career Dr. Chow has always been ready to help when necessary. I would like also to thank Dr. Rick Smith for his time spent teaching the courses in theoretical computer science and mathematical logic which have been invaluable. My special thanks are offered to Dr. Charles Shaffer, Dr. Fred Taylor, and Dr. Joseph Wilson all of whom have assisted me on several occasions during my research. Finally my appreciation is extended to the staff of the Computer and Information Sciences department for their support.

This dissertation was sponsored by SDIO/IST and managed by the department of Navy, Space, and Naval Warfare Systems Command, under contract number N00039-87-C-0221.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TOPOLOGIES AND LINK ASSIGNMENT PROBLEMS IN
LOW ALTITUDE SATELLITE NETWORKS

By

CHRISTOPHER WARD

December 1987

Chairman: Dr. Yuan-Chieh Chow
Major Department: Computer and Information Sciences

In recent years an entirely new branch in communication networks has been evolving; specifically the area of inter-satellite networks. A distinction is made between the satellites currently used, which are simple repeaters for up-link/down-link traffic, and a new generation of networking satellites. It is envisaged that these new satellites will participate in establishing and maintaining a point-to-point store-and-forward communications network in space. By integrating this network with ground based stations the opportunity exists for a flexible global communications system.

Two aspects of inter-satellite communications are satellite topologies (i.e. position and orbit information) and link assignment (i.e. selection of a k-regular communications subgraph). Given the topology, the ground coverage may be determined, the possible communication links identified, and some measure of network survivability established. With a topology selected, the physical communications network must be established; this is known as the "link assignment problem." In a system of homogeneous satellites each satellite will have only a small number of antennae to communicate with a limited number of neighbors. The problem then becomes how to assign the antennae such that acceptable preformance characteristics are obtained.

In this dissertation a model is developed for satellite topologies. Using this, a definition of survivability is formulated which encompasses both communications and coverage. A simulation is used to investigate these satellite topologies and verify coverage formulas. A restrictive satellite topology model, the (N_p, N_s) satellite model, is defined in graph theoretic terms and used to derive a minimum connectivity for (N_p, N_s) systems. With a formal model from which to proceed, the link assignment problem is investigated. The emphasis in this dissertation is to study the effect on propagation delay of changing the following parameters: link retargeting rate, satellite altitude, and the number of antennae. To analyze link assignment a simulation is used which is modeled around the concept of a single shelled multigraph representation for satellite topologies. Using this, results are obtained for different link selection algorithms operating with different values of the parameters. In order to maximize the influence that topologies have on link assignment, the performance criterion was point-to-point propagation delay only.

CHAPTER 1
INTRODUCTION

In recent years an entirely new branch of communication networks has been evolving. Specifically the area of intersatellite networks, which owes its origins to radio mobile networks and satellite cross-links, will soon become a reality. Compared to the other areas of networking, satellite networking is still immature. The reason for its slow growth has been due partly to its limited use and partly to the sophisticated technology necessary to support it. Although still unimplemented, there is great interest in understanding the problems, characteristics, and potential of these networks. This dissertation investigates particular aspects associated with constructing a satellite network.

Before proceeding it is important to define and characterize an intersatellite network. A general satellite network is comprised of satellites in orbit at a number of different altitudes, planes, and inclinations. Each of these satellites possesses the necessary equipment to transmit, receive, store, and process data. A distinction is therefore made between the satellites currently used, which are simple repeaters for up-link/down-link traffic, and this new generation of networking satellites. It is envisaged that these new satellites will participate in establishing and maintaining a point-to-point store-and-forward communications network in space. By integrating this network with ground-based-stations the opportunity exists for a global communications system of great flexibility and high data rates.

With this new field comes an entirely new set of problems, constraints, properties, and research opportunities. Although there are a few "satellite systems" in space (for example the GPS–Global Positioning System) there are at present no intersatellite networks. Therefore issues such as selecting efficient satellite topologies; selecting and maintaining satellite-to-satellite communication links; selecting efficient communication routes; and providing a reliable, secure communications network all need considering. In addition to these problems a new set of constraints exists in satellite networks. Satellites will be in motion relative to one another; therefore directional antennae will be subject to pointing and tracking errors. The power limitations may also influence the design; in satellites the power used by individual components is a major consideration. Similarly each satellite will have a restricted field of view. With the satellite in motion, links will probably exist only for a limited time. Thus at the link level the network will be in a constant state of flux. Besides the new problems and constraints, this new environment also has some characteristics which are completely unique. In particular the motion of a satellite in orbit around the earth is essentially predictable. This single feature makes satellite networks distinct from both the land-based static networks and the random, radio mobile networks. The unique qualities offer the potential for completely new link assignment schemes and routing strategies. In addition the relationship between altitude and satellite visibility presents a new view on hierarchical networks. To summarize: the study of intersatellite networking offers exciting research potential in topologies, link assignment and routing.

In this dissertation two of the previously mentioned problem areas are considered, satellite topologies and link assignment. The term "satellite topology" in this context refers to the position and orbital information of a number of satellites which are to communicate via point-to-point communications links. The topology may be considered the lowest level in networking. It was realized early in the research that the selection of topology greatly influences the characteristics and performance of the network. In particular given the topology, the ground coverage may be determined, the possible communication links identified, and some measure of network survivability established. Assuming that the network is for global communications, the ground coverage characteristics clearly play an important role in selecting appropriate topologies. Once a topology has been selected, the physical communications network must be established. Establishing an effective physical network is known as the "link assignment problem." In a system of homogeneous satellites each satellite will have only a small number of antennae to communicate with a limited number of neighbors. This will be particularly true if highly directional point-to-point links are used (for example laser links). The problem then becomes how to assign the antennae such that acceptable preformance characteristics are obtained using the resulting subgraph. In addition since the topology is constantly changing, the antennae will be reassigned periodically. However in the interests of preformance it may be beneficial to reassign links frequently even though the topology does not demand it. Finally it is noted that the link assignment scheme will influence the routing algorithm. In fact the routing algorithm and link assignment algorithm should be designed together so use may be made of the unique predictability characteristics of satellite networks.

In Chapter 2 background information related to satellite networks is given. First a general overview explaining the rationale for computer networks is presented. Following this a description of the more typical network measures is introduced; these measures form the basis for objective functions used in Chapter 4. As a prologue to satellite networks, the area of mobile communications is explored. As can be seen later, many of the ideas used in satellite networks arise from packet radio mobile systems. Another important area for background research is the physical characteristics of satellite networks. Included in this are Newtonian mechanics, the physical communication options, links and targeting, antennae pointing and tracking, and network predictability. Finally in Chapter 2 two proposed satellite network schemes are discussed.

In Chapter 3 the issue of satellite topologies is addressed. First a model is defined. Although simplifying assumptions are made, the result is a model which may be manipulated with ease yet provides a good approximation to the real situation. To the model are added the constraints specifying a legal model. The model is legal in the sense that the research emphasis is directed toward global communications. With a definition for a legal satellite system established, the relationship between two of the primary constraints, coverage and connectivity, is examined. In order to permit the analysis of coverage characteristics a major part of the chapter is devoted to an explanation of a ground coverage simulation developed specifically for investigating satellite topologies. The final part of Chapter 3 examines a particular class of satellite model, the (N_p, N_s) model. The model is first introduced, then a formal specification of the model is derived and some of its associated properties are shown. In particular the relationship between the (N_s, N_p) model and its minimum connectivity is derived.

In Chapter 4 the area of link assignment is investigated. The problem may be stated formally as follows, "given a network, select a spanning subgraph in which each of the vertices has at most k edges connected to it, such that it minimizes some objective function." In fact since the number of antennae on a satellite remains fixed, once installed, the communications network will improve by using as many of these antennae as possible. Thus the ideal link assignment would have each node with exactly k edges connected to it (where k is the number of antennae). This is the k-regular graph problem and is known to be NP Complete.

In this dissertation no attempt has been made to find the optimal solution to the link assignment problem. The solution techniques must of necessity rely on heuristic approaches which will yield only results which are suboptimal. In addition the criterion for optimal link assignment is open to debate; there is no ideal objective function with which to measure it. Instead the goal of this dissertation is to examine the effect that varying different satellite parameters has on the performance of two reasonable but simple algorithms. The parameters investigated were the altitude, the link assignment strategy, the value of k (i.e. the connectivity), and the objective function. First the link assignment problem is formally specified. A method for complexity reduction is introduced in which the satellite topology is viewed as a multigraph. An algorithm is then introduced based on multigraphs which may be used to compare these parameters. An interesting feature of this link assignment algorithm is its relationship with the (N_p, N_s) model. The final part of Chapter 4 summarizes the results of the analysis which reveals some interesting findings.

Finally in Chapter 5 the more important results derived in this dissertation are described. In concluding a discussion extending the scope of the research is made, with reference to topics already under investigation.

So far there has been little evidence that a demand for satellite networks even exists. One might argue that research without some expectation of usefulness is futile. This comment is particularly valid when the research thrust is directed at problems associated with connecting together satellites, each of which would cost many millions of dollars. In this section the two principal applications for such a system are discussed in an attempt to justify research efforts in this area.

The first area in which a satellite network will be of use is clearly for global voice-data communications. Already civilian communications make extensive use of satellites with which to transfer data. In fact satellite broadcast of television has become an important medium to reach the rural and isolated regions of the United States and the world in general. The more specialized low altitude satellite systems which would participate in satellite networks are ideally suited to one form of traffic, voice data. At present the geosynchronous satellites used to relay voice data are unpleasant to use due to the long propagation delays (typical voice satellites have a propagation delay in the order of 240 ms.). A low altitude satellite network has an average propagation delay less than 100 ms, which is significantly lower. The other obvious advantage with a satellite network is that it has ability to reach isolated regions and serves the mobile communications community.

The other major area in which low altitude satellite networks could be of use is in military operations. Already there are several military satellite systems, the most prominent being GPS and MILSTAR. The flexibility derived from voice links using satellites would be highly desirable. More recently the Strategic Defense Initiative (SDI) places a specific emphasis on using space-based platforms to provide sensor, management, and weapons capability. With such a system there would clearly be a considerable amount of traffic to be communicated between different nodes. In addition the cost of placing such a system in space would almost certainly demand that it serve a useful peacetime function, for example relaying voice and data.

CHAPTER 2
BACKGROUND

In this chapter we examine the background leading to intersatellite communications. The following paragraphs give a broad overview of communication networks in general. The first section presents the layered ISO/OSI model. Following this a section is devoted to each of the three important topics which influence the design and analysis of satellite networks: measures of network efficiency, current research in mobile communications, and physical considerations for satellite networks. An overview of the appropriate work in each of these areas is presented along with references to the authors. Finally two proposed satellite network algorithms are discussed, Brayers "Low Altitude Satellite Network," [8] and the NRL/Harris Corporation "SDI Network."

Initial studies in communication networks were stimulated in response to rapidly increasing demands of the telephone system. For many years the problem of permitting many users to communicate with one another has been well researched and understood. The facilities required by the original telephone system demanded an end-to-end virtual circuit be established between the users (i.e. the users "see" the connection as a direct link). The originating user places a call to the receiving user, waits for the link to be established and commences talking. With the advent of multiprocessing computers it became desirable to permit these computers to communicate with many users and to exchange information with one another.

What then is the difference between telephone networking principles and computer communications? The answer, it turns out, is very little. Today we continue to see the two areas merging together so that we might expect them to one day become indistinguishable from one another. Already the telephone system makes extensive use of computers for routing messages between different junctions. In fact one of the leading telephone communication companies, AT&T, has for a long time written much of its communication software in a general purpose high level language called "C". Similarly, many large corporations now find that their widely distributed sites require access to non-local data and must communicate across the telephone exchange.

Having established that telecommunications and computer networks run along parallel lines, what are the differences? The principal difference is the structure of the data in the two systems. In telecommunications the data are essentially voice traffic. [1] The end-to-end communications protocol must therefore provide a virtual circuit (i.e. behave as though there were a direct communication link). That is, the originating user selects the destination address (dials), the receiving user acknowledges the call (lifts up the handset) and until one or the other terminates the call (replaces the handset) the bidirectional link is maintained by the telephone company. Beyond this the telephone exchange is free from responsibility. It is now the users who must operate under the same (assumed) protocol, sharing a common language, and pausing to permit the other user to begin talking. In contrast to this, in computer communications there may be many different end user services. Consider some possibilities: simple terminal communications, database access by some utility, file transfer, and mail transfer. The effect of these services on the communication network is to impose a much greater degree of structure than is necessary in the telecommunications network. The services require not only virtual circuit, but also

[1] According to plans laid out by ISDN, the telecommunications system will soon support services like image transfer, digital data, etc.

message broadcasting (for mail), unordered packet delivery (for file transfer and random access), encryption etc. In addition, the protocols to support these services are communicating with computers and must therefore be completely unambiguous and self-correcting.

From earlier paragraphs it should be clear that both the telecommunications services and computer networks are similar at the physical level. Both are responsible for passing the raw packets of data between end users. The differences emerge when the higher level functions are considered (e.g. natural language vs computer requests). The task of permitting computers to communicate with one another turns out to be so complex that the problem is divided into a number of layers. Each layer has a well defined function and consists of a set of primitives and data structures. The primitives are made available to the higher layers for their communication needs, while the layer itself makes use of the primitives presented by the layer below. At the lowest layer the communication is achieved by hardware (i.e. physical communication links). A detailed discussion of function and application of each layer is presented in the section "Networking Overview" of this dissertation.

One of the key ideas behind layering the network is that physical characteristics may be hidden from the higher layers. For example one layer may offer primitives to open a "direct" connection with any other node on the network. Thus layers higher than this are relieved of the responsibility of routing packets through the network.

An important aspect when designing communication algorithms is that some criteria exist by which the different algorithms may be compared. In this dissertation only a small part of the overall communication package will be considered. In particular only effective satellite topologies and communication links will be considered. However it is necessary to establish appropriate measures. The section "Measures of Network Efficiency" describes the standard measures and tools used in network design.

Within the general field of communications many different classes of network exist. By far the largest class is land-based point-to-point communication networks. These networks consist of stationary nodes (e.g. computers, workstations etc) which are connected together by a combination of local area networks and long haul networks. Although the network topology may change as nodes switch in and out of operation, none of the nodes moves. Typically the network topology changes very slowly compared to the average time a communication link is maintained. Another type of network used is the mobile network. This network class represents only a few percent of the communications, yet, because of its unique properties, it presents many interesting issues which have to be resolved. In particular the nodes are mobile and communicate with one another using omnidirectional radio broadcasting. Mobility within the network suggests that the network topology changes quite rapidly with respect to time. To maintain efficient communication between all nodes in the network involves the use of several specialized routing algorithms which are discussed in detail in the section "Mobile Communication Systems" of the present study.

Another class of network which may be expected to differ from the previous network classes is a satellite network. At present there are no implementations of inter-satellite networks; they may therefore be considered an extremely small percentage of the total network communications used. Again, however, the unique properties of satellite networks suggest that new routing algorithms may be appropriate to satisfy the demands of the new class. Because this class is new there is little background material on the subject. The section "Satellite Network Theory" discusses the physical aspects of orbiting bodies, the communication options available for satellite networking, and elaborates on the problem of link selection and targeting within a satellite network. Finally the major communication satellite systems are also reviewed, including the military communications system proposed by Brayer [7,8], and the SDI satellite system by the NRL/Harris Corp. in the section entitled "Proposed Satellite Networks."

2.1 Networking Overview

The function of a computer communications network is to provide a mechanism by which many end-users can exchange data in a secure, reliable and efficient manner. In general end-users are uninterested in the way in which this communication takes place; i.e. they require *transparent* communications. To facilitate this the end-user is presented with a communications model, and the task of communicating data is reserved to a specific suite of programs and hardware to implement this model. Due to the complexity of writing communications software, and because of the differing levels of service required by different users, the problem is subdivided into a number of distinct, functional layers. Each functional layer may be accessed by the layer immediately above using a set of well defined procedures. The layer in turn will provide some specified function and may use the procedures of the layer immediately below it to achieve this function. One might expect that each network would use completely different layers and functions; in fact, there is broad agreement on the number of and types of these layers. The layers and their related functions as defined by the International Standards Organization (ISO) Open System for Interconnection (OSI) are listed below. The seven layers in increasing abstraction are as follows:

- The physical layer is concerned with the transmission of unstructured data over a physical medium. The specifications deal with the electrical, mechanical, and electronic hardware. In addition they specify the functional control of the devices and the programming characteristics of accessing the devices. This is the lowest level and is the only place at which data "physically" pass. The physical layer is described in detail by McClelland [39].

- The data link layer offers reliable data transfer across the physical link. The layer sends blocks of information between two adjacent nodes. Nodes are adjacent if they are in direct communication with one another. Included in this layer is link error detection/correction. Additionally this layer ensures that all blocks are delivered and enforces link flow control. Details of the Data Link Layer may be found in Conrad [10].

- The network layer provides the upper layers with a logical view of the network which is independent of the physical connections and switching technologies. This layer controls the opening, closing, and maintaining of logical communication channels. Notice it is this layer that is responsible for breaking a message into packets, routing the packets through the sub-net, and combining the packets at the destination. Thus the selection of routing protocols pertains to this layer. Details of this layer are described by Ware [53].

- The transport layer provides reliable transfer of end-user data between end-users. Additionally it offers end-users error recovery and flow-control. It is from this layer and above that the user is effectively insulated from the underlying network. Two basic types of service are provided to higher layers, connection-oriented and connectionless. A connection-oriented service provides for the connection, maintenance and termination of a logical communication stream between transport users. In a connectionless (or Datagram) service packets of information between users are guaranteed delivery, but not necessarily in order. Details of the Transport layer are given in Knightson [35].

- The session layer provides the control structures for communicating between applications and establishes, manages and terminates connections between co-operating applications. This is perhaps the least useful layer in the ISO model [49] because many of its functions are frequently done in the transport layer. Details may be found in Emmons and Chandler [16].

- The presentation layer provides independence of data representation (syntax) between similar applications. Typical applications are a virtual terminal protocol (VTP), teletex, video text, encryption etc. Clearly there is not one application; rather the layer may be thought of as a suite of different standards for different high level functions. Information on the presentation layer may be found in Hollis [26].

- The application layer provides distributed information services and an end-user entry point to the layered model. The entry points are designed for application processes to make use of the OSI model. In particular, mechanisms such as virtual file protocols for remote file access and job control protocols, are to be supported at this layer. Details of this layer may be found in Bartoli [3].

A pictorial diagram demonstrating the layered network concept is shown in Figure 2.1

For the interested reader, there are several good books on General Computer Networks. In particular Tanenbaum's Computer Networks [51], and Stalling's Data and Computer Communications [49] are both excellent.

2.2 Measures of Network Efficiency

In order to evaluate the effectiveness of different satellite topology selections and link assignment algorithms, it is necessary to possess measures to quantify network

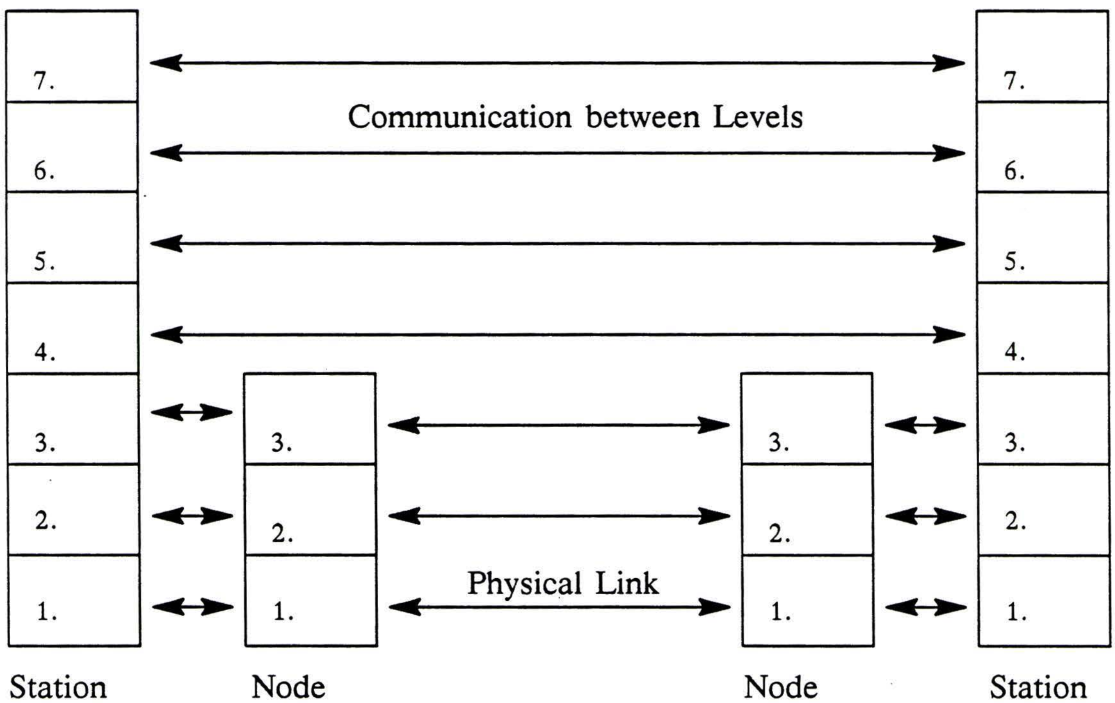


Figure 2.1. The ISO Layered Reference Model

selection. The selection of the measure depends on which aspect of the network is of interest and on the complexity of the communications model. In order to discuss graph properties we informally define some graph terms now and note that a graph model for communications will be formally defined in Chapter 3. A *graph* consists of a set of vertices (or nodes) which are connected together by edges (or arcs.) A graph may be either directed or undirected; a directed graph indicates that information flow may travel in one direction only (the direction must then be specified). In an undirected graph the information may flow bidirectionally. The *degree* of a vertex is the number of edges entering it. A *sequence* is a sequence of vertices in which adjacent edges in the sequence are connected by edges. A *cycle* is a sequence starting and finishing at the same vertex. A *path* is a sequence between two vertices which contains no cycles, and the *length* of the path is the number of edges it contains. In general there may be several paths between two vertices.

2.2.1 Network Models

A communications network may be viewed as five increasingly complex models, as outlined below. From these models particular algorithms are introduced which will be discussed in greater detail in the following section.

- Unit cost communications model (no routing). In the simplest model only the point-to-point graph is considered; either an edge exists or it does not; no additional information is kept. Given the vertices to be networked, selected edges are required to satisfy certain graph properties. In particular *connectivity* is an essential requirement of any network. Connectivity is important because it gives some indication of reliability, an important consideration in network design. There are two forms of connectivity, node-connectivity and edge-connectivity.

- Node-connectivity is the number of nodes that must be removed to disconnect nodes x and y in the graph (i.e. so that no path exists between them). A network which is n connected can, in general, suffer the loss of $n-1$ nodes and remain (one) connected.
- Edge-connectivity is the number of edges that must be removed to disconnect nodes x and y in the graph.

We note without proof that Node-connectivity is a stronger statement than Edge-connectivity [52]; i.e. if C is connectivity, then $C_{node} \leq C_{edge}$. Classical algorithms for connectivity analysis have been proposed by Even and Tarjan [17] and Kleitman [34].

- Unlimited capacity communication models (no routing). In this model the topology of the network and the cost assignment for the communications links alone are considered. Thus the model is an extension of the previous by the addition of a cost associated to each link. In real network terms the cost might be line cost, a propagation delay or some other quantity to be minimized. Given such a model the following measures are typically of interest.

  - A <u>Minimum-cost-spanning-tree algorithm</u> is defined as a selection of the available edges to construct a tree (a connected graph with unique paths between all pairs of points) such that the sum of the costs of the edges is a minimum. Minimum spanning trees play an important role in network topology selections; in particular they make excellent starting points for the construction of the final network. Two algorithms which may be used to find a minimum cost spanning tree of a graph are by Dijkstra [12] and Kruskal [36]. Other algorithms also exist [54].

- An <u>All-pairs-least-cost-path algorithm</u> is another criterion for selecting subgraphs may be based on the value of the all-pairs least cost path. As the name suggests, the measure depends on the longest source-destination geodesic (i.e. shortest path) of the graph. Clearly a graph with a low value will be compact in the sense that no vertices will be unnecessarily distant from the remainder of the network. The problem of solving the all-pairs least cost path has been solved by Floyd [18] and has a close correspondence with Kleene closure [32].

- Capacity constrained communication models (no routing). Another addition to the communications model is to consider that each of the edges of the graph has not only a cost associated with it but also a *capacity.* By including a capacity the model becomes closer to a real network. With a capacity assigned to each edge, the questions become the following:

  - What is the maximum flow through the system given a source vertex and a sink vertex?
  - Which edges are necessary to permit this flow?

  Notice that satisfying the maximum network flow may cause a conflict with measures on the previous models. That is, a graph which gives maximum flow may not yield a good least cost path. The max-flow/min-cut problem has been solved by several algorithms, most notably by Malhotra [38], Fratta, Gerla, and Klienrock [21], and others [14,19,27,45,31,23].

- Queueing theory based models (with routing.) Within a store and forward communications network, source to destination delay may be caused by two reasons. The first is propagation delay, which has been effectively included as a cost associated with each edge. Another source of delay in a store and forward network is the queueing delay that packets undergo within a node. In networks with short propagations delays and in networks with high traffic, the queueing delay dominates. To measure networks with queueing delays traditional queueing models are used. Two implications of queueing theory based models are that the load on the network is predicted (usually a stochastic process, with a Poisson distributed arrival time for packets to each node) and a *routing algorithm* exists. Clearly this model is closer to reality than the previous; however many new assumptions are required. In particular a routing algorithm has to be selected and implemented. To summarize, the queueing model permits analysis of not only the network topology but also the routing and congestion control techniques used. However these benefits are offset by the complexity of the model and associated assumptions. Much of the work done in this area has been by Kleinrock [33].

- Simulation models (with routing). The most sophisticated criteria for network performance would be the results (throughput, utilization, delay, etc) from a simulation of the network. Typically a simulation includes realistic generation of data to the network (perhaps by simulated job loads etc) and a simulation of the algorithms and control mechanisms used to pass the data around. These simulations are typically complex and highly specific to the individual problem.

Perhaps the most significant detail to notice from the five criteria is that as the model becomes more complex, many factors other than the topology and link cost are involved in the selection of the communication links. In fact it is typical to regard the problem of selecting the physical network independently from the selection of the routing algorithm and system performance analysis.

2.2.2 Selected Network Algorithms

The following are a selection of algorithms appropriate to the analysis and measurement of networks in which only the topology is considered (i.e. restricted to types 1 and 2 above). The order of the algorithms presented differs from the previous order due to algorithmic dependencies.

All-pairs least cost path

Multiple applications of Dijkstra's algorithm. One method of finding the all-pairs least cost path of a graph is to use an algorithm which finds the shortest paths between two specific nodes and apply it to all pairs in the graph. One such algorithm for node to node shortest path is by Dijkstra [12]. Each node is labeled with its distance from the source along the best known path (which is maintained by each node having a *prev* pointer, to the previous node in the path). Initially all distance labels are set to infinity. A label may be either tentative or permanent; initially all labels are tentative. As the algorithm proceeds, paths are found and marked tentative; when it is discovered that no shorter path exists to a node, it is marked as permanent. The algorithm step is:

> Start out by marking the source node permanent, then for each node adjacent to it, examine the node and check to see if a new shortest path exists by going through the original node. If one exists, mark the cost of this new path (tentative) and update the *prev* pointer. Then review all the tentatively marked nodes, select the one of least cost and label it permanent.

Kleene algorithm. The all pairs problem can also be solved by representing the graph as a connectivity matrix. Let the communications graph be represented by a vertex set V and an edge set E. Define the cost of an edge $v, w \in E$ to be $cost(v, w)$. Let $A_G = (a_{ij}), 1 \leq i, j \leq n$ be the following matrix:

$$a_{ij} = \begin{cases} cost(v_i, v_j) & \text{if } (v_i, v_j) \in \text{E} \\ 0 & \text{otherwise} \end{cases}$$

The sum of the powers of A_G is the closure of A_G. The time complexity for this is $O(n^3)$. The closure of A_G yields directly the all pairs least cost paths between source and destination pairs. For a detailed description of the relationship between path problems and matrix multiplication see Melhorn [43].

Minimum cost spanning tree algorithms

The problem may be stated as follows. "Given a connected undirected graph whose edges have a real valued cost function *cost*, find a spanning tree of the graph whose total edge cost is minimum." In general the approach used is called the *greedy method*: we build up the spanning tree edge by edge, selecting at each iteration the smallest appropriate edge. To explain the following algorithms we introduce an edge-coloring process [52]. Initially all edges of the graph are uncolored. Edges are colored one at a time, with either blue (edge is accepted for spanning tree), or red (edge is rejected). The edges are colored by two rules which maintain color invariance.

- *Blue Rule.* Select a cut (an imaginary line cutting the graph) that crosses no blue edges. Amongst the uncolored edges crossing the cut, select the one of minimum cost and color it blue.

- *Red Rule.* Select a simple cycle containing no red edges. Amongst the uncolored edges on the cycle, select the one of maximum cost and color it red.

The method of selecting the rule and edges is nondeterministic; we are free to apply either rule in arbitrary order until all the edges are colored. We state without proof the following theorem.

The greedy method colors all edges of any connected graph and maintains color invariance[52].

Kruskal's algorithm consists of applying the following step to the edges sorted by ascending order.

If the current edge e has both ends in the same blue tree, color it red; otherwise color it blue.

We note Kruskal's algorithm builds random blue trees which eventually connect together to form a single spanning tree. The running time for this algorithm is $O(m \log n.)$

Prim & Dijkstra's algorithm uses an arbitrary starting vertex s and consists of repeating the following step $n - 1$ times.

Let T be the blue tree containing s. Select a minimum-cost edge incident to T and color it blue.

Notice that Prim's algorithm builds a single connected tree. The running time for this algorithm using an efficient heap representation may be reduced to $O(m \log_{2+m/n} n)$.

Max flow algorithms

Several algorithms exist to compute the maximum flow between two points on a graph. It is interesting to note that these same algorithms serve to compute the connectivity of a graph. This becomes apparent once the max-flow/min-cut theorem is understood. This theorem is defined as follows:

The maximum flow between any two arbitrary nodes in a graph is equal to the capacity of the minimum cut separating those two nodes [20].

Table 2.1. History of Maximum Flow Algorithms

Date	Discoverer	Running Time
1956	Ford and Fulkerson	-
1969	Edmonds and Karp	$O(nm^2)$
1970	Dinic	$O(n^2m)$
1974	Karzanov	$O(N^3)$
1978	Malhotra, et al.	$O(n^3)$
1977	Cherkasky	$O(n^2m^{1/2})$
1978	Galil	$O(n^{5/3}m^{2/3})$
1979	Galil and Naamad, Shiloach	$O(nm(\log n)^2)$
1980	Sleator and Tarjan	$O(nm \log n)$

The history of the maximum flow algorithm is outlined in Table 2.1.

In this section only two of the above algorithms will be discussed, the Ford-Fulkerson algorithm and the Malhotra algorithm. In order to discuss these we need to define a few terms. A flow on a graph G is a real valued function with the following properties:

- *skew symmetry* $f(v,w) = -f(w,v)$. If $f(v,w) > 0$ we say there is a flow from v to w.

- *Capacity Constraint* $f(v,w) \leq cap(v,w)$. If e_{vw} is a edge such that $f(v,w) = cap(v,w)$, we say the flow saturates e_{vw}.

- *Flow conservation.* For every vertex v other than the source and termination $\sum f(v,w) = 0$.

In addition we define the following. The *residual Capacity* for a flow f is a function on the vertex pairs given by $res(v,w) = cap(v,w) - f(v,w)$. The *residual graph* R of a flow f is the graph with vertex set V, source s, and sink t, and a edge e_{vw} of Capacity $res(v,w)$ for every pair v,w such that $res(v,w) > 0$. An *augmenting path* for f is a path p from s to t in R. The residual Capacity of p denoted by $res(p)$ is the minimum value of $res(v, w)$ for all the edges in p. We can increase the value of f by any amount Δ up to $res(p)$ by increasing the flow on every edge of p by Δ.

Ford-Fulkerson algorithm. Begin with a flow of zero on all the edges in the graph. Repeat the following step until obtaining a flow without an augmenting path.

> Find an augmenting path p for the current flow. Increase the value of the flow by pushing $res(p)$ units along p.

Knowing a maximum flow, we can compute a minimum cut in $O(m)$ time.

Malhotra's algorithm. A flow f is a *blocking flow* if every path of from s to t contains a saturated edge. Notice that the value of a blocking flow cannot be increased by pushing additional flow along any path, although it may be possible to increase the flow value by rerouting, i.e. decreasing the flow on some edges and increasing it on others. Malhotra's algorithm is based on using this blocking flow idea. Initially we delete from the original graph every vertex and edge not on a path from s to t. We maintain for each vertex v the *potential throughput* of v, as defined by

$$throughput(v) = \min\left\{ \sum_{e_{uv}\in E} (cap(u,v) - f(u,v)), (cap(v,w) - f(v,w)) \right\} \quad (2.1)$$

To define $throughput(s)$ and $throughput(t)$ we define dummy edges of infinite capacity from s to t. To find a blocking flow we repeat the following step until t is not reachable from s:

> *saturating step.* Let v be a vertex of minimum potential throughput. Send $throughput(v)$ units of flow forward from v to t by scanning the vertices

in topological order, and backward from v to s by scanning the vertices in reverse topological order. Update all throughputs, delete any newly saturated edges from G, and delete all vertices and edges not on a path from s to t.

Although this method is simple it has two drawbacks. When actually implemented it is quite complicated. Furthermore it preferentially sends flow through narrow bottlenecks, causing an unnecessary number of augmenting steps.

Connectivity algorithms

Now that Maximum Flow algorithms have been introduced, the algorithms for connectivity may be described. The k-connectivity problem may be specified as follows: "what is the minimum number of node (edge) disjoint paths between all source/destination pairs in the given graph." The maximum flow algorithm may be used to solve this problem in the following way. Label each edge of the graph with a capacity of 1. Now between any two vertices s and t calculate the maximum flow (using a previous algorithm), then by the Max flow/Min cut theorem; this is the minimum cut also. But since each path contributes only 1 to the flow, the max flow is equal to the number of edge disjoint paths between them (i.e. the connectivity). Using this principle the following algorithms are discussed.

Kleitman's algorithm. The algorithm due to Kleitman [34] is as follows: to find if a graph is k node connected, pick a node at random and call it N_1. Verify that the node connectivity between N_1 and all other nodes is at least $k+1$. Now delete N_1 and all its associated edges from the graph, and choose another node N_2. Check that this node is k connected to all other nodes. Remove N_2 and choose another edge N_3; check whether this node is $k-1$ connected to all other nodes etc. Continue until you have checked that some node is 1 connected to all nodes in the remaining graph. If this is the case, the algorithm terminates successfully. The complexity of this algorithm is given by kn applications of the max-flow algorithm. Thus connectivity is a computationally complex problem.

Even's algorithm. Another algorithm by Even and Tarjan [17] may be used to check if a graph is k connected. Label all the nodes $1 \cdots n$ and perform the following.

1. Form the subset with nodes $1 \cdots k$ and check that within this subset there are k node-disjoint paths between all pairs. If this is successful perform step 2.

2. For each node $j, k < j \leq n$ perform the following.

   (a) form the subset $L = 1 \cdots j - 1$

   (b) add a new node X and connect it to each node in L.

   (c) verify that there are k node-disjoint paths between j and X

The complexity of this algorithm is given by n applications of the max-flow algorithm; thus Even's and Tarjan's algorithm is usually 3 to 4 time faster than Kleitman's.

2.3 Mobile Communication Systems

2.3.1 Characteristics of Radio Mobile Systems

One of the few fields which offers any insight into intersatellite communications is that of packet radio communications as used by the military [9]. This should not be confused with the cellular mobile phone system [15,30] which is currently in commercial use. The use of mobile communication systems plays a key role in Command, Control, and Communications (C^3) as envisaged in the Battle 2000 scenario [50] in which military nodes are highly mobile. In these situations the network topology is rapidly changing as nodes move in and out of communication range of one another. There have been several different systems designed for radio communications; two of the prominent ones are PRNET [28] and the Intra Task Force Network [1]. These both make use of omni-directional radio broadcasting as the physical communication medium. Although different from anticipated satellite communication links, the network routing strategies required to keep track of a continuously changing topology are similar. In the following section the aspects associated with changing topologies are considered. From this a few general algorithmic principles are discussed followed by case studies of the two previously mentioned systems.

2.3.2 Random Topology Changes

A typical mobile network consists of many highly maneuverable nodes which are able to communicate with one another by the use of packet radio. The most obvious application for mobile networks is to support mobile military communications. In this case the nodes would either be land vehicles, ships or planes. Each node is equipped with a radio transmitter/receiver which has some communication range (which may vary depending on atmospheric propagation etc.). Thus for a particular instant in time, the system may be considered a communications graph in which the vertices are vehicles and the edges are communication paths between nodes that are in range of the transmitter.

In this scenario as time progresses the network of communication paths will change as nodes move in and out of range of one another. Clearly these changes to the network will be random (or at least unpredictable). Thus over time the communications network may be viewed as a graph with nodes and edges appearing and disappearing. As the motion of the vehicles is random, so too will be the changes in the graph.

2.3.3 General Algorithm Principles

An algorithm which is to support source to destination communications over such a changing topology must have the following characteristics.

- Permit new nodes to enter network. Provision should be made to permit nodes to reenter the network as quickly as possible.

- Recognize the loss of failed nodes. When a node fails, all routes passing through it will be disrupted, and global routing tables will be incorrect. Thus prompt dissemination of this knowledge is important.

- Maintain timely communications between nodes. Messages should be delivered fairly, avoiding starvation.

- Support priority messages. Frequently it is considered essential that particular messages arrive at their destination. Typically this is achieved by assigning a priority to messages.

- Congestion control and avoidance. Even when under heavy load, the network should operate effectively. In particular algorithms should prevent individual nodes (and the network as a whole) from being swamped with packets.

2.3.4 Case Study–PRNet

PRNET makes use of two routing protocols [29], a primary and a backup. The primary algorithm is point-to-point and operates on a multi-station architecture, although it was originally designed for a single station. An alternate algorithm, working at lower utilization, provides packet broadcast and will function in a stationless environment. Operation in this environment is achieved by the "flooding" algorithm. PRNET was selected as typical of the single station class of network, so attention centers on the station control aspects of the primary algorithm with only an overview of the alternate.

To briefly review the terminology used in multistation architectures three elements are available, Stations, Repeaters, and PR-nodes. Although there is no explicit mention of the processing capabilities, it is assumed that there are only a few, well spaced stations. These represent highly computational and limited mobility nodes. Surrounding them are unmanned, low mobility repeaters, and many highly mobile PR-nodes (perhaps attached to vehicles).

The algorithm is based upon the generation of a backbone network of repeaters and stations. This links PR-nodes with stations and represents the mainstream communication path between repeaters. Two data structures are required to coordinate the routing. A neighbor table accumulates data on surrounding nodes for periodic transmission to the central control station and a routing table, in which there is one entry of each source/destination route actually going through the node, which is used to forward packets. This table will change frequently as routes are established and broken. Details of routing table entries are provided by the control station in response to network setup requests.

Point-to-point routing

Prior to discussing details of network initialization and operation, the technique used to pass a message in point-to-point routing is discussed. Two distinct operations are involved: 1) route finding, in which control stations and the receiving node are made aware that a node is trying to connect, and 2) route setup, in which a path between the two nodes is established. These operations will be explained later. A packet is composed of two parts, header and data. The header is made up of an immediate destination field, a source/destination identifier, and possibly several alternate immediate destinations. The data field contains the message to be transferred. In point-to-point routing a logical circuit is represented as a table entry in selected repeaters between the two end nodes. An element of the table would contain two fields, a source/destination identifier, and the next few downstream repeaters. When a repeater recognizes a packet (by the identifier), the header is modified to contain the new immediate destination set and retransmitted. Repeaters ignore packets for which they have no table entry. An example of point-to-point routing is shown in Figure 2.2.

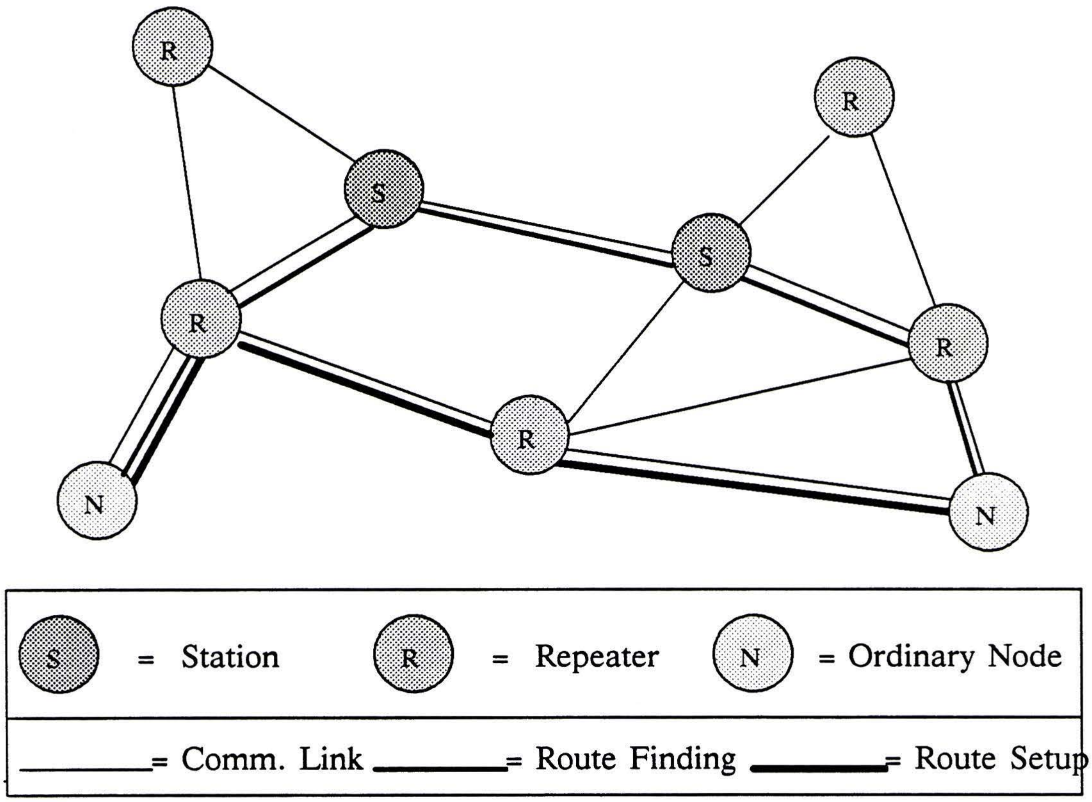


Figure 2.2. Point-to-point Routing

Additional flexibility may be gained by including several repeater addresses. This allows a detour from the primary route should a node become inoperative, as shown in Figure 2.3.

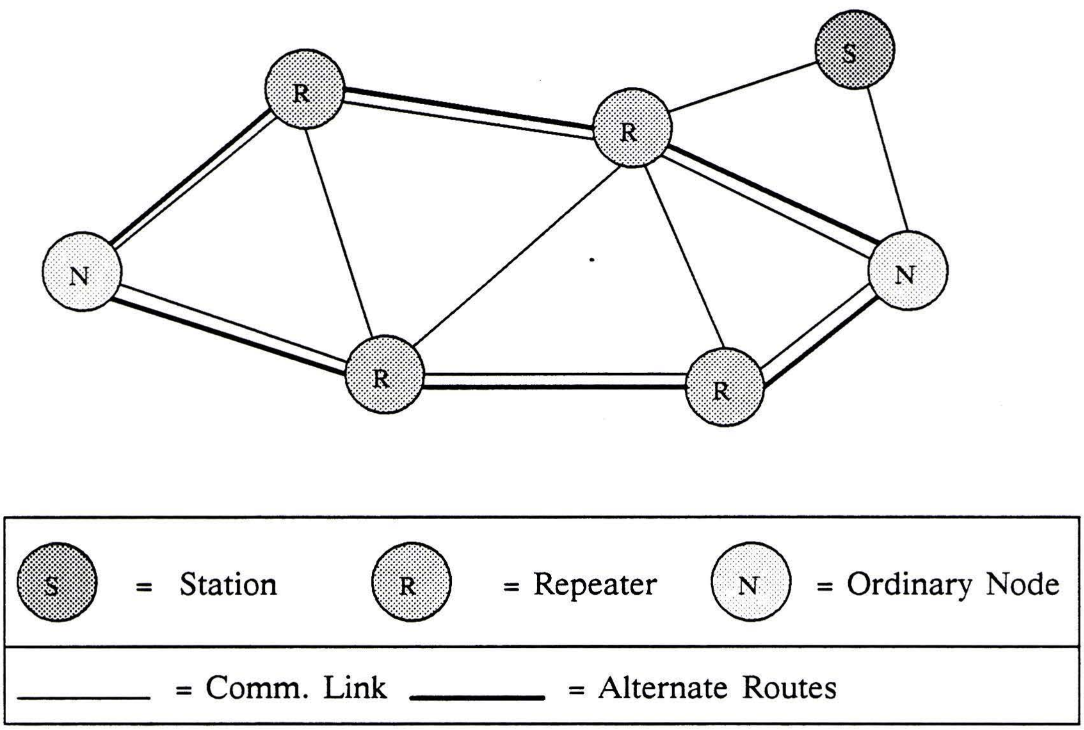


Figure 2.3. Alternate Routes in Point-to-point

Once established, PR-nodes communicate between one another using the same route for the duration of the session, although the route could be changed if desired.

Backbone creation

It is important, when developing the backbone, that links are reliable and the routing information accurate. This enables the control station to assign efficient links. Information on the connectivity of the network is gathered by having each node build a neighbor table (which includes details of link quality) and send it to the local station. From this information a station may use a labeling algorithm to develop the network. The labeling proceeds as follows.

When a node first becomes on-line it will pause for a short while (to prevent system overload at startup), then periodically broadcast a Radio-on-Packet (ROP). The rate at which the packet is transmitted depends on the type of node. For example a fixed radio might send one every thirty seconds while a mobile unit might send one every five seconds. These ROPs contain information about the node and its neighbor table (i.e. about the surrounding nodes). From the ROPs the station may start to label nodes. The labeling convention identifies the minimum number of hops a node is from the station. A node in direct communication is said to be at level 0. A node in communication with a level 0 node but not with the station is said to be at level 1. This continues for levels 2, 3 etc, as demonstrated in Figure 2.4. The label scheme can be implemented in the following fashion. Entries are made in a table for each node a station can hear directly. These are labeled by the station as level 0 nodes. Once labeled a node continues to send periodic ROPs and will forward ROPs of surrounding nodes to the station. Now level 1 nodes may be found directly from the ROPs received by the station from level 0 nodes. The backbone is calculated by the station and sent to the appropriate nodes. Using level 1 nodes, the process may be repeated, gathering ROPs from the next level out, computing the backbone, labeling nodes, etc. until all nodes are labeled. Once completed the station possesses the global connectivity map and is in a position to compute the best route between any source and destination; "best" is established by some criterion such as minimum hop-count.

In a multistation environment there will be nodes labeled by several stations. Two stations which have labeled a common repeater are neighbor stations. The summary

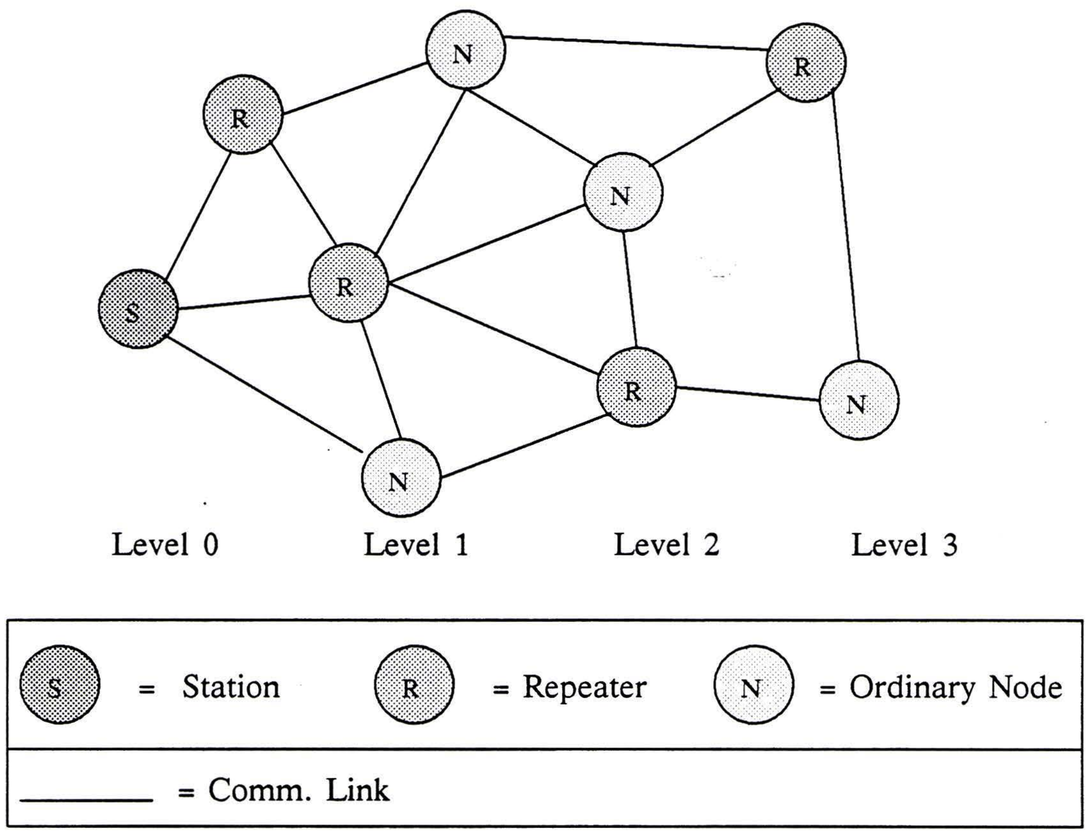


Figure 2.4. Labeling Nodes in PRNET

ROP sent by the repeater will include sufficient information so that each station becomes aware of all links to its neighbor stations. This is the only information required by the the station to permit multistation communication.

As previously mentioned, each node is responsible for sending periodic ROPs even after labeling is completed. These are used by the station to maintain a correct, accurate picture of the network. If a station should cease to hear from a node, the station will reconfigure the local area. It should be clear that repeaters failing to respond will cause all nodes at a lower level that make use of them to be rerouted. Conversely, the station is responsible for periodic relabeling of each node and refreshing routing table entries within the node. Should a table entry go unrefreshed, the link is assumed to have expired. Entries are timed out by the node, with the age of each entry reported in the ROP. An expired slot is not used in routing but neither is it erased. This allows stations to quickly resume labeling should a connection resume communication. If the routing slots become filled at the node, the expired entries may be overwritten. In the event a node fails to hear from the station, the alternate (stationless) algorithm is initiated.

To summarize this section, all nodes are labeled corresponding to their distance from the station. Nodes periodically send ROPs which contain status information regarding neighbors and routing table entry times. Stations respond to this by reaffirming labels and sending updates to the routing tables. The multistation design achieves several desirable goals: 1) complete redundancy among multiple stations, such that should one station become inoperative, a neighbor will automatically take control; 2) all devices are shielded from knowledge about source to destination routing and stations; and 3) PR nodes are kept simple, improving the economy and reliability of the system.

Route finding/packet setup

Assuming the network is established and in steady state, a node may communicate with another by establishing a virtual circuit for subsequent packets. The circuit is constructed in two phases, route finding (locating the destination) and packet setup (installing routing entries in repeaters on route). These are explained below.

Assume node i wishes to communicate with node k and has not done so in the immediate past. The first step is to find a route between the two nodes. In multistation architecture the route may either be directly from node to node, or indirectly, between the local stations of nodes i and k. The decision depends on whether the nodes are fixed or highly mobile. In the former case, routing is node to node; in the latter, routing is station to station with the initial and final steps (which may change frequently) controlled by the local stations. In the situation of a mobile node routing to a fixed node, the route would be from station to fixed node. The route is established by node i sending a request for a route to its local station giving the destination address (logical naming is not considered at this level; it is assumed that the node somehow knows the address of the destination). The action of the station depends on whether the destination is known (local) or unknown (local to some other station).

- Destination node unknown. The source station sends a request to each of its neighbor stations via the shared repeaters passing the destination ID and its own ID. Neighbors append this information and initiate a similar sequence. Eventually a route selector set is delivered to the destination and processed as described below.

- Destination node known. The source station selects the best route between the two nodes and generates a selector set, i.e. a list of intermediate repeaters through which packet setup should proceed. The list may contain several alternate routes to provide security and flexibility. Finally the route-finding packet reaches the destination node (or station of fixed nodes). The system is now in a position to begin route setup.

The actions of route setup proceed as follows. The route packet containing selector sets is passed from the destination to the previous repeater. Upon receipt, the repeater makes an entry in its routing renewal table containing the source/destination key and the next few repeaters in the chain. The repeater may update the route setup packet if a more efficient route is available based on local knowledge of the surrounding nodes. By repeating this process the packet will eventually be passed back to the originator, establishing the route.

Once a route is allocated by the repeaters, packet forwarding is quite simple. The packet needs only the source/destination key as identification; and the immediate destination address. Repeaters receiving a packet which they recognize use table look up and transmit the packet to the next node. To ensure packet transmission, positive acknowledgement is used between each link.

Routing in stationless mode

When a station becomes inoperative, PRNET provides an alternate routing mechanism which takes over automatically. It is assumed that circumstances surrounding stationless operation represent a highly disrupted system. An important criterion in this scenario is that the routing algorithm be as reliable as possible with minimum cost to network efficiency. The stationless operation in PRNET is similar to multi-stationed routing in that a two phase operation is used, route discovery followed by route setup. Route discovery involves limited flooding. Each repeater, on hearing a route discovery request, will either pass it to the destination node if known, or append its own identity and broadcast it. The request thus produces a wavefront spreading across the network. Using this technique the destination is guaranteed to receive the packet. With the selector set now available, route setup may proceed as previously. Repeaters will ignore subsequent copies of the same packet, so eventually all route discovery packets will expire. Stationless routing provides an effective backup against network damage while maintaining the requirements of network security at only slightly decreased performance.

Conclusions

The primary routing algorithm is extremely flexible, with stations able to allocate "best" routes between nodes based on current information. Routing within a single station may be made optimal while multistation routing introduces only limited overheads. Building a route is a complex procedure but once established is efficient for the remainder of the session. The alternate algorithm offers a similar service to the primary with the following differences: 1) stations can collect connectivity information and make more efficient routing decisions, 2) stations can evaluate all routes when making a selection, 3) stations can detect changes in connectivity quality and compensate by reducing the routing load, and 4) stations can perform global congestion control. In all, PRNET provides an extremely effective communication algorithm based on a multistation environment.

2.3.5 Case Study–Intratask Force Network Algorithm

The Linked Cluster Algorithm developed by Baker and Ephremides [2] is the second routing algorithm to be considered. The work is built on the HF (High Frequency) Intra Task Force (ITF) Network [1] which is a general purpose network offering extended line of sight communications for naval units. The algorithm is designed for rapidly changing network topologies, with operational requirements emphasizing survivability. It is intended that the algorithm be fully distributed, with automatic network reorganization. The restricted information transfer between nodes, and the development of cluster heads, places this algorithm as a multistation architecture.

The algorithm relies on the creation and maintenance of interconnected clusters. A cluster consists of similar nodes which operate in one of three roles; a cluster head (or "head node"), possibly a few gateway nodes, and several ordinary nodes. The head may be considered a station, through which nodes communicate, while gateways act as links between clusters. Local communication takes place between ordinary nodes, or via the cluster head. Gateways enable communication with distant clusters based on a backbone established between cluster heads. With the network so far described each node may communicate (by some means) with any other node. However, this algorithm is unsuited for military purposes. There are no duplicate paths between nodes, and cluster heads leave the network routing vulnerable to global disruption.

The algorithm is significantly improved by using *several* overlaid connectivity maps, each connecting all nodes into a network, yet each using different cluster heads and routes. This is accomplished by splitting the HF Band into M sub-bands, each with $1/M$ of the available bandwidth. Communication takes place over $M-1$ of the sub-bands while the remaining sub-band has a new connectivity map constructed. Additionally one of the communication bands is reserved for network control. The maps are built from data received since the previous reconstruction. The network reconstruction algorithm is required to run once for each sub-band and cycles around all the sub-bands. The algorithm works on different sub-bands in a time division multiple access (TDMA) fashion; it is therefore important that all nodes maintain an accurate global time. It should be noted that transmission properties of different sub-bands produce varied connectivity maps. In particular, as the frequency increases, the broadcast range decreases [1], reducing the number of connections in the network.

During the reconstruction of a connectivity map all packet routing is evenly distributed across the other sub-bands. The algorithm consists of two major parts, network creation and routing within the network; these are discussed below.

Connectivity map construction

This section is concerned primarily with the creation of the connectivity maps performed by the channel engaged in network control of its sub-band. The algorithm is explained and a brief discussion of those qualities which are desirable (e.g. distributed control and link activation) and those which pose problems (e.g. the quantity of information required by each node, and network security) is presented.

Channel control proceeds in the following manner. The control schedule consists of M epochs, in which at any interval, epoch i will be engaged in reconstructing its connectivity maps while epochs $1 \ldots i-1, i+1 \ldots M$ will be used for normal message communication. Epoch i is divided into two frames. Each node constructs its own frames which are divided into a number of slots. The size is an integer power of two, for convenience. The first slot is assigned for communication between the node and its cluster head. The remaining slots are assigned for listening to adjacent nodes and for transmitting to neighbors. A diagram showing the control channel timing schedule is given in Figure 2.5.

For an epoch, the outline of the algorithm proceeds as follows. In frame 1 each node builds partial connectivity maps from the nodes *it* can hear. In frame 2 each node broadcasts the information to its immediate neighbors. Thus at the end of frame 2 all bidirectional links between nodes are established. With a little additional information it is possible to establish cluster heads and gateways as frame 2 progresses. By the end of frame 2 the network connectivity map is built allowing several routing algorithms to be used. The suggested routing technique is discussed later. First an

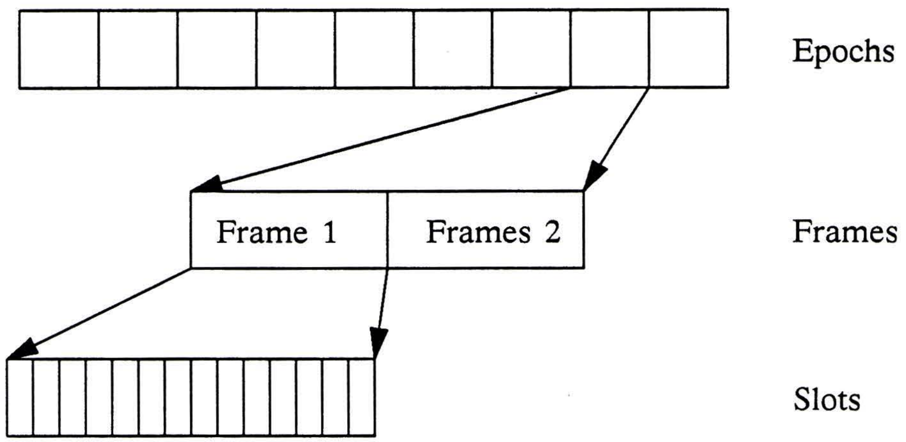


Figure 2.5. Control Scheduling

epoch is considered in more detail. To maintain connectivity information each node is required to hold the following data structures. Three tables, *HeadsOneHopAway, HeadsTwoHopsAway, NodesHeard,* are required along with *Connectivity*, a matrix.

Frame 1. As previously stated, for each node, the frame consists of 2^k slots. Initially each node will attempt to assign adjacent nodes to particular slots, forming a schedule. Likewise, surrounding nodes will label their slots giving rise to inconsistencies. These are resolved by labeling the slots systematically, with nodes having lowest IDs in the first free slot etc. Although this will reduce inconsistencies, the tables will contain errors until the completion of frame 1. During slot i, node i broadcasts its own identity and the identity of the nodes it has heard during the previous slots (i.e its *NodesHeard* list). Other nodes $j = 1 \cdots k, j \neq k$ listen, and are able to determine nodes that they can hear and have heard them; these nodes are marked in the ith row of the *Connectivity* matrix. By the completion of frame 1, all nodes are aware of their connectivity with nodes having later slots, and each node has a slot allocated with its neighbors.

Frame 2. In its assigned slot each node broadcasts row i of its *Connectivity* matrix. Using this, surrounding nodes are able to establish the connectivities missed in frame 1 and also determine those nodes which are two hops away (i.e. nodes connected to the broadcasting node yet not connected to the receiving node).

During this frame node i also transmits a node status. Since the node is aware of all its connectivities it is now in a position to determine whether it has the highest identity of the adjacent nodes. The status may be cluster head, gateway, or ordinary node. At the end of Frame 2 each node should be able to complete its own table of *HeadsOneHopAway* as well as table of *HeadsTwoHopsAway*, identifying the cluster heads.

One effect of the algorithm is that a cluster head may "cover" the same nodes as another node as shown in Figure 2.6. In this case a redundant cluster head exists which may be deleted. It is unclear whether it is desirable to remove these heads, which reduces the connectivity of the backbone, or leave them as additional routes.

If redundancy is to be removed, heads may check to see if all their adjacent nodes are connected to another node in the *HeadsTwoHopsAway* table with a higher ID than themselves. In this case the node remains an ordinary node. In a similar manner, ordinary nodes are able to distinguish between redundant heads and essential ones.

The final stage in establishing the connectivity map is linking up the cluster heads. This proceeds as follows.

- If two heads are in direct communication, the link is established automatically; see Figure 2.7.

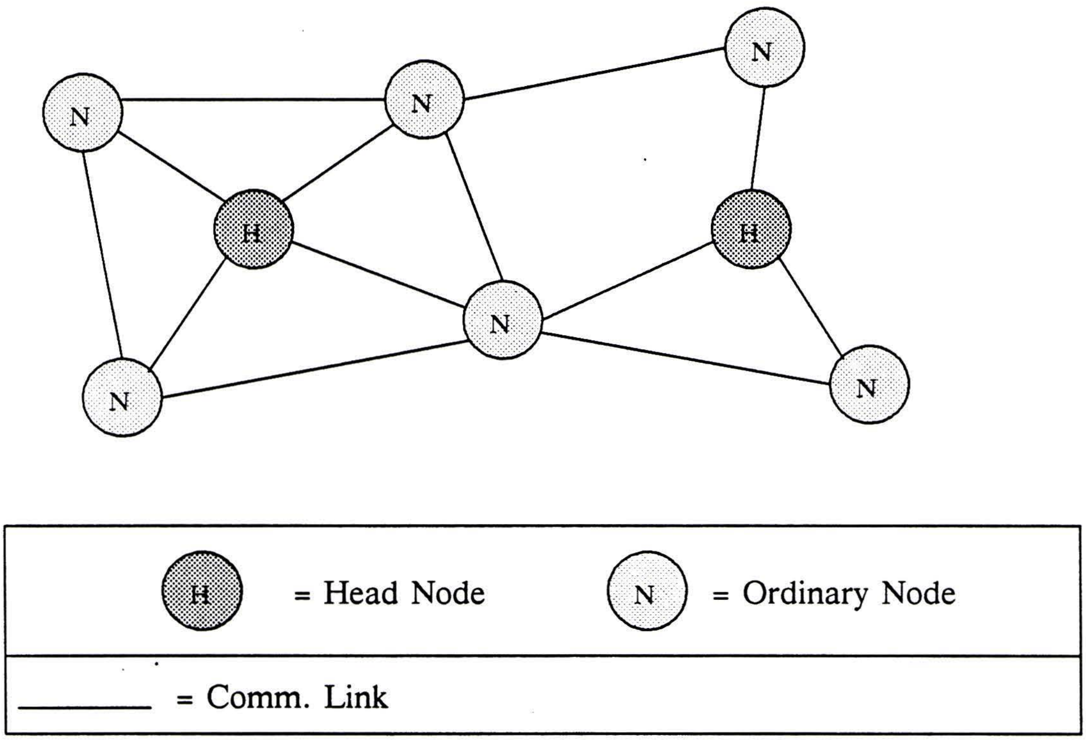


Figure 2.6. Head Cover

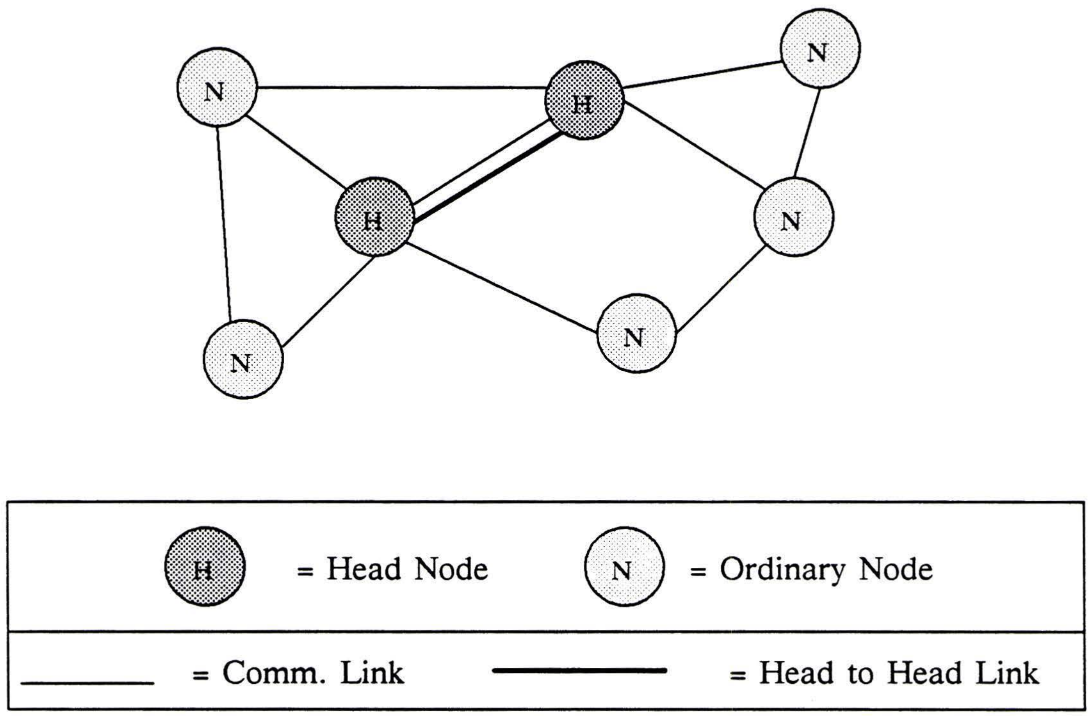


Figure 2.7. Direct Head Linkage

- When exactly one node is required to connect the two separate clusters, the node which has the highest ID, and is able to hear both heads, is selected as the gateway; see Figure 2.8.

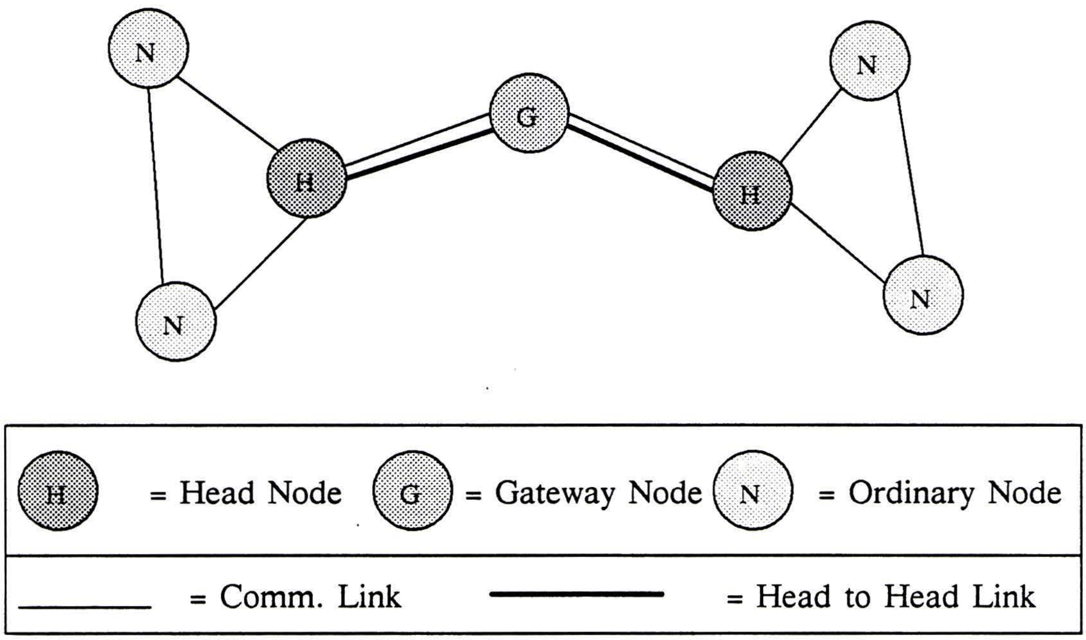


Figure 2.8. Two Head Hops Apart

- If more than one node is required to connect the separate clusters, each node must check all pairs in its *HeadsOneHopAway* and *HeadsTwoHopsAway* tables where the first node is its own cluster head, and the second is in its *HeadsTwoHopsAway* list; see Figure 2.9.

Routing

As with PRNET the most appealing routing strategy is to use the cluster heads as local controllers. Route selection is managed by the heads and may be either direct communication or through the head. Intercluster communication can be coordinated by cluster heads and gateways operate in a fashion similar to those in PRNET.

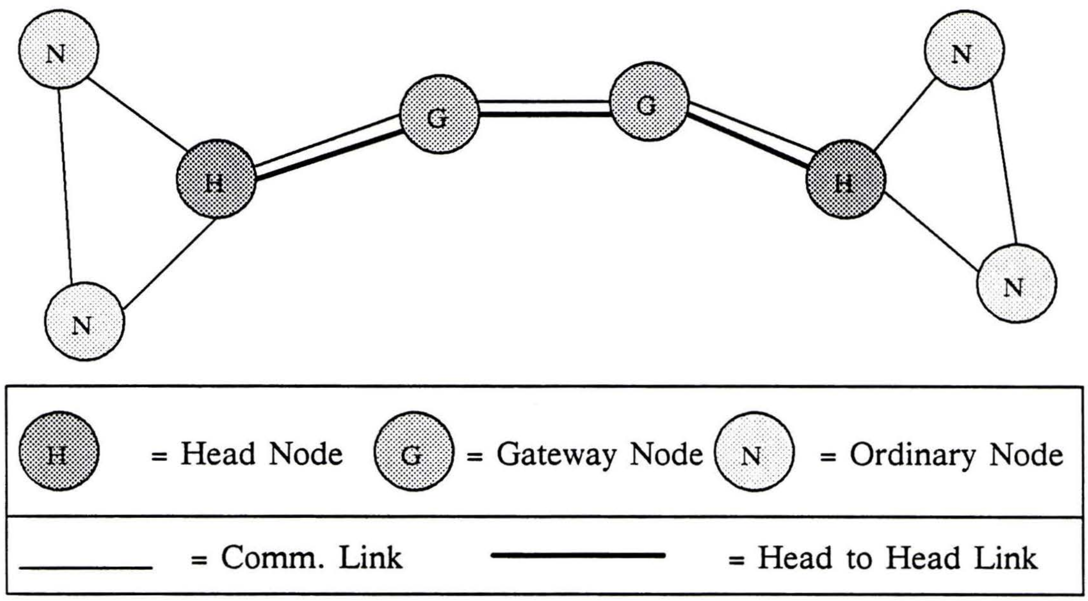


Figure 2.9. Heads Many Hops Apart

The network exhibits many of the features desired in military communications. The use of multiple sub-bands, each with different transmission characteristics and network topologies, yields a resilient network. Several issues are not addressed in the Baker's paper [2] which should be mentioned. The control channel is based on a time slotted algorithm and requires that each node broadcast in a given slot. Such time critical elements are awkward in radio broadcast systems where signal fading, imaging etc are frequent. In a similar fashion, there is no mention of maintenance of the control channel, and questions regarding the addition and deletion of nodes are unasked. Additionally, the resolution of slot assignments may cause wasted table space since we require that each node reserve the same slot value for other nodes. A node in limited communication, with access to only a few other nodes, may have large gaps in its table.

2.4 Satellite Network Theory

2.4.1 Orbital Mechanics

The motion of an artificial satellite in orbit about the earth may be described by Newton's laws of motion [46]. Consider the system of two bodies with masses M and m as illustrated in Figure 2.10. The position vectors of the bodies M and m are given by $\mathbf{r}_M$ and $\mathbf{r}_m$ respectively. The forces acting on the bodies are given by

$$\begin{aligned}\mathbf{F}_M &= M\frac{d^2\mathbf{r}_M}{dt^2}\\ \mathbf{F}_m &= m\frac{d^2\mathbf{r}_m}{dt^2}\end{aligned}$$

According to Newton's law of gravitation, the attractive force between two bodies is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Thus

$$\mathbf{F}_M = -\mathbf{F}_m = g\frac{Mm}{r^2}\left(\frac{\mathbf{r}}{r}\right)$$

where g is the earth's gravitational constant. From the above three equations and $\mathbf{r} = \mathbf{r}_m - \mathbf{r}_M$ we obtain

$$\begin{aligned}\frac{d^2\mathbf{r}}{dt^2} &= -g(M+m)\frac{\mathbf{r}}{r^3}\\ \frac{d^2\mathbf{r}}{dt^2} &= -\mu\frac{\mathbf{r}}{r^3}\end{aligned} \tag{2.2}$$

where $\mu = g(M+m) \approx gM$, since the mass of the satellite is negligible compared to that of the earth. Equation 2.2 completely describes the motion of a satellite orbiting the earth.

A satellite may be in either a circular or an elliptical orbit. The characteristics of the satellite are described by Kepler's laws published in 1609 and 1619 (third law).

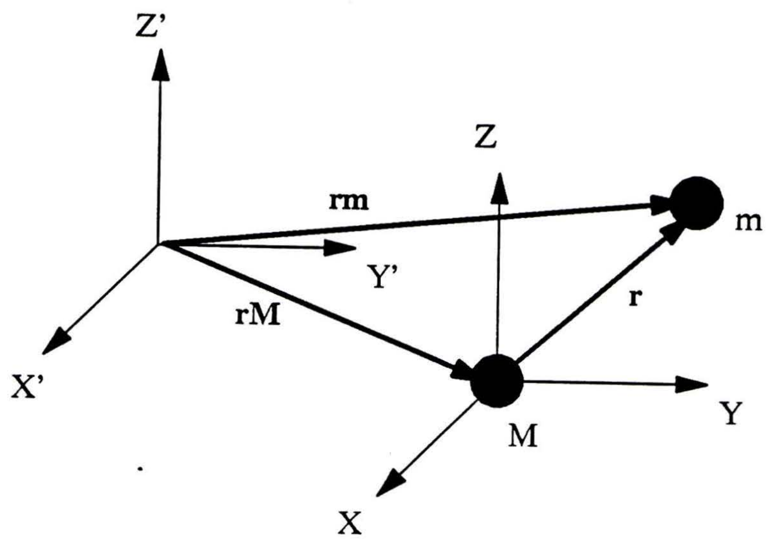


Figure 2.10. Relative Motion of Two Bodies

First Law: The orbit of each planet is an ellipse, with the sun at a focus.

Second Law: The line joining the planet to the sun sweeps out equal areas in equal times.

Third Law: The square of the period of a planet is proportional to the cube of its mean distance from the sun.

Using the following notation, Kepler's equation is stated below without derivation.

r	=	distance of satellite from the primary focus F (i.e. center of the earth)
ν	=	true anomaly, angle measured from F in the direction of motion, starting from the perigee (point of minimum altitude)
a	=	semimajor axis of ellipse
b	=	semiminor axis of ellipse
e	=	eccentricity
E_a	=	eccentric anomaly, angle measured from the center of the ellipse

O in the direction of motion, starting from the perigee

p	=	semiparameter
q	=	perigee distance, the point on the orbit closest to F
Q	=	apogee distance, the point on the orbit furthest from F
t_0	=	some initial time instance
t	=	some period after t_0
M	=	mean anomaly

Kepler's equation:

$$M = E_a - e\sin(E_a) = -\frac{\sqrt{\mu}}{a^{3/2}}(t - t_0) \tag{2.3}$$

By setting E_a to 2π and $T = t - t_0$ for the satellite period;

$$T = 2\pi\frac{a^{3/2}}{\sqrt{\mu}} \tag{2.4}$$

To find the instantaneous velocity of the satellite, use the scaler product of equation 2.2 and $d\mathbf{r}/dt$, to get

$$\frac{d\mathbf{r}}{dt} \bullet \frac{d^2\mathbf{r}}{dt^2} = -\frac{\mu}{r^3}\left(\frac{d\mathbf{r}}{dt} \bullet \mathbf{r}\right) = -\frac{\mu}{r^2}\left(\frac{dr}{dt}\right)$$

Integrating this, and using the boundry conditions $dr/dt = 0$, and $r = q$ at the perigee, the orbital velocity is given by

$$V = \sqrt{\mu\left(\frac{2}{r} - \frac{1}{a}\right)} \tag{2.5}$$

Note that a circular orbit is just a special case of the elliptical orbit where $a = b = r$.

Although these equations give the period and velocity of a satellite they are in general insufficient to permit calculating the position of a satellite given initial conditions and some time interval (commonly referred to as the prediction problem). In fact the problem of finding ν given a, e, ν_0 and $t - t_0$ is complex; however solutions exist, the most popular being the *Universal Variable Formulation.* The equations involved in this are complex and may be found in Bate et al. [5]. It is sufficient to say that the problem is solved using iteration on a set of equations to converge on the solution. No closed form solution exists, to quote Small [48], in *An Account of the Astronomical Discoveries of Kepler*:

> This problem has, ever since the time of Kepler, continued to exercise the ingenuity of the ablest geometers; but no solution to it which is rigorously accurate has been obtained. Nor is there much reason to hope that the difficulty will ever be overcome. . . . [5, p. 193]

Similarily orbit determination from two positions and the time between them is complex and has no closed form solution. It would appear therefore that computations involving satellite systems (i.e. tracking, routing etc. within a satellite network) will involve a great deal of numeric calculation to achieve even the simplest of results. In fact this is the case for elliptical orbits; however by restricting the problem to circular orbits the difficulties vanish, and this is the approach used in this dissertation. In restricting the network to circular orbits one may wonder what the effect will be on the overall solution. This issue is addressed further along.

Perturbing accelerations on satellites

Having briefly discussed the path of a satellite in terms of the "two-body" problem, it is worth mentioning that real-world problems cannot be solved using such naïve techniques [5]. The actual path of a satellite will deviate from the theoretical due to perturbations from other masses and additional forces from non-Keplerian sources. In the interests of completeness the most important of these are mentioned below; however no further reference will be made to them. There are two main categories of perturbation, *predictable perturbations* and *unpredictable perturbations.*

The most important predictable perturbations are 1) the presence of other attracting bodies (e.g. the moon), 2) atmospheric drag, 3) the oblateness of the earth, 4) solar radiation, 5) magnetism, and 6) relativistic effects. Fortunately analytic formulations exist for these and may be included in orbital calculations (at the expense of computation time). Unpredictable perturbations can come from many sources; however the principal ones are 1) meteor collisions and 2) wind gusts. Because of their nature these must be handled in a stochastic manner. One advantage with satellite orbits is that they may be observed; thus there is the potential for feedback in the system. This property is extremely important in keeping track of real satellites. It is quite usual to make 50 position sightings of a satellite per day so that the model of it is kept accurate.

Equations for circular orbits

Using the fact $a = b = r$ for circular orbits, the equation for the period of the orbit simply becomes

$$T_c = 2\pi \frac{r^{3/2}}{\sqrt{\mu}} \tag{2.6}$$

The speed required for a circular orbit of radius r may be derived from the energy equation

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a} \tag{2.7}$$

and is given by

$$V_c = \sqrt{\frac{\mu}{r}} \tag{2.8}$$

Notice that the greater the radius of a circular orbit, the less is the speed required to keep the satellite in orbit. For low altitude orbits the speed ranges from 7.35 km/sec (altitude of 1000 km) to 4.93 km/sec (altitude of 10000 km).

3-Dimensional rotation

An object may be rotated about a given frame of reference by a set of orthogonal rotation matricies. Let the three dimensional rotation matrices be given by

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & \sin(\theta_x) \\ 0 & -\sin(\theta_x) & \cos(\theta_x) \end{bmatrix} \quad R_y = \begin{bmatrix} \cos(\theta_y) & 0 & -\sin(\theta_y) \\ 0 & 1 & 0 \\ \sin(\theta_y) & 0 & \cos(\theta_y) \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos(\theta_z) & \sin(\theta_z) & 0 \\ -\sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Using these, and selecting the center of the earth as our inertial frame of reference with arbitrary but consistent x, y, z axes, a body may be rotated about any of the axes.

Circular satellite system

A system of satellites may be described by shells of satellites, with each shell at a particular altitude. The satellites on a shell will have the same velocity and period of revolution (from equations 2.8 and 2.6). Within a shell, satellites may lie in different orbital planes; those satellites which lie on the same plane are *co-planer* and form a *plane of satellites.* Typically satellites will be equi-spaced in planes (i.e. the angle between successive satellites is the same and constant over time) and planes will be arranged to achieve some particular function. Individual satellite systems and models will be discussed in the following chapter. The positioning of planes of satellites into the 3-dimensional coordinate system may be achieved by maintaining the individual values of ν for each satellite with reference to its own plane (planes have some initial configuration of satellites), called the perifocal system. It is then a simple matter to rotate the x_{2d}, y_{2d} values of the satellite from its own frame to the x_{3d}, y_{3d}, z_{3d} coordinate of the real system using the following equation (derived from 2.4.1).

$$\mathbf{B} =$$

$$\begin{bmatrix} N_1^2 + (1 - N_1^2)C_\delta & N_1N_2(1 - C_\delta) + N_3S_\delta & N_1N_3(1 - C_\delta) - N_2S_\delta \\ N_1N - 2(1 - C_\delta) - N_3S_\delta & N_2^2 + (1 - N_2^2)C_\delta & N_2N_3(1 - C_\delta) + N_1S_\delta \\ N_1N_2(1 - C_\delta) + N_2S_\delta & N_2N_3(1 - C_\delta) + N_1S_\delta & N_3^2 + (1 - N_3^2)C_\delta \end{bmatrix}$$

$$\begin{bmatrix} x_{3d} \\ y_{3d} \\ z_{3d} \end{bmatrix} = \mathbf{B} \begin{bmatrix} x_{2d} \\ y_{2d} \\ 0 \end{bmatrix}$$

where

$$\begin{aligned} C_\delta &= \cos(\delta) \\ S_\delta &= \sin(\delta) \\ N_1 &= \cos\left(\frac{\pi}{2} - \alpha\right) \\ N_2 &= \cos(\alpha) \\ N_3 &= 0 \end{aligned}$$

and α is a rotation of the satellite plane about the z-axis followed by δ, an inclination of the rotated plane. Thus $0 \leq \alpha < 2\pi$ and $-\frac{\pi}{2} \leq \delta \leq \frac{\pi}{2}$.

<u>Justification of circular orbits</u>

The topologies selected for analysis in terms of coverage and connectivity representation in the following chapters share a common characteristic in that they all use only circular orbits. This restriction requires some comment. In section 2.4.1 it was stated that the equations necessary to predict the position of a satellite in elliptical orbit were both complex and iterative. In contrast, the equations to predict the position of a satellite in circular orbit (section 2.4.1) are quite easy and have a closed form. This difference is manifest in almost all the calculations pertaining to orbital mechanics. While restricting satellite topologies to circular orbits permits a number of interesting measures to be derived, the use of elliptical orbits effectively prohibits such measures to exist without much iterative computation. The reason for this is simply that all analysis in a satellite system makes use of the instantaneous positions of the satellites; if there is no closed form for even this, then clearly further results will be iterative.

In addition to the complexity of dealing with elliptical orbits, it is unlikely that any topological benefit will emerge from using these orbits with the theoretical model used later. The advantages associated with elliptical orbits are derived from the increased energy of the system which is used to counter "real" problems such as atmospheric drag, solar radiation etc. The model used omits all these conditions in the interest of analyzing the routing issues without geometry "noise."

In concluding this section it should be noted that realism would not be difficult to incorporate into the problem. Elliptical orbits and the effects of all the perturbative accelerations could be included into the analysis. In fact, in the calculation of real satellite orbits these details are included into the problem. As a consequence no closed form analysis would be possible, and computation of any measures would be *much* more lengthy. The actual techniques developed in this project, however, would remain the same.

2.4.2 Physical Communication Options

There are essentially two communications options which may be employed between satellites: microwave communication links and laser communication links. Both of these are electromagnetic radiation; the difference is in the frequency of the carrier. It should be noted that the properties associated with the two are quite different. At present there are no "Satellite Networks" as such; there are, however, intersatellite links (ISL) which are used to relay data around the globe. The cross link microwave frequencies are in the K and V bands as follows:

K	band	20	-	30 GHz
V	band	54	-	58 GHz
		59	-	64 GHz

The V bands have transmission band-widths of 3.5 and 5 GHz. One reason for selecting such high carrier frequencies is due in part to the constraints placed on satellite transmission power and receiver dish size. A more recent candidate for intersatellite communications is the use of the laser beam, which has the advantage of operating at much higher frequencies than millimeter radio (e.g., 3×10^5 GHz). Looking at the effect of beam divergence in lasers, we see the beam is considerably more focused. Consider two satellites each at altitude A above the Earth, separated by an angle ϕ_s. If they have transmitter diameters d and are communicating at wavelength λ, we know from Gagliardi [22] that beam divergence (e.g. half power beam-width) ϕ_b, is given by

$$\phi_b = k\frac{\lambda}{d} \tag{2.9}$$

where k is an efficiency factor.

For an antenna dish this approximates to $1.02\frac{\lambda}{d}$. For a satellite at an altitude of 1300 km we can calculate the beam-width, ϕ_b, for both radio (1-meter dish) and laser (6-inch lens) and compare them:

$$\text{radio} : \phi_b \quad = \quad \frac{1.02 * 5.5 \times 10^{-3}}{1} \tag{2.10}$$

$$
\begin{aligned}
&= 5.6 \times 10^{-3} \text{ radians} \\
\text{laser} : \phi_b &= \frac{1.02 * 1 \times 10^{-6}}{0.15} \qquad (2.11) \\
&= 1 \times 10^{-5} \text{ radians}
\end{aligned}
$$

Thus the laser gives a beam-width which is 250 times smaller than that of the radio. Assuming that the two satellites are stationed one radian apart, and the radius of the earth is R_e, then the distance between them, z is given by

$$
z = 2(R_e + A)\sin(\phi_s) \qquad (2.12)
$$

$$
\begin{aligned}
&= 2(6378 + 1300)\sin(1) \\
&= 1.29 \times 10^4 \text{ km}
\end{aligned}
$$

So, using the two communication media the beam divergence, x, is approximately

$$
x = z\sin(\phi_b) \qquad (2.13)
$$

$$
\begin{aligned}
\text{radio} : x &= 1.29 \times 10^4 \sin(5.6 \times 10^{-3}) \\
&= 72,000 \text{ meters} \\
\text{laser} : x &= 1.29 \times 10^4 \sin(1 \times 10^{-5}) \\
&= 129 \text{ meters}
\end{aligned}
$$

Assuming that both methods exceed the requirements in data band-width, which is the most effective? There are tradeoffs associated with each. Radio is expensive on power and antenna size; however, the exact position of the target satellite is not required; i.e., tracking is not too difficult. For laser, the power and antenna requirements are low; however, the tracking requirements are very high.

Typical communication distances between satellites are likely to be at least 1500 km (e.g., 20-30 satellites in orbit 1000 km above the earth). This distance causes typical propagation delays on the order of 6 msec, which is slow compared to the data band-width.

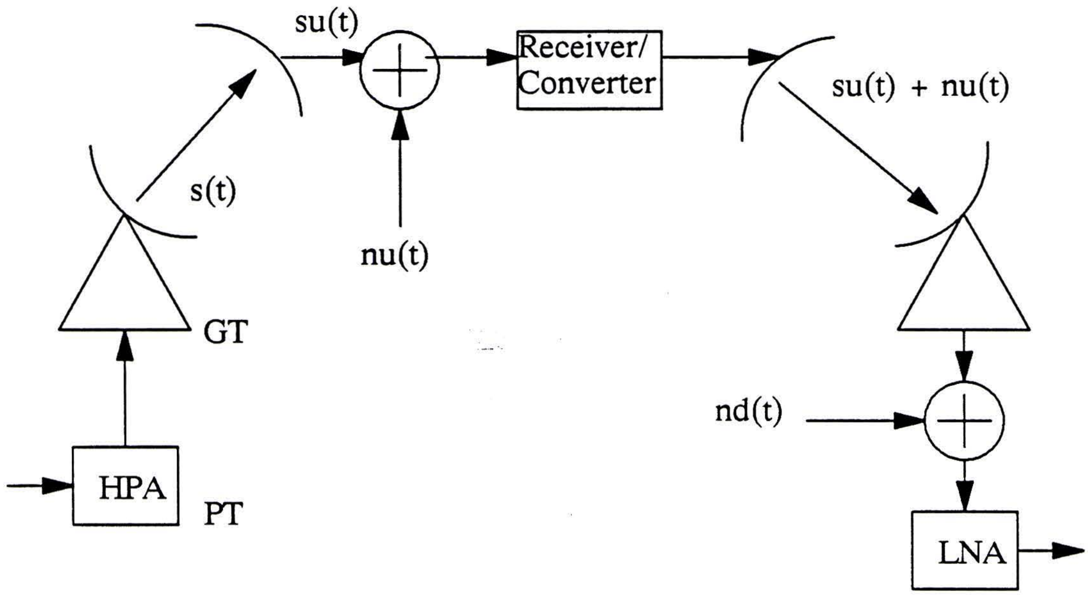
Since the SDI project calls for directed beams to pinpoint a moving target (e.g., ground-launched missile), we may assume that tracking is an issue which can be solved. From a communication viewpoint, it would, therefore, seem that laser communications offer the most effective medium.

2.4.3 Establishing Links and Targeting

One important assumption necessary for intersatellite communications is that the technology exists to permit satellite communication links to be established and maintained until no longer required. This dissertation uses this assumption; however in this section some justification is given to verify that the assumption is reasonable. The communication may be considered in two distinct parts, 1) Earth/Satellite Links and 2) Satellite/Satellite links. If both these issues are resolved, then the link establishment and maintenance are possible.

Earth/Satellite links

An earth/satellite link consists of an up-link and a down-link, the characteristics of which are different. The signal quality of the up-link depends on how strong the signal is when it leaves the earth, and how the satellite receives it. Conversely, the down-link signal quality depends on the satellite signal strength and the earth station receiver. Since ground stations and satellites have very different characteristics, the up-link and down-link signal qualities will also be different. As with satellite/satellite communications there are two options: 1) use microwave communications, and 2) use laser communications. In either case a simple satellite link would be as shown in Figure 2.11. Signal attenuation may be considered the combination of three effects: 1) free space loss, 2) atmospheric attenuation, and 3) rain-induced attenuation.



PT = Power Transmitted
GT = Gain, Transmitter
su(t) = Received signal
nd(t) = Noise, Detector

HPA = High Power Amplifier
s(t) = Signal Transmitted
nu(t) = Noise
LNA = Low Noise Amplifier

Figure 2.11. Basic Satellite Link

Recall from the previous section that the characterists of electromagnetic radiation are functions of the wavelength. In particular consider a simple model in which the transmission occurs between the satellite and ground station and that the signal attenuation is caused by free space loss, and atmospheric attenuation only (i.e no antenna tracking loss, reflected signal interference, satellite interference, etc). The free space loss in this system is given by [25]

$$L_u = \left(\frac{4\pi d_u}{\lambda_u}\right) \tag{2.14}$$

The atmospheric attenuation L for microwave radiation consists of atmospheric absorption which is significant only in bands centered at 22.2GHz (water vapor), 60GHz and 118Ghz (oxygen), and rain attenuation. Fortunately rain is not a serious problem at 6/4GHz band; however it is a major problem above the 10GHz region. The atmospheric attenuation may be simply modeled by the formula

$$L_a(dB) = \frac{L'_a(dB) + (\rho_0 - 7.5g/m^3) + c_T(21\deg C - T_0)}{\sin E} \tag{2.15}$$

where

L'_a	=	zenith one-way attenuation for a moderately humid atmosphere ($7.5g/m^3$ surface water vapor) and a surface temperature of 21 deg. Given in Table 2.2
b_p	=	water vapor density correction coefficient; given in Table 2.3
c_T	=	temperature correction coefficient; given in Table 2.3
ρ_0	=	surface water vapor density

The equations for rain-induced attenuation are derived from statistical models and yield results which agree with experimental observations. The four main models are the Rice-Holmberg [47], Dutton-Dougherty [13], Lin [37], and Crane [11]. These models are complex and will not be further discussed.

The effects of the rain attenuation depend greatly on the ground area to be covered. We note that even at the worst condition the signal attenuation is about 10dB/km. Looking at the atmospheric attenuation on the light wave spectrum, we see from the graph in Figure 2.12 below that severe signal attenuation takes place.

Table 2.2. Attenuation for Moderately Humid Atmosphere

Frequency (GHz)	Altitude (km) 0	0.5	1.0	2.0	4.0
10	0.053	0.047	0.042	0.033	0.02
15	0.084	0.071	0.061	0.044	0.023
20	0.28	0.23	0.18	0.12	0.05
30	0.24	0.19	0.16	0.10	0.045
40	0.37	0.33	0.29	0.22	0.135
80	1.30	1.08	0.90	0.62	0.30
90	1.25	1.01	0.81	0.52	0.22
100	1.41	1.14	0.92	0.59	0.25

Table 2.3. Water Vapor Density and Temperature Correlation Coefficient

Frequency (GHz)	Water vapor density correction b_p	Temperature correction c_T
10	2.10×10^{-3}	2.60×10^{-4}
15	6.34×10^{-3}	4.55×10^{-4}
20	3.46×10^{-2}	1.55×10^{-3}
30	2.37×10^{-2}	1.33×10^{-3}
40	2.75×10^{-2}	1.97×10^{-3}
80	9.59×10^{-2}	5.86×10^{-3}
90	1.22×10^{-1}	5.74×10^{-3}
100	1.50×10^{-1}	6.30×10^{-3}

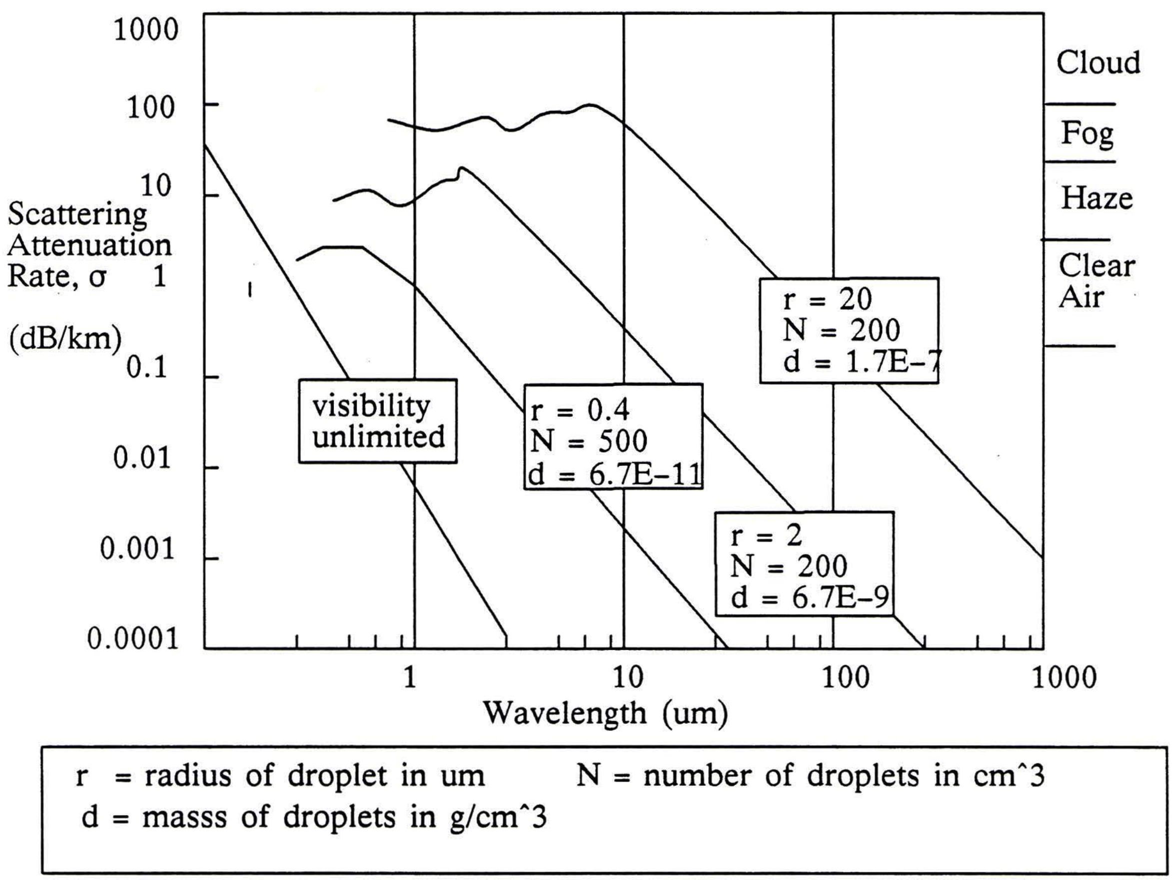


Figure 2.12. Atmospheric Attenuation of Optical Wavelengths

Notice that even with only hazy conditions the signal attenuation is 12dB/km. From the previous discussion it should be apparent that the major problem in up-link/down link communications is the signal attenuation caused by atmospheric effects. We emphasize that the microwave communication links are at a substantial advantage in this area, particularly below the 10GHz frequency. In conclusion it is clear that the with regard to up-link/down-link the microwave communication link is superior.

Satellite/Satellite links

As with Earth/Satellite links the option exists to have either microwave communications or laser communications. Since the communications are outside the atmosphere the signal attenuation models used in the previous section are inappropriate; in fact the attenuation is once again controlled by the *free space loss.* Thus laser communication is far superior.

$$L_u = \left(\frac{4\pi d}{\lambda}\right)$$

In this case the problems associated with severe laser attenuation as a result of the atmosphere are removed. The feasibility of using optical communications for low altitude satellite links has been investigated by Biederman et al. [6]. Their report discusses and compares different implementations of transmitting and receiving using lasers. Although analysis of this area is outside the scope of this dissertation, a few of the more salient observations and results are provided below. The interested reader is referred to the original report. The typical laser communication model is described pictorially in Figure 2.13.

The laser source may be either multiple mode, noncoherent transmission in the 133-665 GHz range, or may be a single mode laser permitting coherent systems. Modulation to the transmission link may be by either an external waveguide or by

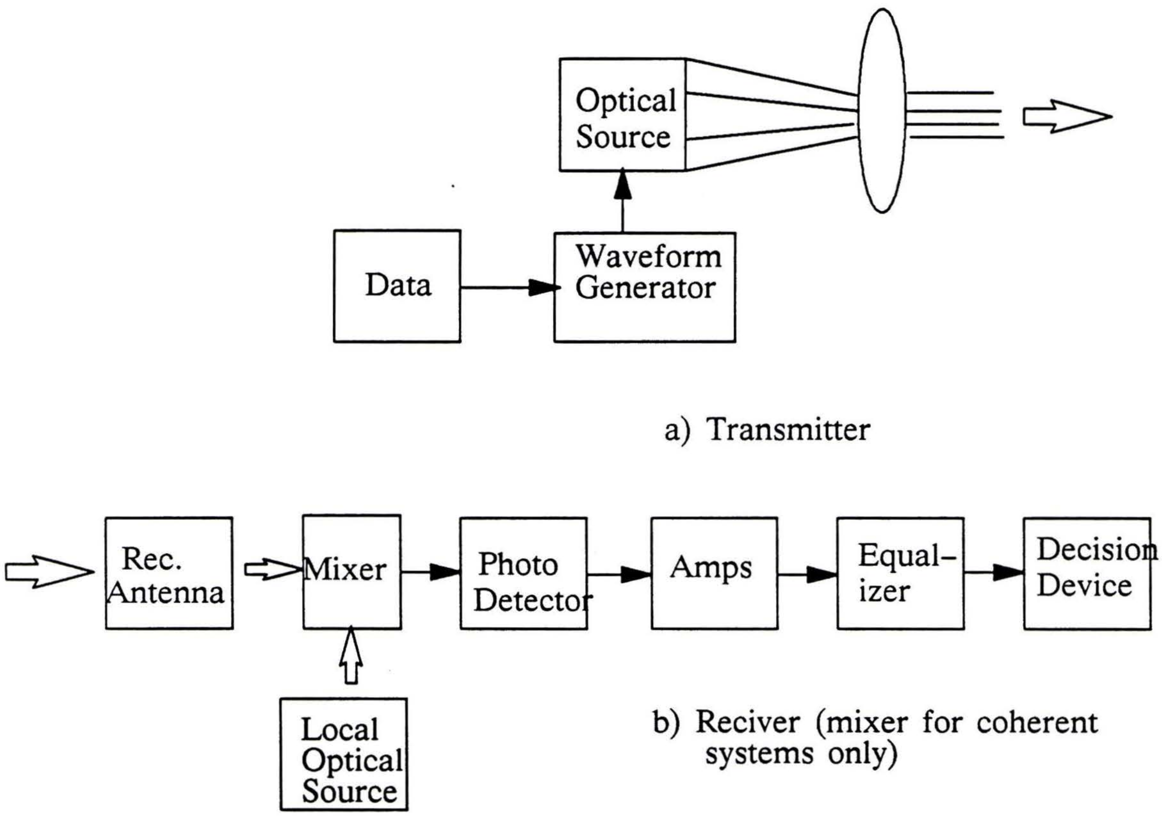


Figure 2.13. Diagram of digital Optical Communication System

changing the laser drive current [4]. The transmitter antenna is usually a Cassergrain telescope which has a far-field on-axis gain given by [44]

$$G = 2\left(\frac{2\pi R_o}{\lambda^2}\left(e^{-\left(\frac{b}{R_o}\right)^2} - e^{-\left(\frac{a}{R_o}\right)^2}\right)^2\right)$$

where

R_o	=	$1/e^2$ power point of the feed beam at the exit aperture of diameter D, D $> 3R_o$
a	=	radius of primary mirror
b	=	radius of secondary mirror
λ	=	carrier wavelength

At the other end of the system, a photodetector is used to convert the absorbed optical field collected by the optical antenna into an electrical signal. In addition the photodetector may also amplify the primary photocurrent so that a sufficient signal-to-noise ratio is obtained. Typically photodetectors are photomultiplier, a p-i-n, or an avalanche photodiode.

Signaling

The choices of coding techniques for laser communications using direct detection are either NRZ coding or Manchester coding. The disadvantage with NRZ code is that its performance depends on threshold and gain of the receiver which are in turn functions of the signal and noise. The alternative (Manchester coding) has several advantages over NRZ: 1) the optimal threshold is always zero, independent of the SNR; 2) the laser power pattern for long strings of ones or zeros which occur with NRZ does not exist with the Manchester code. These patterns degrade receiver sensitivity and simplify power stability circuitry; 3) baseline wander is not a problem since there is zero DC component. The price paid for these advantages is that the transmitted pulse rate is effectively double that of the NRZ code scheme.

Multiple access communications

We note in passing that it may be desirable for several satellites to communicate with one receiver by using multiple access systems. Figure 2.14 shows the way in which several systems may send data which is focused onto a single photodetector array so the the individual messages may be separated. Clearly the effects of the signals on one another (i.e. crosstalk) degrade the individual receiver elements; however there are large cost benefits with such a device.

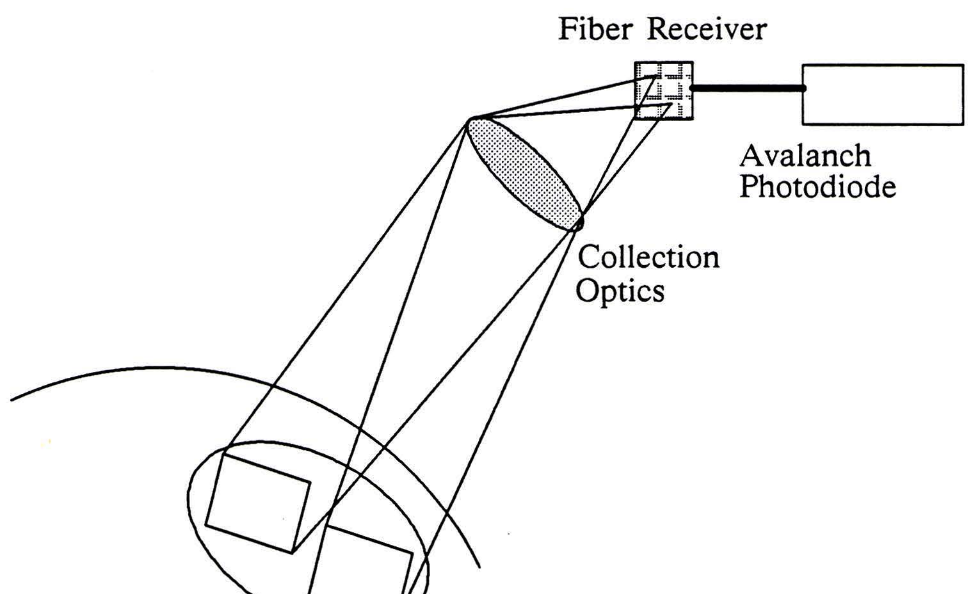


Figure 2.14. A Multiple Direct Detection Receiver

2.4.4 Antenna Pointing and Acquisition

Before data communication can occur in a cross-link it is necessary for the source to point the laser transmitter to the destination, and the destination to align the receiving equipment. These operations are called pointing and spatial acquisition, respectively. In section 2.4.2 it is mentioned that an optical beam will diverge on the order of a few micro radians. Thus the pointing mechanism must be able to point to an accuracy of half this distance. Inaccuracies in directing a communication beam over large distances are caused by several factors:

- The difficulties that exist in establishing the exact direction in which to send the beam. The error arises from inaccuracy in both the location of the transmitting satellite and the destination satellite. Establishing global coordinates for satellites may be achieved either by using the relative positions of fixed stars or by the use of a global positioning system such as GPS.

- The errors inherent in the pointing equipment are another source for pointing errors. It is expected that the antennae will be pointed by either electronic, electrical, or mechanical devices. Errors from these devices will be accentuated by stress, fatigue etc.

- Finally there will always be some error due to the motion of the satellites relative to one another. This effect will occur even if the motions are accurately predicted by celestial mechanics (see section 2.4.1) due to coefficient errors and unpredictable perturbations.

Establishing an exact direction may be effectively controlled by using a system such as GPS. The essential concept of GPS is that each satellite broadcasts a carrier imposed with a high data rate binary code. The primary carrier is at 1575.42 MHz, with a secondary at 1227.60 MHz. The data consist of a coarse/acquisition code (C/A) and a precise (P) code. Each satellite transmits a different code, designed to provide minimum interference with each other. The C/A code is repeated frequently and is easy to identify while the P code is a long code. The continuous tracking of a single satellite enables the receiver to place itself within a shell from the satellite. Measurements from two satellites place the receiver in a circle, and a third measurement places the receiver within a small segment. GPS is designed to provide global coverage (with 18 satellites); however the accuracy of the system varies with both receiver position and with time. The position of the satellites (i.e. the satellite geometry) determines the geometric dilution of precision (GDOP) as indicated in Figure 2.15. This same effect will be prevalent in a satellite communication geometry. The accuracy of the experimental GPS is exceptionally high with the precision code accurate to 7 m. In addition to GPS providing accurate position information, each satellite is equipped with a cesium clock and is able to provide GPS absolute timing information to a receiver having accuracy within 100 nsec.

The other open loop pointing errors (those from equipment and orbital motion) are expressed in Table 2.4. The important point to note is that the cumulative error is on the order of 2 μrad, recalling from the previous section the beam-width at the destination is ≈ 5 μrad; thus laser pointing to a satellite is an achievable goal.

Tracking satellites

Once the receiving antenna has acquired the direction of the transmitting antenna, it is a simple matter to establish a tracking loop between the two satellites. The function of this loop is to maintain the receiving antenna properly oriented relative to the arriving optical signal field. The operation of tracking loops can be found in

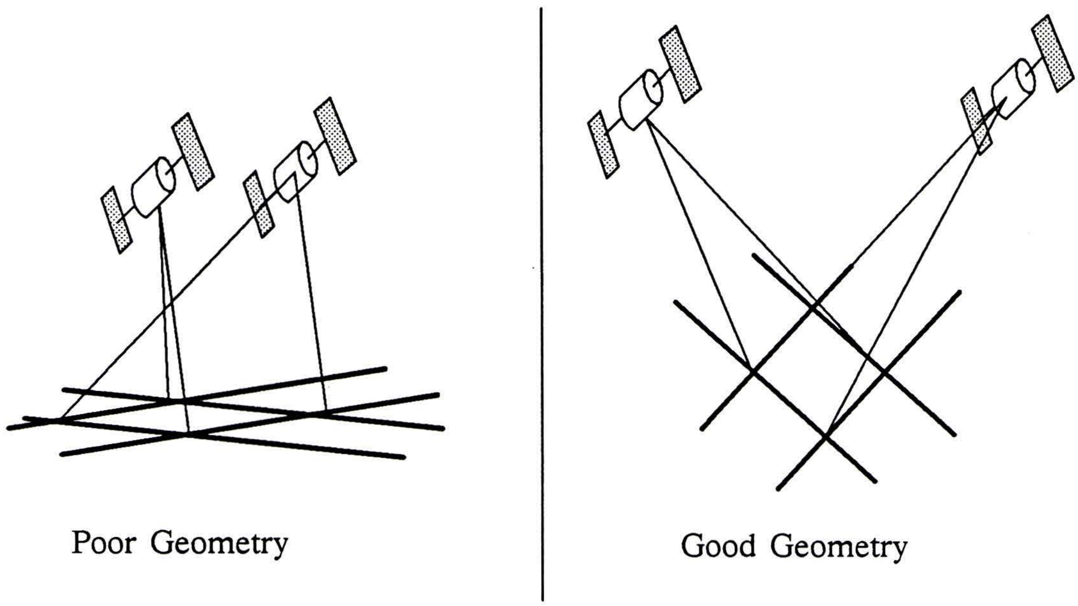


Figure 2.15. Geometric Dilution of Precision

Table 2.4. Open Loop Pointing Errors

Source	Value (μrads)	Type
Spatial Position	0.3	Random
Attitude	0.6	Random
Mounting Error	1.2	Static
Attitude Control	0.2	Random
Gimbal Control	0.1	Random
Total	2.4	($\sum R$)

reference [6]. We note that even with current technology tracking errors less than 1 micro-radian are attainable.

Once a satellite has stopped communicating and the cross-link has been disconnected there may be a period during which the satellite is in communication isolation. Assuming that no broadcast facility is available, the satellite must rely on positional information regarding other satellites to permit antenna re-acquisition. Providing the period is short, a satellite may use orbital mechanics to predict the position of another satellite's ephemeris. The largest relative angular velocity of a satellite is 200 micro-radians/sec [6] and may be known to within a fractional error of 10^{-4}. Therefore the error-buildup in open loop tracking is given by

$$\text{Open Loop Tracking Error } = 2 \times 10^{-8} t$$

where t is the time of open loop tracking. So, for 100 seconds the tracking error is 2 micro-radians, which is smaller than the diffraction-limited resolution of the optics, and hence reacquisition will be successful.

If the period of communication isolation is extended beyond this period, several additional factors require consideration:

- The time accuracy of the onboard clocks may drift.
- What is the computation required to accurately calculate the position of the destination satellite using accurate iterative techniques?
- Orbital Tolerance–How effective is the selected representation of satellite motion over time?
- Degradation–What happens to the representation as time progresses? Do the computations fail completely?

Clearly the selection of the hardware and software will greatly affect the performance of satellite reacquisition.

Spatial acquisition

In the same way that targeting addresses the problem of pointing the transmitting antenna toward the destination, so spatial acquisition addresses the problem of locking the receiver antenna to the incoming beam. Assume that two satellites are separated by a distance L; then in Figure 2.16, Ω_{u1} is the solid angle representing uncertainty that satellite 1 has on the position of satellite 2 and Ω_{r1} is the resolution angle of satellite 1. In two-way acquisition both stations possess a transmitter and receiver. Typically one satellite will transmit with a beam sufficiently wide to cover the pointing errors and the second satellite will scan its uncertainty field to locate the transmitted beam. After acquisition the receiver transmits the arrival direction of the original beam. The important aspect of acquisition is the search performed by the receiver over the uncertainty angle Ω_u.

Simulated acquisition times

There are two classes of illumination strategy, parallel and sequential illumination. Similarily there are two classes of receiving strategy, parallel and serial receivers. Without looking at the differences in detail, it should be noticed that in parallel illumination the entire uncertainty region is illuminated while in sequential illumination the region is scanned with a small width beam. The results from a simulation conducted by the Lincom Corporation [6] suggest that sequential scanning of the uncertainty zone improves performance. Similarily parallel receivers make use of an array of detectors (simulation results cover up to 10^4 detectors). These are significantly better than the sequential receivers; however, they are difficult to manufacture. From the results [6] the best strategy appears to be sequential illumination,

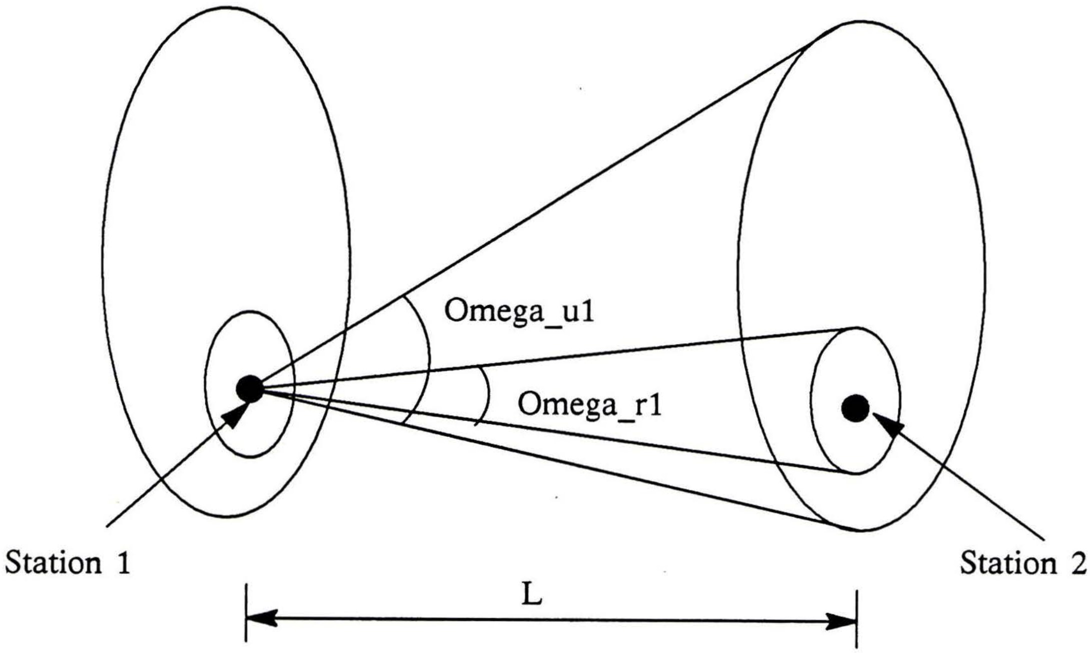


Figure 2.16. Two-way Acquisition Geometry

and photomultiplied direct detectors which perform parallel detection. A graph of the most favorable acquisition system determined in the simulation is given for long and short links in Figures 2.17 and 2.18. Since the transmitter beam is significantly smaller than the uncertainty region, the region may be considered as a number of subregions. Notice from the graph that the acquisition time varies from hundredths of a second to hundreds of seconds. This is an important parameter when analyzing the link assignment problem.

2.4.5 Summary–Network Predictability

This section has introduced the mathematics associated with orbital calculations to demonstrate that a system of satellites in orbit will behave in a predictable manner. Moreover the tools exist to keep track of such a system and are well understood. In addition, given that retargetable laser communication links are feasible and that

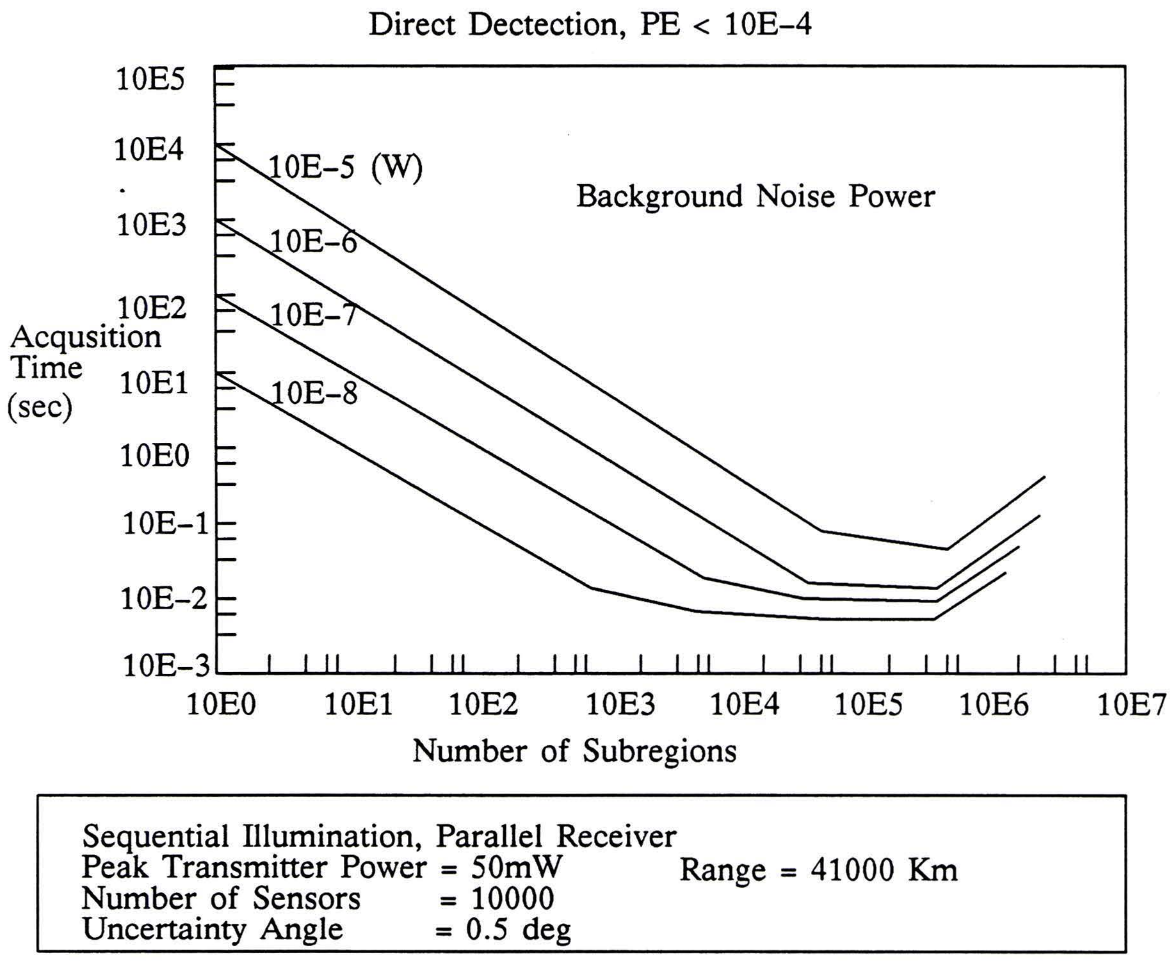


Figure 2.17. Acquisition Time vs. Number of Subregions for Different Background Noise Powers, Long Link

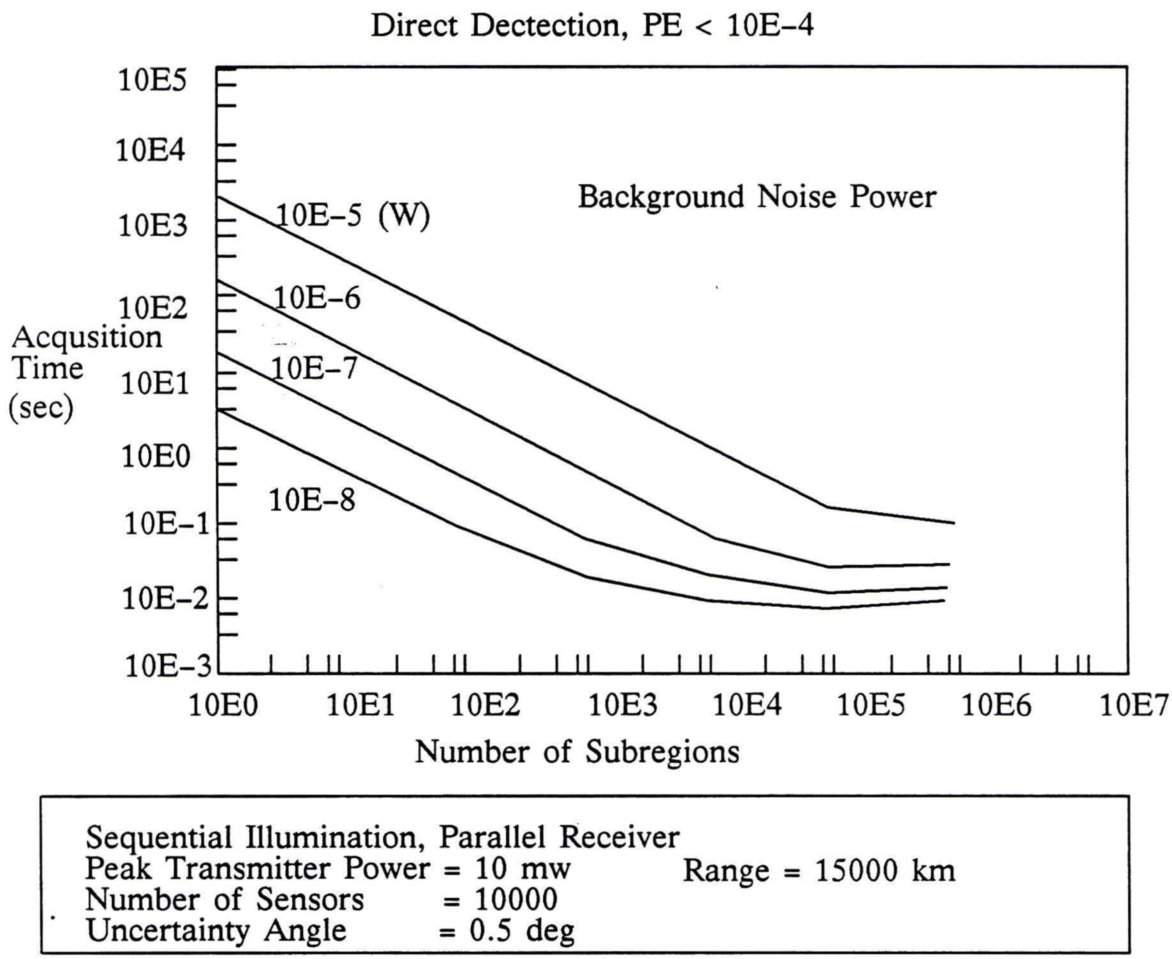


Figure 2.18. Acquisition Time vs. Number of Subregions for Different Background Noise Powers, Short Link

signal attenuation and laser scintillation problems can be overcome, the concept of intersatellite communication is reasonable.

2.5 Proposed Satellite Networks

Until recently there has been little research directed specifically to satellite communication networks. An exception was work done by Brayer [7] in the early 80's supported by the US Air Force. [2] In Section 2.5.1 the satellite topology selected and the routing algorithm used are introduced and discussed. More recently, since the SDI feasibility study into a Ballistic Missile Defense (BMD) system, there have been several government-sponsored research contracts to study satellite communications. Although many of these are classified, and hence unobtainable, one branch of the SDI research program, Innovative Systems and Technology (IST), has been mandated as non-classified. Within this program the Harris Corporation, using a Naval Reseach Laboratory (NRL) model, has developed both link assignment and routing algorithms specifically for the spaced-based fraction of the SDI BMD. In section 2.5.2 a general outline of the Harris research effort is presented and discussed.

2.5.1 Brayer's Low Altitude Satellite Network

In his paper entitled "Packet Switching for Mobile Earth Stations Via Low-Orbit Satellite Network," Brayer describes a distributed control system which dynamically adapts to a changing network environment. The network is intended to permit point-to-point communications between any two places on the earth. The essential aspects of the systems are

- The network continues to operate as nodes enter and leave the network. Thus the network is never brought down for repairs.

[2] US Air Force Electronic System Division under Contract F19628-82-C-0001

- The network is survivable in the sense that there are no key master control stations, the failure of which would cause the network to fail.

- The routing approach provides a large number of alternate paths in a mesh network. In addition the algorithm is such that all nodes act autonomously; thus no node is dependent on any other for control information.

The author divides the network into two sections, the satellite network (which in the terminology of this dissertation includes both the satellite topology, and the link assignment) and the routing concept (i.e. the routing algorithm).

Satellite network

Brayer discusses a very specific satellite architecture. He proposes multiple rings of low altitude (1000-4000 miles) satellites placed in East-to-West orbits at inclined angles of 50° to 60°. The orbits are to be precessed about the equator at separations of 15° to 30°. Each ring has 6 to 12 satellites per orbital ring. With this configuration Brayer states that the full coverage of the air and sea lanes is ensured at all times, with increased coverage toward the poles.

For communications Brayer specifies 10 antennae per satellite (5 transmit antennae and 5 receive antennae). Two antennae point down to the earth (one up-link one down-link), two point forward in the ring, two backward in the ring, two point to the leading ring, and two to the trailing ring. These provide cross-links for satellites both within and between rings.

The following points are noted about the evaluation of Brayer's "Satellite Network":

- From the research conducted for this dissertation, there are many different possible topologies; often these topologies have particular characteristics associated with them which govern the potential connectivity of the system, the propagation delay and the ground coverage. There is no justification of Brayer's selected topology; neither is there any suggestion that the choice is optimal for any objective function, or indeed that there is an intention to be optimal, or that there is even an objective!

- The selection of the cross-links again appears to be arbitrary, without any attempt to make use of the changing topology or network characteristics. In fact a major section of this dissertation is directed at the link assignment problem, a problem that Brayer has completely ignored! We also note that more recent attempts in satellite network communications spend considerable effort in selecting the links to be used (for example the Harris Satellite Network).

- In both the topology selection and the link assignment, the approach by Brayer appears to be an effort to present a very simple, regular structure which may be easily visualized, even though there is no evidence that this selection is justified on any technical merits. We note however, that because the model is simple, it therefore permits the immediate design of a routing algorithm without further attention to the physical side of the problem.

Routing concept

In Brayer's algorithm when a message is received by a node there are two possibilities: 1) a route to the destination is known; 2) the route to the destination is unknown. When the destination is unknown the process of *addressee finding* takes place; when the destination is known, *message routing* is used. Routing is based on shortest path distance between source and destination, where distance is determined by hop count. At startup each node transmits a startup message which identifies the originator (e.g. a node transmits "I am node xx"). From this all adjacent nodes may be identified. Once this startup procedure has been performed the node is ready to engage in communications.

Addressee finding. Assuming that the route to the destination of the message is unknown; the node appends its own ID number to a list which is transmitted along with the original message to some random adjacent node. This sequence is repeated until finally a node is reached which has a route to the destination. Once the destination is reached, an acknowledgement is sent to the originator. This acknowledgement is used by intermediate nodes to update the routing tables associated with that destination. Should a node receive a message which has already passed through it during addressee finding, the message is ignored.

It may be the case that the message fails to reach the destination, perhaps due to node failure. Should this occur, the originator will eventually timeout and try again. If no acknowledgement is received after a predetermined number of attempts; the node notifies the originating sender that the message could not be delivered. It is possible that the sender and the first satellite node are no longer in direct communication (e.g. a plane moving away from a satellite). If this is the case the sender is ensured notification via a special update-table-of-addressees message.

After a short time it should be the case that routing tables between most source and destination nodes will exist; once this stable situation occurs, only a few occasions will merit using the addressee-finding algorithm. It should be noted that, when senders (and receivers) move from the node they were originally connected to, this information is transferred via the update-table-of-addressees message, which is passed from the old originating node to the new one and passed back to the destination to establish new routing entries.

One feature of this system is that congestion of the network is avoided by permitting only addressee-finding *paths* to exist (as opposed to trees); this keeps network traffic at a minimum at the expense of quickly finding the destination.

Message routing. If the route to a destination is known, the algorithm simply passes the message on to the next link in the route. As previously mentioned, routes are maintained by hop count (see PRNET, section 2.3.4) and the routing strategy used by Brayer is that of shortest path. Every message that goes through the network has appended to it the route taken. This information is used to keep the routing tables constantly updated.

Congestion control is implemented by a threshold on node utilization. Once the threshold has been exceeded the node will throttle traffic by rejecting new input traffic from users and discarding messages in transit.

The following points are noted about the "routing concept" in Brayer's network:

- The method employed for establishing point-to-point links is unusual in that only a single path is sought between the source and destination. This approach could lead to long paths between nodes initially.

- The routing is directed by shortest hop. This makes no use of the propagation delay between satellites. In fact, because of the topology of the system, nodes could be geographically close yet not in direct communication and be required to pass messages through several unnecessary hops.

To summarize, the algorithm developed by Brayer, while simple, makes little attempt to derive the maximum potential from the network. In particular none of the special features associated with satellite networks has been used and several of the techniques do not appear to be as robust as more traditional approaches (e.g. modified PRNET).

2.5.2 Harris Corp.'s NRL Model Satellite Network

The second network model reviewed is of research conducted by the Naval Research Laboratory and its contractor the Harris Corporation Government Special Programs Operations. The objective of the project was to design a communications network to support battle management functions for the Strategic Defense Initiative (SDI). The model assumes the existence of space-based weapons, sensors, and managers which participate in distributed communications. As with Brayer's network the model is quite specific and includes approximate numbers and trajectories of the satellite nodes. In keeping with the rest of this dissertation, discussion of the network is subdivided into three sections: the satellite topology, the link assignment algorithm, and finally the routing algorithm.

It is worth noting at the outset that the NRL project has paid particular attention to the unique characteristics associated with satellite networks and has attempted to use them to the advantage of the network. Additionally the architecture has many features in common with Baker's Intra Task Force Network (see section 2.3.5); however the NRL group points out the following differences:

- The SDI system may have thousands of nodes; while the HF-ITFN typically has less than a hundred.
- All nodes in the HF-ITFN are equal and may serve as cluster controllers or cluster gateways. In the HRL SDI model nodes have specific roles, simplifying controller selection procedures.

A description of the satellite topology

The satellite system is arranged into three shells. Each shell (or layer) plays a specific role in the defense system. At the lowest altitude ($\approx$ 500 km) a large number of weapon satellites is arranged into several inclined circular orbits. The second layer at a slightly higher altitude ($\approx$ 1000 km) consists 36 sensor satellites in 6 circular orbits. Finally the third layer is positioned in a far earth orbit ($\approx$ 50000 km) and consists of 12 battle management satellites in 2 circular orbits. Coupled with this physical model is a communication hierarchy. The hierarchy scheme consists of two communication levels. The first level is a backbone network which is composed of point-to-point links between the managers and sensors. The second level is a broadcast system between the backbone complex and the weapons satellites. It is intended that each of the manager satellites be responsible for a number of sensors and weapons geographically beneath it. A coordinated effort by the managers ensures that allocation of resources is timely and efficient. The NRL/Harris researchers believe that the backbone network is sufficiently small to permit frequent reorganization of the links in an effort to keep the system optimal. For this reason work to date has been directed on only backbone link assignment and routing. The broadcast layer has yet to be developed.

The following points are made regarding the Harris satellite topology.

- As with Brayer's work, the topology is detailed. However justification of the topology is not clearly described. The paper asserts that the topology was selected by computer simulation. The details of this simulation, in particular the constraints and objective function of the simulation, are not discussed.

- The topology is interesting because of a close correlation between the functionality of the different layers and the hierarchy of the network. This has elements in common with PRNET (see section 2.3.4).

A description of the link assignment algorithm

The function of the link assignment algorithm is to determine the most effective subset of links to use, given the links possible and antenna restrictions. The NRL/Harris researchers define the algorithm requirements as follows.

- Eliminate fragmentation of nodes into isolated islands (i.e. maintain a connected graph).

- Improve connectivity within islands. In this context improved network connectivity refers to node disjoint paths between all pairs in the island.

As with the ITFN, link assignment is organized into a repeated series of epochs. Each epoch consists of N iteration steps (where N is the number of nodes in the system) with each step in two frames, 1) an add links frame, and 2) a break and add links frame. During the add links frame each node that has unused antennae attempts to create a new link to another node that also possesses an unused antenna provided the link will remain intact for the next iteration. Clearly the node connectivity will be improved by this step. During the break and add links frame there are no uncommmited antennae, therefore links must be disconnected before an improvement is possible. To permit an orderly transition during this frame the node whose ID is the same as the current frame number is used to control the link selection within and between its island. Thus after an epoch, each node will have had the opportunity to improve the network once.

A detailed description of the add links structure is given in Figure 2.19. First, links which are scheduled to fail according to orbital mechanics are disconnected. Second, the join island procedure is activated. Using the broadcast channel and local links each node transmits details of the antennae it has uncommitted. After a suitable interval each node evaluates the information it has received and determines the most desirable links that it can establish. To coordinate the actual selection of links to be assigned the nodes are divided into two classes, inviters and acceptors. The inviters suggest a link assignment to the acceptors which may be either accepted or rejected. The results of the selection are sent back to the inviters who decide with which nodes to establish links. Third, the topology updates are transmitted to the entire network via limited flooding for improved reliability. Fourth, the connectivity of the entire network is improved by checking for node disjoint paths between a sample of randomly selected nodes. The decisions regarding the selection of links are arrived at in a fashion similar to step 3. Finally, the update topology procedure is re-excuted to ensure consistency of the selected links.

The break and add links frame is similar to the add links frame. The same basic procedures, join islands, undate topology, and improve connectivity are executed. The difference between the frames is that during the break and add links frame, links may be readjusted to increase the size or connectivity of islands (recall that only one node in each iteration controls this frame, thus the effects are local) without endangering the network. In particular during the join islands procedure the following steps are undertaken:

- The *broadcast availability* procedure. In this all nodes not in the control node's island send information regarding their uncommitted links.

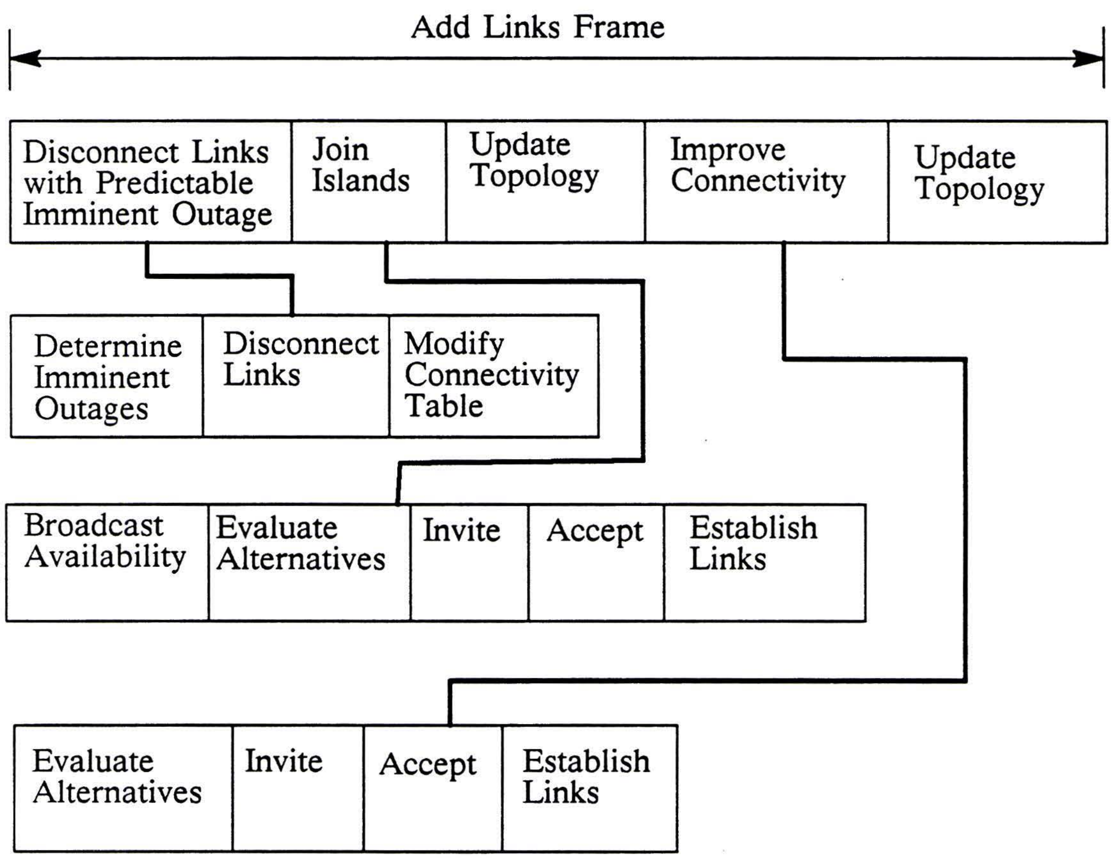


Figure 2.19. Add Links Frame

- The *evaluate alternatives* procedure, where the control node and the nodes within its island evaluate the consequences of modifying links.

- The *select change* procedure. In this the control node decides the best link reassignment policy.

- The *command nodes* procedure. In this the nodes are informed of the decisions made by the control node.

- The *break and establish links* procedure. During this the nodes execute the instructions given to them.

We note that global coordination is required in this frame, necessitating an accurate global clock. One aspect of the NRL/Harris link assignment algorithm particularly liked is the use of orbital mechanics to predict the links which will be disconnected in impending frames, reassigning them elsewhere.

A description of the routing algorithm

The adaptive routing algorithm selected by the NRL/Harris group has a distributed implementation and provides the following general routing functions:

- Generation of routing tables which provide the best routes between the source and destination.

- The algorithm is highly adaptable to changing topology, yet stable to transients in traffic loads.

- The algorithm manages traffic load balancing.

- The algorithm incorporates congestion control.

Coordination between topology updates (link assignment) and routing updates is essential for smooth network operations. With this in mind, routing updates occur immediately after topology updates and are effective from the point at which they are completed until the point at which the next set of routing updates has been completed. Thus the routing table updates occur periodically and last for one iteration of an epoch.

The routing algorithm actually selects several different routing paths to each destination. These paths are maximally node disjoint and are used for load splitting and survivability. The principal path is chosen on the basis of shortest distance, using Dijkstra's Algorithm (see section 2.2.2). Other feasible routes are calculated on the basis of shortest distances to connected neighbors of the destination. In order to achieve a degree of load balancing in the network, traffic is sent along both the primary and alternate feasible routes. The division of traffic to these different paths is based upon routing variables derived from path metrics. The path metric used by NRL/Harris in their adaptive routing algorithm is given by:

$$D_{ik}(f_{ik}) = \frac{[ad_{ik} + 1 + b(f_{ik}/C_{ik})^2]}{C_{ik}} \tag{2.16}$$

where

f_{ik}	is the traffic flow between nodes i and k.
D_{ik}	is the delay metric between nodes i and k, which is a function of the traffic flow f_{ik}.
C_{ik}	is the capacity between the links (in bits or packets per second).
d_{ik}	propagation delay between nodes i and k.
a, b	constants.

The last term in equation 2.16 is used to approximate the queueing delay in a node. The form is used in preference to the more usual $(C_{ik} - f_{ik})^{-1}$ to make the metric more sensitive at lower loads. The constants are used to adjust the relative magnitude of the different terms.

In order to increase the stability of the metric over short time intervals (e.g. transient loads) the values of f_{ik} are averaged over several update intervals i.e.

$$\overline{f_{ik}} = \beta f_{ik} + (1 - \beta)\overline{f_{ik}^{old}} \tag{2.17}$$

where β is the averaging parameter and $\overline{f_{ik}}$ is the average flow based on the most recent value of f_{ik} and the previous averaged value $\overline{f_{ik}^{old}}$.

The principal total path metric M_{ik} between two nodes i and k is the sum of the metrics calculated from equation 2.16. While M_{ik} defines the principal, the alternate feasible total path metrics are given by the following

$$W_{ik}(m) = D_{im}(f_{im}) + M_{ij} \tag{2.18}$$

where m is an adjacent node to i, and D_{im} is the metric for the one hop path from node i to node m.

Load splitting between the principal path and the feasible paths is achieved by a routing variable $\phi_{ij}(m_k)$ which specifies the fraction of the total traffic from node i to node j that will be routed over link k to node m. The routing variable is normalized to unity i.e.

$$\sum_{k=1}^{n} \phi_{ik}(m_k) = 1 \tag{2.19}$$

There are many possible heuristics for allocating the traffic load; the Harris Corporation has set the routing variable as an inverse function of the path metric. This is justified by noting that as the path metric increases, the quality of the path decreases (i.e. it becomes less desirable to use). Specifically the formula they use is

$$\phi_{ij}(m_k) = \frac{1/W_{ij}(m_k)}{\sum_k 1/W_{ij} m_k} \tag{2.20}$$

where the summation is over all feasible paths (i.e. all links into i) and the denominator normalizes the result so that the routing variable satisfies equation 2.19.

As with the link assignment algorithm, the routing algorithm has been carefully designed with military objectives in mind (*viz.* survivability, reliability, multiple redundant paths etc.); therefore attaching too much importance to performance criteria would be unreasonable. Even so, notice that the principal paths are computed based on shortest path by distance, and therefore one would expect the algorithm to perform well. It is noted, in conclusion, that computer simulations of this model have been performed by Harris although the results have not yet been published.

CHAPTER 3
SATELLITE TOPOLOGIES

The first step toward studying satellite communications is a thorough understanding of the inter-relationship between satellite orbits and the properties of the communication links between them.

In the first section of this chapter a simple model for satellite orbitals is introduced. Upon this model are imposed constraints restricting the configurations to those of interest. These constraints are based on an original assumption of the dissertation which stipulates that the analysis be for global satellite communications. Problems encountered when discussing this model are survivability and coverage. These are reviewed, and the inter-relationship between them investigated.

In the second section a simulation package is introduced which may be used to study the coverage of satellite footprints on the earth. This tool is particularly useful in view of the relationship between coverage and connectivity derived in section 3.1.3.

Finally the third section introduces a restricted model of the general satellite topology from which several interesting properties are examined. In particular a graph theoretical model is developed; and a formula for the minimum altitude for complete coverage is given. The ground coverage simulation is used to verify this formula for several sample topologies.

3.1 Analyzing Satellite Topologies

3.1.1 Satellite Topology Model

A *satellite topology* is defined to be a simple model of real satellite systems with the following properties: [1]

- There are S shells of satellites with each shell at a distinct altitude ($S \geq 1$).
- Each shell has P_s planes of satellites ($1 \leq s \leq S,\ P_s > 1$).
- Each plane has $N_{p,s}$ equally spaced satellites ($1 < p \leq P_s,\ 1 < s \leq S,\ N_{p,s} > 2$).
- The topology satisfies some criteria (e.g. complete coverage of a spherical earth).

In addition several simplifying assumptions are made which greatly reduce the complexity of the orbital analysis. These simplifications will not cause significant deviations between the model and the real world system. The assumptions are:

- The earth is assumed to be a solid sphere with a point mass at the center of the sphere.
- Satellites are assumed to be point masses traveling in circular orbits around the earth and are of negligible mass compared to the earth.
- Each satellite is assumed to have a negligible effect on another's orbit (i.e. we may use equations for the two-body problem).
- The satellites are not subject to the perturbative accelerations and random perturberances outlined in the previous chapter.

[1] For more details on Low Altitude Satellite Networks see [40,41,42].

- Communications between satellites depends on satellite-to-satellite visibility only.

Justification for these assumptions in the model may be found in the previous chapter, section 2.4.1. Despite these simplifications the model is still highly flexible and in future sections it will be additionally restricted. In particular the model supports multiple shells; and within each shell multiple planes at different orientations to each other. Observe from equation 2.4 that the period of a particular satellite depends only on the altitude of the shell in which it resides. This implies that for S shells, if the period of each shell is T_i, $i = 1 \cdots S$, the period of the complete system will be the lowest common multiple (LCM) of the periods. Notice that given several shells in low altitudes the expected period of the system is large. For example a system with 3 shells at $500, 1000$, and 1500 km would have a period given by:

$$
\begin{aligned}
T &= LCM(T_1, T_2, T_3) \\
T &= LCM\left(2\pi\frac{6878^{3/2}}{\sqrt{\mu}}, 2\pi\frac{7378^{3/2}}{\sqrt{\mu}}, 2\pi\frac{7878^{3/2}}{\sqrt{\mu}}\right) \\
T &= LCM(94, 105, 116) \text{ minutes} \\
T &= 1144920 \text{ minutes} \\
T &= 795 \text{ days}
\end{aligned}
$$

Clearly using multiple shells causes problems with the periodicity of the system. These problems become important when closed form solutions to different problems are sought. It should be noted for future reference that including the restriction that only one shell be permitted removes this difficulty from the model.

3.1.2 Constraints for a Legal Model

With a satellite model defined, the next step is to define the constraints which specify a *legal* satellite topology. Although no precise definition for the use of the system has been given, a typical application would be to provide global communications or global sensor analysis. The emphasis here is on the concept of a global system. It is also assumed that each satellite has a small number (k) of laser transmitter/receivers. Each of these may be targeted to another satellite based on the visibility condition (i.e. satellites may communicate if they have an unobstructed view of one another). In addition, since the satellites represent the communications network it is essential that each of the satellites be connected to some other satellite by some communications path for all time. To define these requirements formally the following definitions are made:

Coverage definitions

Definition 3.1.1 *The earth is k-covered $\leftrightarrow$ every point on the surface of the earth is visible by at least k distinct satellites for all time.*

For use in conversational speech the following are also defined:

Definition 3.1.2 *The earth has single coverage $\leftrightarrow$ it is 1-covered*

Definition 3.1.3 *The earth is redundantly covered $\leftrightarrow$ it is k-covered, $k > 1$*

Communications definitions

In the following definitions the concept of a satellite communications graph is defined. First general graph properties are formally introduced.

Definition 3.1.4 *Let $G = (V, E, c)$ be the undirected graph with vertices $v_1, \cdots, v_n \in V$ connected by edges $e_1, \cdots, e_m \in E$, where an edge consists of a pair of vertices and a cost between them ($e_i =< \{v_x, v_y\}, c_{xy} > , c_{xy} \geq 0$).*

Definition 3.1.5 *A sequence is an ordered set of vertices* $< v_a, \cdots, v_f >$ *such that adjacent vertices in the set are connected by an edge in* E.

Definition 3.1.6 *A cycle is a sequence of two or more vertices* $< v_a, \cdots, v_f, v_a >$ *with the first and last elements of the set the same and all others different.*

Definition 3.1.7 *A path* p *between vertices* v *and* w *is a sequence of vertices beginning at* v *and ending at* w *containing no cycles.*

Definition 3.1.8 *A minimum cost path* p_{min} *between vertices* v *and* w *is a path where the sum of the edge costs in the sequence is a minimum for all possible paths from* v *to* w. *The cost associated with this path is given by* $mincost(v, w)$.

Definition 3.1.9 *A connected graph is a graph in which there exists a path from each vertex to every other vertex.*

Definition 3.1.10 *A biconnected graph is a graph in which there are two node disjoint paths between all source destination pairs.*

Definition 3.1.11 *An* r*-node-connected graph is one in which there are* r *node disjoint paths between all source destination pairs.*

Minimum constraints for a legal topology

As mentioned in the introduction the satellite topologies under review are those which will permit global communications between all source destination pairs on the earth's surface for all time. With this objective a legal satellite topology is defined.

Definition 3.1.12 *A legal satellite topology is one which possesses the following properties:*

- *The earth is at least 1-covered.*
- *The graph G is connected at all times.*

In future chapters additional constraints will be imposed on the system; e.g. a restriction on the number of communication links.

3.1.3 Survivability and Connectivity

When dealing with actual satellite topologies the specification that the topology be *legal* may be insufficient. This may arise from requiring that the topology remain legal after an anticipated number of failed satellites, or from a desire to offer redundant coverage of the ground. In terms of this study these topologies are defined as being in some way survivable. Considering the question further raises a number of issues which suggest difficulties when dealing with this concept.

One would intuitively believe that a system with 100 satellites could be more survivable than one with 10 satellites. Assuming that the system with 10 satellites is *legal* one could replicate the system 10 times. One would expect that this new system could continue to operate after the failure of several nodes. However a system with 100 satellites is not necessarily more survivable. For example, consider a system with a large percentage of its satellites on one plane. This system may be highly survivable in regions below this plane; but away from this area the failure of one satellite could cause an uncovered region violating the coverage condition.

Survivability over time

The factor of time is another problem when considering survivability of a satellite system. Consider, is a system which covers 95% of the ground all the time survivable? Conversely, is a system which covers the entire ground for 95% of the time survivable? Another point to be made is the importance of maintaining a connected graph. If the network becomes disjoint should it still be considered useful? Initially one might think not; however, consider the case that one of the subgraphs spans a set of satellites which covers the earth for all time. While from a visibility point of view the satellites could form a single spanning graph, with only a few antennae it is possible that the satellites could separate into unconnected groups. Although global communications would still be possible disconnected sub-graphs are undesirable.

From this discussion it would appear that survivability is a difficult concept to isolate with many different view points. In order to proceed we make a simple definition of α-survivability. The definition is strict in the sense that it is probably more restrictive than one would use in practice. The justification for this is that the properties associated with this notion of survivability are easily measurable and relate closely with the simulation described later.

<u>Definition 3.1.13</u> *We say a system is α-survivable if there are α-node disjoint communication paths between all pairs of points on the earth for all time. This implies the following:*

- *All points on the shell are α-covered for all time (coverage condition).*
- *The satellite communications network is α-node-connected (communications condition).*

Both of these conditions have to be met simultaneously by a system of satellites for it to be considered survivable. Notice that both of the conditions are integer functions (coverage and connectivity both range from 0 upwards); this plays an important role later on. Notice also that 1-survivability implies a legal topology.

Derived properties

For two systems of satellites x, y, and a survivability function α then $\forall x, y$ one of the following must hold:

- $\alpha(x) > \alpha(y) \rightarrow x$ is more survivable then y.
- $\alpha(x) = \alpha(y) \rightarrow x$ is as survivable as y.
- $\alpha(x) < \alpha(y) \rightarrow y$ is more survivable then x.

Notice one is able to infer qualitative information from α, i.e., a value of α implies some characteristic of the network. With these points considered, the formal relation for α is given.

Let

$G = (V, E(t))$	be a satellite system at time t ($t \in [0 \cdots T]$).
$E(t)$	be the edge set of visible communication links in G.
Coverage(x)	be the minimum coverage of any point on the earth by the system x.
SatsUsed(x)	be the satellites used in Coverage(x).
KRegSubSet(x)	be the selected edge set $E'(t) \in E(t)$ of communication links used by system x.
NConnectivity(x)	be the node connectivity of the edge set x.

then

$$\alpha = \min \forall t \ (\text{NConnectivity}(\text{KRegSubSet}(\ \text{SatsUsed}(G))),\text{Coverage}(G))$$

Notes:

- It is assumed that nodes (rather than links) will be removed from the satellite system. This assumption is based on the difficulty of intercepting the communication links.

- Since the edge set used in communications is a subset of the possible edges, the survivability of the system includes this requirement.

This definition is complicated in that two different properties are expressed; however in the next section the situation is simplified.

3.1.4 Relationship Between Survivability and Connectivity

Theorem 3.1.1 If a system is c-covered, then it has at least c-node connectivity i.e. c-coverage $\rightarrow$ c-node connectivity.

Proof of Theorem 3.1.1 by construction. Pick two arbitrary points (x, y) on the surface of the earth. By assumption these are both c-covered. Construct a great circle passing through x and y as shown in Figure 3.1. Label the circle as two arcs $\overline{xy}$ and $\underline{xy}$ where $\overline{xy} \geq \underline{xy}$. Recall that the surface is c-covered by satellites; thus there is a communication graph $G = (V, E)$ of vertices (satellites, and the source destination points) and edges (communication links, determined by visibility). There are two cases to consider.

1. Both points are visible to the same c satellites.
2. There are satellites visible to x and not to y and vice-versa.

1) by visibility, the c satellites visible to x are also visible to one another. Thus the satellite subgraph involving these satellites is completely connected. By assumption these satellites are also visible to y. Thus c disjoint paths may be constructed from

x to y by selecting the paths that emanate from x, go to one satellite and then down to y.

2) the following shows that there are at least $(c-1)$ node disjoint paths between x and y via $\underline{xy}$ and there is an additional node disjoint path between x and y along $\overline{xy}$. It is assumed that the source and destination points are not subject to removal.

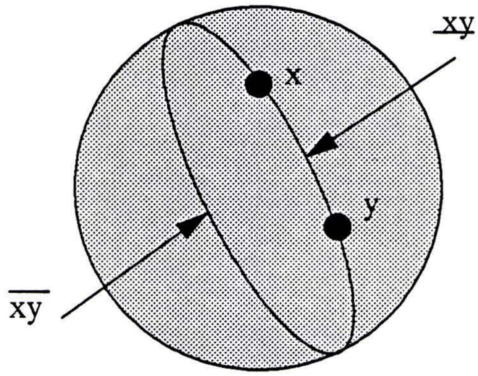


Figure 3.1. Paths between x and y

$\underline{c-1 \text{ paths from } x \text{ to } y \text{ using } \underline{xy}.}$ Consider the worst case, when there are exactly c satellites visible. Start the $c-1$ disjoint paths from x to the $c-1$ satellites closest to y. Proceeding from x along $\underline{xy}$ toward y, let the point on the Earth at which a satellite visible to x moves out of visibility be z_1. At the instant the satellite moves out of visibility a new satellite must become visible to z_1, or z_1 would fail to be c-covered. Again by visibility, since each satellite is visible to z_1, the satellites are all visible to one another. Connect the current furthest (from y) satellite to the new satellite. Repeat this process, moving along $\underline{xy}$ until the point y is reached. Since the satellites visible to y have paths emanating from x and the paths are disjoint (paths were extended to new nodes only, thus the only points shared are the source and destination) there are $c - 1$-node disjoint paths from x to y. This procedure is described graphically in Figure 3.2.

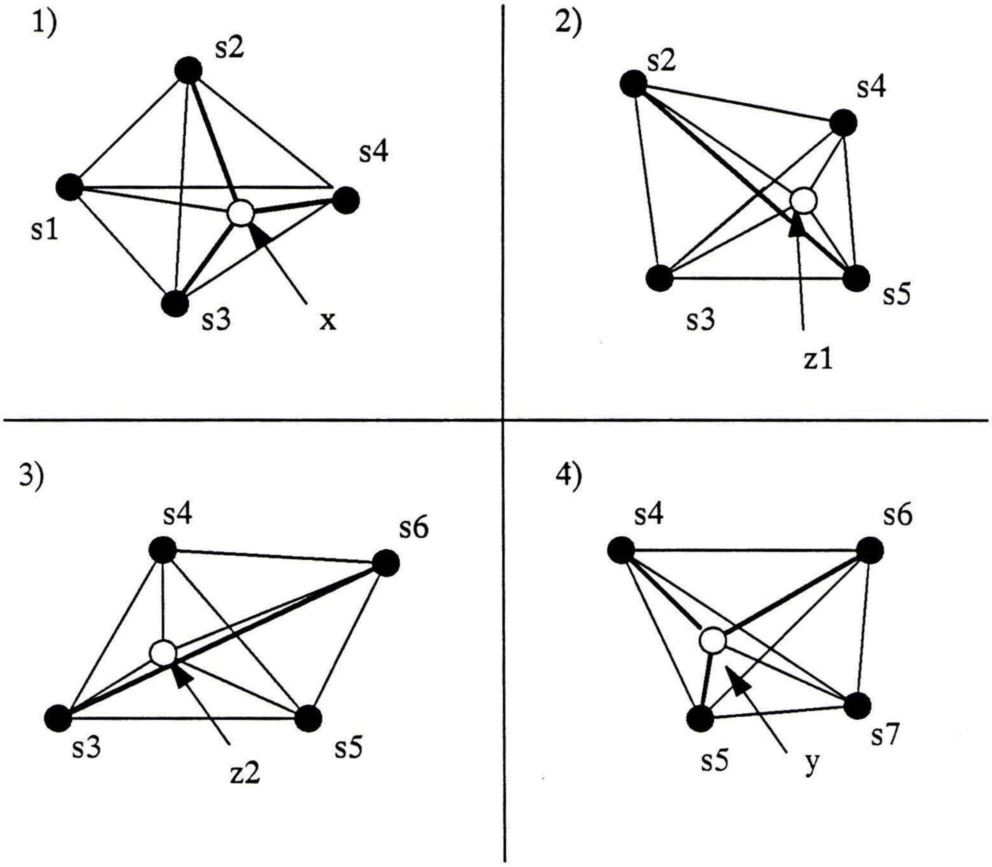


Figure 3.2. Construction of Node Disjoint Paths

<u>One path from x to y using $\overline{xy}$.</u> Consider the path $\overline{xy}$ between x and y. Again starting from x pick the satellite s which is furthest from y (i.e. the one which was not used in the previous section). Again select a point z'_1 which lies on $\overline{xy}$ and is just visible to s. By coverage there is another satellite closer (along $\overline{xy}$ to y which is visible to s). Connect these two satellites and continue to another point z'_2, etc. Eventually there will be a point z'_i which is is visible to the satellite visible to y yet

not used in the previous section. This satellite was the furthest from x along the $\underline{xy}$ path. It is known that the satellites used in this path were previously unused since the path goes in a counter direction to the other paths.

Thus along the $\underline{xy}$ route there are $c-1$ paths, and along the $\overline{xy}$ route there is 1 path. Thus the total number of node disjoint paths is c; therefore the system is c-node connected. □

Corollary 3.1.1 *Since c-coverage → c-node connectivity. The definition for survivability may be reduced to the following:*

Definition 3.1.14 *A system is α-survivable if* $\alpha = \min(\mathrm{Coverage}(S(t)))\,\forall t$

3.2 The Ground Coverage Simulation

To verify that a satellite topology (in the general $S, P_s, N_{p,s}$ model) satisfies the criteria of coverage it was found necessary to simulate the satellite system over the period and extract the pertinent data. The requirements for the simulation are as follows:

- Provide the value of c-coverage.
- Provide the coverage distribution over the earth.
- Offer a statistical summary of the surface coverage to permit comparison between different topologies.

3.2.1 $D_{max}/2$–Satellite Tangential Arc Distance

One value which appears frequently in the following sections is the satellite tangential arc distance. Consider the case of a single satellite in (circular) orbit about the earth, as shown in Figure 3.3. The distance between the satellite and the farthest point visible to is given by the tangent from the earth to the satellite. We call this distance $D_{max}/2$, a function of the earth radius (R_e) and altitude (A) only. By Pythagoras:

$$\begin{aligned}(R_e + A)^2 &= (D_{max}/2)^2 + R_e^2 \\ (D_{max}/2)^2 &= A^2 + 2R_eA \\ (D_{max}/2)^2 &= A^2\left(1 + \frac{2R_e}{A}\right) \\ D_{max}/2 &= A\sqrt{1 + \frac{2R_e}{A}} \qquad (3.1)\end{aligned}$$

Hence a satellite covers a region on the earth with points on the circumference $D_{max}/2$ away from the satellite. The reason for using the label $D_{max}/2$ will become clear in the (N_p, N_s) model to be discussed later.

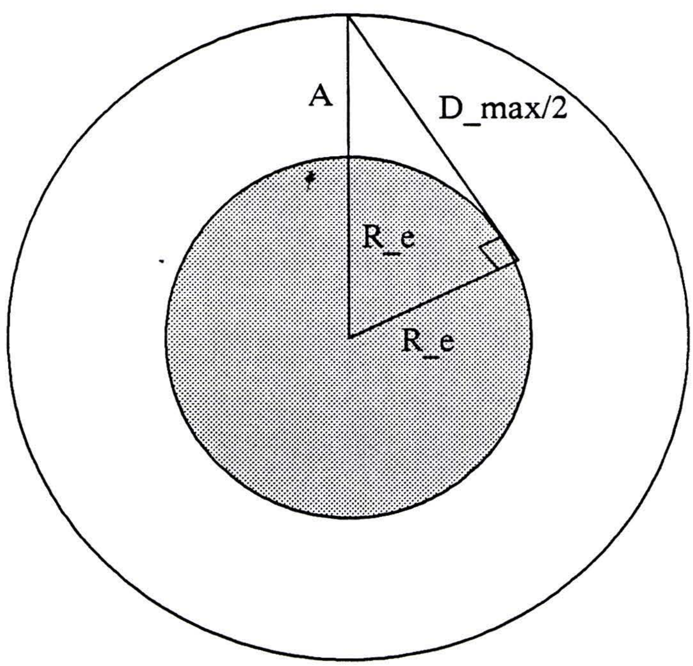


Figure 3.3. $D_{max}/2$ Satellite Arc Distance

3.2.2 Design Requirements

In order to satisfy the simulation requirements the design specifications were as follows:

- Compute ground coverage for a satellite system with up to 3 shells. Each shell to contain at most 6 planes. Each plane to possess at most 9 equally spaced satellites.
- The shell altitudes are to be in the range 500km to 50000km.
- The orbital planes are to be at any azimuth and elevation.
- The satellites are assumed to possess a sensor/communication beam centered directly below them with a half-beam-width ranging from 0° to $\left(\arccos\left(\frac{D_{max}}{A+R_e}\right)\right)^\circ$, the angle whose line projections are tangent to the earth.
- The ground resolution is defined as 1° azimuth (spanning 360°) by 1° elevation (spanning 179°) giving 64440 grid points. This resolution is considered sufficient for verification of topologies.
- The simulation is to run for some user-specified time period starting with the satellites in some initial position.
- The simulation time increment to be user specified, and an integer number of seconds.
- The rotation of the earth is to be optionally included.

3.2.3 Results to be Recorded

The data to be recorded and results obtained are as follows.

- The simulation is to retain the minimum ground coverage for each grid point over all time, (the minimum value on the grid yields the coverage).

- The simulation is to retain the total ground coverage for each grid point over all time. This may be used to compute the average ground coverage.

- A statistical summary of both the minimum and total ground coverage to provide the user with percent single, double, and triple coverage. A user-defined mean deviation cost function is also used to permit the user to emphasize topology characteristics.

- The minimum and total ground coverage grids may be optionally saved at the end of the simulation for further analysis, or as inputs to plot generation packages (see section 3.2.7).

3.2.4 Ground Coverage Algorithm

The algorithm used in determining the ground coverage is given by Algorithm 3.2.1. In essence for each time quantum the regions covered by the individual satellites are marked; these regions are then compared with a minimum coverage region to see if the coverage has been reduced and are incremented into the total coverage region for average case information. At the completion of the simulation interval the tables of statistics are computed. Finally, both the statistics and the grid information are saved to separate files.

Algorithm 3.2.1 *A measure for network reliability via ground coverage*

1. Initialize

   FOR EACH θ, ϕ

   calculate EarthGrid(x,y,z) coords

   initialize CurrentPlane[θ, ϕ], MinPlane[θ, ϕ]

```
            initialize satellite positions
            calculate D_max
    END

2.  FOR i = 1 TO timelimit DO

            Update Satellite Positions
            FOR each satellite DO BEGIN
                θc = Current satellite θ
                φc = Current satellite φ
                WHILE | SatCoords(x,y,z) - EarthGrid(x,y,z) | < D_max/2
                    DO
                    θc = θc± AngleInc
                    WHILE | SatCoords(x,y,z) - EarthGrid(x,y,z) |
                        < D_max/2 DO
                        Increment CurrentPlane Coverage
                        φc = φc± AngleInc
                    END
                END
            END
            { save the minimum ground coverage for this time step }
            FOR EACH grid point
                IF CurrentPlane(θ, φ) < MinPlane(θ, φ) THEN
                    MinPlane(θ, φ) = CurrentPlane(θ, φ)
            END
    END
```

3.2.5 Implementation of the Ground Coverage Algorothm

The algorithm was implemented in the "FORTRAN" programming language. The code size was 1000 lines and took approximately 3 months to write. The program was verified by a graphic front end, verification by hand of simple topologies, and close agreement with the minimum altitude equation given in section 3.3.2. The machine on which the program was developed was a Gould Powernode 9080 supporting the Unix operating system. The typical execution speed for a single topology was 10 minutes.

3.2.6 Statistical Evaluation of Ground Coverage

The statistics gathered are assumed to be over a single period of the system. Without this assumption many of the values are meaningless. For example the ground covered in less than one period will be dependent on the initial conditions of the system and will reveal little useful information. With this in mind an annotated sample of the statistics generated is presented below.

```
SIMULATION STATISTICS FOR GROUND TRACK PROGRAM

         THIS FILE    SD123A0

NUMBER OF PLATFORMS = 1                                    *see note 1

  THERE ARE 2 PLANES AT HEIGHT  11662 (KM)
  PERIOD IS  24113.79 SECS
  USING   20.70 DEG. BEAM-WIDTH (MAX IS 20.70 DEG.)
      ALPHA        DELTA      NUMBER OF SATS     REMOVED SATS
       0.00       -45.00             3
       0.00        45.00             3

HEIGHT OF THE GROUND PLANE ABOVE EARTH = 0 (KM)
ANGULAR RESOLUTION = 1
IGNORE ANGLES FROM 90 TO 90, AND FROM -90 TO -90           *see note 2
TIME INCREMENT    =   121 (SECS)
TOTAL TIME OF RUN = 24114 (SECS)
NUMBER OF ITERATIONS =  199
EARTH ROTATION = 0.00 RAD/SEC

THE FOLLOWING ARE VALID ONLY FOR TOTAL TIME OF ONE PERIOD

MINIMUM COVERAGE OVER THE ENTIRE PERIOD                    *see note 3
    % OF GRID WITH  0 SATELLITES           =     0.00
    % OF GRID WITH >0 SATELLITES           =   100.00
    % OF GRID WITH >1 SATELLITE            =    42.03
    % OF GRID WITH >2 SATELLITES           =     0.00
    % OF GRID WITH >3 SATELLITES           =     0.00

    MEAN COVERAGE                          =     1.42
    VARIANCE                               =     0.24

    NORMALIZED VARIANCE FROM COST FN.                      *see note 4
       0    1    2    3    4    5    6    7
    20.0 0.0  1.0  2.0  4.0  8.0 16.0 20.0
    NORMALIZED VARIANCE                    =     0.02

AVERAGE COVERAGE PER ITERATION                             *see note 5
```

```
% OF GRID WITH  0 SATELLITES            =    0.00
% OF GRID WITH >0 SATELLITES            =  100.00
% OF GRID WITH >1 SATELLITE             =  100.00
% OF GRID WITH >2 SATELLITES            =   58.18
% OF GRID WITH >3 SATELLITES            =    0.00

MEAN COVERAGE PER ITERATION             =    1.93
VARIANCE                                =    0.13

NORMALIZED VARIANCE FROM COST FN.                      *see note 6
   0   1    2    3    4    5    6    7
20.0 0.0  1.0  2.0  4.0  8.0 16.0 20.0
NORMALIZED VARIANCE                     =    0.04
```

Notes on terms used:

1. Information about the satellite topology under consideration is provided for the user, in particular the topology configuration and simulation parameters. The term "platform" has the same meaning as "shell".

2. One option in the input file permits the user to specify that the latitude be constrained to certain bounds during analysis. This feature permits simulations in which only partial coverage of the earth is required.

3. The *minimum coverage over entire period* represents the minimum coverage for each grid point over the duration of the simulation. The surface presented is not a snapshot at any particular time; rather it is the worst coverage for each point. The statistics (0, 1, 2, 3, and >3 satellites coverage) are therefore the minimum for the system.

4. The *normalized variance* is a user-selected weighting function for coverage. Unlike typical statistical analysis the grid does not represent random sampling from a particular distribution. In fact the "distribution" is a direct result of orbital mechanics and it is periodic. Therefore the standard statistical measure

of variance is inappropriate. To offer a similar comparison tool, the simulation computes the following given a grid point (i)

$$v = \sum_i w_i(x)a_i$$

where w_i is the user-defined weight given that the coverage value is x and a_i is the area of the region. This normalized variance permits the user to place emphasis on particular aspects of coverage. In the example 0 coverage is considered highly undesirable (weight 20) while single coverage is considered desirable (weight 0.) By carefully selecting weights the user can immediately see how close the system is to some desired goal. A "good" system will have a normalized variance close to zero.

5. The *average coverage per iteration* represents a time-averaged summary of the grids surface. Unlike the minimum, each grid point could have a non-zero average value and yet have been 0-covered during the simulation at some point. The usefulness of the average coverage is that once constraints have been met (*viz.* minimum coverage) the user may compare average coverages for better overall performance or emphasis in particular areas.

6. As with the minimum coverage, the normalized variance is computed to provide a quick comparison between different systems.

3.2.7 Graphical Representation of Ground Coverage

One observation from using the simulation is the importance of having a visual representation of coverage. To facilitate this the grid data were used as input to a simple two-dimensional density plot program, and to a sophisticated three-dimensional solid object displaying package. Sample outputs from these utilities are given in Appendix B. It should be noted that the three-dimensional object package permits the viewer to rotate the object, and move the viewing perspective interactively. This facility is particularly useful when specific coverage features are sought.

3.3 (N_p, N_s) Topologies, Graph Models, and Multigraphs

3.3.1 Motivation for (N_p, N_s) Topologies

The general satellite topology model described at the start of the chapter is sufficient to represent all satellite topologies which use only circular orbits. This model is, however, in many respects too powerful for several forms of analysis. For example the following problems are encountered when using the general model.

- More than one shell significantly increases the period of the system and causes the number of changes in link connectivity to increase dramatically.
- The model permits non-uniform coverage density. Even assuming that only legal satellite systems are considered; many "unreasonable" systems are included.
- The asymmetry of the networks makes identifying topology characteristics difficult if not impossible, preventing all but the most general properties from being identified.

By making additional restrictions to the general satellite topology model described at the beginning of the chapter, it is possible to derive some interesting results for different satellite topologies. In particular the N_p, N_s satellite model has the following restrictions.

- There is only one (1) shell.
- The shell has N_p planes, with the planes rotated about a single common axis.
- There are N_s equally spaced satellites in each plane.

- The topology satisfies complete coverage of the earth.
- The altitude is the minimum necessary to achieve this coverage.

This model may be additionally restricted to "zero phase" by specifying that, at some time, at least one satellite from each plane are co-incident (i.e. share the same point in space).

3.3.2 Minimum Altitude

Using the N_p, N_s zero phase model the first result that can be derived is the minimum altitude for complete coverage. This is given by:

$$A_{min} = \left(\left[\frac{1}{\cos(\frac{\pi}{N_s})\cos(\frac{\pi}{2N_p})} \right] - 1 \right) R_e \tag{3.2}$$

The proof of this equation is given by McLochlin, et al. [40] The equations for A_{min} and $D_{max}/2$ assume a zero minimum horizon angle (satellite to earth angle) which is unrealistic. These equations can be modified to include a non-zero minimum horizon angle, H_a using the following general transformations for D_{max} and A at $H_a = 0$ to D'_{max} and A' at $H_a \geq 0$

$$\begin{aligned} \theta &= \arctan(\frac{D_{max}}{2R_e}) \\ D'_{max} &= \cos(\theta)D_{max}/\cos(\theta + H_a) && (3.3) \\ A' &= (D'_{max}/2)\sqrt{1 - \cos^2(\theta + H_a)} - R_e + R_e\cos(\theta) && (3.4) \end{aligned}$$

3.3.3 Graph Theoretical Model

The (N_p, N_s) satellite model is defined by using solid geometry in three dimensional space. This property is useful in that it permits a wide body of techniques and mathematical tools to be used. However, the model contains much more information than is required for analyzing satellite connectivity and propagation delays.

Since communications are determined by satellite links and their associated propagation delays, it is appealing to study the problem from a strictly graph theoretic point of view. The goal then is to derive a model for the (N_p, N_s) satellite topologies which is constructed entirely from graph properties. Clearly in order to represent the passage of time, the graph will be defined as a set of time invariant vertices and a function which describes the edge set for t, where t is bounded by the period of the topology in question.

In this section a graph theoretical model for low altitude satellite communications as previously described is developed. First a general multigraph model (G) is constructed suitable for point-to-point satellite communications. Then the model is extended to cover a restricted (weaker) form of the (N_p, N_s) satellite topology (G').

Notation. Since the graph theoretic model is a model of a physical satellite system, the following numeric references to planes are made modulo N_p and references to satellites are made modulo N_s. Thus when discussing a vertex $v_{i,j}$ it is assumed that $v_{i,j} = v_{0,j}$ if $i = N_p$, and $v_{i,j} = v_{i,0}$ if $j = N_s$. Additionally adjacency is defined as follows: $v_{i,j}$ is adjacent to $v_{i,j+1}$, and v_{i,N_s-1} is adjacent to $v_{i,0}$.

General multigraph model

In this section the graph theoretic model presented in section 3.1.2 is extended to permit discussion of a dynamic graph appropriate to the satellite system. First consider the characteristics associated with the satellite system. Let $G = (V, E(t), c)$ be a dynamic communications graph defined over the range $t = [0 \cdots T]$ with some initial time $t_0 = 0$. V is a set of vertices and $E(t)$ is an edge set for time t. Since a single shell satellite system is periodic, the time domain $[0 \cdots T]$ is sufficient to define the system over all time. Each vertex set V_i contains the vertices within a plane. The edge sets $E_{i,i}$ constitute those edge connections within a plane while the edge sets $E_{i,j}, i \neq j$ constitute those edges between planes.

- There exist N_p vertex sets (planes) V_i which define V, with the properties that they sum to V, and are disjoint.

  $$\exists V_0, \ldots, \exists V_{N_p-1} \text{ such that } \left(\bigcup_{i=0,\ldots,N_p-1} V_i \equiv V \right) \& \ \forall i, j,\ i \neq j \ \left(V_i \bigcap V_j \equiv \emptyset \right) \tag{3.5}$$

  In addition, a notation used in the future is $v_{a,b} \in V_a,\ 0 \leq b$.

- There exist edges between the vertices which may be grouped into edge sets which define $E(t)$. The two groups are those vertices within a vertex set, and those between them.

  $$\exists E_{i,j,\ i=0,\ldots,N_p-1,\ j=0,\ldots,N_s-1} \text{ such that } \left(\bigcup_{i,j} E_{i,j} \equiv E \right) \& \ \forall j, k \ \ (E_{i,j} \cap E_{i,k} = \emptyset) \tag{3.6}$$

  where $e \in E_{i,j} \leftrightarrow e = \{u, v, c_{u,v}\}\ u, v \in V, c_{u,v} \geq 0$

- Within every $E_{i,i}$ satellites remain relatively stationary

  $$\forall i \forall t_1, t_2 \forall x, y \ \ (x, y \in V_i \rightarrow c_{x,y}(t_1) = c_{x,y}(t_2)) \tag{3.7}$$

- Within every $E_{i,i}$ each satellite has a nearest neighbor

  $$\forall i \forall x \in V_i\ \exists y \in V_i\ \forall t \forall z c_{x,y}(t) \leq c_{x,z}(t) \tag{3.8}$$

  Define $nearest(x, y)$ to mean y is nearest x, i.e. $nearest(x, y) \rightarrow c_{x,y} \leq c_{x,z}\ \forall z$

The restricted (N_p, N_s) zero phase model

In this section additional constraints are imposed on the general multigraph model to specify a restricted (N_p, N_s) zero phase model, G'. $G' = (V, E'(t)), E'(t) \subseteq E(t)$. The N_p, N_s model is a shell of satellites at a single altitude (S=1). N_s satellites are equally spaced on each of N_p planes which are rotated about a single axis (i.e. $P_s =$ some constant N_p, and $N_{p,s} =$ some constant N_s for all p,s). The properties of the (N_p, N_s) model are formally defined by the following constraints:

- There are N_p planes, with N_s equally spaced satellites in each plane.
- The phase angle between planes is zero, i.e. at time t_0 each plane will have either one (N_p odd) or two (N_p even) satellites coincident with the other planes.
- The N_p planes are equally spaced about a single axis of rotation.
- The topology is placed at the minimum altitude that satisfies the criteria of complete coverage of a spherical earth.

Pictorially the (N_p, N_s) zero phase model is shown in Figure 3.4. In order to characterize these properties for the graph model the additional constraints defined in the next four sections are used. First, however, the satellites within planes are labeled. Let $v_{i,j}$ be the jth satellite in V_i $(0 \leq j < N_s; 0 \leq i < N_p)$. Define adjacent satellites by distance, using the property of minimum cost between adjacent satellites. If u and v are adjacent, then $c_{u,v} \leq c_{x,y} \forall x, y$.

N_p planes with N_s equally spaced satellites

The N_p planes have already been specified by construction of the vertex sets (N_p vertex sets). To specify the satellites, within each $E_{i,i}$ the following holds.

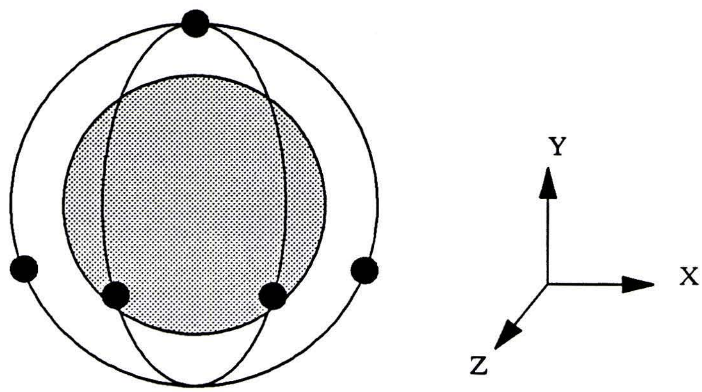


Figure 3.4. N_p, N_s Zero Phase Satellite Topologies

- There are N_s vertices in each vertex set.

$$\forall i \ (|V_i| = N_s) \tag{3.9}$$

- Satellites are equally spaced.

$$\forall x, y, z \in V_i \ (nearest(x, y) \ \& \ nearest(x, z) \rightarrow c_{x,y}(t) = c_{y,z}(t)) \tag{3.10}$$

Graph specification of zero phase–single axis of rotation

- Zero phase between two planes. There exists a time t such that the two planes i & j have elements $v_{i,x}$ & $v_{j,y}$ with the property that there is a zero cost edge.

$$\begin{aligned} \exists i \exists j \ i \neq j \exists t \ \text{ such that } \quad & (v_{i,x} \in V_i) \ \& \ (v_{j,y} \in V_j) \ \& \\ & \left(e = < v_{i,x}, v_{j,y}, c_{v_{i,x}v_{j,y}} > \in E_{i,j}\right) \\ & \text{with } c_{v_{i,x}v_{j,y}} = 0 \end{aligned} \tag{3.11}$$

Proof. By the definition of zero phase in the satellite model for two planes to have zero phase angle implies that at time t, and angular intervals of $2\pi/N_s$

thereafter, a satellite from each plane occupies the same point in space which is at an intersection of the two planes. Note that when N_s is even there will be two points at which satellite locations are coincident. Clearly if two satellites share the same location, the cost $c_{x,y}$ is zero. □

- Zero phase offset. For all time, satellites in adjacent planes are the closest to their equal number in adjacent planes. Let $x \in V_i, y, z \in V_{i+1}$ where $x = v_{i,a}, y = v_{i+1,a}, z = v_{i+1,b}$ $a \neq b$ then:

  - There are nearest neighbors, i.e. $c_{x,y} \leq c_{x,z}$ $\forall z$.
  - At some instant all vertices are co-incident i.e. $\exists t \forall x, y$ $c_{x,y} = 0$.

  Proof. From the (N_p, N_s) model, satellite planes are rotated about a single axis. At time $t = 0$ the vertices v_{i0} for all i occupy the same geographic position. Thus the distance between these satellites is zero. As time progresses these v_{i0} vertices will move equal distances away from the point of coincidence. Exactly $2\pi/N_s$ radians later the same point will be occupied by the v_{i1} vertices. This process continues until the cycle repeats at $t = T$. Thus between any two planes, there are at least N_s instances over the range $[0 \cdots T]$ when the cost between some group of satellites is zero. □

- A single axis of rotation is given by the constraint that all planes are in zero phase.

N_p Equally spaced planes

- Planes equally spaced implies that adjacent nodes between planes have the same cost at each instant of time. Note that this condition holds only in zero phase systems.

$$\forall t, i, j, x, y, z \ \left(i \neq j \ \& \ x, y, z \in E_{i,j}(t) \rightarrow c_{v_{x,y} v_{x+1,y}} = c_{v_{x,z} v_{x+1,z}} \right) \tag{3.12}$$

<u>Graph specification of coverage</u>

Coverage is defined by the following three properties:

- adjacent satellites between planes are always visible to one another.

$$\forall u \exists e =< u, v, c_{u,v} >\in E_{u,u+1} \tag{3.13}$$

<u>Proof.</u> By assumption that the satellite system complies with the (N_p, N_s) model, the visibility condition and the ground coverage conditions have been met. A proof of these characteristics is given in McLochlin et al. [40] Select an arbitrary satellite and rotate the entire system so that Figure 3.5 show the front and end elevations.

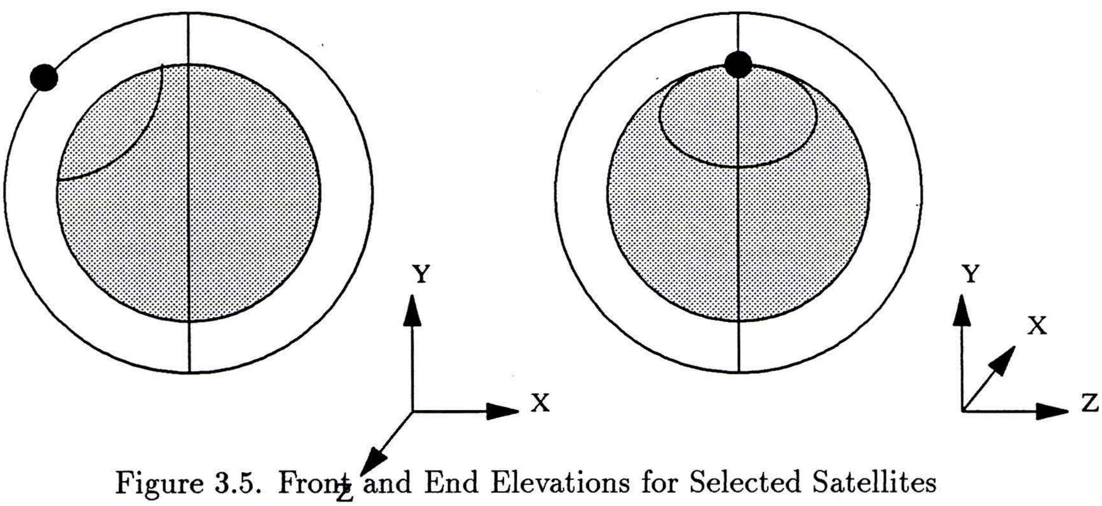


Figure 3.5. Front and End Elevations for Selected Satellites

Let D_{max} be the maximum arc tangent distance between two visible satellites. Consider an arbitrary point x which is just on the boarder of the visibility of this satellite (i.e. $D_{max}/2$ away) and on an imaginary plane perpendicular to the satellite's plane (see Figure 3.6). Then by coverage this point is covered by another satellite at a distance $\leq D_{max}/2$ away. The total distance separating these two satellites is therefore $\leq D_{max}$.

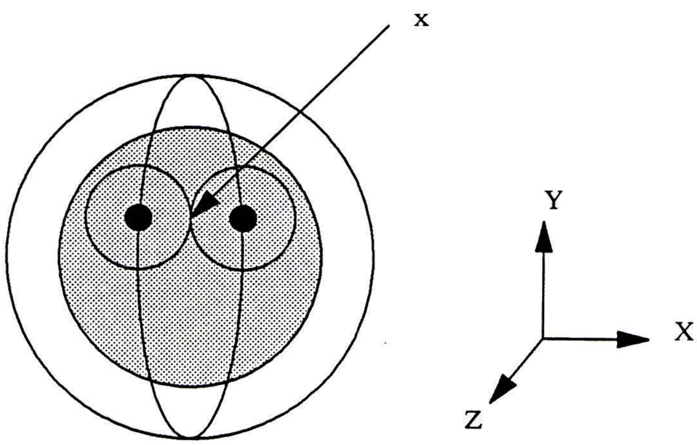


Figure 3.6. Communication Planes

By the visibility condition this implies that the two satellites are visible to one another. Since the plane selected was perpendicular to S_1 there can be no other satellite on the plane of S_1 which covers this point. Thus there is another plane on the near side of the S_1 plane which is in communication with S_1. □

- The minimum costs between N_p successive adjacent vertices from neighboring vertex sets are equal. Let $min(i)$ be the minimum distance from V_i to $V_{i+1} = min(c_{x,y} : \ x \in V_i, \ y \in V_{i+1})$. Then:

$$\forall t \forall i, j \ min(i) = min(j) = const \tag{3.14}$$

Proof. Again by zero phase, at some time one satellite from each plane is coincident. Since the satellites move at constant speed, and the planes are equally spaced the distances between adjacent satellites which are coincident at some time are always equal (see Figure 3.7). □

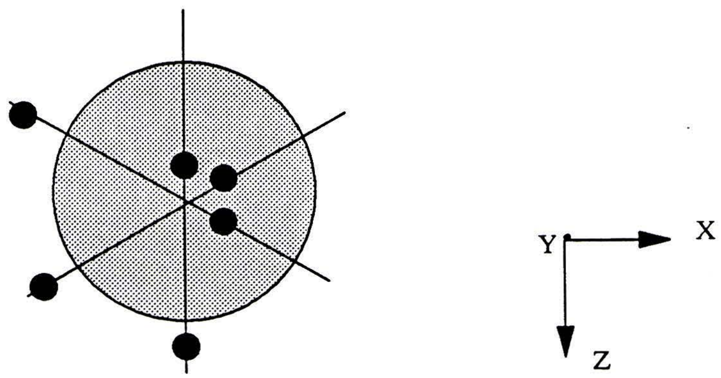


Figure 3.7. Minimum Cost between Successive Planes in Zero Phase System

- Let C be the velocity of light. Then $\forall i, \forall j,\ x = v_{i,j} \in V_i,\ y = v_{i+1,j+1} \in V_{i+1}$

$$max\,(c_{x,y}(t) : t = [0 \ldots T]) = D_{max}/C \tag{3.15}$$

Proof. Again from McLochlin, et al. [40], note that in the (N_p, N_s) model the hardest point to cover is defined by Figure 3.8. The four satellites are on planes $\pm\pi/2N_p$ from the x-z plane, and rotated π/N_s about the y-axis from the t_0 position. At this distance the diagonal satellites are farthest apart and just visible to one another. This distance (defined as D_{max}) is the maximum tangential arc distance between two satellites.

Thus two satellites which are diagonally adjacent (i.e. $x = v_{ij}$ and $y = v_{i+1,j+1}$) have a maximum propagation delay at this farthest point, given by D_{max}/C. In all other instances the distance (and hence propagation delay) is less. So we

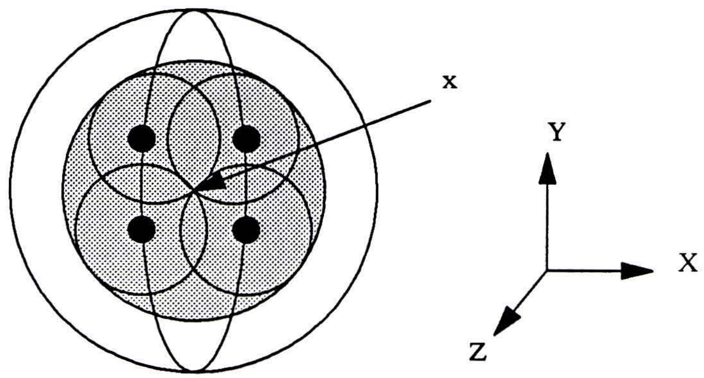


Figure 3.8. Hardest Point to Cover

conclude that for all time, an edge exists between these vertices, with a cost $c_{x,y} \leq D_{max}/C$. □

There may be additional edges; however, these are unnecessary for the proofs to follow. Additionally, this model will be used in Chapter 4 during the analysis of link assignment.

Connectivity in the (N_p, N_s) graph model

Within the confines of the (N_p, N_s) graph model the connectivity may be viewed either empirically or theoretically. From section 3.1.3 it is already known that if the system is c-covered, then it is c-node connected. However this statement underestimated the true connectivity of the satellite graph (i.e. minus the ground points). In the following it is proved that in the (N_p, N_s) model the minimum connectivity is four. Additionally, a table for the connectivities of the lower (N_p, N_s) systems is given.

Theorem 3.3.1 *The minimum node connectivity of an (N_p, N_s) system is four.*

Proof of Theorem 3.3.1 To prove that the graph is four-connected it is necessary to demonstrate the following.

1. At time t_0 the graph is four-connected.
2. From time t_0 to time $t_{2\pi/N_s}$ the graph remains four-connected.

In order to show these properties the concept of a level is introduced.

Lemma 3.3.1 *For all time, a subset of the satellite links defined by visibility, form dynamic rings about the Earth which may be divided into levels.*

Proof of Lemma 3.3.1 The rings are defined by induction at time t_0 and are shown to exist in a dynamic situation until $t_{2\pi/N_s}$. At time t_0 the rings are defined as follows:

1. Select one of the points at which the satellites are coincident and define this to be level 0.
2. Label with the value $n + 1$ all unlabeled satellites adjacent and on the same plane of satellites labeled n.

The number of levels will be $\lfloor N_s/2 \rfloor + 1$, labeled $0 \cdots \lfloor N_s/2 \rfloor$. Labeling the level 0 satellites as the top yields Figure 3.9. Within each of these levels links exist between adjacent planes, forming a ring.

For level 0 and level $\lfloor N_s/2 \rfloor$ this is trivially true. When the satellites are coincident they are clearly completely connected, thus there is a ring. For the case when N_s is odd the nodes on level $\lfloor N_s/2 \rfloor$ also form a clique. Since the planes are set at equal angles about the axis of rotation, the satellites on level $\lfloor N_s/2 \rfloor$ will be at an equal distance from one another and from the point at which the axis of rotation passes through the surface of the earth nearest them (call this point y). Call these satellites

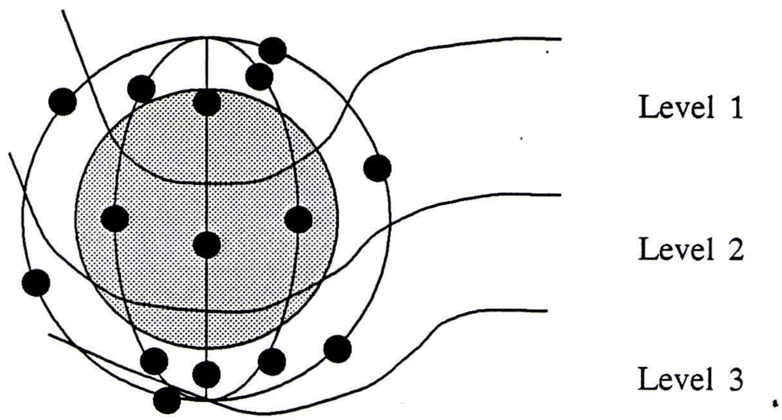


Figure 3.9. Links Used in (N_p,N_s) Model

$n_1, \cdots n_b$. By coverage this point will be visible to all of the satellites on the level. By triangles the distance $(n_i, y) + (y, n_j)$ will be $\geq$ than (n_i, n_j) for all i, j. Thus if the point y is covered, the nodes form a completely connected graph.

For the levels between 0 and $\lfloor N_s/2 \rfloor$ it is shown that the coverage condition ensures that adjacent links on the same level are connected. For two adjacent nodes n_1 and n_2, select the (unique) great circle which passes through both nodes. Consider the point on the earth which is equidistant from n_1 and n_2. By coverage the point of intersection is covered. Assume by way of contradiction that the two nodes in question do not cover this point. By coverage it is known that *some* satellite x covers this point. Consider the distance between x and y; since the lines (n_1, y) and (n_2, y) are on a great circle the distance (x, y) will be greater or equal to either of these. However all satellites have the same maximum satellite-earth visibility. Thus if (x, y) is less than $D_{max}/2$ then so will be (n_1, y); so y will be visible to n_1 #. To conclude the proof it is necessary only to note that the distance $(n_1, y) + (y, n_2)$ will always be $\leq D_{max}$, thus if y is covered then n_1 and n_2 are visible to one another.

Corollary 3.3.1 A subset of the links in the (N_p, N_s) model may be viewed as $\lfloor N_s/2 \rfloor + 1$ rings in which the first and last rings are cliques.

1. Links assignments at time t_0. From corollary 3.3.1 notice that regardless of whether the topology is N_s odd or N_s even, the top and bottom levels are cliques. Assume without loss of generality that there are s satellites per plane and p planes. From lemma 3.3.1 the system may be divided into levels in which satellites within a ring form a level. From section 3.3.3 adjacent satellites are also visible to one another.

For 4-connectivity between a and b there are three cases to prove: 1) communication from non-apex to non-apex, 2) communication from apex to non-apex, and finally 3) communication from apex-to-apex, where apex refers to levels 0 and $\lfloor N_s/2 \rfloor$. These cases are proved below.

- Non-apex to non-apex communication. If both a and b are on the same plane the following four routes may be used to establish four independent paths.

    1. A path clockwise from a to b.
    2. A path anti-clockwise from a to b.
    3. A path which goes up one level, clockwise to node connected to b, then down to b.
    4. A path which goes down one level, clockwise to node connected to b, then up one level.

    If the two satellites are on different planes, project the paths to the same level and proceed as before.

- Non-apex to apex communication. Since the minimum satellite topology is $N_p = 2, N_s = 3$ there are four links from the apex to the adjacent level. Assume that the apex in question is level 0. The following four routes may be selected.

1. A path from the apex a directly down to b.
2. A path from the apex directly down to the node on the same level as b and adjacent to it clockwise.
3. A path from the apex directly down to the node on the same level as b and adjacent to it anti-clockwise.
4. A path from the apex directly down to a node one level below the level of b not using previous routes, then clockwise around plane until a node connected to b is found. Use this link.

- Apex to apex communication. Use any four paths directly down from apex at level 0 to apex at level $N_s/2$.

This concludes the proof for time t_0. $\square$

2) Links assignments at time $> t_0$. To make link assignments at any time greater than t_0 it is only necessary to notice that the proof for constructing the levels will hold at all times. The difference occurs in selecting the vertices to be at level 0. For $t_n > t_0$ these are defined as the N_p satellites closest to one of the points of intersection where the axis of rotation passes through the surface of the earth. With these satellites labeled as level 0 the construction and proof are as before.
This concludes the proof for $t_n > t_0$ $\square$.

Using the links formed by the proof a set is selected which, when used in conjunction with the in plane links, will give a mesh which is four-connected as shown above. This mesh property of the (N_p, N_s) model (and satellite topologies in general) has been found to be particularly useful for reliability.

The actual connectivities of different topologies are presented in Table 3.1. Notice that the minimum is indeed four; however the number is typically greater. In addition, notice that for $N_p = 3, N_s = 3$ the network is completely connected. The results were derived from a variation of the ground coverage simulation in which the connectivity was computed for successive time steps over the period of the topology. Once computed these are of course periodic.

Table 3.1. Satellite-to-Satellite Connectivity of Selected Configurations

$N_p * N_s$ satellites	alt. (km)	k-connectivity link	k-connectivity node	minimum degree	% fully connected
2X3 = 6	11662	4.5	4.5	4.5	96.7
2X4 = 8	6378	4.5	4.0	4.5	74.3
2X5 = 10	4771	4.9	4.0	4.9	64.2
3X5 = 15	2725	6.5	6.0	6.5	52.1

* k-connectivity, degree, and fully connected percentages are time-averaged using 0 and $2\pi/(N_sN_p)$ phase offsets on the satellite network only.

3.3.4 Simulation Results for the (N_p, N_s) Model

The results from the coverage simulation of different (N_p, N_s) topologies are given in the following tables. In Table 3.2 coverage information is given for the satellite systems at their minimum altitude (as given by equation 3.2). The table includes information on the worst case situations of single and double satellite failure. In Table 3.3 coverage information is given for redundantly covered systems. All systems are at an altitude of 11662 km, which is the minimum altitude for single coverage by the 2 planes of 3 satellites. The table includes information on worst case situations of single, double, and triple satellite failure. These values are determined by running each topology with different combinations of failed satellites and selecting the minimum coverage from the results.

The following should be noted from Tables 3.2 and 3.3. From the single coverage table notice that the effect of removing a single satellite is to increase the overall

Table 3.2. Satellite Failure with Single Coverage

Satellite Topology	Alt (km)	0 sats. removed % Cover by x sats. $x=0$	$x \geq 1$	$x \geq 2$	1 sat. removed % Cover by x sats. $x=0$	$x \geq 1$	$x \geq 2$	2 sats. removed % Cover by x sats. $x=0$	$x \geq 1$	$x \geq 2$
2X3 = 6	11662	0	100	42	28	72	0	71	29	0
2X4 = 8	6378	0	100	43	28	72	0	57	43	0
2X5 = 10	4771	0	100	44	27	73	0	56	44	0
2X6 = 12	4037	0	100	46	27	73	8	56	45	7
2X7 = 14	3633	0	100	46	26	73	19	53	43	18
2X8 = 16	3385	0	100	82	9	91	52	18	82	33
2X9 = 18	3221	0	100	90	5	95	61	10	91	37
3X3 = 9	8351	0	100	56	13	87	24	37	63	9
3X4 = 12	4037	0	100	45	17	83	24	41	59	10
3X5 = 15	2725	0	100	41	19	81	23	40	60	15
3X6 = 18	2126	0	100	40	19	81	24	39	61	15
3X7 = 21	1796	0	100	40	19	81	24	39	61	16
3X8 = 24	1593	0	100	41	19	81	25	39	61	16
3X9 = 27	1459	0	100	41	19	81	26	39	61	17
4X3 = 12	7429	0	100	65	6	94	53	17	83	34
4X4 = 16	3385	0	100	51	12	88	41	25	75	24
4X5 = 20	2155	0	100	40	14	86	33	27	73	22
4X6 = 24	1593	0	100	39	15	85	30	28	72	18
4X7 = 28	1284	0	100	40	14	86	29	28	72	16
4X8 = 32	1094	0	100	42	14	86	29	27	73	15
4X9 = 36	969	0	100	42	14	86	28	27	73	14
5X3 = 15	7034	0	100	72	3	97	58	10	90	50
5X4 = 20	3106	0	100	60	5	95	43	14	86	33
5X5 = 25	1911	0	100	48	8	92	33	18	82	27
5X6 = 30	1366	0	100	45	9	91	32	19	81	22
5X7 = 35	1065	0	100	44	10	90	32	20	80	20
5X8 = 40	881	0	100	43	11	89	31	21	79	19
5X9 = 45	759	0	100	43	11	89	31	21	79	18

Table 3.3. Satellite Failure with Double Coverage

		0 sats. removed			1 sat. removed		
Satellite	See	% Cover by x sats.			% Cover by x sats.		
Topology	Note	$x = 0$	$x \geq 1$	$x \geq 2$	$x = 0$	$x \geq 1$	$x \geq 2$
2X5	(1)	0	100	84	8	92	38
2X6	(2)	0	100	100	0	100	91
3X3	(3)	0	100	100	0	100	44
3X3	(4)	0	100	100	0	100	44
3X4	(5)	0	100	100	0	100	79
4X3	(7)	0	100	100	0	100	83
4X3	(8)	0	100	100	0	100	83
		2 sats. removed			3 sats. removed		
Satellite	See	% Cover by x sats.			% Cover by x sats.		
Topology	Note	$x = 0$	$x \geq 1$	$x \geq 2$	$x = 0$	$x \geq 1$	$x \geq 2$
2X5	(1)	37	63	7	62	37	0
2X6	(2)	9	91	72	14	86	13
3X3	(3)	25	75	2	79	21	3
3X3	(4)	21	79	4	76	24	1
3X4	(5)	0	100	66	35	65	16
4X3	(7)	0	100	55	26	74	5
4X3	(8)	0	100	55	20	80	3

Notes:

1. Single axis of rotation. Alt = 11662.
2. Single axis of rotation. Alt = 11662.
3. Single axis of rotation. Alt = 11662.
4. Three axis, inclined at 60° rotated by 120°. Alt = 11662.
5. Single axis of rotation. Alt = 11662.
6. Single Axis of rotation. Alt = 11662.
7. Two axis, set 90° to one another. Alt = 11662
8. Single axis of rotation. Alt = 11662.

miniumum coverage to about 20% (ranging from 3% to 28%). Although the number of satellites has an effect (with fewer satellites giving less coverage) it is not proportional to the extra satellites. Notice also that some configurations perform more effectively as satellites fail. In particular the systems 4X3,5X3, and 2X9 offer good degradation to satellite failure.

From the double coverage table, notice that all systems have been selected with approximately the same number of satellites (9 to 12). The configuration 2X5 is not really a double coverage configuration, as can be seen from the data; it is included to demonstrate the value of this simulation at identifying poor topologies. Comparing the results of systems with the same N_p, N_s, yet differing orbital configurations, notice that there is only moderate change (15%) in the percentage of uncovered regions. Finally notice that with 12 satellites the surface is triple covered (i.e. can withstand the loss of 2 satellites) a great improvement for only 2 extra satellites.

CHAPTER 4
LINK ASSIGNMENT

In Chapter 2 it was shown that one of the characteristics of a satellite communications network was the satellite-to-satellite link technology. It was argued that two communications options are available, radio broadcast and point-to-point laser communications. Of these, it was concluded that for high data-rates, the only practical solution was the laser communications. Also it was established that each satellite would possess a fixed number of transmitter/receiver pairs with which it could communicate to some subset of the visible neighbors. Thus at each instant in time the satellite system may be viewed as a graph of potential communication links with only some of the links realized. Recalling that the propagation delay between satellites depends on the distance between them; the question arises "what is the *best* assignment of communication links given the graph of possible links?" This is called the link assignment problem.

There are essentially two views of link assignment. The simplest view considers the links as possessing infinite capacity, and makes no assumptions about traffic across the network. With this model measures of link assignment strategies are based only on the propagation delays of the individual links. This model is therefore a function of the *satellite topology* only. The optimal solution for this approach corresponds to the minimizing the propagation delay between all satellite pairs (see Floyd [18].) The other view of link assignment is to consider each link as possessing a finite capacity. In this model assumptions regarding link capacities, the message traffic, and the routing algorithm are required. Given these the network flow may be analyzed (see Fratta, et al. [21]) to determine the most appropriate selection of links. Although this technique is much more powerful than the first (in terms of providing a realistic model of the final system) there are difficulties in justifying the assumptions. In particular the analysis becomes a study including the expected traffic and routing strategies; this is significantly different from link assignment alone. Additionally the computational complexity of analyzing flow-deviation is much greater. For these reasons the model used in the research was the simpler unlimited capacity model.

4.1 Points to Consider in the Link Assignment Problem

In dealing with the link assignment problem several points require consideration:

- What are the constraints of the problem?
- What does *best* assignment mean? That is to say, what objectives are to be satisfied.
- Since the satellite system varies with time, the link assignment computed for time t_1 may be different from that for time t_2. What effect does this have on the link assignment problem?
- How do the physical considerations of the system (e.g. time delays caused by mechanical motion etc.) affect the problem?

Each of these issues is now considered separately.

4.1.1 The Constraints of the Problem

Since this is a communication network, the primary constraint is that the graph of selected edges be spanning over the satellites. If the graph fails to span all satellites it is not possible to communicate between all pairs of points on the ground. Thus the network fails to meet the design specifications. Once this objective has been met there are additional requirements considered desirable. The selected graph should have a high node-connectivity. A link assignment such that the removal of a few satellites renders the graph into disjoint segments is undesirable from survivability considerations. The selected graph should satisfy some performance criterion, as discussed later.

4.1.2 Selection of an Objective Function

There are several possible objective functions which could be considered. The function should reflect the graph property of interest, e.g. the propagation delay between nodes over the entire network. The two objective functions considered here are:

- *MaxMinOF* in which the worst case propagation delay between all satellites is considered.
- *MeanOF* in which the sum of the propagation delays between all source destination pairs divided by the number of links (i.e. the mean cost) is used.

It is required that the link assignment algorithm *minimize* the selected objective function.

4.1.3 The Effect of Changing Topologies

In Chapter 2 the issues associated with changing topologies were discussed. In particular it was noted that the topology will change periodically. This influences link assignment in the following way. Consider the case in which the topology changes quickly; if link assignment is performed only once, an optimal (in the sense of some objective function) assignment for time t_1 may rapidly become inappropriate at a later time. Additionally, with satellite motion, some of the links selected for link assignment may disappear as satellites move apart. Thus a link assingment may be valid for only a limited time. From this observation the following two options appear feasible:

- Perform link assignment whenever necessary (i.e. when either a possible link appears or disappears).
- Perform link assignment at frequent intervals, so the that the link assignment is always optimal.

Clearly the latter of these two will give a better link assignment over the period; however, this is offset by the computation required to find the link assignment, and the time to retarget satellite laser communication links (during which no data may flow) as explained in the next paragraph.

4.1.4 The Physical Considerations of the Problem

In section 2.4.3 the problems associated with retargeting a laser communication link were presented. In particular it was stated that for small angles the time to retarget a laser communication link would be short; however for large angles, the time might be quite long. Thus the time to change link assignments depends not only on the number of links that are to be retargeted, but also on the angle over which they are to be moved. Clearly during the reassignment of a communication laser neither the old link nor the new link may be used to carry data. Therefore a penalty is associated with every link reassignment which depends on the angle to be traversed.

4.2 Formal Problem Specification

The formal requirements are to find a k-regular [1] spanning subgraph of the communication links with the following constraints:

[1]The use of the term k-regular is not strictly correct since a limited numer of vertices with less than k edges are acceptable as discussed later.

- $G' = (V, E', c) \;\; E' \subseteq E$
- G' is k-regular

4.3 Objective Function

Finally, as mentioned previously, we require that G' *minimizes* some objective function. The selection of the objective function should reflect some property of the graph of interest. The example of objective functions discussed earlier is

$$\text{MaxMinOF} = \max\left(mincost(v,w)\right) \;\forall v, w \;\; v \neq w \;\in V \tag{4.1}$$

in which the worst case of the minimum propagation delay between all pairs is considered or the objective function

$$\text{MeanOF} = \frac{\sum_{\forall v,w,\; v\neq w\; \in V}\left(mincost(v,w)\right)}{\text{number of links}} \tag{4.2}$$

in which the mean value of the propagation delays over all pairs of vertices is used as the criterion.

It is noted that these are not the only possible functions which could have been selected. In the results presented in this paper the MaxMinOF function is given extra emphasis. The justification for this is that the function provides the maximum propagation delay across the entire network, an important consideration for communications.

We know from graph theory that the k-regular graph problem is NP complete (see Garey & Johnson [24]). This realization prevents any practical algorithms from obtaining solutions for a large number of satellites. In order to analyze these topologies the constraint is relaxed to permit a small number (typically one) of links to be missing from graph. We are therefore interested in effective algorithms which will give non-optimal but satisfactory solutions.

4.4 Objectives of Analysis

When considering link assignment the large number of parameters in a satellite topology suggest many possible observations. In this section the data collected to analyze link assignment are discussed and justified.

- The overall performance of k-regular graphs compared with unrestricted graphs. Clearly it is important to ascertain the effects of restricting the edge set on propagation delay. The expected result from these observations would be some degradation in the propagation delay when using k-regular graphs. The percentage degradation will be an important fact.

- The stability of propagation delay with altitude. Another important aspect when dealing with k-regular graphs is the effect of changes in altitude on the propagation delay. This analysis is useful for two reasons: 1) the altitude of the satellites will naturally vary slightly, and 2) if there is only slight increase in the propagation delay for small changes in altitude, it may be advisable to place satellites in higher orbits to ensure reliable communications.

- The effect of updating link assignment quickly vs slowly. There are essentially two philosophies toward link assignment: 1) working from the assumption that link retargeting has no cost associated with it. The entire satellite topology is frequently (e.g. every few seconds) reviewed to ascertain the most effective link assignment. 2) Conversely, assume that link retargeting has a high associated cost. The network should be reviewed only when necessary (i.e. when there is a change in the satellite topology connectivity). It is expected that the mean propagation delay for topologies using quickly changing link assignment would be less than the corresponding slow changing link assignment.

- The effect of changing the link selection algorithm. As mentioned previously there are many different methods of assigning links. Reviewing all the different strategies is impractical. Instead two greedy algorithms are chosen, both are simple and closely reflect the objective functions which are to be minimized. Details of the algorithms are given later in this chapter.

- The effect of changing k, the regularity of the graph. The final results recorded are the effects of changing the regularity of the graph. Particular interest is given to the effect on propagation delay and connectivity by increasing the number of links. For example, it may be that an increase of one antenna per satellite makes a significant improvement in the propagation delay (or connectivity) of the network.

4.5 Complexity Reduction Using Multigraphs

At first glance the satellite k-regular graph problem appears to be the same as the graph theory problem. Notice however that there is typically a high degree of symmetry in satellite systems (e.g. satellites will be equally spaced in planes) which may be effectively exploited to greatly reduce the problem space. In particular, for a single shell system, let us specify that within a plane each satellite must communicate with its nearest two in-plane neighbors. This reduces the effective value of k by 2 (e.g. a 4-regular problem now becomes a 2-regular one). This is permissible because within planes the visibility will remain constant and the propagation delay between adjacent nodes is a constant (and minimum). Using this assumption, consider each plane as a single vertex in a multigraph; with the unassigned edges of each vertex in the plane available to connect to other multigraph vertices. The constraint is now that the multigraph be $(2 * N_{p,s})$-regular. An example of this is given in Figure 4.1.

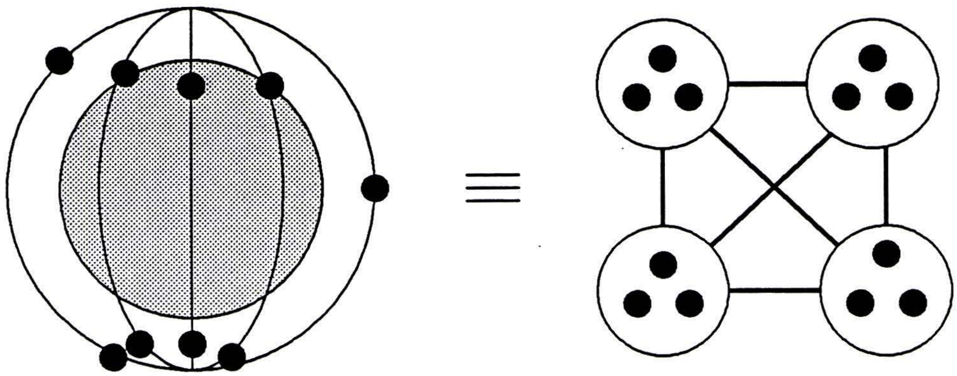


Figure 4.1. Graph to Multigraph Mapping

The problem complexity is now significantly reduced. For example, a satellite configuration with 1 shell of 4 planes of 3 satellites (12 satellites in total) would have an effective complexity of one multigraph link assignment for a 4-node graph as opposed to one link assignment for a 12-node graph.

4.6 The Multigraph Algorithm

4.6.1 Design Requirements

The design specifications for the simulation are as outlined below.

- Using a simple algorithm determine a solution to the k-regular graph problem (where by k-regular it is intended that all of the nodes have at most k edges, and most of them have exactly k edges). [2]
- The user is permitted to specify a set of edges which will be used (regardless of the their cost) for all time provided that they exist.
- The user may specify the following options:
  - The algorithm selects links assignments either when the connectivity changes, or at a fixed interval (e.g. every 5 seconds).

- – The connectivity of the resulting selection be determined.

- The program be modular to permit the selection of different objective functions, and different link selection strategies.

4.6.2 Input Data Format

The multigraph algorithm takes as input two files: a connectivity file which contains link connectivities for a particular topology over time; and a preload file which contains a list of preferred links (which will be used regardless of their cost).

The connectivity file is generated by a program closely related to the ground coverage program of Chapter 3 and is known as the network generator program. The input to the network generator is the same as that of the ground coverage (see section 3.2). An annotated sample of the network generator output (i.e. link assignment input) is described below.

```
file = LD123A0                                            *see note 1
totalsats =    6                                          *see note 2
vertex  platform  plane  satellite
     1        1      1         1                          *see note 3
     2        1      1         2
     3        1      1         3
     4        1      2         1
     5        1      2         2
     6        1      2         3

changes :  15 time =      5                               *see note 4
        105  1  2                                         *see note 5
        105  1  3
         86  1  4

        ...

        105  4  6
        105  5  6
inc     :  15 time =     10                               *see note 6
```

[2] It is a well known fact that for particular graphs and values of k it is not possible to select *exactly* k edges for each vertex. Also, as was noted previously, to find solutions (if they exist) with *exactly* k edges is an NP-complete problem and the performance of algorithms would be unacceptable.

```
    105  1  2
    ...
    105  5  6
...
changes :    0 total topologies =     1                *see note 7
```

Notes on terms used:

1. The file name. The name of the input file to be processed.

2. The number of satellites used. The number of satellites used in the simulation.

3. A description of the satellite orbit information. For each satellite the relative position of its orbit is given. This information is necessary to construct the multigraph. The data provided are a vertex number (for reference), the satellite platform (shell), plane, and satellite number within the plane.

4. Topology change control information for the connectivity map. The first field may either be "changes" or "inc" (see note 6). The "changes" field implies that the network to follow is the result of a change in the satellite topology. After the "changes" field is the number of connectivity pairs to follow. Finally the "time" entry is the time from t_0 that the topology applies.

5. A description of the connectivity information. Connectivity information is denoted by a set of three integers per link. The first integer indicates the link cost in milliseconds. The second and third integers give the indices of the vertices used. Note that all links in the model are assumed to be bidirectional.

6. Time increment control information for the connectivity map. The first field is "inc" which implies that the connectivity remains unchanged from the previous data set. The remaining fields are consistent with note 4.

7. The end of file mark. A value of "changes" equal to zero is used to indicate the end of the input file. The second field contains the number of different topology changes which occurred during the time period.

Associated with each connectivity file is a user-defined preload file. The intended use of the preload file is to permit the user to specify a particular set of links to be used preferentially. This is particularly useful for considering different in-plane link assignment strategies (e.g. a ring). An annotated example of the preload file is presented below.

```
size = 6 degree = 5                              *see note 1
6                                                *see note 2
1     1 1 1    1 1 2                             *see note 3
1     1 1 2    1 1 3
1     1 1 1    1 1 3
1     1 2 1    1 2 2
1     1 2 2    1 2 3
1     1 2 1    1 2 3
```

Notes on terms used:

1. A description of the control information. The "size" field specifies the number of satellites (used for consistency checks). The "degree" field indicates the value of k which is to be used in the simulation.

2. An indicator of the number of edges to be preloaded (0 or more).

3. For each edge to be preloaded the source and destination satellites are specified (by shell,plane, and satellite). The triples must be from lowest satellite to highest by numeric order.

4.6.3 Details of the Algorithm

With a formal definition of multigraphs given in section 3.3.3, the multigraph algorithm used to analyze the k-regular graph problem is introduced. The pseudo code for the algorithm is provided below, however a general synopsis is presented here. There are three inputs to the program:

- The graph representing satellite communication links.
- The map between the vertices of the graph and the position of the satellites within the planes.
- A file which specifies which of the possible communication links is to be assigned *a priori*. In this paper the preassigned links are those adjacent to satellites within a plane (i.e. rings).

From the input a *multigraph distance table* is generated. Each element of the table is a pointer to a stack of structures which represent those links which join the two multigraph vertices together. Included in the structure are the source, destination vertices, and the propagation delay between them. The multigraph is created by sorting the edges by decreasing propagation delay and mapping each edge in the original graph to a multigraph pair by pushing the element onto the appropriate stack (see Figure 4.2). The result is a table in which the top element of each stack represents the shortest path between the two multigraph vertices (i.e. planes).

A link selection algorithm (see section 4.6.4) is then used to construct a sparse graph on the multigraph vertices visible (i.e. those at the top of the stack). Since the number of nodes in the multigraph is small, the chances of selecting more than k edges emanating from a given vertex is low. The links selected by selection algorithm are used as a template to add edges into the desired k-regular graph. Since all vertices within a plane are connected (via preselection), the edges selected from the multigraph will connect these planes together forming a connected graph. Should the link selection algorithm be unable to span the multigraph the largest subtree is used. Finally the elements used in the tree are popped off their respective stacks, and the process repeated (now using edges with slightly longer delays). If at any time a vertex is connected to k other vertices, it is ignored during future iterations (i.e. the top of stack elements are tested prior to being passed to the link selection algorithm and popped if unsuitable).

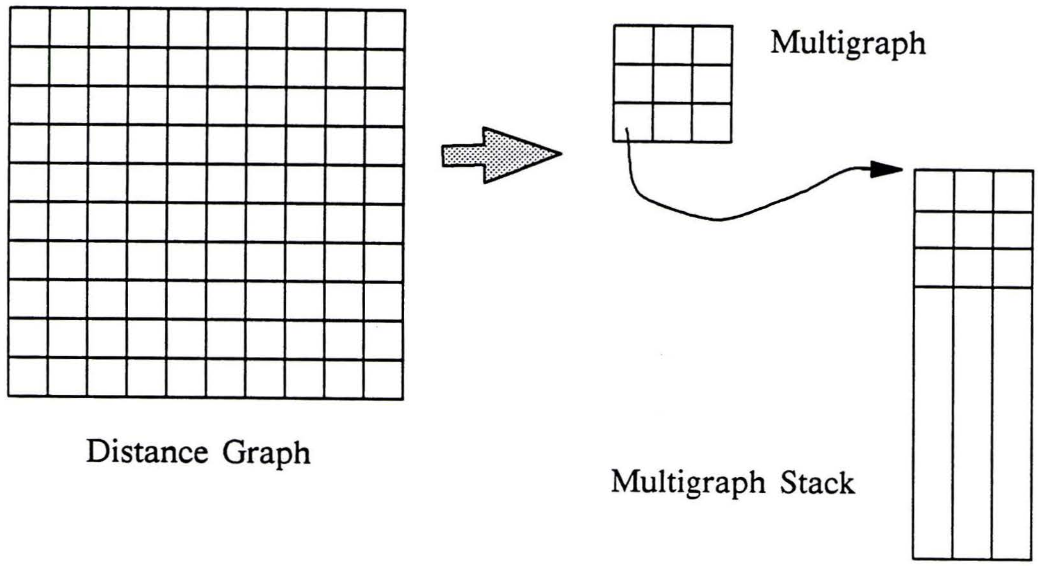


Figure 4.2. Distance Graph to Multigraph Transformation

The final result is a graph which may be thought of as having been constructed in the following manner:

- Connect all adjacent nodes in each satellite plane.
- Attempt to place layers of spanning trees between the planes until each vertex in the graph has k edges.

Finally the objective function is applied to the selected graph in order to generate the statistics given in the next section.

Algorithm 4.6.1 An algorithm to compute link assignments for a given satellite topology.

```
InitializeGraphInfo;
ReadPreSelectInfo;                          /* Reads links to be used a priori */
WHILE (GraphsLeftToProcess) DO
BEGIN
    InitializeMultigraphStacks;
    ReadGraphInfo;                          /* Reads vertices and plane maps */
    BuildLinkCostList;                      /* place links in a stack */
    SortLinkCostList;                       /* sort links by increasing delay */
    /* Place Links in MultiGraph */
    FOR i = 1 TO NumberOfLinks DO
        IF (MGMap(SourceEdge) <> MGMap(DestEdge)) THEN
            MultiPush(MGMap(SourceEdge),MGMap(DestEdge),
                SourceEdge,DestEdge,Delay);
    /* Place repeated Spanning Trees */
    WHILE (EdgesLeftToExamine && (EdgesUsedInOutputGraph< kn/2)) DO
    BEGIN
        CopyTopsToTempGraph;                /* Copy the Stack Tops to a */
                                            /* temp graph */
        SelectEdges;                        /* Link Selection Algorithm */
        CopyResultsToOutputGraph;
        PopAppropriateStackElements;
    END
    IF (EdgesUsedInOutputGraph ≠ kn/2) THEN
        InformUser("Couldn't Make k-regular Graph");
    InformUser(LinkAssignmants,CostOfGraph);
END
```

In concluding this section on the multigraphs we discuss how the algorithm would be incorporated into the typical network selection procedure. The basic method for generating acceptable solutions proceeds in an iterative fashion. An iteration step consists of:

- Generate a network topology.

- Check that it meets the necessary constraints.

- Apply local perturbations and see if a better topology is generated.

After attempting a sufficient number of starting topologies the user selects the best found so far and uses it. It is intended that the first step (generate a network topology) be achieved using the multigraph technique. Comparing the execution speed of this technique with non-multigraphs the following points are noted. Finding a solution to the k-regular graph problem could require $O(2^e)$ operations. Successive applications of the Minimum Cost Spanning Tree (MCST) algorithm could take $O(e^2 \log n)$ (i.e. e applications of MCST), while the multigraph algorithm takes $O(ee' \log n')$ where e' is the edge count of a single layer of the multigraph, and n' is the vertex count of the multigraph. For the example with 3 planes of 4 satellites, these figures are (assuming $e = n^2$, $e' = n'^2$) 2^{144},51527, and 1423 respectively. Thus the algorithm is efficient and produces *initial* networks with desirable characteristics.

4.6.4 Link Selection Algorithms Considered

In each iteration of the multigraph algorithm the top plane of the multigraph is passed to a link selection algorithm which is responsible for selecting those links which will be added to the k-regular graph for the current iteration.

As with objective functions, there are many different link selection algorithms which could be applied in the multigraph algorithm. Two algorithms were used in order to permit some degree of comparison. Those selected were as follows:

- Kruskal's algorithm [36] for finding a minimum cost spanning tree. The use of this algorithm is justified by the observation that the spanning tree uses the fewest number of edges to connect the graph ($n-1$ edges for n nodes). Another reason for the interest in this algorithm is that, when used in conjunction with the (N_p, N_s) topologies, it is provable that a connected graph will result after the first iteration (see section 4.6.1).

- Floyd's algorithm [18] for finding the all pairs least cost paths. The primary purpose of this algorithm was to offer a contrast with the minimum spanning tree. The characteristics of all pairs least cost paths lead one to expect the following properties:
  - The resulting k-regular graphs will have slightly lower propagation delays.
  - There is a higher probability that the graphs will be disconnected at low values of k.

4.6.5 Implementation of the Link Assignment Algorithm

This algorithm was implemented as two programs. The first program (a front end program) was written in the "FORTRAN" programming language; and only generated the connectivity maps for use by the second program. The code size was 600 lines and took approximately one month to write. The second program was written in the "C" programming language. The source code size was 3000 lines and took six months to write. Both programs were verified by comparison with a simple topology and by results obtained from work done by McLochlin, et al. [42]. The machine on which the program was developed was the Gould Powernode 9080 supporting the Unix operating system. The typical execution speed for a single topology was 10 seconds.

4.6.6 Statistical Evaluation of Link Assignment

The output from the link assignment algorithm consists of run information, a histogram, and statistics accumulated on the connectivity maps processed. An annotated sample of output is given below describing the statistics gathered.

```
File = OD133A0, using preload TP123 and data LD133A0 *see note 1
Sim = slow change, Opt = multigraph, Graph = min span tree : sorted,
OF = max(min (s,d) pair)
```

```
Number of sats = 9, MaxDegree = 4                    *see note 2

Histogram of O.F. frequency                          *see note 3
[ 100 -  109]  179 ******************************
[ 110 -  119]  142 ********************
[ 120 -  129]  148 *********************
[ 130 -  139]  158 ************************

...
[ 490 -  499]    0
[ 500 - INF]     0
                                                     *see note 4
Net Cngs Conn Cngs  Samples   TotalD  MinD  MaxD  MeanD    SD   CPU
   5          3       1288    159533   90    161  123.8   22.2   25

MinC   MaxC   MeanC    SD                            *see note 5
  4      4     4.0    0.0

*** warning 5 graph(s) failed to use all links,      *see note 6
    mean of 1 link(s) unused per failed graph

*** warning 3 graph(s) were unconnected              *see note 7
```

Notes on terms used:

1. A description of the control information. The head of each file consists of the file names used (output file name, and the input preload and data file names) and information regarding the simulation run. The information consists of:

   - Sim: the type of simulation, currently either "slow" or "fast", indicating the frequency at which links are retargeted. "Slow" implies that retargeting is done only when the satellite topology changes; "fast" implies that retargeting is performed at every time step.

   - Opt: the optimizing algorithm selected. Details of the different algorithms attempted are listed in section 4.7.2. Currently the algorithms investigated are unbounded, optimal, and multigraph.

- Graph: for the multigraph algorithm there is additionally: a per-iteration link selection algorithm, which may be either min-span-tree or all-pairs-least-cost-path; and the link selection order, which may be either random or sorted.
- OF: the objective function used to measure the algorithm. The objective functions are discussed in detail in section 4.3 and may be either MaxMinOF or MeanOF.

2. A system summary. A brief note of the number of satellites used in the simulation and the maximum degree of any link (i.e. the regularity of the graph).
3. A histogram of the simulation. As the simulation proceeds a histogram is generated which maps the number of networks by objective function (therefore the units are milliseconds). The histogram is useful to identify unusual characteristics of different topologies.
4. A summary of simulation statistics. The most important results from the simulation are the statistics printed which are as follows:
   - Net Chgs: records the number of times the physical network topology changes over the simulation (i.e. the number of time-instances links either break or are made).
   - Conn Cngs: records the number of times that the k-regular graph network topology changes over the simulation.
   - Samples: provides the number of networks used to generate these results. These networks are typically a time sequence for a single satellite topology.
   - TotalD: the total delay from the objective function over the entire simulation (in milliseconds).

- MinD: the minimum delay from the objective function over the entire simulation (in milliseconds).
- MaxD: the maximum delay from the objective function over the entire simulation (in milliseconds).
- MeanD: the time-averaged mean delay of the objective function over the entire simulation (in milliseconds). The average is obtained by the dividing the total delay by the number of samples analyzed.
- SD: the standard deviation for the time-averaged mean above. Obtained by the second moment.
- CPU: the number of seconds the entire simulation took to run. This is used for comparison and efficiency purposes.

5. Connectivity statistics. If the user selects the option the connectivity of each k-regular graph is computed. These statistics are displayed after the general statistics. The values printed are as follows:

   - MinC: the minimum node connectivity for the entire simulation.
   - MaxC: the maximum node connectivity for the entire simulation.
   - MeanC: the mean node connectivity of the graphs over the entire simulation.
   - SD: the standard deviation of the mean connectivity.

6. An optional field indicating those networks in which some links were unused in the k-regular graph. The number of graphs in which this occurred and the average number of unused links per network are indicated. A value of 1 unused link per network is considered acceptable since this is often a function of the graph rather than the link selection.

7. An optional field indicating that some k-regular graphs were actually unconnected. Each network which was unconnected is printed (not shown here) for evaluation. Clearly graphs producing this warning should be investigated.

4.6.7 Relationship between Algorithm and (N_p, N_s) Model

In order for the multigraph algorithm to be useful, it is imperative that it satisfy the original problem constraints, i.e. that it generate a spanning regular subgraph of the original graph; which minimizes the objective function.

Theorem 4.6.1 *Given an (N_p, N_s) model with satellite planes in a single axis of rotation, and with zero phase offset, the multigraph algorithm using Kruskal's algorithm for link selection constructs a connected subgraph of the original graph.*

Lemma 4.6.1 *The edges at each instant t, $E(t)$, such that there is a minimum cost $c_{v_i,v_{i+1}}$ for each i in the restricted (N_p, N_s) zero phase model, are selected by the multigraph algorithm. Furthermore, these edges form a spanning tree on the multigraph i.e. across the vertex sets.*

Proof of Lemma 4.6.1 from the first clause in the graph specification of coverage. Notice that between neighboring vertex sets adjacent satellites are visible. Let V_i be any vertex set; pick the shortest edge to a neighboring vertex set V_{i+1}, call this edge $e_{v_{i,j},v_{i+1,k}}$. From the second clause in the graph specification of coverage similar edges may be selected between adjacent vertex sets working in a cyclic fashion $(i, i+1, \ldots)$. Notice that these edges form a path in the multigraph. After $N_p - 1$ edges have been selected every vertex is connected. Because each edge is a minimum the path is also minimum for time t. Since at most 2 edges are used in each plane (V_i to V_{i+1}, and V_i to V_{i-1}) even if they emanate from the the same satellite, only 2 edges have been used.

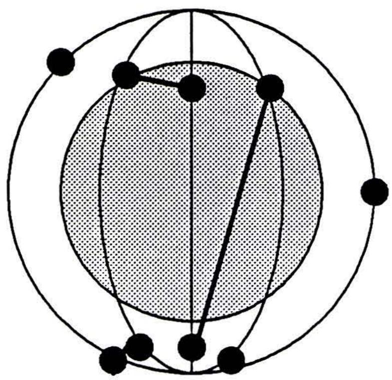


Figure 4.3. Sample Spanning Path between Planes

Proof of Theorem 4.6.1 by considering the multigraph algorithm. The edges for each multigraph edge are stored in a stack sorted by minimum cost. Thus the top of the stack contains the minimum cost element between two planes. The *first application* of Kruskal's algorithm to the top-of-stack elements will result in the path given by lemma 4.6.1 since *there is no other shorter path between all planes.* According to lemma 4.6.1 this path has at most two edges per satellite. Each satellite has at least two free edges before the path is selected ($k \geq 4$) therefore none of the edges selected by Kruskal's algorithm will be rejected when added to the final graph. Since all satellites within planes are connected and all planes are connected by Kruskal's algorithm, the entire graph is connected after one application of the iteration step. □

Theorem 4.6.2 *In the first iteration of the multigraph algorithm a minimum cost spanning tree is constructed.*

Proof of Theorem 4.6.2 from theorem 4.6.1. We notice that the edges selected by Kruskal's algorithm form a minimum cost spanning tree over the multigraph. Since none of the edges is rejected the planes are indeed connected in one iteration.

4.7 Results from Multigraph Algorithm

In keeping with section 4.4 the results are in five parts corresponding to the objectives of the analysis. Prior to this a description of the algorithm parameters and tables of all the data collected are presented.

4.7.1 Algorithm Parameters

In this section we consider the performance of the algorithm. The results are based on data generated by the multigraph simulation program for low altitude satellite networks. In these tests the (N_p, N_s) zero phase satellite model was used (see section 3.3 for details). The propagation delay figures in Tables 4.5 to 4.16 are given in milliseconds.

4.7.2 Table of Data and Keys

Algorithm analyzed and labeling key

The algorithms in Table 4.1 were investigated. The terms used are as follows.

- "Unbounded search": no restrictions are made to the value of k. Clearly, without restricting the graph to k, the performance of the graph should be improved.

- "Exhaustive search": a variation of the "eight queens" algorithm is used to find the optimal solution to the k-regular graph problem. This approach has the unfortunate property of being NP-complete.

- "k-regular multigraph": the graph is k-regular within the design specifications outlined in earlier sections.

List of parameter combinations simulated

Table 4.2 gives the relationship between the parameters which could be varied and the simulation algorithms. The parameters considered were as follows:

- "Altitude": the effects of an increase in the altitude were examined. The altitudes reviewed were 101% and 110% of the base altitude.
- "Slow/Fast change in the link assignment": the links are either assigned whenever the topology changes (i.e. slow) or at frequent intervals (i.e. fast) of 5 seconds.
- "k-values": the algorithms were tested with the same topologies at several different values of k, the regularity of the graph.

General topology information. For each topology there are several statistics described in section 4.6.6 which remain unchanged as the parameters vary. In Table 4.3 the values of these statistics are recorded.

Results–control algorithms. Table 4.4 contains the results for the control algorithms using the slow update speed; and Table 4.5 contains the results of the control algorithms using the fast update speed.

Results–algorithm statistics, altitude. Tables 4.6,4.7,4.8, and 4.9 present the results for the propagation delays using different multigraph algorithms at different altitudes.

Results–algorithm statistics, connectivity. Table 4.10 contains the connectivities of the three main algorithms (Alg1, Alg4, and Alg2) for the base 101 altitude.

Results–algorithm statistics, general. Tables 4.11,4.12,4.13, and 4.14 contain the general propagation delay statistics for the algorithms Alg2, Alg1, and Alg4 together, and Alg3 and Alg5 together. Notice that unconnected graphs are marked by "–".

Results–effect of k. In Tables 4.15,4.16 the effect of changing the value of k (the regularity constant) is recorded. The tables represent the fast and slow link assingment updates.

4.7.3 Analysis of Results

This section discusses the observations derived from the previous tables.

Performance of k-regular graphs

The effectiveness of the multigraph algorithm at obtaining link assignments with low propagation delays may be obtained by comparing Tables 4.4 with 4.11, and 4.5 with 4.12. The objective function used for comparison was the MaxMinOF function. As might be expected Con1 (the unbounded algorithm) performs consistently better than all the other algorithms. The Con3 algorithm represents the optimal solution to the problem (Con3 uses exhaustive search to explore all possibilities of k-regular graphs). The difficulty with Con3 is that the algorithm runs in exponential time. Because of this, analysis of complex topologies is intractable. In order to gain some small insight into the sorts of figures which should be obtained for k-regular graphs the results from Con3 for six and eight satellites are given. As with the multigraph, the user has the option to preload links. The results for Con3 are given for the "no preload" and the "ring preload" situations. Notice that, although the number of samples is small, the increase in propagation delay from using the optimal with no preload to using ring preload is about 7%.

The relative increase in performance of Alg2, Alg1, and Alg4 are listed in Table 4.17. Alg2 (in which the multigraph algorithm is used but the links are selected at random) is significantly worse than the others, which is as expected. Notice however, that the value is only $\approx 29\%$, which suggests that although attempts to optimize link assignment gives rise to improvements, the improvements are not "orders of magnitude." The other point of interest detailed in the tables is that there is a marginal improvement in using Alg1 over Alg4. Recall that Alg4 performs link selection using the all-pairs least cost graph, while Alg1 performs link selection using a minimum cost spanning tree. Since Alg4 is attempts to use more links per iteration than Alg1 the results are consistent with expectations. Additionally Alg4 has a major disadvantage which is discussed in section 4.7.3.

The other point to note from Table 4.17 is that for the less complex topologies there is no difference between Alg1 and Alg4. Again this agrees with the anticipated results, since for small graphs the spanning tree and the all-pairs least cost path will be very similar, giving rise to similar results.

Stability of propagation with altitude

In Tables 4.6,4.7,4.8, and 4.9 the effects of increasing the altitude are investigated. The last three columns of each table consist of the percentage increase in propagation delays from increasing the altitude. The altitudes selected were 101% and 110% of the base altitude (where the base altitude is the minimum altitude for the complete coverage of the earth using the (N_p, N_s) zero phase model). Notice that for the different objective functions and link selection algorithms the increase in propagation delay is consistently in the range 2 to 6 percent, with only a few exceptions. This fact is quite surprising since the increase over the base altitude is 10%.

While it was anticipated that the percentage increase in propagation delay would depend on the topology, a surprise result was that for a few systems the propagation delay actually went down. This counter-intuitive situation may arise because an increase in altitude may cause an increase in the network connectivity permitting better routes to be selected. As can be seen from the results though, the effects of this mechanism are usually much smaller than the increase caused by altitude.

Updating links quickly vs slowly

In the following table (Table 4.18) the effect of updating the link assignment periodically is compared with that of updating it only when the satellite topology changes. The former method of updating is called "fast," while the latter is know as "slow." The results are the percentage improvement (i.e. (fast/slow-1)*100) for the two multigraph link assignments (minimum spanning tree and all-pairs graph) and the two objective functions (MaxMinOF and MeanOF). Thus the results are for four algorithms, Alg1, Alg4, Alg3, and Alg5.

Contrasting algorithms Alg1, Alg4, Alg3 and Alg5, the first point to notice is that there is no significant improvement between the slow and fast methods. The average improvement is about two percent although there is a considerable improvement for the topology with two planes of three satellites. Comparing the objective functions (i.e. comparing Alg1 and Alg4 with Alg3 and Alg5) suggests that the maximum propagation delays are affected more by fast topology updates than the mean propagation delays are. In fact looking at Alg3 and Alg5 the improvement of the fast algorithms over the slow algorithms is on the average only one percent. When this result is combined with the observation that performance of a complete network is better represented by the mean propagation delay the conclusion emerges that there is little point in investing computation time and sophisticated hardware to implement a link assignment scheme which attempts to keep globally optimal link assignments. Although there are other interesting findings in this report, this result is particularly significant in view of the original research objectives.

Effect of different link selection algorithm

In Tables 4.19 and 4.20 the percentage improvement by using the all-pairs least cost graph link selection over the minimum cost spanning tree link selection is shown. In the first table the objective functions MaxMinOF and MeanOF at the slow rate are detailed. In the second table the objective functions are detailed at the fast rate. Notice that for most of the topologies the improvement is small. The exception is the single axis four planes of four topology. In this case for both update rates and both objective functions there is an average increase in the propagation delay of fifteen percent by using the all-pairs least cost graph over the spanning tree. This result could well be caused by failings inherent in the greedy algorithm. Notice also that particular topologies have similar effects on the different comparisons, again suggesting that the greedy algorithm is less efficient through selecting too many edges per iteration.

Effect of changing k

The final observation made is an analysis of the effect of changing the value of k, the regularity of the graph. This analysis was performed for Alg1 only (i.e. a link selection of minimum spanning trees and the MaxminOF objective function). The first point to notice is that no tables are given for the ratio of propagation delays for $k = 3$ / $k = 4$ the reason for this is that according to Table 4.15 many of the graphs are unconnected, and the propagation delays for the more complex topologies are significantly longer. These results are consistent with expectations, since the multigraph algorithm uses two edges from each satellite to connect the rings within planes. This leaves only one edge per satellite for interplane communication. The only topology which emerges uneffected is the two planes of three, in which only one link is required to connect the graph.

Considering Table 4.21, notice that there is significant advantage in many of the topologies from using 5 antennae per satellite instead of only 4. Additionally it is interesting that the most benefit is received by the maximum propagation delays. This is explained by noting that additional links between those satellites which cause the maximum propagation delay, and other planes will have a large impact on the all-time maximum propagation delay.

4.8 Chapter Conclusions

In this chapter a satellite multigraph has been formally introduced and and algorithms have been presented with which to study the properties of this structure. For the analysis the (N_p, N_s) model has been used with both single and double axes of rotation. The results from running the multigraph algorithm with different objective functions and *per iteration* link selections are given in the previous section.

Table 4.1. Labeling Scheme for Algorithms

Label	Title of Algorithm
Con1	Unbounded search for optimal solution, MaxMinOF
Con2	Unbounded search for optimal solution, MeanOF
Con3	Exhaustive search for k-regular optimal soln, MaxMinOF
Con4	Exhaustive search for k-regular optimal soln, MeanOF
Alg1	k-regular multigraph using min spanning trees, MaxMinOF
Alg2	k-regular multigraph using random edge selection, MaxMinOF
Alg3	k-regular multigraph using min spanning trees, MeanOF
Alg4	k-regular multigraph using all pairs min cost path edges, MaxMinOF
Alg5	k-regular multigraph using all pairs min cost path edges, MeanOF

Table 4.2. Algorithms used under Different Tests

Label	$k=3$		$k=4$				$k=5$	
	Altitude 101%		Altitude 101%		Altitude 110%		Altitude 101%	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
Con1			Yes	Yes				
Con2			Yes	Yes				
Con3			Yes					
Con4			Yes					
Alg1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Alg2			Yes	Yes				
Alg3			Yes	Yes	Yes	Yes		
Alg4			Yes	Yes				
Alg5			Yes	Yes				

Table 4.3. General Topology Information

Satellite Topology	See Note	Base Altitude	101% Altitude	110% Altitude	Network Changes	Number Samples
2X3 A0 = 6	1	11662	11778	12828	1	1623
2X4 A0 = 8	1	6378	6442	7016	3	722
2X5 A0 = 10	1	4771	4819	5248	5	471
3X3 A0 = 9	1	8351	8435	9186	5	1196
3X4 A0 = 12	1	4037	4077	4441	7	532
3X5 A0 = 15	1	2725	2752	2998	13	347
4X3 A0 = 12	1	7429	7503	8172	13	1085
4X4 A0 = 16	1	3385	3419	3724	11	482
3X5 C0 = 15	2	2725	2752	2998	13	347
3X6 C0 = 18	2	2126	2147	2339	9	261
4X3 C0 = 12	3	7429	7503	8172	17	1085
4X4 C0 = 16	3	3419	3453	3761	7	482
4X5 C0 = 20	3	2155	2177	2371	5	315

Notes:

1 Single axis of rotation.

2 2 planes about x axis, 1 plane about y axis.

3 2 planes about x axis, 2 planes about y axis.

Table 4.4. Propagation Delay (ms) for Control Algorithms, Slow Update

Satellite Topology	Con1				Con3 - No Preload				Con3 - Ring Preload			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	105	113	110	2.5	135	179	154	13.2	148	179	165	9.1
2X4 A0	117	120	119	1.2	117	121	119	1.5	117	134	125	6.1
2X5 A0	88	97	93	2.6	Runtime Complexity Prevents Analysis							
3X3 A0	86	114	100	8.6								
3X4 A0	98	98	98	0.0								
3X5 A0	72	80	77	2.5								
4X3 A0	80	115	98	10.3								
4X4 A0	92	92	92	0.0								
3X5 C0	72	80	77	2.5								
3X6 C0	83	84	84	0.1								
4X3 C0	99	107	103	2.4								
4X4 C0	92	92	92	0.0								
4X5 C0	76	80	78	1.3								

Table 4.5. Propagation Delay (ms) for Control Algorithms, Fast Update

Satellite Topology	Con1			
	Min	Max	Mean	SD
2X3 A0	105	113	110	2.5
2X4 A0	117	120	119	1.2
2X5 A0	88	97	93	2.6
3X3 A0	86	114	100	8.6
3X4 A0	98	98	98	0.0
3X5 A0	72	80	77	2.5
4X3 A0	80	115	98	10.3
4X4 A0	92	92	92	0.0
3X5 C0	72	80	77	2.5
3X6 C0	83	84	84	0.1
4X3 C0	99	107	103	2.4
4X4 C0	92	92	92	0.0
4X5 C0	76	80	78	1.3

Table 4.6. Effect of Altitude on Propagation Delays using MaxMinOF, Slow Update

Satellite Topology	Results at 101 % Base			Results at 110 % Base			% Increase		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	179	210	192	189	222	203	5.5	5.7	5.7
2X4 A0	120	120	120	126	126	126	5.0	5.0	5.0
2X5 A0	88	104	96	92	109	101	4.5	4.8	5.2
3X3 A0	86	155	120	90	161	124	4.6	3.8	3.3
3X4 A0	98	140	112	102	144	115	4.1	2.9	2.7
3X5 A0	72	90	83	74	93	84	2.8	3.3	1.2
4X3 A0	93	141	111	96	146	115	3.2	3.5	3.6
4X4 A0	92	126	107	96	125	108	4.3	-0.8	0.1
3X5 C0	72	90	83	74	93	84	2.8	3.3	1.2
3X6 C0	84	104	90	87	106	93	3.6	1.9	3.3
4X3 C0	116	192	123	121	208	158	4.3	8.3	28.5
4X4 C0	92	125	110	96	130	115	4.3	4.0	4.5
4X5 C0	80	93	86	84	93	90	5.0	0.0	4.7

Table 4.7. Effect of Altitude on Propagation Delays using MaxMinOF, Fast Update

Satellite Topology	Results at 101 % Base			Results at 110 % Base			% Increase		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	105	191	128	111	202	136	5.7	5.8	6.3
2X4 A0	120	120	120	126	126	126	5.0	5.0	5.0
2X5 A0	88	104	96	92	109	101	4.5	4.8	5.2
3X3 A0	86	139	111	90	146	116	4.7	5.0	4.5
3X4 A0	98	128	108	102	133	112	4.1	3.9	3.7
3X5 A0	72	89	80	74	91	83	2.8	2.2	3.8
4X3 A0	93	191	108	96	201	114	3.2	5.2	5.6
4X4 A0	92	118	104	96	123	108	4.3	4.2	3.8
3X5 C0	72	89	80	74	91	83	2.8	2.2	3.8
3X6 C0	84	100	90	87	103	93	3.6	3.0	3.3
4X3 C0	116	189	145	121	198	153	4.3	4.8	5.5
4X4 C0	92	125	108	96	130	112	4.3	4.0	3.7
4X5 C0	77	114	86	78	104	86	1.3	-9.6	0.0

Table 4.8. Effect of Altitude on Propagation Delays using MeanOF, Slow Update

Satellite Topology	Results at 101 % Base			Results at 110 % Base			% Increase		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	97	118	108	102	124	114	5.2	5.1	5.6
2X4 A0	73	76	74	76	82	79	4.1	7.9	3.9
2X5 A0	63	64	64	66	68	67	4.8	6.3	4.7
3X3 A0	72	90	81	75	94	87	4.2	4.4	7.4
3X4 A0	54	71	62	56	72	64	3.7	1.4	3.2
3X5 A0	50	55	53	51	57	54	2.0	3.6	1.9
4X3 A0	66	84	74	69	88	78	4.5	4.8	5.4
4X4 A0	50	64	57	52	64	59	4.0	0.0	3.5
3X5 C0	50	55	53	51	57	54	2.0	3.6	1.9
3X6 C0	47	53	49	49	54	51	4.3	1.9	4.1
4X3 C0	78	88	81	81	94	88	3.8	6.8	8.6
4X4 C0	58	69	61	60	72	64	3.4	4.3	4.9
4X5 C0	49	52	51	51	53	52	4.1	1.9	2.0

Table 4.9. Effect of Altitude on Path Delays using MeanOF, Fast Update

Satellite Topology	Results at 101 % Base			Results at 110 % Base			% Increase		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	90	104	97	95	110	103	5.5	5.7	6.1
2X4 A0	73	76	74	76	80	78	4.1	5.2	5.4
2X5 A0	62	64	64	65	67	66	4.8	4.7	3.1
3X3 A0	72	90	80	75	94	84	4.1	4.4	5.0
3X4 A0	54	62	61	56	69	63	3.7	11.2	3.2
3X5 A0	50	55	52	51	57	54	2.0	3.6	3.8
4X3 A0	64	84	74	65	88	78	1.5	4.7	5.4
4X4 A0	50	61	57	52	62	59	4.0	1.6	3.5
3X5 C0	50	55	52	51	57	54	2.0	3.6	3.8
3X6 C0	47	52	49	49	53	51	4.2	1.9	4.1
4X3 C0	78	89	83	81	93	86	3.8	4.5	3.6
4X4 C0	58	67	60	60	70	62	3.4	4.5	3.3
4X5 C0	48	57	51	48	55	51	0.0	-3.6	0.0

Table 4.10. Analysis of Connectivity by Topology, Base101 Altitude.

Satellite Topology	Algorithm Alg1			Algorithm Alg2			Algorithm Alg4		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	4	4	4.0	3	3	3.0	4	4	4.0
2X4 A0	4	4	4.0	3	4	3.7	4	4	4.0
2X5 A0	3	3	3.0	3	4	3.4	3	3	3.0
3X3 A0	4	4	4.0	2	4	2.8	4	4	4.0
3X4 A0	4	4	4.0	2	4	3.0	4	4	4.0
3X5 A0	3	4	3.9	2	4	2.9	3	4	3.6
4X3 A0	3	4	3.7	2	4	2.8	–	–	–
4X4 A0	2	4	3.5	2	3	2.6	2	4	2.8
3X5 C0	3	4	3.9	2	4	2.9	3	4	3.8
3X6 C0	4	4	4.0	2	3	2.9	4	4	4.0
4X3 C0	2	4	3.4	2	4	2.7	2	3	2.3
4X4 C0	3	4	3.8	2	3	2.3	2	4	3.3
4X5 C0	3	4	3.4	2	3	2.6	2	4	2.6

Table 4.11. Propagation Delay (ms) using Multigraph with MaxMinOF, Slow Update

Satellite Topology	Alg2				Alg1				Alg4			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	105	179	146	21.9	179	210	192	9.9	179	210	192	9.9
2X4 A0	120	134	129	4.2	120	120	120	0.0	120	120	120	0.0
2X5 A0	96	129	112	10.3	88	104	96	4.7	88	104	96	4.7
3X3 A0	164	246	182	23.7	86	155	120	21.7	86	155	120	21.7
3X4 A0	119	153	136	8.4	98	140	112	13.1	98	140	112	13.1
3X5 A0	100	127	108	5.8	72	90	83	5.3	72	90	83	5.3
4X3 A0	167	242	210	17.0	93	141	111	14.9	-	-	-	-
4X4 A0	135	175	149	12.2	92	126	106	8.2	101	144	121	10.3
3X5 C0	100	127	108	5.8	72	90	83	5.3	72	90	83	5.3
3X6 C0	89	112	103	7.7	84	104	90	6.2	84	104	90	6.2
4X3 C0	162	240	213	19.4	116	192	126	18.5	161	193	186	8.3
4X4 C0	126	171	147	16.0	92	125	110	7.8	99	125	113	5.5
4X5 C0	103	123	110	7.0	80	93	86	3.7	80	118	97	14.3

Table 4.12. Propagation Delay (ms) using Multigraph with MaxMinOF, Fast Update

Satellite Topology	Alg2				Alg1				Alg4			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	106	210	183	18.3	105	191	128	16.2	105	191	128	16.2
2X4 A0	117	178	124	8.1	120	120	120	0.0	120	120	120	0.0
2X5 A0	88	165	107	11.5	88	104	96	4.7	88	104	96	4.7
3X3 A0	130	258	182	22.5	86	139	111	16.3	86	139	111	16.3
3X4 A0	98	172	137	13.1	98	128	108	10.0	98	128	108	10.0
3X5 A0	86	144	108	7.9	72	89	80	4.8	72	89	80	4.8
4X3 A0	155	321	208	27.6	93	191	108	14.5	-	-	-	-
4X4 A0	97	212	145	15.8	92	118	104	6.3	92	185	121	11.5
3X5 C0	86	144	108	7.9	72	89	80	4.8	72	89	80	4.8
3X6 C0	84	124	101	8.9	84	100	90	5.5	84	100	90	5.5
4X3 C0	-	-	-	-	116	189	145	28.4	117	193	180	12.8
4X4 C0	102	209	143	16.0	92	125	108	7.5	93	138	112	4.5
4X5 C0	87	161	113	12.4	77	114	86	5.7	80	126	94	14.7

Table 4.13. Propagation Delay (ms) using Multigraph with MeanOF, Slow Update

Satellite Topology	Alg3				Alg5			
	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	97	118	108	7.0	97	118	108	7.0
2X4 A0	73	76	74	0.8	73	76	74	0.8
2X5 A0	63	64	63	0.3	63	64	64	0.3
3X3 A0	72	90	81	5.5	72	90	81	5.5
3X4 A0	54	71	62	4.1	54	71	62	4.1
3X5 A0	50	55	53	1.6	50	55	53	1.7
4X3 A0	66	84	74	5.2	-	-	-	-
4X4 A0	50	63	57	2.9	60	75	67	3.3
3X5 C0	50	55	53	1.6	50	56	53	1.7
3X6 C0	47	53	49	1.2	47	53	49	1.2
4X3 C0	78	88	81	2.7	83	89	88	1.4
4X4 C0	58	69	61	1.7	59	68	62	2.4
4X5 C0	49	52	51	0.7	51	57	54	2.3

Table 4.14. Propagation Delay (ms) using Multigraph with MeanOF, Fast Update

Satellite Topology	Alg3				Alg5			
	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	90	104	97	4.2	90	104	97	4.2
2X4 A0	73	76	74	0.7	73	76	74	0.8
2X5 A0	62	64	64	0.4	62	64	64	0.4
3X3 A0	72	90	80	6.2	72	90	80	6.2
3X4 A0	54	62	61	2.9	54	66	61	2.9
3X5 A0	50	55	52	1.7	50	55	53	1.8
4X3 A0	66	84	74	5.1	–	–	–	–
4X4 A0	50	61	57	2.6	50	80	66	3.5
3X5 C0	50	55	52	1.7	50	55	53	1.8
3X6 C0	47	52	49	1.0	47	52	49	1.0
4X3 C0	77	89	82	4.2	78	89	87	1.7
4X4 C0	58	67	60	1.3	58	72	62	1.5
4X5 C0	48	57	51	1.5	51	58	53	2.0

Table 4.15. Propagation Delay (ms) using Multigraph with MaxMinOF and Different k values, Slow Update

Satellite Topology	Alg1, $k = 3$				Alg1, $k = 5$			
	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	148	179	168	9.1	105	113	110	2.5
2X4 A0	120	163	143	12.8	117	120	119	1.1
2X5 A0	104	119	112	4.3	88	104	94	3.9
3X3 A0	–	–	–	–	86	155	114	18.6
3X4 A0	182	197	191	5.3	98	112	101	4.0
3X5 A0	–	–	–	–	72	88	80	4.7
4X3 A0	–	–	–	–	92	141	112	13.3
4X4 A0	–	–	–	–	92	126	105	8.2
3X5 C0	–	–	–	–	72	88	80	4.7
3X6 C0	112	169	142	20.0	84	85	84	0.1
4X3 C0	196	239	212	16.7	116	123	119	2.5
4X4 C0	112	191	134	23.9	92	116	106	6.7
4X5 C0	110	145	118	12.6	80	87	83	1.9

Table 4.16. Propagation Delay (ms) using Multigraph with MaxMinOF and Different k values, Fast Update

Satellite Topology	Alg1, $k = 3$				Alg1, $k = 5$			
	Min	Max	Mean	SD	Min	Max	Mean	SD
2X3 A0	148	179	165	9.1	105	113	110	2.5
2X4 A0	120	163	143	12.4	117	120	120	1.1
2X5 A0	94	119	110	5.8	88	104	93	2.7
3X3 A0	–	–	–	–	86	129	108	12.4
3X4 A0	–	–	–	–	98	110	99	2.3
3X5 A0	–	–	–	–	72	88	80	4.8
4X3 A0	–	–	–	–	92	136	113	10.5
4X4 A0	–	–	–	–	92	118	104	6.3
3X5 C0	–	–	–	–	72	88	80	4.8
3X6 C0	–	–	–	–	84	86	84	0.5
4X3 C0	196	238	205	18.4	115	123	119	2.5
4X4 C0	112	209	141	33.8	92	117	108	5.3
4X5 C0	110	148	114	9.5	80	87	83	1.9

Table 4.17. Percentage Increase in Mean Values of Algs. over Con1

Satellites	Alg2		Alg1		Alg4	
	Slow	Fast	Slow	Fast	Slow	Fast
2X3 A0	32	66	74	16	74	16
2X4 A0	8	4	1	1	1	1
2X5 A0	20	15	3	3	3	3
3X3 A0	82	82	20	11	20	11
3X4 A0	114	39	14	10	14	10
3X5 A0	40	40	7	4	7	4
4X3 A0	114	112	13	10	–	–
4X4 A0	61	57	15	13	31	31
3X5 A0	40	40	7	4	7	4
3X6 A0	22	20	7	7	7	7
4X3 A0	116	–	22	40	80	74
4X4 A0	59	55	19	17	22	19
4X5 A0	41	44	10	10	24	20

Table 4.18. Percentage Improvement of Fast over Slow for Different Algs.

Satellite Topology	Alg1			Alg4			Alg3			Alg5		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
2X3 A0	71	10	50	70	10	50	7	14	11	8	13	11
2X4 A0	0	0	0	0	0	0	0	0	0	0	0	0
2X5 A0	0	0	0	0	0	0	2	0	0	0	0	0
3X3 A0	0	12	8	0	11	8	0	0	1	0	0	1
3X4 A0	0	9	4	0	9	4	0	14	2	0	8	2
3X5 A0	0	1	4	0	1	4	0	0	2	0	0	0
4X3 A0	0	7	3	–	–	–	0	0	0	–	–	–
4X4 A0	0	7	2	10	8	0	0	3	0	20	-6	2
3X5 C0	0	1	4	0	1	4	0	0	2	0	1	0
3X6 C0	0	4	0	0	4	0	0	2	0	0	1	0
4X3 C0	0	2	-15	37	0	3	1	-1	-1	6	0	1
4X4 C0	0	0	2	7	-10	1	0	3	1	2	-6	0
4X5 C0	4	8	0	0	-7	3	2	-9	0	0	-2	2

Table 4.19. Percentage Improvement of All Pairs Least Cost Graph vs Minimum Cost Spanning Tree, Slow Update

Satellite Topology	Alg4/Alg1			Alg5/Alg3		
	Min	Max	Mean	Min	Max	Mean
2X3 A0	0	0	0	0	0	0
2X4 A0	0	0	0	0	0	0
2X5 A0	0	0	0	0	0	1
3X3 A0	0	0	0	0	0	0
3X4 A0	0	0	0	0	0	0
3X5 A0	0	0	0	0	0	0
4X3 A0	–	–	–	–	–	–
4X4 A0	10	14	14	20	19	17
3X5 C0	0	0	0	0	1	0
3X6 C0	0	0	0	0	0	0
4X3 C0	38	1	47	6	1	9
4X4 C0	8	0	-3	2	-2	2
4X5 C0	0	26	12	4	10	6

Table 4.20. Percentage Improvement of All Pairs Least Cost Graph vs Minimum Cost Spanning Tree, Fast Update

Satellite Topology	Alg4/Alg1			Alg5/Alg3		
	Min	Max	Mean	Min	Max	Mean
2X3 A0	0	0	0	0	0	0
2X4 A0	0	0	0	0	0	0
2X5 A0	0	0	0	0	0	0
3X3 A0	0	0	0	0	0	0
3X4 A0	0	0	0	0	7	0
3X5 A0	0	0	0	0	0	0
4X3 A0	–	–	–	–	–	–
4X4 A0	0	56	16	0	31	16
3X5 C0	0	0	0	0	0	1
3X6 C0	0	0	0	0	0	0
4X3 C0	1	2	24	1	0	6
4X4 C0	1	10	4	0	8	3
4X5 C0	4	10	9	6	2	4

Table 4.21. Percentage Improvement of Alg1 for $k = 5$ over $k = 4$ for Slow and Fast Updates

Satellite Topology	$k = 5$ / $k = 4$, Slow			$k = 5$ / $k = 4$, Fast		
	Min	Max	Mean	Min	Max	Mean
2X3 A0	70	86	75	0	69	16
2X4 A0	3	0	1	3	0	0
2X5 A0	0	0	2	0	0	3
3X3 A0	0	0	5	0	8	3
3X4 A0	0	25	10	0	16	16
3X5 A0	0	2	4	0	1	0
4X3 A0	1	0	0	0	40	-5
4X4 A0	0	0	0	0	0	0
3X5 C0	0	2	4	0	1	0
3X6 C0	0	22	7	0	16	7
4X3 C0	0	56	6	1	53	22
4X4 C0	0	8	4	0	7	0
4X5 C0	0	7	4	-4	31	4

CHAPTER 5
CONCLUSIONS

5.1 Importance of Research Topics

In this dissertation two specific areas associated with intersatellite communication have been examined. The first area, that of satellite topologies, is particularly important for the construction of a satellite network. The selection of a topology greatly affects the characteristics of the network. The altitude of the system will influence the network propagation delay, the time which satellites are visible to a point on the ground, the coverage properties, and the number of satellites required to cover the earth. The orientation of the satellite topology affects both the connectivity and coverage characteristics. Since the communication links are determined by the topology it is clear that selection of any topology will influence link selection and therefore the effectiveness of routing. The second area, a study of link assignment properties, is the other major feature of satellite networks. Unlike stationary networks and packet radio networks the predictable nature of satellite topologies offers a new dimension to satellite networking. The first level at which predictability may be used is at the link assignment level. It is therefore important to ascertain the effects of using this unique property.

5.2 Major Results

In the area of satellite topologies this dissertation has made the following contributions:

- A strong relationship is established between coverage and connectivity. In particular an integer-valued function is defined for network survivability which incorporates both ground coverage and network connectivity. By establishing that coverage is a stronger statement than connectivity the definition for survivability is reducible to one measure.

- A graph theoretic formal model for a restricted class of networks known as the (N_p, N_s) zero phase topology is constructed. The (N_p, N_s) topologies have the characteristics that all satellite planes are rotated about one axis only, and the satellites are well spaced away from the poles. The significance of the graph model is that by using only a subset of the visible edges a satellite topology may be reduced to a set of formal constraints on a time-varying graph in which only the edges and propagation delays are used.

- A relationship between the node-connectivity of the (N_p, N_s) graph model and coverage thus is established. It is proved that the the minimum connectivity of the (N_p, N_s) zero phase system is at least four. This result is especially significant in view of the previous observations between coverage and connectivity. It states, in effect, that a single coverage N_p, N_s satellite topology will be at least four-node connected.

- The construction of a simulation to characterize ground coverage for a general multilayered satellite topology. Using this simulation the coverage characteristics of some simple satellite topologies are investigated. The simulation is used to verify that the connectivity and coverage of the (N_p, N_s) system behave as expected.

In the area of link assignment this dissertation has made the following contributions.

- The development of a link assignment algorithm in which the complexity of the original problem is significantly reduced by using properties inherent in satellite topologies to limit the number of links to be assigned.
- A close relationship between the multigraph algorithm and the graph theoretical model are established by proving that the multigraph algorithm constructs a connected graph in the first iteration when applied to the (N_p, N_s) graph models.
- The results of running the multigraph algorithm on a number of topologies are used to determine the following effects: the change in propagation delay of the network as the altitude increases; the difference in propagation delay between computing the link assignment only when necessary, as opposed to computing it frequently; the change in performance of different link selection procedures; and the effect of changing k, the maximum degree of the graph, on connectivity and propagation delay.

From the results of the link assignment study the following conclusions are drawn with respect to the topologies analyzed.

- There is little performance improvement by frequently retargeting the communications links. It is anticipated that the overheads associated with link retargeting would remove any improvement which might have resulted.
- The effect of increasing or decreasing k by one from its base value of four has a considerable effect on both the connectivity and the performance.

- Small increases in altitude cause only few percent increase in the propagation delay. In fact it is possible that due to increased satellite visibility the propagation delay may improve when the altitude is increased.

5.3 Significance of this Work

Technology has advanced to the point at which communication systems using satellite networks may become a reality in the near future. In this new field there are many unanswered questions regarding the best topologies, physical communication media, and networking strategies to use. The research in this dissertation is directed toward investigating aspects of both the topologies and networks. The results in the last section suggest there are significant performance differences, depending on the assumptions made, and strategies selected. In particular the designers of any satellite system should pay careful attention to the measures analyzed in this work.

Apart from the area of satellite networking, the technique of using multigraph reduction to reduce the complexity may be applied to other networks in which the motion of the nodes is periodic and predictable. This may possibly have applications in robot control within a factory.

5.4 Future Research Directions

A problem that has hampered progress in satellite networking is that the subject matter is so new that the number of research options is overwhelming. Even using the models presented in this paper there are many options open for further study. The effects of using different preload structures (i.e. structures other than a ring) were not investigated, even though these would probably improve the performance. No attempt was made to use perturbation heuristics to improve the starting networks which would also have improved the network performance. Another direction which would yield fruitful results would be to incorporate into the model some measure of the interruption caused to the network during link reassignment.

Using this same multigraph structure the analysis could be extended by assigning capacities to the links and using network flow algorithms to measure the performance. This method would clearly require studies on the expected traffic into the network, the capacity of the links, and a queueing model for the satellites.

Another branch of satellite networking would be to use a completely different model. One model under review divides the satellite shell into regions. By doing this, and using the regions as local communication boundaries, a natural two-level hierarchy is established. This approach will certainly give different characteristics, although one might expect the propagation delays to be similar.

To summarize, the study of satellite networks is still very much in its infancy. The unique features of satellite topologies (*viz.* predictability, and the relationship between coverage and connectivity) offer many new research opportunities in this field of communications.

APPENDIX A
KEY TO GROUND COVERAGE ALGORITHM

A.1 Description of Structures

EarthGrid: three 1° azimuth by 1° elevation arrays giving the x, y, z coordinates of a grid covering the surface of the earth.

CurrentPlane: an array with the same dimension as above. It is used to maintain the number of times that each grid point has been covered during the current iteration.

MinPlane: an array with the same dimensions as above. It is used to maintain the minimum coverage over all time for each grid point.

A.2 Explanation of Procedures

InitializeSatellitePositions: this procedure uses the description of a satellite topology contained in the input file to determine the initial positions of each of the satellites.

UpdateSatellitePlanes: this procedure moves each satellite in space by the distance it would travel in one time increment.

IncrementCurrentPlaneCoverage: this procedure increments the current coverage statistics for the particular region considered at that point.

APPENDIX B
GROUND COVERAGE DISPLAYS

In order to gain a conceptual feel for different satellite topologies the coverage program discussed in Chapter 3 permits the user to save the final coverage grids as a computer file. Two different *backend programs* were written which would take this file and generate graphical representations of the data. These programs are briefly discussed below along with annotated output from different satellite topologies.

B.1 2D Density Plots

One backend program generates ordinary text files which may be printed on a line printer. A complete picture is generated on 9 sheets of 132 column paper. The grid is unscaled; however the printable characters for each grid point are selected according to their density to allow the user to "see" the areas of differing density. In figure B.1 a sample is given (reduced to 17.85% of the original) of $2/9^{th}$s of a full plot, in which the density contours may be seen.

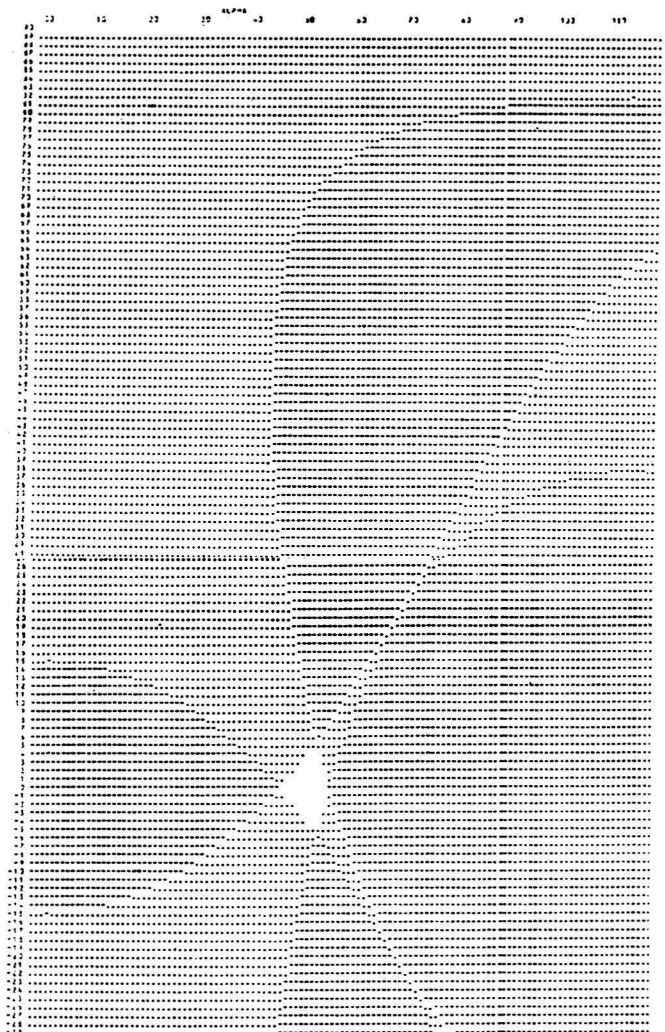


Figure B.1. Reduced Partial Plot of the 2D Density Contours for $N_p = 2, N_s = 3$, Alt. = 6378 km

B.2 3D Solid Objects

A more sophisticated backend program was written to take the output from the coverage program and generate a solid object rendition of the earth. The sphere is colored according to the density of coverage ranging from either red (0 coverage) or from green to blue in different shades proportional to the coverage. Sample figures are not included in this dissertation.

APPENDIX C
KEY TO LINK ASSIGNMENT ALGORITHM

C.1 Description of Structures

GraphsLeftToProcess: a variable used to determine if the input file has finished.

NumberOfLinks: a variable containing the number of links (i.e. edges) which are to be used in the construction of the graph.

EdgesLeftToExamine: a variable containing the number of links which remain to be used.

EdgesUsedInOutputGraph: a variable containing the number of edges which have been assigned to the graph under construction. This will differ from (NumberOfLinks - EdgesLeftToExamine) by the number of edges that have been discarded.

SourceEdge: a variable containing the source satellite for the current iteration.

DestEdge: a variable containing the destination for the current iteration.

Delay: a variable containing the propagation delay between the source and destination satellites.

C.2 Explanation of Procedures

InitializeGraphInfo: this procedure is responsible for initializing the system variables and data structures.

ReadPreSelectInfo: in this procedure details describing the user-defined links are read in.

InitializeMultiGraphStacks: this procedure is responsible for setting the multigraph stacks to zero.

ReadGraphInfo: in this procedure the next graph from the input stream is read in. The source-destination cost matrices are updated accordingly.

BuildLinkCostList: using the current graph, a list is constructed of the edges and associated propagation delays.

SortLinkCostList: the list constructed in BuildLinkCostList is sorted according to user specifications (usually sorted by least cost).

MGMap: a procedure which takes as input a satellite vertex number and determines the plane number to which it belongs.

MultiPush: a procedure which takes as input two satellite vertex numbers, two planes, and an associated propagation delay between them and incorporates the appropriate data in the multigraph structure.

CopyTopsToTempGraph: copies the "top layer" of the multigraph structure to a cost matrix.

SelectEdges: an algorithm which selects from the current cost matrix, some subset of the edges; these edges will be added to the output graph if possible.

CopyResultsToOutputGraph: using the edges selected by SelectEdges this procedure updates the output graph ensuring that no violations take place.

PopAppropriateStackElements: removes the edges used in the output graph and those edges which would be unusable in the future due to degree constraints.

InformUser: a procedure which informs the user of the status of the multigraph algorithm.

REFERENCES

[1] D. J. Baker. Distributed control of broadcast radio networks with changing topologies. *Proc. of the IEEE Infcom'83, San Diego, CA*, 49–55, 1983.

[2] D. J. Baker and A. Ephremides. The architectural organization of a mobile radio network via a distributed algorithm. *IEEE Trans. on Commun.*, COM-29(11):1694–1701, November 1981.

[3] P. D. Bartoli. The application layer of the reference model of open systems interconnection. *Proc. of the IEEE*, 71(12):1404–7, December 1983.

[4] E. E. Basch and T. G. Brown. Introduction to coherent optical fiber transmission. *IEEE Communications Magazine*, 23:22–3, May 1985.

[5] R. R. Bate, D. D. Mueller, and J. E. White. *Fundamentals of Astrodynamics.* Dover Publications, New York, NY, 1971.

[6] L. Biederman, F. Ghazvinian, Yu-Teh Su, M. Shiva, and T. Schonhoff. *Final Report–Evaluation of Optical Communication Link for Crosslink Communication.* Lincom Corp., Los Angeles, CA, 1986.

[7] K. Brayer. Adaptive networking of variable topology satellite networks. *Sixth International Conference on Digital Satellite Communications, Phoenix, AZ*, VIII.14–VIII.20, September 1983.

[8] K. Brayer. Routing in a mobile network–fact or fantasy. *Data Communications*, 145–159, August 1983.

[9] V. G. Cerf and R. E. Lyons. Military requirements for packet-switched networks and their implications for protocol standardization. *Computer Networks*, 7(5):296–307, October 1983.

[10] J. W. Conard. Services and protocols of the data link layer. *Proc. of the IEEE*, 71(12):1378–83, December 1983.

[11] R. K. Crane. Prediction of attenuation by rain. *IEEE Trans. on Commun.*, COM-28(9):1717–35, September 1980.

[12] E. W. Dijkstra. A note on two problems in connection with graphs. *Numerical Mathematics*, 1:269–71, October 1959.

[13] E. J. Dutton and H. T. Dougherty. *Modelling the Effects of Cloud and Rain upon Satellite- to-Ground System Performance.* Technical Report, Office of Telecommunications, Boulder, CO, March 1973.

[14] I. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency of network flow problems. *Journal of the ACM*, 19:248–64, 1972.

[15] N. Ehrlich. The advanced mobile phone service. *IEEE Communication Magazine*, 17(3):9–15, March 1979.

[16] W. F. Emmons and A. S. Chandler. OSI session layer: services and protocols. *Proc. of the IEEE*, 71(12):1397–400, December 1983.

[17] S. Even and R. E. Tarjan. Network flow and testing graph connectivity. *SIAM, Journal of Computing*, 4:307–518, 1975.

[18] F. W. Floyd. Algorithm 97: shortest path. *Communications of the ACM*, 5:345, 1962.

[19] L. R. Ford and D. R. Fulkerson. *Flows in Networks.* Princeton University Press, Princeton, NJ, 1962.

[20] H. Frank and L. Fish. *Communication, Tranmission and Transportation Networks.* Addison-Wesley, Reading, MA, 1971.

[21] L. Fratta, M. Gerla, and L. Kleinrock. The flow deviation method: an approach to store-and-forward communication network design. *Networks*, 3:97–133, 1973.

[22] R. M. Gagliardi. *Satellite Communications.* Lifetime Learning Publications, Belmont, CA, 1984.

[23] Z. Galil and A. Naamad. Network flow and generalized path compression. In *11th ACM Symposium on Theory of Computing*, 1979.

[24] M. R. Garey and D. S. Johnson. *Computers and Interactability - A Guide to NP Completenss.* Freedman & Co., New York, NY, 1979.

[25] Tri. T. Ha. *Digital Satellite Communications.* Macmillan Publishing Co., New York, NY, 1986.

[26] L. Hollis. OSI presentation layer activities. *Proc. of the IEEE*, 71(12):1401–3, December 1983.

[27] A. Itai and Y. Shiloach. Maximum flow in planer graphs. *SIAM Journal of Computing*, 8:135–50, 1979.

[28] R. E. Kahn. The organization of computer resources into a packet radio network. *IEEE Trans. on Commun.*, COM-25(1):169–178, January 1977.

[29] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman. Advances in packet radio technology. *Proc. of IEEE*, 66(11):1468–96, November 1978.

[30] K. Kammerlander. C900- an advanced mobile radio telephone system with optimum frequency utilization. *IEEE Jounal on Selected Areas in Communications*, SAC-2(4):589–597, July 1984.

[31] A. V. Karzanov. Determining the maximum network flow in a network by the method of preflows. *Soviet Mathematics*, 15:434–7, 1974.

[32] S. C. Kleene. *Representation of events in nerve nets and finite automata, Automata Studies.* Princeton University Press, Princeton, NJ, 1956.

[33] L. Kleinrock. *Queuing Systems, Vol. 2: Computer Applications.* John Wiley, New York, 1976.

[34] D. Kleitman. Methods for investigating the connectivity of large graphs. *IEEE Trans. on Circuit Theory*, CT-16:232–233, May 1969.

[35] K. G. Knightson. The transport layer standardization. *Proc. of the IEEE*, 71(12):1394–96, December 1983.

[36] J. B. Kruskal. On the shortest spanning subtree of a graph and the travelling salesman problem. *Proc. of the Amer. Math. Society*, 7:48–50, 1956.

[37] S. H. Lin. Empirical rain attenuation model for earth-satellite paths. *IEEE Trans. on Commun.*, COM-27(5):812–7, May 1979.

[38] V. M. Malhotra, M. Pramodh Kumar, and S. N. Maheshwari. An $o(n^3)$ algorithm for finding maximum flows in networks. *Information Processing Letters*, 7:277–8, 1978.

[39] F. McClelland. Services and protocols of the physical layer. *Proc. of the IEEE*, 71(12):1372–7, December 1983.

[40] C. McLochlin, C. Ward, Y. C. Chow, J. N. Wilson, and R. Newman-Wolfe. *Closed Forms and Algorithms for Determining Propoagation Delay of Low Altitude Satellite Networks.* TR 87-2, University of Florida, Gainesville, FL, 1987.

[41] C. McLochlin, C. Ward, Y. C. Chow, J. N. Wilson, and R. Newman-Wolfe. Determining the delay and reliability of low altitude satellite network topologies using simulation. In *Symposium on the Simulation of Computer Networks*, Colorado Springs, CO, August 4-7 1987.

[42] C. McLochlin, C. Ward, Y. C. Chow, J. N. Wilson, and R. Newman-Wolfe. Optimizing the delay and reliability of low altitude satellite network topologies. In *Milcom'87*, Washington, DC, October 19-22 1987.

[43] K. Melhorn. *Data Structures and Algorithms 2: Graph Algorithms and NP-Completeness.* Springer-Verlag, New York, NY, 1984.

[44] R. Monte. Space optical communications with the Nd: YAG laser. *Proc. of the IEEE*, 66(3), March 1978.

[45] W. S. Newell. *Optimal Flow through Networks.* Technical Report, MIT, Cambridge, MA, 1958.

[46] Sir Isaac Newton. *Principia–The motion of bodies.* University of California Press, Berkeley, CA, 8th edition, 1974. Translated by Andrew Motte in 1729.

[47] P. L. Rice and N. R. Holmberg. Cumulative time statistics of surface point-rainfall rates. *IEEE Trans. on Commun.*, COM-21(10):1075–91, October 1973.

[48] R. Small. *An Account of the Astronomical Discoveries of Kepler.* University of Wisconsin Press, Madison, WI, 1963.

[49] W. Stallings. *Data and Computer Communications.* Macmillan Publishing Co., New York, NY, 1985.

[50] Tactical Information Exchange. Tactical information exchange (tei) framework development for 1990's frame. *Final Report Tactical Information Exchange (TIE) Working Group and Senior Review Group and Senior Review Group Participants R&D Assoc.*, October 1981.

[51] A. S. Tanenbaum. *Computer Networks.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.

[52] R. E. Tarjan. *Data Structures and Network Algorithms.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.

[53] C. Ware. The OSI network layer: standards to cope with the real world. *Proc. of the IEEE*, 71(12):1384–87, December 1983.

[54] A. C. Yao. An $o(|e| loglog |v|)$ algorithm for finding minimum spanning trees. *Information Processing Letters*, 4:21–3, 1975.

BIOGRAPHICAL SKETCH

Christopher Ward was born in London, England, in 1959. He received a B.S. degree in computer science from the Hatfield Polytechnic, Hatfield, Hertfordshire, England, in 1983. He enrolled in the M.S. degree in computer science at the University of Florida, Gainesville, Florida, in 1983 where he specialized in mobile communications. He received his M.S. degree in 1984 and has since been working toward a Ph.D. At present he is involved in research with Randy Chow on a funded project to investigate link assignment strategies and routing algorithms in inter-satellite networks.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Y.C. Chow

Yuan-Chieh Chow, Chairman
Associate Professor of
Computer and Information Sciences

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles V. Shaffer

Charles V. Shaffer
Professor Emeritus of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Fredrick J. Taylor

Fredrick J. Taylor
Professor of Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Rick L. Smith
Associate Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph N. Wilson
Joseph N. Wilson
Assistant Professor of
Computer and Information Sciences

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1987

Herbert A. Bevis
Dean, College of Engineering

Dean, Graduate School