Capacity Dimensioning and Routing for Hybrid Satellite and Terrestrial Networks

by

Lillian Lei Dai

B.Sc., Electrical Engineering and Computer Science
University of Calgary, Canada (2000)

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2002

[June 2002]

Author_______________________________

Department of Electrical Engineering and Computer Science
May 10, 2002

Certified by_______________________________

Vincent W. S. Chan
Joan and Irwin Jacobs Professor of Electrical Engineering
& Computer Science and Aeronautics & Astronautics
Director, Laboratory for Information and Decision Systems
Thesis Supervisor

Accepted by_______________________________

Arthur C. Smith
Chairman, Department Committee on Graduate Students

Capacity Dimensioning and Routing for Hybrid Satellite and Terrestrial Networks

by

Lillian Lei Dai

Submitted to the Department of Electrical Engineering and Computer Science on May 10, 2002, in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering and Computer Science

Abstract

For future commercial broad-band data satellite networks, satellite network topology, link capacity, and routing have major impacts on the cost of the network and the amount of revenue the network can generate. In this thesis, we establish a mathematical framework, using two-stage stochastic programming, to assist network designers in selecting the most cost-effective network topology for data applications. The solution to the stochastic programming formulation gives optimal link capacities and an optimal routing strategy for different network topologies, taking into account uncertainties in long-term aggregate traffic statistic estimation. For several classes of satellite topologies of interest, analytical solutions have been derived. In particular, we give the optimal topology selection criteria for a general GEO satellite network and identify regions where a hybrid satellite-terrestrial topology is the most cost effective. For LEO and MEO satellite networks, comparisons between different routing strategies for polar constellations with and without seam are presented. The analytical solutions we have derived capture the salient network design parameters and their relationships. These analyses offer much insight into the design tradeoffs for a hybrid satellite-terrestrial network.

Thesis Supervisor: Vincent W. S. Chan
Joan and Irwin Jacobs Professor of Electrical Engineering
& Computer Science and Aeronautics & Astronautics
Director, Laboratory for Information and Decision System

Acknowledgements

I would like to express my sincere gratitude towards my advisor, Professor Vincent W.S. Chan, for giving me guidance and encouragements over the past two years. His insights nurtured my approach to research and helped to refine my thesis. Professor Chan's vision and commitment for excellence in all aspects of engineering continue to be a tremendous source of inspiration for me. I am also grateful to DARPA and MIT for providing me with financial support, without which, this work would not be possible.

I would like to thank all my friends[1] for their companionship and support; in particular, I thank my friends at LIDS for the many insightful discussions and for making my life enjoyable at MIT. My gratitude also goes towards my professors at MIT and at the University of Calgary. I am forever indebted to their guidance, support, and encouragements.

Lastly, I would like to thank my family - mom, dad, and Linda - for believing in me and always being there for me. No words can express my gratitude for the unconditional love they have given me throughout my life. I would also like to thank my grandparents, uncles, aunts, and cousins for their love and support.

[1] Special thanks to: Robert Chiew for love and encouragements; Lalitha Sankaranarayanan for mentorship; Jason Bau, Serena Chan, Li-Wei Chen, Patrick Choi, Todd Coleman, Bishwaroop Ganguly, Anand Ganti, Kyle Guan, Danielle Hinton, Ari Libarikian, Chunmei Liu, Desmond Lun, Irina Medvedev, Etty Shin, Jun Sun, Richard Thommes, Frank Wang, Hungjen Wang, Guy Weichenberg, Yong-Gang Wen, Edmund Yeh, Won Yoon; SEBC committee members; Bin Hu, Janice Lee, Vinson Lee, Brenda Ng, Steve Wang, Yi Yu, Noreen Zaman, Johnny Zhang.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Satellites systems have been used to provide various telecommunication services for more than half of a century. The first generation systems are geosynchronous earth orbit (GEO) satellites providing TV broadcasting and telephony trunking services between large, fixed Earth stations. Second generation systems have added the element of mobility, employing large mobile terminals for communication between ships, aircrafts and land vehicles. In the 1990s, there has been a surge of interest in developing third generation personal mobile satellite systems. These systems are intended to provide personal communications services to users with small, hand-held terminals. Many third generation satellite systems have been proposed in the USA to provide global telephony and data services [1]. The Iridium and Globalstar systems, designed to provide global, personal mobile telephony services, came into service in the late 1990s. Unfortunately, both businesses have declared bankruptcy within one or two years of service. Since the failure of the two businesses, many other proposed systems have either been cancelled or have been put on hold.

Despite commercial failure of the two satellite businesses, it is foreseeable that satellite communication systems can play an important role as a part of the Next Generation Internet (NGI), particularly in providing ubiquitous communication for multimedia and high data rate networking applications. The large coverage areas, rapid deployability, and inherent support for mobile services, make satellite systems a likely candidate to be integrated into the future telecommunication infrastructure. It has also become clear, however, that in order to realize future commercial broad-band data satellite networks, it is critical for satellite network designers to consider both technical and economical design factors to ensure that the satellite network is technically robust and economically competitive in the commercial market.

Several satellite network design factors have major impacts on its commercial success: physical topology, link capacities, and routing strategy. These design factors greatly influence the cost of the system as well as the amount of revenue the network can generate. Recent research activities have placed much emphasis on the design of the satellite segment, with a noticeable gap on a complete hybrid satellite-terrestrial network design methodology [2], [3], [4]. There is a need to study how satellite and terrestrial networks can complement each other to improve the efficiency of both networks.

In general, given a particular market the satellite network serves, there exist many different satellite network topologies capable of providing the same type of services, but at very different costs. A satellite network topology specifies satellite altitude, constellation, the number of satellites and gateways, and how communication nodes are interconnected. For example, the Iridium and Globalstar satellite systems both serve the satellite mobile telephony market, but Iridium uses inter-satellite links (ISLs) whereas Globalstar does not. One of the most important goals for satellite network design is to identify the most economical satellite network topology for the intended application.

Clearly, the cost of gateways and ISLs play an important role in determining the optimal network topology. There are three scenarios: 1. if the cost of providing ISLs is very low, then these links should be dimensioned to carry the worst-case traffic demand. 2. if the ISLs are very expensive, then it would be more economical to not use ISLs at all. In this case, all of the traffic must be routed by the terrestrial fiber network at an extra cost. 3. if neither of the previous two extreme scenarios holds, then a hybrid satellite-terrestrial network utilizing ISLs and the terrestrial fiber network is more economical. The ISLs may be dimensioned to support average traffic demand in order to achieve high utilization of the ISLs, while the excess traffic may be routed by the terrestrial fiber network at a cost. The problem of optimal satellite topology selection, taking into account system cost, capacity dimensioning and routing, has not been studied previously.

In this thesis, we establish a mathematical framework to analyze and compare different satellite network topologies from an economic perspective. We assume that the network designer has a reasonable prediction of future traffic demand, albeit with some uncertainty. For a particular satellite network topology, two-stage stochastic programming is used to find optimal link capacities and an appropriate routing strategy to minimize an effective system cost [5]. The effective system cost is defined to be the sum of satellite network investment cost, cost of routing using terrestrial links, and an opportunity cost for rejecting excess input traffic, subject to quality of service (QoS) constraints. Once effective system costs are obtained for different topologies, the most cost-effective topology may be selected for deployment. We derive analytical solutions for several GEO, medium earth orbit (MEO), and low earth orbit (LEO) topologies and obtain criteria for optimal topology selection.

The contents of this thesis are organized as follows: In Chapter 2, we present some background on the current research in the field. Chapter 3 presents simple parametric link cost models used for system cost optimization. In Chapter 4, the stochastic optimization problem is formulated for general satellite networks. This formulation is useful for obtaining optimal link capacities and an optimal routing strategy for different satellite topologies. In Chapter 5 and 6 we analyze GEO, MEO and LEO system design in detail. In particular, we show optimal topology selection criteria based on link cost parameters. Chapter 7 summarizes the major results and presents a discussion for possible future research directions.

Chapter 2

Background

The primary satellite design issues we wish to study are: optimal satellite network topology selection based on system cost, link capacity dimensioning for the various links in a satellite network, and routing for different classes of traffic under uncertain input traffic conditions. In this chapter, we first give a summary of some of the existing and proposed satellite network topologies. Next, we discuss some of the current system cost modeling techniques. Finally, we discuss some of the link capacity dimensioning and routing techniques in literature. These discussions will set the stage for the analysis we present in subsequent chapters.

2.1 Satellite Topologies

As mentioned before, satellite network topology is a critical design factor that influences the cost of the satellite network. By satellite network topology, we mean satellite altitude, constellation, the number of satellites and gateways, and how communication nodes are interconnected. A summary of some of the commercial mobile satellite systems (in service and proposed) is given in Table 2-1.

Table 2-1 Commercial Satellite Systems (In service and proposed)

Primary Service	Satellite Systems	Coverage	Altitude	Trajectory	Satellite ISL	# Sats	# GWs	Cost ($B)
Voice	Iridium[2]	Global	LEO 780 km	Circular Polar	RF 4-way ISL	66	12	4.7
	Globalstar[3]	Global	LEO 1414 km	Circular Walker	None	48	60*	2.5
	ACeS[4]	Regional	GEO	Stationary	None	1		
	Thuraya[5]	Regional	GEO	Stationary	None	2		
Data	Teledesic[6]	Global	LEO	Circular Polar	Optical 8-way ISL	288		
	Spaceway[7]	Regional	GEO	Stationary	None	2		5
	Astrolink[8]	Global	GEO	Stationary	Optical	9		4

From Table 2-1, it is clear that there exist many different satellite topologies capable of providing similar types of services. Given the numerous possibilities for satellite system topology, a system designer must identify the components that are most costly and optimize the system to provide desired quality of service at low costs. Aside from cost, there are also numerous other factors that may influence topology selection. Some of the non-technical factors include: data security (may favor the use of ISLs since signals can bypass some countries), gateway placement constraints due to political reasons, inability to service some countries due to governments protecting its own local providers, lawful intercept requirement imposed by some countries, etc [6]. Although these factors are extremely important, and must be considered when designing a satellite system, we ignore their effects in our analysis in order to focus on the technical problem at hand.

---

[2] Iridium LLC began service in May 1998 and declared bankruptcy in Aug. 1999. Iridium Satellite LLC acquired all operating assets of Iridium LLC in March 2001 and resumed commercial service.

[3] Globalstar began service in Oct. 1999 and filed for bankruptcy protection in Feb. 2002. *approximately 60 gateways proposed; 25 gateways in service as of April 2002.

[4] Asia Cellular Satellite System began service in Sept. 2000. It services 24 countries in Southeastern Asia.

[5] Thuraya began service in April 2001. It intends to serve 99 countries in Europe, North and Central Africa, the Middle East, Central Asia and the Indian Subcontinent.

[6] Teledesic is targeted to begin service in 2005. The numbers provided in Table 2-1 are for its original plan. Teledesic has recently changed the design of its constellation to 30 MEO satellites, after acquiring the ICO system. No other details are publicly available.

[7] Spaceway system is targeted to begin service in 2002.

[8] Astrolink system is targeted to begin service in 2003.

Several authors have studied specific satellite network topologies and some have proposed new topologies; however there lacks a concrete guideline on selecting the optimal topology. In [7], the author gives a concise summary of the basic constellation types: Geosynchronous, Walker, Streets-of-Coverage, and Elliptical. Although orbital mechanics impose constraints on satellite constellations, there remain a large number of satellite constellations to select from. In theory, there is a continuum of possibilities for satellite network topology with satellites in different constellations either interconnected with one another or not. Service requirements such as delay, path loss, coverage area, availability, minimum number of satellites for global coverage and satellite life time in different orbits serve to reduce the space of topologies. The remaining satellite topologies can be used to satisfy all of the service requirements. The problem is then to select the optimal topology from this feasible topology set. We address the optimal topology selection problem in this thesis from both technical and economical perspectives.

2.2 Cost Analysis

System cost analysis is critical for any commercial system design. Typically, there are three ways to estimate system cost:

1. Analyze Component Costs – this method is applicable after a preliminary design has been done and the actual critical components have been identified.
2. Comparing Similar Systems and Extrapolate – this method is valid only if similar systems exist and can be use as design guidelines. For broad-band data satellite networks, this method is not applicable since there are no available systems to extrapolate costs from.
3. Parametric Analysis – this method formulates mathematical relationships between salient technical parameters and cost. It can be used effectively to study system tradeoffs. Since the design of a broad-band data satellite network is still in the research stage, this is the most logical cost analysis method for this study.

Cost studies have been performed extensively in industry. NASA has developed an Aerospace Small Satellite Cost Model (SSCM) to evaluate the cost of designing, building, and testing of a modern small satellite. There are also handbooks providing detailed guidelines on parametric cost estimation [8]. The cost models used typically involve a large number of parameters and produce numerical results that depend heavily on input parameters. Since broad-band satellite system will utilize new technological building blocks, the cost models developed in previous studies will not be suitable for our study.

Several cost and system optimization studies have been done for satellite planning. In [9], the authors formulated cost models based on actual costs from several commercial, military and special purpose satellites. Optimal technical parameters such as satellite power, mass, and number of transponders have been used in the cost model. With the objective of profit maximization subject to technical parameter and cost constraints, a geometric programming algorithm was used to generate numerical results for specific input parameters. This study was performed for a single satellite system and does not generalize to a network of satellites. Furthermore, the cost models used may not be applicable to future broad-band satellite systems since ISL cost was not considered in their model.

In [4], ISL cost was formulated as a function of link distance and capacity. This model was used to optimize interconnections between satellites and to study the effect of different traffic distributions on the space segment. For the purpose of the current study, we would like to take a similar approach as used in [4] whereby critical cost components are identified and characterized based on simple models with a few salient parameters. This is considered to be a desirable approach since the goal of the study focuses on network aspects which depend on the cost of interconnections between satellites and the gateways. By formulating cost as a function of link distance and capacity, much insight may be gained on optimal system topology.

2.3 Link Capacity Dimensioning

There have not been many studies on link capacity dimensioning for a data satellite network. In [2], some techniques for spot beam and ISL capacity dimensioning are presented. Relative link capacity and optimal capacity dimensioning for the entire network are not considered. Moreover, none of the dimensioning techniques examine relative monetary costs between links which we consider to be a critical factor to network capacity design.

For general network dimensioning problems, there are two main formulations. For circuit-switched telephony networks, designers traditionally use Poisson input model and Erlang formulas to compute the number of circuits required on a link subject to probability of blocking constraints [10]. In a network setting, fixed routing strategy and link independence are often assumed to keep the problem tractable. For packet-switched networks, flow models are often used for link dimensioning. For a known input flow demand, the goal is to optimally route the flow in order to minimize system cost subject to delay constraints. As described in [11], optimizing network with respect to both link capacity and routing is difficult. Hence, heuristic methods are often used to iteratively design link capacity and routing in order to find the optimal system topology.

Both formulations above are used to obtain a single-point optimal solution for a known traffic pattern. In reality, traffic pattern is usually not known accurately in advance. For terrestrial networks, this does not present a significant problem since there is flexibility to modify the existing systems, upgrade components, and to expand the capacity of a network when traffic pattern changes or if the network is not utilized efficiently. For a satellite network, however, it is very difficult to modify the satellite components once satellites are launched into orbit. Hence, when designing a satellite system, it is critical to consider the uncertainties in traffic pattern. This problem may be addressed by two-stage stochastic programming [12] as we will explore in this thesis.

2.4 Optimal Routing Over Hybrid Networks

Recent research in routing over satellite networks have focused on routing algorithms for LEO satellite networks utilizing ISLs. Most of the works consider static or topology adaptive routing over the space segment assuming that the link capacities are given [2]. As we have described in Section 2.3, routing and capacity dimensioning are two tightly coupled problems that strongly influence each other. Hence to study routing over satellite networks, one should consider the impact of routing schemes on link capacities and the cost of the entire satellite network.

For a satellite network interconnected with the terrestrial network, the terrestrial facilities may be used to carry some of the satellite traffic. This allows the ISLs to be dimensioned with lower capacity which reduces the cost of the space segment while improving the utilization on these links. Aside from interface protocol issues at the satellite-terrestrial network boundary, there are fundamental routing issues that need to be addressed for a hybrid satellite-terrestrial network including routing strategies for different types of traffic and the impact of different routing strategies on satellite network cost. There appears to be a research void on the topic of routing over hybrid satellite and terrestrial networks. This thesis study is intended to fill that void.

Chapter 3

Parametric Cost Functions

Satellite system cost plays a major role in determining the success or failure of a satellite business. In this chapter, we formulate simple parametric link cost models based on fundamental relationships among link data rate, antenna aperture, link distance, and cost. These cost models will be useful for satellite network topology comparison and for dimensioning link capacities.

There are several elements contributing to the overall satellite system cost. We assume that the user terminal cost and the cost of user terminal connection to the gateway are born by the user and the terrestrial carrier respectively. In any case, these costs are common to all satellite network topologies and are thus inconsequential to the outcome of our study. The satellite system cost consists of the following factors:

Initial investment

- Space segment: satellites, spectrum cost, launch cost (plus insurance).
- Ground segment: gateways, extension from the gateways to the terrestrial backbone, control centers.
- Additional interconnection: fibers that interconnect gateways and control centers

Monthly fixed cost

- Space segment: satellite control.

- Ground segment: operation of gateways and control centers.
- Interconnection: maintenance.

Monthly variable cost

- Amount paid for routing traffic on the terrestrial fiber network. We assume that the service is provided by a commercial terrestrial network service provider and a per-flow cost is incurred whenever a unit flow of traffic is routed on a fiber link.

The parametric cost models we will formulate are derived in a similar fashion as those in [4]. The cost elements listed above are grouped into individual link costs as functions of link capacity and link distance. These cost functions will serve as bases for comparison between different network topologies.

For a satellite network interconnected with the terrestrial network, the possible communication links are: up and down links to user terminals (access links), ISLs, up and down links to the gateways (gateway links), and terrestrial links. The communication nodes and different links are shown in Figure 3-1. Depending on the satellite network topology, some of these links may not be used.

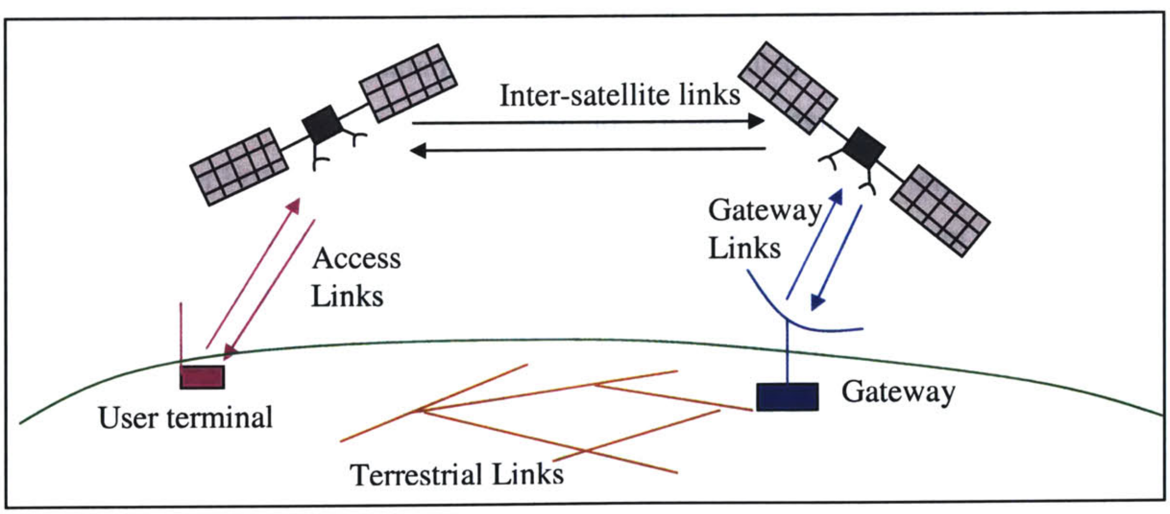


Figure 3-1 Communication Links in a Hybrid Satellite-Terrestrial Network

For ease of analysis, we make the assumption that all satellites in the network are identical. This is a reasonable assumption for satellite networks since mass-producing identical satellites can greatly reduce the cost and risk of the system. The advantage is especially apparent for low earth orbit (LEO) systems where tens, if not hundreds, of satellites are needed for global coverage. For geosynchronous satellite (GEO) systems, the identical satellites assumption may not always hold but is amenable for analysis.

The ISLs are likely to use free-space optical systems as these systems have been identified to be more cost-effective over the traditional radio frequency (RF) links for data rates above 10 Mbps [13]. The gateway links and access links will likely be RF based. For ease of analysis, all gateway links for a particular topology are assumed to be identical in capacity and cost. In reality, depending on the traffic distribution, some gateways may be dimensioned to carry more traffic than others. The access links are assumed to be identical in terms of cost and capacity for all satellite topologies with equal number of satellites and with satellites at the same altitude. In our analysis, we restrict ourselves to compare only those topologies with equal number of satellites and with satellites at the same altitude. Essentially, we compare three satellite topologies: satellites interconnected by ISLs, satellites interconnected by gateways, and satellites interconnected by both ISLs and gateways. Since the access links are assumed to be identical in all three cases, they do not contribute to the cost difference between topologies and will not be included as a part of the formulation. The terrestrial links are assumed to have much higher capacity compared to ISLs and gateway links. A single strand of terrestrial fiber link can carry approximately 100 Tbps ($\times 10^{12}$ bits per second) of traffic for long distance transmission while the satellite gateway link capacity is largely constraint by RF spectrum allocations to a few hundred Mbps. Due to capacity limitation, satellites will not satisfy all or a significant fraction of current or future traffic demands; however, it can be used to complement the terrestrial fiber network. Link cost functions for ISLs, gateway links, and terrestrial links will be developed in the following sections.

3.1 Inter-Satellite Link Cost Estimate

Inter-satellite links (ISLs) can employ either laser or radio frequency (RF) equipments. For completeness, both laser and RF ISLs will be analyzed. Block diagrams for the ISL subsystem are shown in Figures 3-2 and 3-3 for laser and RF ISL respectively. The difference in cost between laser and RF ISLs stems from the difference in the operating frequency. RF ISLs are most likely to operate in the 23GHz and 60GHz range. These frequencies correspond to atmospheric absorption peaks and are best suited for ISL use. Laser ISLs operate on the order of 10^{14} Hz. This difference in operating frequency impacts beam divergence angles and antenna size (weight). Operating at higher frequencies requires smaller antennas but more complex pointing and tracking systems due to smaller beam divergence angle. In this section, the word antenna will be used for both RF antenna and laser communication telescope.

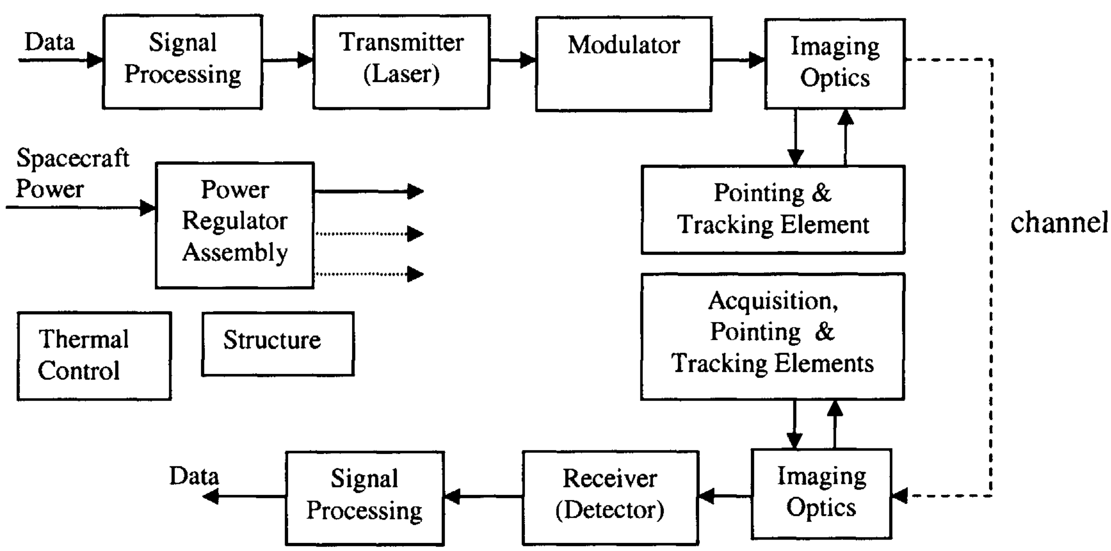


Figure 3-2 Laser ISL Components [14]

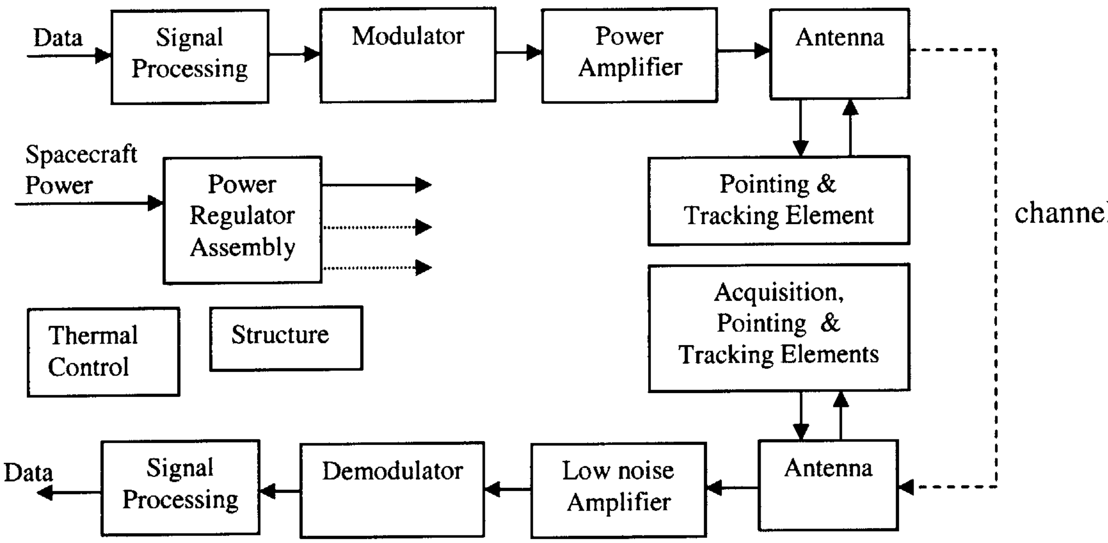


Figure 3-3 RF ISL Components [14]

In the ISL link cost model, we are interested in expressing cost as a function of distance and link capacity. The cost of an ISL depends strongly on the weight of the ISL subsystem and the amount of power it requires. Increase in satellite payload weight leads to higher satellite bus cost and launch cost. Since majority of the weight of the ISL subsystem comes from antenna, tracking system, and the supporting structures, these elements directly influence the cost of an ISL. The amount of power an ISL needs to establish a communication link influences the size of the solar panel and on-board batteries which, in-turn, influences the weight of the power subsystem.

Assume that the onboard power allocated for each ISL is limited and is held constant. Then the cost of each ISL, $\$_{ISL}$, depends mainly on the weight of the ISL subsystem, W_{ISL}. Typically, launch cost for satellites is expressed in terms of dollars per pound. This is a linear relationship relating cost and weight of a satellite. The ISL cost (for transmitter and receiver pair) may be approximated as $\$_{ISL} = k_1 W_{ISL}^{\alpha} + k_2$, where $1 \le \alpha \le 2$; k_1 and k_2 are constants for a particular ISL technology and a particular launch vehicle. The cost of building an ISL subsystem as well as the increased complexity in the satellite bus contribute to both k_1 and k_2.

As mentioned before, W_{ISL} is directly related to the antenna aperture diameter, D_{ISL}. For laser ISLs, the primary telescope mirror constitutes a large portion of the weight. Telescope mirror weight is approximately proportional to the area of the mirror. Together with the telescope enclosure and the supporting structures, the weight of an ISL subsystem is proportional to ${D_{ISL}}^{\beta}$, where $2 \le \beta \le 3$. For RF ISLs, the antenna constitutes a large portion of the weight and the weight of the ISL subsystem is also roughly proportional to ${D_{ISL}}^{\beta}$. Overall, the weight of an ISL can be approximated as $W_{ISL} = a{D_{ISL}}^{\beta} + b$, where a and b are constants that depends on technology.

The ISL cost can be written as

$$\begin{aligned} \$_{ISL} &= k_1 W_{ISL}^{\alpha} + k_2 = k_1 \left(a{D_{ISL}}^{\beta} + b \right)^{\alpha} + k_2 \\ &\approx k_1' {D_{ISL}}^{\alpha\beta} + k_2' \end{aligned} \tag{3.1}$$

The approximation in Equation 3.1 is made assuming the parameter b is relatively small. We would like to relate this cost function to the capacity and distance of the link. This can be done by examining the link equation. Let the power available to establish one ISL link to be P_T.

At the receiving satellite, the received power is

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2 \left(\frac{1}{L_{others}} \right) \tag{3.2}$$

where

d = physical distance of the ISL

G_T = transmission antenna gain

G_R = receiving antenna gain

λ = operating wavelength

L_{others} = Losses due to antenna pointing, transmission line, and coupling.

Typically, the transmit and receive antennas for an ISL are of the same size. The antenna gains may be expressed as $G = \frac{4\pi\eta A}{\lambda^2} = \frac{\eta\pi^2 D^2}{\lambda^2}$ where η represents antenna efficiency, A represents antenna aperture area, and D represents antenna aperture diameter.

Equation 3.2 can be rewritten as $P_R = P_T\left(\frac{\pi^2\eta(D)^2}{\lambda^2}\right)\left(\frac{\pi^2\eta(D)^2}{\lambda^2}\right)\left(\frac{\lambda}{4\pi d}\right)^2\left(\frac{1}{L_{others}}\right)$. At the decoder, the carrier power to noise power ratio C/N is:

$$\frac{C}{N} = \frac{P_R}{kT_s B} = \left(\frac{P_T}{kT_s B}\right) D^4 \left(\frac{\pi\eta}{4\lambda d}\right)^2 \left(\frac{1}{L_{others}}\right) = \frac{E_b R}{N_o B} \tag{3.3}$$

where k = Boltzmann's constant, T_s = receiver noise temperature in Kelvin, B = bandwidth in hertz, and R = data rate in bits per second.

Given a specific E_b/N_o required for decoding, and assuming all other variables are held constant, the antenna aperture diameter must change in order to obtain different data rates. The relationship between D and R can be expressed as:

$$D = \left(\frac{Rd^2}{\left(\frac{\pi\eta}{4\lambda}\right)^2 \left(\frac{1}{L_{others}}\right)\left(\frac{P_T}{kT_s(E_b/N_o)}\right)}\right)^{\frac{1}{4}} \tag{3.4}$$

Defining a constant:

$$\xi = \frac{1}{\left(\frac{\pi\eta}{4\lambda}\right)^2 \left(\frac{1}{L_{others}}\right)\left(\frac{P_T}{kT_s(E_b/N_o)}\right)} \tag{3.5}$$

we can rewrite Equation 3.4 as

$$D = \xi^{\frac{1}{4}} R^{\frac{1}{4}} d^{\frac{1}{2}} \tag{3.6}$$

Given this relationship, we can rewrite Equation 3.1 in terms of R and d.

$$\$_{ISL} = k_1' D_{ISL}^{\alpha\beta} + k_2' = k_1' \left(\xi^{\frac{1}{4}} R^{\frac{1}{4}} d^{\frac{1}{2}} \right)^{\alpha\beta} + k_2' \tag{3.7}$$

Equation 3.7 is applicable to both laser and RF ISLs. We expect $k_{1,RF}' \geq k_{1,laser}'$ due to the much higher carrier frequency used for laser systems, which results in smaller telescope aperture required. The constant term, k_2', includes all pointing and tracking hardware. We expect $k_{2,RF}' \leq k_{2,laser}'$ since higher precision is required to perform laser beam tracking due to the narrower beam divergence angle for laser beams.

If link distance is kept constant, we can express Equation 3.7 by:

$$\$_{ISL} = m_1 R^{\frac{\alpha\beta}{4}} + m_2, \text{ where } 0.5 \leq \frac{\alpha\beta}{4} \leq 1.5 \tag{3.8}$$

m_1 and m_2 are constants for a given technology.

3.2 Gateway Link Cost Estimate

In this section, parametric cost model for a gateway link is established. Figure 3-4 shows the downlink subsystem on the satellite as well as the ground station components. We assume that the cost of the uplink is the same as that of the downlink. Hence, in the subsequent development, only downlink is considered.

The cost of the downlink consists of both satellite and ground component costs. For the satellite downlink subsystem, majority of the cost comes from the RF antenna and the supporting structures. The derivation for this cost is similar to the derivation of cost for RF ISLs. For the ground station, majority of the cost comes from the antenna and supporting structures. There is also a fixed cost component that includes base band and control equipment, physical facility, electrical power, temperature and humidity control, and connection to the terrestrial network.

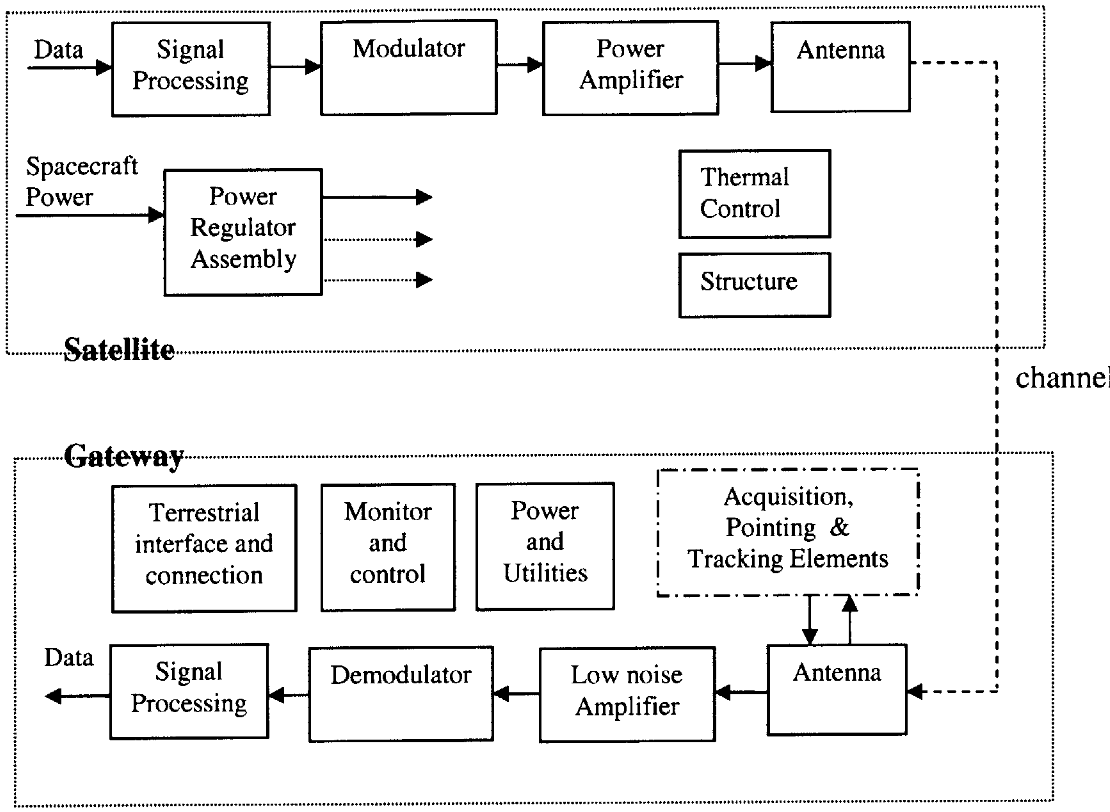


Figure 3-4 Satellite Downlink Subsystem [14]

The cost of the satellite component has the form $\$_{sat} = k_3 D_T^{\alpha\beta} + k_4$ where $1 \le \alpha \le 2$ and $2 \le \beta \le 3$; k_3 is a proportionality multiplier, and k_4 is a constant cost. The cost of the ground component has the form $\$_{ground} = k_5 D_R^{\gamma} + k_6$, where $2 \le \gamma \le 3$. Note that the antenna on-board the satellite and the antenna at the gateway are generally not the same size. Together, the total cost of the downlink has the form

$$\$_{down} = k_3 D_T^{\alpha\beta} + k_5 D_R^{\gamma} + k_7 \tag{3.9}$$

As before, we would like to express the cost function in terms of link capacity and link distance. For the downlink, assume that the on-board transmitted power is held constant at P_T. At the gateway, the received power is: $P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \left(\frac{1}{L_{others}(f)}\right)$, where

d = physical distance of the downlink

$L_{others}(f)$ = path loss due to precipitation, scintillation, atmospheric absorption, antenna pointing loss, multipath/shadowing, etc. This loss factor is a function of signal carrier frequency since atmospheric absorption and attenuation are frequency dependent.

At the decoder, the carrier power to noise power ratio is

$$\frac{C}{N}=\frac{P_R}{kT_sB}=\left(\frac{P_T\eta_T\eta_R}{kT_sB}\right)D_T^{\,2}D_R^{\,2}\left(\frac{\pi}{4\lambda d}\right)^2\left(\frac{1}{L_{others}(f)}\right)=\frac{E_bR}{N_oB} \tag{3.10}$$

Rearranging Equation 3.10 yields

$$D_T^{\,2}=\frac{Rd^2}{\left(\frac{P_T\eta_T\eta_R}{kT_s(E_b/N_o)}\right)D_R^{\,2}\left(\frac{\pi}{4\lambda}\right)^2\left(\frac{1}{L_{others}(f)}\right)} \tag{3.11}$$

Defining a constant:

$$\chi=\frac{1}{\left(\frac{\pi}{4\lambda}\right)^2\left(\frac{1}{L_{others}(f)}\right)\left(\frac{P_T\eta_T\eta_R}{kT_s(E_b/N_o)}\right)} \tag{3.12}$$

We can write the above cost function as

$$\$_{down}=k_3\left(\frac{\chi Rd^2}{D_R^{\,2}}\right)^{\frac{\alpha\beta}{2}}+k_5D_R^{\,\gamma}+k_7=\left(k_3\chi^{\frac{\alpha\beta}{2}}\right)\left(\frac{R^{\frac{\alpha\beta}{2}}d^{\alpha\beta}}{D_R^{\,\alpha\beta}}\right)+k_5D_R^{\,\gamma}+k_7 \tag{3.13}$$

We can find the optimal D_R to minimize the cost function.

$$\frac{d(\$_{down})}{dD_R}=0$$

$$\frac{d(\$_{down})}{dD_R}=\left(-\alpha\beta k_3\chi^{\frac{\alpha\beta}{2}}\right)\left(\frac{R^{\frac{\alpha\beta}{2}}d^{\alpha\beta}}{D_R^{\,\alpha\beta+1}}\right)+k_5\gamma D_R^{\,\gamma-1}=0$$

$$D_R = \left(\frac{\alpha\beta k_3 \chi^{\frac{\alpha\beta}{2}} R^{\frac{\alpha\beta}{2}} d^{\alpha\beta}}{k_5 \gamma} \right)^{\frac{1}{\gamma+\alpha\beta}} \tag{3.14}$$

Substituting this expression into the original cost equation 3.13 and obtain

$$\$_{down} = \left[\left(k_3 \chi^{\frac{\alpha\beta}{2}} \right) \left(\frac{\alpha\beta k_3 \chi^{\frac{\alpha\beta}{2}}}{k_5 \gamma} \right)^{\frac{-\alpha\beta}{\gamma+\alpha\beta}} + k_5 \left(\frac{\alpha\beta k_3 \chi^{\frac{\alpha\beta}{2}}}{k_5 \gamma} \right)^{\frac{\gamma}{\gamma+\alpha\beta}} \right] \left(R^{\frac{\alpha\beta}{2}} d^{\alpha\beta} \right)^{\frac{\gamma}{\gamma+\alpha\beta}} + k_7 \tag{3.15}$$

Finally, the cost of the downlink, with optimized ground station antenna size, is

$$\$_{down} = k_8 \left(R^{\frac{\alpha\beta}{2}} d^{\alpha\beta} \right)^{\frac{\gamma}{\gamma+\alpha\beta}} + k_7 \tag{3.16}$$

If link distance is kept constant, we can express Equation 3.16 by:

$$\$_{down} = m_3 R^{\frac{\gamma\alpha\beta}{2(\gamma+\alpha\beta)}} + m_4, \text{ where } 0.5 \le \frac{\gamma\alpha\beta}{2(\gamma+\alpha\beta)} \le 1 \tag{3.17}$$

m_3 and m_4 are constants for a given technology.

3.3 Terrestrial Link Cost

For a satellite system manufacturer, aside from the cost of constructing gateway interconnection to the terrestrial network, no other cost is generated for the ground links. However, for a satellite network operator, in order to ensure connectivity, terrestrial networks belonging to other operators must be used. Usually the satellite network operator is charged a fee for leasing some fiber capacity or is charged based on the volume of traffic traversing the terrestrial links. In the design of a satellite network then, it is not sufficient to only account for the capital cost of building a system, but also costs that are generated after the system is in service. In our study, we assume that a cost is incurred for routing on each terrestrial link based on the amount of satellite traffic traversing that link. The terrestrial link cost is

$$\$_{fiber} = \zeta y \tag{3.18}$$

where ζ is the per-unit flow cost on a link of fixed distance, and y is the amount of flow on a link. From a first order analysis, we assume that ζ varies linearly with the distance of a link. The longer the distance of the link, the higher the per-unit cost.

3.4 Summary

In this chapter, we have developed parametric link cost functions for ISL, gateway links, and terrestrial links. The ISL and gateway link cost functions are expressed in terms of link capacity for fixed link distance and power. We assume a per-unit flow cost is incurred whenever satellite traffic is routed on each terrestrial link. These parametric cost functions will be used for capacity dimensioning and optimal topology selection in the subsequent chapters.

Chapter 4

Two-Stage Stochastic Programming

We are interested in the solution to the problem: given parametric cost models for the various building blocks of a satellite network, what are the optimal topology, link capacity, and the associated routing strategy? In this chapter, we formulate the satellite link dimensioning and routing problem (subsequently termed optimization problem) using two-stage stochastic programming. This is an important variant to the basic link capacity dimensioning problem that faces terrestrial network designers. Unlike terrestrial networks where the planning horizon is fairly short and terrestrial facilities may be incrementally upgraded and changed to reflect changes in traffic demand pattern and emergence of new technologies, satellite network designers are faced with the unique challenge of a long-term planning horizon on the order of 5 to 10 years, during which, the physical components on the satellites are unalterable. Moreover, because of the high investment cost and the relatively short duration of system life time, the satellite system must be designed to be as cost-effective as possible in order to generate the desired profit.

The basic intuition that forms the basis of the mathematical formulation for the optimization problem is as follows: Given a particular physical network topology and input traffic matrix, traffic must be routed on the various links from source to destination according to a routing algorithm. The total traffic carried by each link must be less than the link capacity. As shown in Chapter 3, the cost of each link can be expressed as a function of link capacity. Hence, in the network design stage, link capacities should be optimally dimensioned to carry the input traffic at the lowest system cost. The optimal routing strategy is one which allows the input traffic to be carried by the network such that the links are optimally dimensioned.

One of the inputs to the optimization problem is satellite network topology. Some of the possible topologies are shown in Chapter 2. The formulation developed in this chapter may be used for any satellite topology. In Chapter 5 and 6, we analyze and compare several representative topologies in detail.

The second input to the optimization problem is the input traffic matrix. Conventional optimization techniques (non-linear programming) rely heavily on the accuracy of the traffic matrix prediction for link dimensioning and give results that could be far from optimal when the actual input parameters differ slightly from the predicted. A more useful solution should explicitly express the tradeoffs between link capacities and system cost under uncertain traffic demands. The optimization problem with uncertain traffic demand may be addressed using two-stage stochastic programming.

A two-stage stochastic programming formulation naturally incorporates into the optimization link capacity dimensioning, routing, and uncertainty in input traffic demand. In the first stage, the formulation aims to find optimal link capacities in order to minimize an effective system cost. The effective system cost is defined to be the sum of satellite network investment cost, cost of routing using terrestrial links, and an opportunity cost for rejecting excess input traffic, subject to quality of service (QoS) constraints. The cost for rejecting excess input traffic may be viewed as a loss in revenue. In the second stage, link capacities are assumed to be given and the formulation aims to maximize the utilization of satellite links through optimal routing given a realization of the input traffic demand.

Mathematically, a satellite network can be represented as a graph $G(V,E)$, where V is the set of vertices representing satellites and ground stations and E is the set of directed edges representing the communication links between vertices. In the satellite network we consider, there are two types of vertices (satellite and ground station) and three types of edges (ISL, gateway links, and terrestrial links).

The first-stage link dimensioning problem is formulated as follows:

$$J(N) = c(N) + E\left[G(N,\tilde{r}) + D(N,\tilde{r})\right] \tag{4.1}$$

$$\Im = \min_{N}\left\{J(N)\right\} \tag{4.2}$$

where

$J(N)$: Effective system cost function.

$\Im$: Optimal effective system cost.

N_{ISL}: Capacity of an ISL (design variable).

$N_{up/dn}$: Capacity of a gateway link (design variable).

$N = \begin{bmatrix} N_{ISL} \\ N_{up/dn} \end{bmatrix}$: Link capacity vector.

$c(N)$: Total investment cost function of satellite ISL and gateway links.

$\tilde{r}$: Traffic demand matrix for OD pairs. Each element in the matrix is a random variable with known probability density.

$E\left[G(N,\tilde{r})\right]$: Expected cost of routing traffic on the terrestrial links.

$E\left[D(N,\tilde{r})\right]$: Expected cost of rejecting traffic.

The major difficulty in solving this problem lies in the evaluation of the expectation function. Here, the argument inside the expectation function is the effective routing cost incurred after the network is built. This is considered to be the second stage problem. As mentioned before, all link capacities are assumed to be known for this stage. For a specific realization of the input demand matrix, the goal is to optimally route the input demand to minimize the amount of traffic on the terrestrial links (since routing on terrestrial links incurs incremental costs while routing on other links generate no additional cost) and the amount of demand rejected while satisfying some QoS requirement, typically in the form of delay for data traffic.

Denote

W = Set of all OD pairs.

P_w = Set of all directed paths for each OD pair w.

The second stage problem is formulated as follows:

$$G(N,r)+D(N,r)=\min\left\{\left(\zeta \sum_{\substack{\text{all ground} \\ \text{station pairs } (i,j)}} \sum_{\substack{\text{all paths } p \\ \text{containing } (i,j)}} x_p\right)+\left(\kappa \sum_{w\in W} s_w\right)\right\} \tag{4.3}$$

s.t.

$$\sum_{\substack{\text{all paths } p \\ \text{containing link } (m,n)}} x_p \leq N_{(m,n)}, \text{ (maximum flow constraint)} \tag{4.4}$$

$$\left(\sum_{p\in P_w} x_p\right)+s_w = r_w, \text{ (conservation of flow constraint)} \tag{4.5}$$

$$\sum_{\substack{\text{all links } (m,n) \\ \text{on a path}}} \left(\frac{F_{(m,n)}}{N_{(m,n)}-F_{(m,n)}}+T_{(m,n)}\right) \leq T_{\max}, \text{ (delay constraint)} \tag{4.6}$$

$$x_p \geq 0,\ \ \forall p\in P_w, \forall w\in W, \text{ (non-negative flow)} \tag{4.7}$$

$$s_w \geq 0,\ \forall w\in W, \text{ (non-negative rejected flow)} \tag{4.8}$$

In words, the second stage formulation finds an optimal routing strategy for a specific realization of the input demand r in order to minimize the effective routing cost. For each OD pair $w\in W$, the total input traffic rate is r_w. This traffic may either be routed or rejected. If the input traffic is routed, there are many paths this traffic can traverse. The total flow on a particular path $p\in P_w$ is x_p. The total amount of rejected traffic for an OD pair w is s_w. A per unit cost of ζ is incurred when routing on the terrestrial links. We assume fully connectedness between all gateways. Thus, flow between two gateways (i,j) is $F_{(i,j)} = \sum_{\substack{\text{all paths p} \\ \text{containing } (i,j)}} x_p$. Similarly, a per unit cost of κ is incurred when traffic is rejected.

Equation 4.4 constrains the maximum flow on links. The sum of routed and rejected traffic for each OD pair must equal to the input traffic demand as shown in Equation 4.5. Let the total flow on a link (m,n) be $F_{(m,n)} = \sum_{\substack{\text{all paths } p \\ \text{containing link } (m,n)}} x_p$. Using the *M/M/1* delay formula, Equation 4.6 imposes a QoS constraint in terms of delay, where $T_{(m,n)}$ is the link propagation delay and $T_{\max}$ is the maximum path delay tolerable.

The second stage problem may be solved using well known multi-commodity flow algorithms [15]. This problem needs to be solved for all possible realizations of the input demand matrix to find the expected routing cost. Moreover, in order to solve the first stage problem, the second stage problem needs to be solved for all possible link capacity values.

In general, it is difficult to solve this problem computationally for large networks without further simplifications to the problem setup. However, for certain classes of interesting network topologies and traffic demand matrix, this problem may be solved analytically. These analytical solutions offer much insight into the topology selection problem, as well as give optimal link capacity and routing strategies. The analysis for GEO and LEO networks are given in Chapter 5 and 6 respectively.

Chapter 5

GEO Satellite-Terrestrial Network

Traditionally, GEO satellites have been used for TV broadcast, marine communications, and long distance telephony services. In the 1970s, commercial use of satellites for mobile communications first began with the launch of the COMSAT/Marisat system. Recently, GEO satellites such as the Asian Cellular Satellite (ACeS) system and the proposed Thuraya system found new applications in serving regional mobile telephony markets. It has been recognized that GEO satellites will play an important role for future broad-band data oriented applications. Due to its inherent star architecture and long propagation delay, a GEO system is most suitable for broadcast, multicast, delay-insensitive, and non-real time data applications. It also has the distinct advantage of requiring only a few satellites for large coverage areas, making the cost of communication essentially independent of distance. Another significant advantage of GEO satellites is that services can begin with only one satellite in orbit; therefore it is not necessary to launch the entire constellation of satellites to begin generating revenue for the satellite business.

With hybrid satellite-terrestrial networks, there is increased flexibility for a satellite system to efficiently utilize terrestrial network links to alleviate congestion and to achieve high utilization of the satellite links. In current literature, much work has been done to study the interoperability of satellite and terrestrial networks in terms of network protocols. We assume that interoperability will not be an issue in the future and the interface between the satellite system and terrestrial networks will essentially be transparent. Hence we can focus on the satellite link dimensioning and routing aspect of network design and treat the satellite and terrestrial networks as one entity. By using the terrestrial network, the satellite links may not have to be designed to carry the worst-case traffic, resulting in significant satellite system cost reductions. However, as mentioned in Chapter 3, the terrestrial network typically charges a fee for using the fiber links to carry satellite traffic. During the satellite network design process, one needs to balance the cost of providing capacity on the various links with the costs that may incur during the operation of the satellite system.

The stochastic programming formulation outlined in Chapter 4 can be applied to designing a GEO satellite system interconnected with the terrestrial fiber network. Given a prediction of the future aggregate traffic demand statistic, the stochastic programming formulation gives optimal link dimensions and an optimal routing strategy for each satellite topology under consideration. The most cost-effective satellite topology can then be chosen for implementation.

In general, three different types of user terminals can be used to access the satellite network: 1. terminals that have direct access to the satellite network but no access to the ground network (Satellite Terminals), 2. terminals that can access the satellite network only via terrestrial networks interconnected with gateways (Ground Terminals), and 3. terminals that can access both the satellite network and the terrestrial network directly (Dual Terminals). Satellite Terminals are especially useful in remote regions where no ground infrastructures exist. These terminals can range from mobile handheld devices to fixed units. Ground Terminals can be mobile devices with access to the terrestrial wireless infrastructure or fixed units with access to the ground network. Dual Terminals have the maximum flexibility in terms of access but are more complex to build.

For these terminals, we consider three different traffic classes: Satellite Terminal to Satellite Terminal traffic (ST-ST), Satellite Terminal to Ground Terminal traffic (ST-GT), and Ground Terminal to Satellite Terminal (GT-ST) traffic. A Dual Terminal may be considered as a combination of Satellite Terminal and Ground Terminal. Traffic generated by and destined for a Dual Terminal can then be categorized into the above three classes of traffic.

We will use the formulation in Chapter 4 to analyze optimal topology selection, optimal link capacity dimensioning and routing for GEO satellite networks. In Section 5.1 and 5.2, we derive analytical solutions for a single class of traffic under uncertain input traffic. In Section 5.3, we analyze a general GEO satellite network using worst-case, uniform, all-to-all traffic for multiple classes of traffic. Finally, in Section 5.4, we derive the optimal number of satellites for a general satellite network.

5.1 Topology Selection for Satellite Terminal to Satellite Terminal (ST-ST) Traffic

We study the topology selection problem for ST-ST traffic between terminals under the coverage of two different satellites. The results we obtain here are independent of the actual traffic distribution in the network and allow easy comparison between different topologies. For ST-ST traffic, there are three different ways the traffic can be routed: 1. Route all traffic on ISLs, 2. Route all traffic on the terrestrial links via gateway links, or 3. Route some traffic on ISLs and some on the terrestrial links. Clearly, these routing strategies influence optimal link dimensions on the various links. In this section, we analyze the routing and capacity dimensioning problem for two satellites incorporating uncertain input demand. In particular, we solve the effective system cost minimization problem for special input density functions. The solutions obtained in this section apply to the case where the satellite network is intended to serve only ST-ST traffic.

5.1.1 ST-ST Traffic Carried by ISL Only

For ST-ST traffic, there is no need to use gateways as all traffic can be routed using only the ISLs. Using this topology, the input traffic is either carried by the ISLs or rejected as depicted in Figure 5-1. We formulate and solve the optimization problem for this topology below.

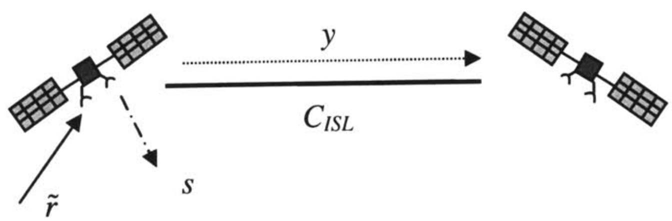


Figure 5-1 ISL Only Topology for ST-ST Traffic

Parameters

C_{ISL}: Capacity on an ISL link (design variable)

$\tilde{r}$: Random demand (with known probability density function (pdf), $p_{\tilde{r}}$)

r: A realization of the random demand

$E\left[D\left(C_{ISL},\tilde{r}\right)\right]$: Expected cost for rejecting traffic

$m_1 C_{ISL} + m_2$: Cost function of a link. We use linear cost function here to make analytical solutions possible. As shown in Chapter 3, linear cost functions are within the range of estimated parameter values.

s: Amount of rejected flow

y: Flow on the link

κ: Per-unit cost of rejected flow

T_{ISL}: Propagation delay on the ISL

$T_{\max}$: Maximum delay tolerable

Formulation

$$\Im = \min_{C_{ISL}} \left\{ m_1 C_{ISL} + m_2 + \underset{\tilde{r}}{E}\left[D\left(C_{ISL},\tilde{r}\right)\right] \right\}$$

$$D(C_{ISL}, r) = \min_{s} (\kappa s)$$

s.t.

$$\begin{aligned} y &\le C_{ISL} && \text{capacity constraint} \\ y + s &= r && \text{flow constraint} \\ \frac{y}{C_{ISL} - y} + T_{ISL} &\le T_{\max} && \text{delay constraint} \\ y &\ge 0 \\ s &\ge 0 \end{aligned}$$

Analytical Solution

Solving the second stage problem, the largest flow that can be carried by the link is:

$$y = C_{ISL}\left(\frac{1}{1 + 1/(T_{\max} - T_{ISL})}\right) = \tau C_{ISL} \tag{5.1}$$

$$\text{where } \tau = \frac{1}{1 + 1/(T_{\max} - T_{ISL})} < 1 \tag{5.2}$$

Thus, the amount of overflow on the link given a realization of input traffic demand, *r*, is:

$$s = \begin{cases} r - \tau C_{ISL} & \text{if } r \ge \tau C_{ISL} \\ 0 & \text{otherwise} \end{cases}$$

Then the second stage cost is:

$$D(C_{ISL}, r) = \begin{cases} \kappa(r - \tau C_{ISL}) & \text{if } r \ge \tau C_{ISL} \\ 0 & \text{otherwise} \end{cases}$$

Having obtained an expression for the second stage problem as a function of C_{ISL}, the first stage problem may be solved.

$$\Im = \min_{C_{ISL}} \left\{ m_1 C_{ISL} + m_2 + \int_{\tau C_{ISL}}^{r_{\max}} \kappa(r - \tau C_{ISL})\, p_{\tilde{r}}(r)\, dr \right\}$$

$$J(C_{ISL}) = m_1 C_{ISL} + m_2 + \int_{\tau C_{ISL}}^{r_{\max}} \kappa(r - \tau C_{ISL})\, p_{\tilde{r}}(r)\, dr$$

$$\frac{dJ(C_{ISL})}{dC_{ISL}} = m_1 - \kappa\tau \int_{\tau C_{ISL}}^{r_{\max}} p_{\tilde{r}}(r)\, dr \tag{5.3}$$

$$\frac{d^2 J(C_{ISL})}{dC_{ISL}^{\;2}} = \kappa\tau^2 p_{\tilde{r}}(\tau C_{ISL}) \geq 0 \tag{5.4}$$

As shown in Equation 5.4, for any input demand pdf, the second derivative of $J(C_{ISL})$ is always non-negative, hence the effective system cost is a convex function with respect to C_{ISL} for any input pdf. If a local minimum exists, it is also the global minimum. A minimum exists for this function since the function is convex on the closed interval $0 \leq C_{ISL} \leq \frac{r_{\max}}{\tau}$.

The optimal link dimension can be found by setting Equation 5.3 to zero. In particular, the following condition must be satisfied:

$$\int_{\tau C^*_{ISL}}^{r_{\max}} p_{\tilde{r}}(r)\,dr = \frac{m_1}{\kappa\tau} \leq 1 \tag{5.5}$$

with strict inequality when $C^*_{ISL} > 0$. Hence, this topology is feasible only if $\frac{m_1}{\tau} < \kappa$.

Analytical Solutions for Families of Input Demand pdfs

To further study properties of the effective system cost, we examine some families of input demand pdfs. In particular, we study uniform and shifted Gaussian pdfs.

1. Uniform Distribution

In this case, $\tilde{r}$ is uniformly distributed over $[0, r_{\max}]$. This traffic statistic represents high uncertainty in aggregate traffic statistic prediction. The effective system cost is formulated as:

$$\Im = \min_{C_{ISL}} \left\{ m_1 C_{ISL} + m_2 + \frac{1}{r_{\max}} \int_{\tau C_{ISL}}^{r_{\max}} \kappa (r - \tau C_{ISL})\,dr \right\}$$

$$J(C_{ISL}) = m_1 C_{ISL} + m_2 - \kappa\left(\tau C_{ISL} - \frac{r_{\max}}{2}\right) + \frac{\kappa\tau^2 C_{ISL}^{\;2}}{2r_{\max}}$$

$$\frac{dJ(C_{ISL})}{dC_{ISL}} = m_1 - \frac{\kappa\tau}{r_{\max}}(r_{\max} - \tau C_{ISL})$$

The optimal link capacity and optimal effective system cost are:

$$C_{ISL}^{*}=\frac{r_{\max}}{\tau}\left(1-\frac{m_1}{\kappa\tau}\right) \tag{5.6}$$

$$\Im=\frac{m_1 r_{\max}}{\tau}+m_2-\frac{m_1^{2} r_{\max}}{2\kappa\tau^2} \tag{5.7}$$

It is clear from Equations 5.6 and 5.7 that the optimal link capacity depends on the relative values of m_1 and κ. If the cost of rejecting traffic is high, then the link must be dimensioned to accommodate the worst-case traffic, $r_{\max}$. As $\kappa\to\infty$, $C_{ISL}^{*}\to\frac{r_{\max}}{\tau}$, and $\Im\to\frac{m_1 r_{\max}}{\tau}+m_2$. If the cost of rejecting traffic is low, a lower dimension on the link will reduce the effective system cost.

If the link is dimensioned for the worst-case traffic, $r_{\max}$, under a static dimensioning method, we obtain $J\left(\frac{r_{\max}}{\tau}\right)=\frac{m_1 r_{\max}}{\tau}+m_2$. This is strictly greater than the effective system cost shown in Equation 5.7. Hence, stochastic dimensioning results in a lower effective system cost compared to worst-case dimensioning.

2. Shifted Gaussian-distributed Demand Function

In this case, $\tilde{r}$ has a shifted Gaussian density function with parameters: $\mathrm{N}(\bar{r},\sigma)$. We restrict the realization of $\tilde{r}$ within the interval $[0,r_{\max}]$. For distributions that have large mean and small variance, the tails of the distribution are negligible. This traffic statistic represents little uncertainty in traffic prediction if the variance is small.

$$\Im=\min_{C_{ISL}}\left\{m_1C_{ISL}+m_2+\kappa\int_{\tau C_{ISL}}^{r_{\max}}\left(r-\tau C_{ISL}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(r-\bar{r}\right)^2}{2\sigma^2}\right)dr\right\}$$

$$J\left(C_{ISL}\right)=m_1C_{ISL}+m_2-\kappa\tau C_{ISL}Q\left(\frac{\tau C_{ISL}-\bar{r}}{\sigma}\right)+\kappa\bar{r}\left[Q\left(\frac{\tau C_{ISL}-\bar{r}}{\sigma}\right)-Q\left(\frac{r_{\max}-\bar{r}}{\sigma}\right)\right]$$
$$+\frac{\kappa\sigma}{\sqrt{2\pi}}\left[\exp\left(\frac{-\left(\tau C_{ISL}-\bar{r}\right)^2}{2\sigma^2}\right)-\exp\left(\frac{-\left(r_{\max}-\bar{r}\right)^2}{2\sigma^2}\right)\right]$$

$$\frac{dJ\left(C_{ISL}\right)}{dC_{ISL}} = m_1 - \kappa\tau\left(Q\left(\frac{\tau C_{ISL} - \bar{r}}{\sigma}\right) - Q\left(\frac{r_{\max} - \bar{r}}{\sigma}\right)\right)$$

To achieve minimum, $Q\left(\frac{\tau C^*_{ISL} - \bar{r}}{\sigma}\right) - Q\left(\frac{r_{\max} - \bar{r}}{\sigma}\right) = \frac{m_1}{\kappa\tau}$. Hence, the optimal link capacity and the optimal effective system cost are:

$$C^*_{ISL} = \frac{\sigma}{\tau}Q^{-1}\left(\frac{m_1}{\kappa\tau} + Q\left(\frac{r_{\max} - \bar{r}}{\sigma}\right)\right) + \frac{\bar{r}}{\tau} \tag{5.8}$$

$$\Im = \frac{m_1\bar{r}}{\tau} + m_2 + \frac{\sigma\kappa}{\sqrt{2\pi}}\left[\exp\left(\frac{-1}{2}\left(Q^{-1}\left(\frac{m_1}{\kappa\tau} + Q\left(\frac{r_{\max} - \bar{r}}{\sigma}\right)\right)\right)^2\right) - \exp\left(\frac{-1}{2}\left(\frac{r_{\max} - \bar{r}}{\sigma}\right)^2\right)\right] \tag{5.9}$$

If the cost of rejecting traffic is high, $C^*_{ISL} \rightarrow \frac{r_{\max}}{\tau}$, as $\kappa \rightarrow \infty$. Then the optimal effective cost becomes $\Im \rightarrow \frac{m_1 r_{\max}}{\tau} + m_2$ as expected. For the shifted Gaussian density function, if the variance is low, the optimal link capacity is very close to the mean of the input traffic for any cost parameters.

<u>Observations</u>

Using some hypothetical parameter values, we plot optimal link capacity with respect to the ratio between marginal cost of an ISL and the cost of rejecting traffic. This is shown in Figure 5-2. It can be observed that for uniform distribution, optimal link capacity is very sensitive to the link cost ratio while for Gaussian distribution, optimal link capacity is fairly insensitive to the link cost ratio. Hence, in general, sensitivity of the optimal link capacity with respect to relative link cost depends on the input traffic demand probability density function.

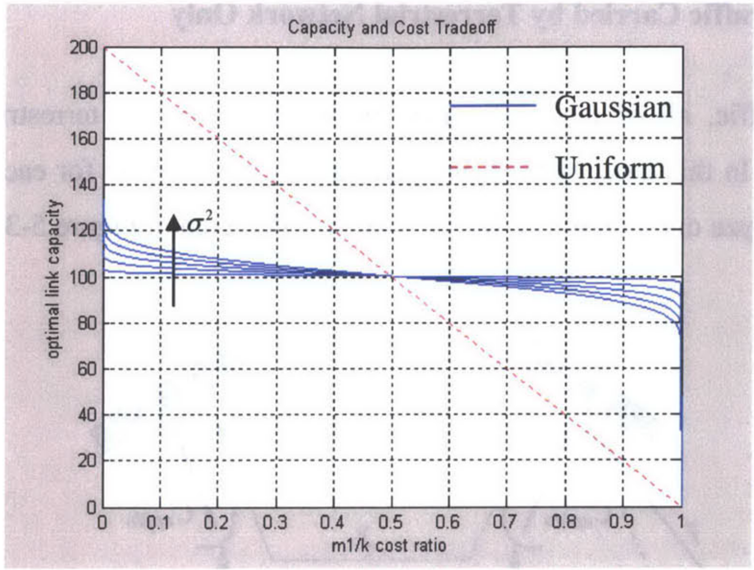


Figure 5-2 Optimal Link Capacity vs. Link Cost Ratio

Define an expected link utilization factor, $\bar{\rho}$, that measures the expect link flow over the total capacity of the link:

$$\bar{\rho} \triangleq \frac{E[Y]}{C_{ISL}^*} = \frac{1}{C_{ISL}^*}\left(\int_0^{\tau C_{ISL}^*} r p_{\tilde{r}}(r)\,dr + \int_{\tau C_{ISL}^*}^{r_{\max}} \tau C_{ISL}^* p_{\tilde{r}}(r)\,dr \right) \tag{5.10}$$

Equation 5.11 below shows that $\bar{\rho}$ is a strictly decreasing function of C_{ISL}^* as the first derivative $\frac{d\bar{\rho}}{dC_{ISL}^*}$ is strictly less than zero. This implies that higher link utilization may be achieved for lower C_{ISL}^*. We will show in Section 5.1.3 that by using a hybrid topology with integrated satellite and terrestrial network, ISL link utilization is higher than the ISL only topology.

$$\frac{d\bar{\rho}}{dC_{ISL}^*} = -\frac{1}{{C_{ISL}^*}^2} \int_0^{\tau C_{ISL}^*} r p_{\tilde{r}}(r)\,dr < 0 \tag{5.11}$$

5.1.2 ST-ST Traffic Carried by Terrestrial Network Only

For ST-ST traffic, all of the traffic may be routed using the terrestrial network via satellite relay. In this case, at least one gateway must be visible for each satellite at all times. We analyze one-hop traffic for the topology depicted in Figure 5-3.

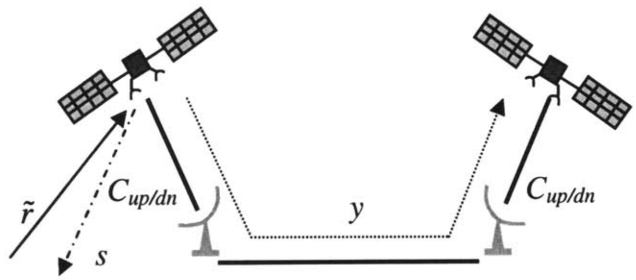


Figure 5-3 Ground Only Topology for ST-ST Traffic

<u>Parameters</u>

$C_{up/dn}$: Capacity on either the up or down link (design variable)

$\tilde{r}$: Random demand (with known probability density function, $p_{\tilde{r}}$)

r: A realization of the random demand

$D\left(C_{up/dn}, \tilde{r}\right)$: Cost function for not satisfying demand

$G\left(C_{up/dn}, \tilde{r}\right)$: Cost function for routing on the terrestrial links

$m_3 C_{up/dn} + m_4$: Cost function of a link (Assuming linear cost function)

s: Unsatisfied flow demand

y: Flow on the link

κ: Per-unit cost of rejected flow

ζ: Per-unit cost of flow on a terrestrial link

$T_{up/dn}$, T_{gnd}: Propagation delay on the links

$T_{\max}$: Maximum delay tolerable

Formulation

$$\Im = \min_{C_{up/dn}} \left\{ 2\left(m_3 C_{up/dn} + m_4\right) + E\left[D\left(C_{up/dn}, \tilde{r}\right) + G\left(C_{up/dn}, \tilde{r}\right) \right] \right\}$$

$$D\left(C_{up/dn}, r\right) + G\left(C_{up/dn}, r\right) = \min\left(\zeta y + \kappa s\right)$$

s.t.

$$y \le C_{up/dn} \quad \text{capacity constraint}$$

$$y + s = r \quad \text{flow constraint}$$

$$\frac{2y}{C_{up/dn} - y} + 2T_{up/dn} + T_{gnd} \le T_{\max} \quad \text{delay constraint}$$

$$y \ge 0$$

$$s \ge 0$$

Analytical Solution

In this problem, there are two scenarios:

1. If $\zeta > \kappa$, then traffic should always be rejected and optimal gateway link capacity is just zero. This case is not very interesting as the satellite system should not be built.
2. If $\zeta < \kappa$, then traffic should always be routed if possible. In this case, the maximum flow is

$$y = C_{up/dn} \left(\frac{1}{1 + 2 \big/ \left(T_{\max} - 2T_{up/dn} - T_{gnd}\right)} \right) = \tau_2 C_{up/dn} \tag{5.12}$$

$$\text{where } \tau_2 = \frac{1}{1 + 2 \big/ \left(T_{\max} - 2T_{up/dn} - T_{gnd}\right)} \tag{5.13}$$

$$\text{Thus, } s = \begin{cases} r - \tau_2 C_{up/dn} & \text{if } r \ge \tau_2 C_{up/dn} \\ 0 & \text{otherwise} \end{cases}$$

The second stage cost is:

$$D\left(C_{up/dn}, r\right) + G\left(C_{up/dn}, r\right) \begin{cases} = \zeta r & \text{if } r < \tau_2 C_{up/dn} \\ = \zeta \tau_2 C_{up/dn} + \kappa\left(r - \tau_2 C_{up/dn}\right) & \text{otherwise} \end{cases}$$

$$\Im = \min_{C_{up/dn}} \left\{ 2\left(m_3 C_{up/dn} + m_4\right) + \int_0^{\tau_2 C_{up/dn}} \zeta r p_{\tilde{r}}(r)\,dr + \int_{\tau_2 C_{up/dn}}^{r_{\max}} \left(\zeta \tau_2 C_{up/dn} + \kappa\left(r - \tau_2 C_{up/dn}\right)\right) p_{\tilde{r}}(r)\,dr \right\}$$

$$J\left(C_{up/dn}\right) = 2\left(m_3 C_{up/dn} + m_4\right) + \int_0^{\tau_2 C_{up/dn}} \zeta r p_{\tilde{r}}(r)\,dr + \int_{\tau_2 C_{up/dn}}^{r_{\max}} \left(\zeta \tau_2 C_{up/dn} + \kappa\left(r - \tau_2 C_{up/dn}\right)\right) p_{\tilde{r}}(r)\,dr$$

$$\frac{dJ\left(C_{up/dn}\right)}{dC_{up/dn}} = 2m_3 - (\kappa - \zeta)\tau_2 \int_{\tau_2 C_{up/dn}}^{r_{\max}} p_{\tilde{r}}(r)\,dr \tag{5.14}$$

$$\frac{d^2 J\left(C_{up/dn}\right)}{dC_{up/dn}^{\;2}} = (\kappa - \zeta)\tau_2^{\;2} p_{\tilde{r}}\left(\tau_2 C_{up/dn}\right) \geq 0 \tag{5.15}$$

The optimal link dimension can be found by setting Equation 5.14 to zero. In particular, the following condition must be satisfied:

$$\int_{\tau_2 C^*_{up/dn}}^{r_{\max}} p_{\tilde{r}}(r)\,dr = \frac{2m_3}{(\kappa - \zeta)\tau_2} \leq 1 \tag{5.16}$$

with strict inequality when $C^*_{up/dn} > 0$. Hence, this topology is feasible only if $\frac{m_3}{\tau_2} < \frac{\kappa - \zeta}{2}$ and $\zeta < \kappa$.

Analytical Solutions for Uniform Input Density Function

Here we provide analytical solution to the case where $\tilde{r}$ is uniformly distributed over $[0, r_{\max}]$.

$$\Im = \min_{C_{up/dn}} \left\{ 2\left(m_3 C_{up/dn} + m_4\right) + \frac{1}{r_{\max}} \int_0^{\tau_2 C_{up/dn}} \zeta r\,dr + \frac{1}{r_{\max}} \int_{\tau_2 C_{up/dn}}^{r_{\max}} \left(\zeta \tau_2 C_{up/dn} + \kappa\left(r - \tau_2 C_{up/dn}\right)\right) dr \right\}$$

$$\frac{dJ\left(C_{up/dn}\right)}{dC_{up/dn}} = 2m_3 - (\kappa - \zeta)\frac{\tau_2}{r_{\max}}\left(r_{\max} - \tau_2 C_{up/dn}\right)$$

The optimal link capacity and optimal effective system cost are:

$$C^*_{up/dn} = \frac{r_{\max}}{\tau_2}\left(1 - \frac{2m_3}{(\kappa - \zeta)\tau_2}\right) \tag{5.17}$$

$$\Im = 2\left[\frac{m_3 r_{\max}}{\tau_2}\left(1-\frac{m_3}{\tau_2(\kappa-\zeta)}\right)+m_4\right]+\frac{r_{\max}\zeta}{2} \tag{5.18}$$

If the cost of rejecting traffic is high, then the link must be dimensioned to accommodate worst-case traffic. $C^*_{up/dn} \to \frac{r_{\max}}{\tau_2}$, as $\kappa \to \infty$. Then the optimal effective cost becomes $\Im \to 2\left[\frac{m_3 r_{\max}}{\tau_2}+m_4\right]+\frac{r_{\max}\zeta}{2}$.

The expected link utilization on the gateway link is:

$$\bar{\rho} = \frac{E[Y]}{C^*_{up/dn}} = \frac{1}{C^*_{up/dn}}\left(\int_0^{\tau_2 C^*_{up/dn}} r p_{\tilde{r}}(r)\,dr + \int_{\tau_2 C^*_{up/dn}}^{r_{\max}} \tau_2 C^*_{up/dn} p_{\tilde{r}}(r)\,dr\right) \tag{5.19}$$

Equation 5.20 shows that $\bar{\rho}$ is a strictly decreasing function of $C^*_{up/dn}$.

$$\frac{d\bar{\rho}}{dC^*_{up/dn}} = -\frac{1}{{C^*_{up/dn}}^2}\int_0^{\tau C^*_{up/dn}} r p_{\tilde{r}}(r)\,dr < 0 \tag{5.20}$$

5.1.3 Hybrid Satellite-Terrestrial Link Dimensioning for ST-ST Traffic

For a hybrid network, ST-ST traffic can be carried by either ISL or terrestrial links, or rejected. The topology we analyze is shown in Figure 5-4. The amount of traffic carried by different links determines the capacity of the links which, in turn, determines the cost of the network.

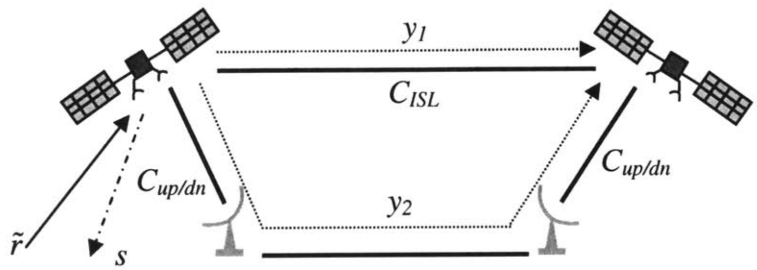


Figure 5-4 Hybrid Topology for ST-ST Traffic

Parameters

C_{ISL} : Capacity on ISL (design variable)

$C_{up/dn}$: Capacity on gateway link (design variable)

$\tilde{r}$: Random demand (with known probability density function, $p_{\tilde{r}}$)

r : A realization of the random demand

$D\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)$: Cost function for overflow traffic

$G\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)$: Cost function for routing on the ground link

$m_1 C_{ISL} + m_2$: Cost function of an ISL link (Assuming linear cost function)

$m_3 C_{up/dn} + m_4$: Cost function of a gateway link (Assuming linear cost function)

y_1 : Flow on ISL

y_2 : Flow on ground path

s : Rejected flow

κ : Per-unit cost of rejected flow

ζ : Per-unit cost of flow on a terrestrial link

T_{ISL}, $T_{up/dn}$, T_{gnd} : Propagation delay on the links

$T_{\max}$: Maximum delay tolerable

Formulation

$$\Im = \min\left\{m_1 C_{ISL} + m_2 + 2\left(m_3 C_{up/dn} + m_4\right) + E\left[D\left(C_{ISL}, C_{up/dn}, \tilde{r}\right) + G\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)\right]\right\}$$

$$D\left(C_{ISL}, C_{up/dn}, r\right) + G\left(C_{ISL}, C_{up/dn}, r\right) = \min\left(\zeta y_2 + \kappa s\right)$$

s.t.

$$\begin{aligned}
y_1 &\le C_{ISL} \\
y_2 &\le C_{up/dn} \\
y_1 + y_2 + s &= r \\
\frac{y_1}{C_{ISL} - y_1} + T_{ISL} &\le T_{\max} \\
\frac{2y_2}{C_{up/dn} - y_2} + 2T_{up/dn} + T_{gnd} &\le T_{\max} \\
y_1, y_2 &\ge 0 \\
s &\ge 0
\end{aligned}$$

Analytical Solution

In the second stage problem, there are two scenarios:

1. If $\zeta > \kappa$, then traffic should not be routed to the terrestrial links and the optimal gateway link capacity is just zero. This case is identical to the ISL only case analyzed in Section 5.1.1.
2. If $\zeta < \kappa$, then traffic should always be routed if possible. In this case, ISL links will always be used first to avoid the routing charge on the ground links. If the ISL links are full, some traffic will be routed on the ground. Excess input demand will be rejected if the gateway links are full.

The maximum flow on the ISL link is:

$$y_1 = C_{ISL}\left(\frac{1}{1 + 1/\left(T_{\max} - T_{ISL}\right)}\right) = \tau C_{ISL} \tag{5.21}$$

The maximum flow on the ground path is:

$$y_2 = C_{up/dn}\left(\frac{1}{1 + 2/\left(T_{\max} - 2T_{up/dn} - T_{gnd}\right)}\right) = \tau_2 C_{up/dn} \tag{5.22}$$

Thus,

$$s = \begin{cases} r - \tau C_{ISL} - \tau_2 C_{up/dn} & \text{if } r \ge \tau C_{ISL} + \tau_2 C_{up/dn} \\ 0 & \text{otherwise} \end{cases}$$

$$D\left(C_{ISL},C_{up/dn},r\right)+G\left(C_{ISL},C_{up/dn},r\right)=\begin{cases}0 & \text{if } r\le \tau C_{ISL}\\ \zeta\left(r-\tau C_{ISL}\right) & \text{if } \tau C_{ISL}< r<\tau C_{ISL}+\tau_2 C_{up/dn}\\ \zeta\tau_2 C_{up/dn}+\kappa\left(r-\tau C_{ISL}-\tau_2 C_{up/dn}\right) & \text{otherwise}\end{cases}$$

This gives the first stage function:

$$J\left(C\right)=m_1 C_{ISL}+m_2+2\left(m_3 C_{up/dn}+m_4\right)+\int_{\tau C_{ISL}}^{\tau C_{ISL}+\tau_2 C_{up/dn}}\zeta\left(r-\tau C_{ISL}\right)P_{\tilde r}\left(r\right)dr$$
$$+\int_{\tau C_{ISL}+\tau_2 C_{up/dn}}^{r_{\max}}\left(\zeta\tau_2 C_{up/dn}+\kappa\left(r-\tau C_{ISL}-\tau_2 C_{up/dn}\right)\right)P_{\tilde r}\left(r\right)dr$$

$$\frac{\partial J\left(C\right)}{\partial C_{ISL}}=m_1-\tau\left(\zeta\int_{\tau C_{ISL}}^{\tau C_{ISL}+\tau_2 C_{up/dn}}P_{\tilde r}\left(r\right)dr+\kappa\int_{\tau C_{ISL}+\tau_2 C_{up/dn}}^{r_{\max}}P_{\tilde r}\left(r\right)dr\right) \quad (5.23)$$

$$\frac{\partial J\left(C\right)}{\partial C_{up/dn}}=2m_3-\left(\kappa-\zeta\right)\tau_2\int_{\tau C_{ISL}+\tau_2 C_{up/dn}}^{r_{\max}}P_{\tilde r}\left(r\right)dr \quad (5.24)$$

$$\frac{\partial^2 J\left(C\right)}{\partial C_{ISL}{}^2}=\tau^2\left(\kappa-\zeta\right)P_{\tilde r}\left(\tau C_{ISL}+\tau_2 C_{up/dn}\right)+\tau^2\zeta P_{\tilde r}\left(\tau C_{ISL}\right)\ge 0 \quad (5.25)$$

$$\frac{\partial^2 J\left(C\right)}{\partial C_{up/dn}{}^2}=\tau_2{}^2\left(\kappa-\zeta\right)P_{\tilde r}\left(\tau C_{ISL}+\tau_2 C_{up/dn}\right)\ge 0 \quad (5.26)$$

In order for $C^*_{up/dn}>0$ and $C^*_{ISL}>0$, from Equation 5.24,

$$0<\int_{\tau C^*_{ISL}+\tau_2 C^*_{up/dn}}^{r_{\max}}P_{\tilde r}\left(r\right)dr=\frac{2m_3}{\left(\kappa-\zeta\right)\tau_2}<1 \quad (5.27)$$

From Equations 5.23 and 5.24,

$$0<\int_{\tau C^*_{ISL}}^{\tau C^*_{ISL}+\tau_2 C^*_{up/dn}}P_{\tilde r}\left(r\right)dr=\frac{m_1}{\tau\zeta}-\frac{2m_3\kappa}{\left(\kappa-\zeta\right)\zeta\tau_2}<1-\frac{2m_3}{\left(\kappa-\zeta\right)\tau_2} \quad (5.28)$$

Hence, for positive link capacities, the following conditions must be satisfied:

$$\frac{2\kappa m_3}{\left(\kappa-\zeta\right)\tau_2}<\frac{m_1}{\tau}<\frac{2m_3}{\tau_2}+\zeta,\ 0<\frac{m_3}{\tau_2}<\frac{\left(\kappa-\zeta\right)}{2},\text{ and } \zeta<\kappa$$

If these conditions are not satisfied, then this topology should not be used.

Analytical Solutions for Uniform Input Density Function

We provide analytical solution to the case where $\tilde{r}$ is uniformly distributed over $[0, r_{\max}]$.

$$\Im = \min\left\{ \begin{array}{l} m_1 C_{ISL} + m_2 + 2\left(m_3 C_{up/dn} + m_4\right) + \dfrac{1}{r_{\max}} \displaystyle\int_{\tau C_{ISL}}^{\tau C_{ISL}+\tau_2 C_{up/dn}} \zeta\left(r - \tau C_{ISL}\right) dr \\ + \dfrac{1}{r_{\max}} \displaystyle\int_{\tau C_{ISL}+\tau_2 C_{up/dn}}^{r_{\max}} \left(\zeta \tau_2 C_{up/dn} + \kappa\left(r - \tau C_{ISL} - \tau_2 C_{up/dn}\right)\right) dr \end{array} \right\}$$

$$\frac{\partial J(C)}{\partial C_{ISL}} = m_1 - \kappa\tau + \frac{\kappa\tau^2 C_{ISL}}{r_{\max}} + \frac{(\kappa - \zeta)\tau_2 \tau C_{up/dn}}{r_{\max}}$$

$$\frac{\partial J(C)}{\partial C_{up/dn}} = 2m_3 - (\kappa - \zeta)\tau_2 + \frac{(\kappa - \zeta)\tau_2 \tau C_{ISL}}{r_{\max}} + \frac{(\kappa - \zeta)\tau_2^{\,2} C_{up/dn}}{r_{\max}}$$

The optimal link capacity and optimal effective system cost are:

$$C^*_{ISL} = \frac{r_{\max}}{\tau}\left(\frac{2m_3}{\tau_2\zeta} - \frac{m_1}{\tau\zeta} + 1\right) \tag{5.29}$$

$$C^*_{up/dn} = \frac{r_{\max}}{\tau_2\zeta}\left(\frac{m_1}{\tau} - \frac{2\kappa m_3}{\tau_2(\kappa - \zeta)}\right) \tag{5.30}$$

$$\Im = r_{\max}\left[\frac{m_1}{\tau}\left(1 - \frac{m_1}{2\zeta\tau}\right) + \frac{2m_3}{\zeta\tau_2}\left(\frac{m_1}{\tau} - \frac{\kappa m_3}{\tau_2(\kappa - \zeta)}\right)\right] + m_2 + 2m_4 \tag{5.31}$$

If the cost of rejecting traffic is high, then the link must be dimensioned to accommodate worst case traffic. It is interesting that for this distribution, the ISL link capacity is independent of κ. $C^*_{ISL} = \frac{r_{\max}}{\tau}\left(\frac{2m_3}{\tau_2\zeta} - \frac{m_1}{\tau\zeta} + 1\right)$, $C^*_{up/dn} \to \frac{r_{\max}}{\tau_2\zeta}\left(\frac{m_1}{\tau} - \frac{2m_3}{\tau_2}\right)$, as $\kappa \to \infty$.

Then the optimal effective cost becomes:

$$\Im = \frac{r_{\max} m_1}{\zeta\tau}\left[\zeta - \left(\frac{m_1}{2\tau} - \frac{2m_3}{\tau_2}\right)\right] + m_2 + 2m_4$$

The combined capacities on ISL and gateway link can support worst case traffic

$$C^*_{ISL} + C^*_{up/dn} = r_{\max}\left(\frac{1}{\zeta}\left(\frac{2m_3}{\tau_2} - \frac{m_1}{\tau}\right)\left(\frac{1}{\tau} - \frac{1}{\tau_2}\right) + \frac{1}{\tau}\right) \ge \frac{r_{\max}}{\tau_2}$$

The expected link utilizations on the links are:

$$\bar{\rho}_{ISL} = \frac{E[Y_1]}{C^*_{ISL}} = \frac{1}{C^*_{ISL}}\left(\int_0^{\tau C^*_{ISL}} r p_{\tilde{r}}(r)\,dr + \int_{\tau C^*_{ISL}}^{r_{\max}} \tau C^*_{ISL} p_{\tilde{r}}(r)\,dr\right)$$

$$\bar{\rho}_{up/dn} = \frac{E[Y_2]}{C^*_{up/dn}} = \frac{1}{C^*_{up/dn}}\left(\int_{\tau C^*_{ISL}}^{\tau C^*_{ISL}+\tau_2 C^*_{up/dn}} r p_{\tilde{r}}(r)\,dr + \int_{\tau C^*_{ISL}+\tau_2 C^*_{up/dn}}^{r_{\max}} \left(\tau_2 C^*_{up/dn}\right) p_{\tilde{r}}(r)\,dr\right)$$

$$\frac{d\bar{\rho}_{ISL}}{dC^*_{ISL}} = -\frac{1}{C^{*\,2}_{ISL}} \int_0^{\tau C^*_{ISL}} r p_{\tilde{r}}(r)\,dr < 0$$

$$\frac{\partial \bar{\rho}_{up/dn}}{\partial C^*_{up/dn}} = \left(\frac{\tau\tau_2 C^*_{ISL}}{C^*_{up/dn}}\right) p_{\tilde{r}}\left(\tau C^*_{ISL} + \tau_2 C^*_{up/dn}\right) - \frac{1}{C^{*\,2}_{up/dn}} \int_{\tau C^*_{ISL}}^{\tau C^*_{ISL}+\tau_2 C^*_{up/dn}} r p_{\tilde{r}}(r)\,dr$$

$$\frac{\partial \bar{\rho}_{up/dn}}{\partial C^*_{ISL}} = \frac{\tau\tau C^*_{ISL}}{C^*_{up/dn}}\left(p_{\tilde{r}}\left(\tau C^*_{ISL} + \tau_2 C^*_{up/dn}\right) - p_{\tilde{r}}\left(\tau C^*_{ISL}\right)\right)$$

$\bar{\rho}_{ISL}$ is a decreasing function of C^*_{ISL}. Compared with the ISL only topology, C^*_{ISL} for the hybrid case is smaller. Hence the expected ISL link utilization for the hybrid topology is higher.

5.1.4 Topology Comparison for ST-ST Traffic

The optimal effective system costs for the three topologies with uniform input demand density are summarized in Table 5-1. Note that the amount of traffic carried by the three topologies under optimal link capacities is not the same.

Table 5-2 Optimal Capacity and Effective System Cost for Uniform Distributed ST-ST Traffic

Topology	Optimal Link Capacity and Effective System Cost
ISL Only	$C^*_{ISL} = \frac{r_{\max}}{\tau}\left(1-\frac{m_1}{\kappa\tau}\right)$ $\Im = \frac{m_1 r_{\max}}{\tau} + m_2 - \frac{m_1^2 r_{\max}}{2\kappa\tau^2}$ $\frac{m_1}{\tau} < \kappa$
Ground Only	$C^*_{up/dn} = \frac{r_{\max}}{\tau_2}\left(1-\frac{2m_3}{(\kappa-\zeta)\tau_2}\right)$ $\Im = 2\left[\frac{m_3 r_{\max}}{\tau_2}\left(1-\frac{m_3}{\tau_2(\kappa-\zeta)}\right)+m_4\right]+\frac{r_{\max}\zeta}{2}$ $\frac{m_3}{\tau_2} < \frac{(\kappa-\zeta)}{2}$, and $\zeta < \kappa$.
Hybrid	$C^*_{ISL} = \frac{r_{\max}}{\tau\zeta}\left(\frac{2m_3}{\tau_2}+\zeta-\frac{m_1}{\tau}\right)$ $C^*_{up/dn} = \frac{r_{\max}}{\tau_2\zeta}\left(\frac{m_1}{\tau}-\frac{2\kappa m_3}{\tau_2(\kappa-\zeta)}\right)$ $\Im = r_{\max}\left[\frac{m_1}{\tau}\left(1-\frac{m_1}{2\zeta\tau}\right)+\frac{2m_3}{\zeta\tau_2}\left(\frac{m_1}{\tau}-\frac{\kappa m_3}{\tau_2(\kappa-\zeta)}\right)\right]+m_2+2m_4$ $\frac{2\kappa m_3}{(\kappa-\zeta)\tau_2} < \frac{m_1}{\tau} < \frac{2m_3}{\tau_2}+\zeta < \kappa$, $0 < \frac{m_3}{\tau_2} < \frac{(\kappa-\zeta)}{2}$, and $\zeta < \kappa$

Depending on the actual values of cost function parameters and link delay values, one may compare the three effective costs and select the one with the lowest cost for implementation. Since the optimal link cost and effective system cost depend heavily on the input traffic statistic, one must take some care in estimating the long term aggregate traffic statistic.

We may gain some perspective on the topology selection problem by examining the feasible region for each topology which is independent of the input traffic statistic. The feasible regions for the three topologies are shown in Figure 5-5. These feasible regions are defined by marginal link costs. Note that in some regions, only one topology is feasible. There also exists a region such that a satellite system should not be built. Within the overlapping regions, an optimal topology should be selected based on the effective system costs.

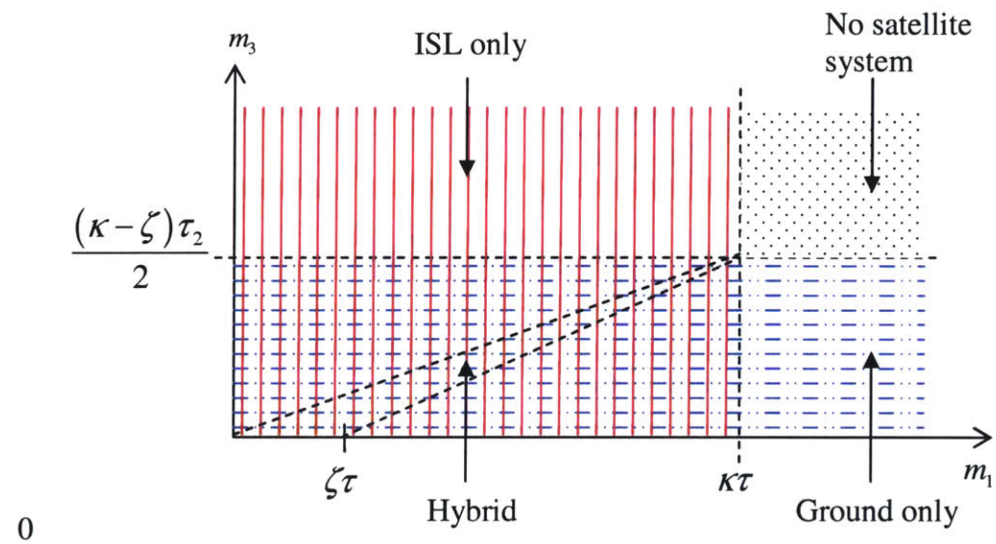


Figure 5-5 Feasible Regions for Different Topologies Using ST-ST Traffic

It can be observed from Figure 5-6 that there exist more possible marginal cost pairs (m_1, m_3) that make the ISL only topology feasible. There are relatively fewer marginal cost pairs (m_1, m_3) that make the hybrid topology feasible. Furthermore, within the region where a hybrid topology is feasible, there may exist cost parameters that make the ISL only or the ground only topology more cost-effective.

This result is quite intuitive. For ST-ST traffic, traffic traverse fewer links if only ISLs are used for routing. Hence, one would expect the ISL only topology to have cost advantages over the other topologies. However, if the ISLs are very expensive, a ground only topology may be more cost-effective. If the marginal cost of a gateway link is approximately $(\kappa-\zeta)/(2\kappa)$ times that of an ISL, then a hybrid topology may be more cost-effective.

5.2 Topology Selection for Satellite Terminal to Ground Terminal (ST-GT) and GT-ST Traffic

For ST-GT and GT-ST traffic between two satellite coverage areas, traffic must be routed on at least one gateway link to access the Ground Terminal. In this section, we analyze the routing and capacity dimensioning problem incorporating uncertain input demand. We assume that there is equal amount of ST-GT and GT-ST traffic to be routed. The detailed derivations are shown in Appendix A. Here we present the main results of that analysis.

5.2.1 ST-GT and GT-ST Traffic Carried by ISL Only

If only ISLs are used for routing, the input demand traffic is either carried by the ISL or rejected. This scenario is depicted in Figure 5-6.

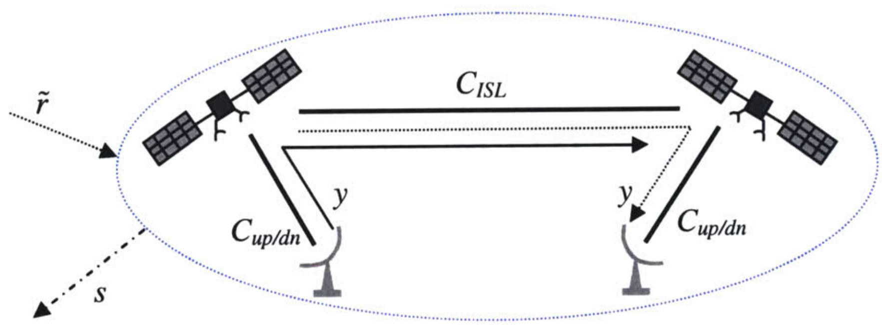


Figure 5-6 ISL Only Topology for ST-GT and GT-ST Traffic

For positive link capacity, the following condition must be satisfied:

$$\frac{m_1 + m_3}{\tau_3} < \kappa \tag{5.32}$$

$$\text{where } \tau_3 = \frac{1}{1 + 2/\left(T_{\max} - T_{ISL} - T_{up/dn}\right)} \tag{5.33}$$

5.2.2 ST-GT and GT-ST Traffic Carried by Terrestrial Network Only

If all input demand traffic is routed on the terrestrial network, at least one gateway must be visible for each satellite at all times. This scenario is depicted in Figure 5-7.

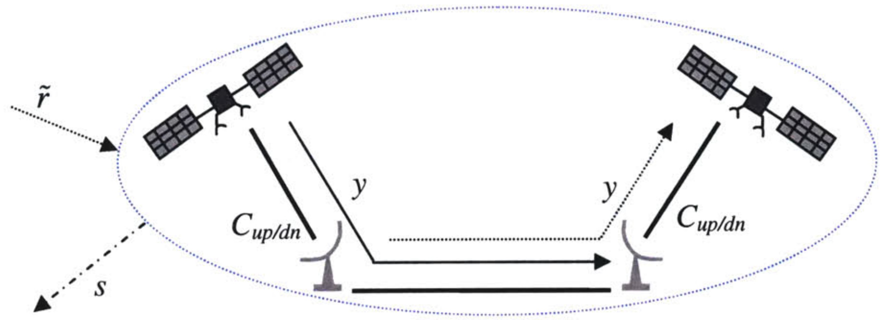


Figure 5-7 Ground Only Topology for ST-GT and GT-ST Traffic

For positive link capacity, the following condition must be satisfied:

$$\frac{m_3}{(\kappa-\zeta)} < \tau_4,\ \zeta < \kappa \tag{5.34}$$

$$\text{where } \tau_4 = \frac{1}{1 + 1/\left(T_{\max} - T_{up/dn} - T_{gnd}\right)} \tag{5.35}$$

5.2.3 Hybrid Satellite-Terrestrial Link Dimensioning for ST-GT and GT-ST Traffic

For a hybrid network, traffic can by carried on one of two paths as shown in Figure 5-8.

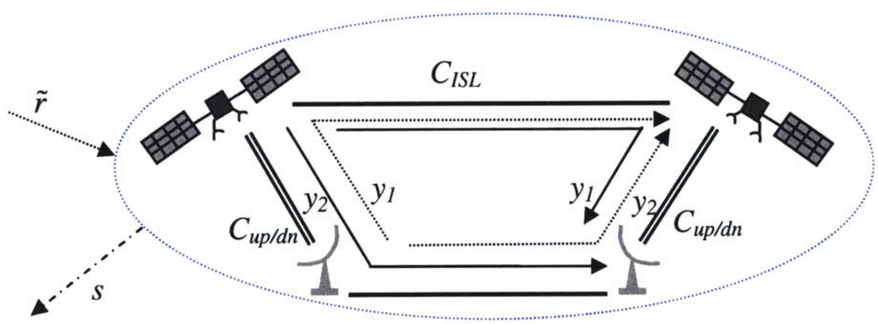


Figure 5-8 Hybrid Topology for ST-GT and GT-ST Traffic

The main results from the stochastic programming formulation are summarized below. For positive link capacities, the following conditions must be satisfied:

$$m_3\left(\frac{\kappa}{\tau_4(\kappa-\zeta)}-\frac{1}{\tau_3}\right)<\frac{m_1}{\tau_3}<m_3\left(\frac{1}{\tau_4}-\frac{1}{\tau_3}\right)+\zeta \tag{5.36}$$

$$0<\frac{m_3}{\tau_4}<(\kappa-\zeta),\ \zeta<\kappa \tag{5.37}$$

5.2.4 Topology Comparison for ST-GT and GT-ST Traffic

The feasible regions for the three topologies are shown in the diagram below. Within the overlapping regions, an optimal topology should be selected based on the effective system costs.

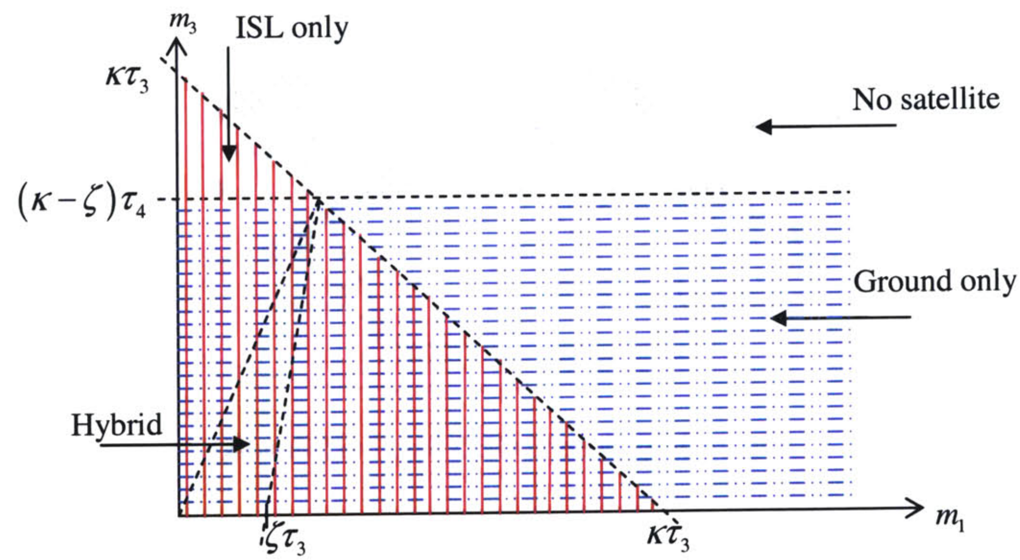


Figure 5-9 Feasible Regions for Different Topologies Using ST-GT and GT-ST Traffic

The feasible regions in Figure 5-9 are very different from the feasible regions for ST-ST traffic in Figure 5-5. It can be observed from Figure 5-9 that there exist more possible marginal cost pairs (m_1, m_3) that make the Ground only topology feasible.

5.3 GEO Satellite Link Capacity Dimensioning and Routing

In Sections 5.1 and 5.2 we analyzed the topology selection problem for a single class of traffic and for traffic between one origin and destination satellite pair. We now consider a general GEO hybrid satellite-terrestrial network with multiple classes of traffic and multiple origin and destination satellite pairs.

A GEO satellite-terrestrial network can be modeled, to a first order approximation, as two interconnected rings. The network we analyze has N equally spaced satellites interconnected in a ring and N gateways, each connected to one satellite. This topology is shown in Figure 5-10. We choose this model for analysis in order to yield analytical solutions to give some insights into the network design problem.

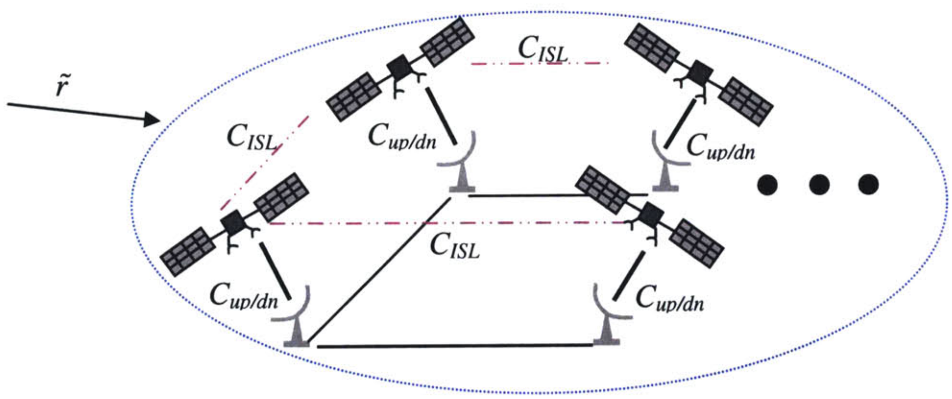


Figure 5-10 General GEO Satellite Network

Obtaining analytical solutions for a general GEO network with multiple traffic classes using stochastic programming is difficult. For each $\left(C_{ISL}, C_{up/dn}\right)$ pair, the second stage problem needs to be solved for all realizations of the random traffic demand $\tilde{r}$ distributed between 0 and $r_{\max}$. With the solution from the second stage problem, an effective system cost must be obtained for each $\left(C_{ISL}, C_{up/dn}\right)$ pair. The optimal effective system cost can then be found by optimizing over all $\left(C_{ISL}, C_{up/dn}\right)$ pairs. This problem may be solved computationally for large networks. To keep the problem tractable, simplifications must be made by limiting the number of possible $\left(C_{ISL}, C_{up/dn}\right)$ pairs and by discretizing input demand distributions. [12] details computational issues for general stochastic programming problems. Since numerical solutions require known cost parameters and input traffic statistic, the results obtained can only represent special cases but will not offer much insight on the design problem. Instead of running simulation on large combinations of inputs, we analyze a special case where analytical solution may be obtained.

We assume that the network is dimensioned to carry the worst-case traffic demand. This represents the case where rejection cost is extremely high and it is never optimal to reject any traffic. Instead of finding the expected routing cost over all possible realizations of input traffic, we find the routing cost for the worst-case traffic, $r_{\max}$. Based on these assumptions, the ISLs will be dimensioned to carry higher capacity than otherwise, representing worst-case dimensioning for ISLs.

Our main goal in this section is to establish optimal off-line routing rules for multiple traffic classes. With off-line routing, the input traffic is assumed to be known and optimal routing can be achieved through a centralized routing decision. This is done for satellite capacity dimensioning during the network design stage. Off-line routing is contrasted with on-line routing where routing decisions are typically distributed among network nodes and future traffic demands are unknown. As the name suggests, on-line routing is performed after the system is put into service. On-line routing decisions depend on network link capacities and the current state of the network. Since optimal off-line routing rules result in optimal link capacities, on-line routing decisions are directly influenced by optimal off-line routing rules.

Throughout our analysis, we assume that each class of traffic contributes to a fraction of the total input traffic demand. Define three fractions R_{ST-ST}, R_{ST-GT}, and R_{GT-ST} to represent the fraction of total traffic for each traffic type such that $R_{ST-ST} + R_{ST-GT} + R_{GT-ST} = 1$. For ease of analysis, we assume that $R_{ST-GT} = R_{GT-ST}$.

Let the total amount of traffic demand in the network be $r_{\max}$. There are $r_{\max} R_{ST-ST}$, $r_{\max} R_{ST-GT}$ and $r_{\max} R_{GT-ST}$ total amount of ST-ST, ST-GT, and GT-ST traffic respectively. The traffic demand matrix in the network is assumed to be uniform, all-to-all. Under this assumption, equal amount of traffic, $\frac{r_{\max}}{N}$, originates from each satellite footprint and equal amount of traffic, $\frac{r_{\max}}{N^2}$, flows from one satellite footprint to every satellite footprint in the network (including the footprint in which the traffic originates).

Since we assume that all traffic demand must be satisfied for this analysis, given C_{ISL} and $r_{\max}$, $C_{up/dn}$ is known. This simplifies computation significantly as the problem is now reduced to static traffic, worst-case dimensioning. In the following sections, we will compute effective system costs for all possible values of C_{ISL}. The optimal effective system cost is obtained by optimizing over C_{ISL}. We relax the delay constraint to simply the analysis. This has the effect of assuming all paths can satisfy delay constraints.

1. All Traffic Carried by ISLs

For the worst-case traffic realization, r_{max}, let all traffic be routed using only the ISLs and gateway links. Since all ISLs must have the same link dimension (identical link assumption), the optimal routing strategy needs to minimize the load on the worst link. We will show that minimum hop (min-hop) routing is optimal in Theorem 5.3.1. Min-hop routing is a strategy that routes traffic along the path that has the least number of links between the source and destination. If there exist more than one such paths, traffic is split equally and carried on all min-hop paths.

***Theorem 5.3.1** (Optimality of min-hop routing):*

For the GEO satellite network shown in Figure 5-10, and under the uniform all-to-all traffic assumption, if all of the traffic can be accommodated by ISLs, min-hop routing results in the lowest amount of traffic on the most heavily loaded link and is the optimal routing strategy. In fact, all links are equally loaded under min-hop routing.

Proof:

The proof will be presented in two parts. First, we derive the minimum amount of traffic required to flow on any link in order to satisfy the traffic demand. Next, we show that min-hop routing achieves the minimum flow on all of the links.

Consider a ring network with N nodes. Making a cut across any two links results in two disjoint networks, A and B, as shown in Figure 5-11.

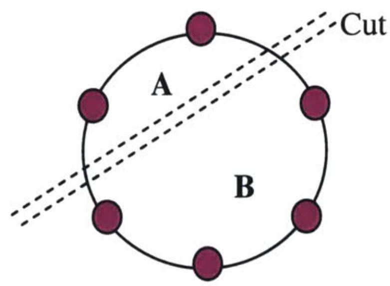


Figure 5-11 Cut Set for a N node Satellite Network

For uniform all-to-all traffic, the total amount of traffic flow from A to B is $\frac{r_{max}}{N^2}(N-M)M$, where M is the number of nodes in A. This is identical to the total amount of traffic flow from B to A. Since all links must be dimensioned to the most heavily loaded link, it is optimal for the links that have been cut to carry equal amount of traffic. Hence, each link that has been cut should carry:

$$L = \frac{r_{max}}{2N^2}(N-M)M \tag{5.38}$$

Now we identify the integer number M^* such that Equation 5.38 is maximized. Although M can only take on integer values, we try to minimize Equation 5.38 over continuous values for M.

$$\frac{dL}{dM} = \frac{r_{max}}{2N^2}(N - 2M) = 0 \tag{5.39}$$

$$M^* = \begin{cases} \frac{N}{2} & \text{for N:even} \\ \frac{N+1}{2} & \text{for N:odd} \end{cases} \tag{5.40}$$

Substituting Equation 5.40 into Equation 5.38 results in the maximum cut-set flow on a link that has been cut:

$$L_{\max} = \begin{cases} \frac{r_{max}}{8} & \text{for N:even} \\ \frac{r_{max}\left(N^2 - 1\right)}{8N^2} & \text{for N:odd} \end{cases} \tag{5.41}$$

Since the ring network is symmetrical, the maximum cut-set flow, $L_{\max}$, represents the minimum amount of traffic each link must carry.

We will now show that min-hop routing of uniform all-to-all traffic achieves this minimum. Using min-hop routing, the total number of hops, T_{hop}, from all OD pairs is:

For N:odd

$$\begin{aligned} T_{hop} &= N\sum_{i=1}^{\frac{N-1}{2}} 2i = \frac{2N\left(\frac{N-1}{2}+1\right)\left(\frac{N-1}{2}\right)}{2} \\ &= \frac{N(N^2-1)}{4} \end{aligned} \tag{5.42}$$

and for N:even

$$\begin{aligned} T_{hop} &= N\left(\sum_{i=1}^{\frac{N}{2}-1} 2i + \frac{N}{2}\right) = N\frac{2\left(\frac{N}{2}-1+1\right)\left(\frac{N}{2}-1\right)}{2} + \frac{N^2}{2} \\ &= \frac{N^3}{4} \end{aligned} \tag{5.43}$$

The amount of traffic carried on each link can then be calculated as:

$$L_{\text{min-hop}} = T_{hop} \times \frac{r_{\max}}{N^2} \times \frac{1}{\text{total number of links}}$$

$$L_{\text{min-hop}} = \begin{cases} \dfrac{N^3}{4}\dfrac{r_{\max}}{N^2}\dfrac{1}{2N} = \dfrac{r_{\max}}{8} & \text{for N:even} \\ \dfrac{N(N^2-1)}{4}\dfrac{r_{\max}}{N^2}\dfrac{1}{2N} = \dfrac{r_{max}\left(N^2-1\right)}{8N^2} & \text{for N:odd} \end{cases} \tag{5.44}$$

It can be seen that $L_{\text{min-hop}} = L_{\max}$. Hence min-hop routing is the optimal routing strategy.

□

Using Theorem 5.3.1, under min-hop routing, the optimal unidirectional ISL link capacity is:

$$C_{ISL} = \begin{cases} \dfrac{r_{\max}\left(N^2-1\right)}{8N^2} & \text{if N:odd} \\ \dfrac{r_{\max}}{8} & \text{if N:even} \end{cases} \tag{5.45}$$

The unidirectional gateway link capacity required to accommodate all traffic is:

$$C_{up/dn} = \frac{r_{\max} R_{ST-GT}}{N} \tag{5.46}$$

This is obtained by observing that only ST-GT and GT-ST traffic require routing over the gateway links.

In this case, since all traffic is routed using the ISLs, there is no routing cost. The resulting effective system cost is:

$$J = \begin{cases} m_1 \dfrac{r_{\max}\left(N^2-1\right)}{4N} + 2m_3 r_{\max} R_{ST-GT} + 2N\left(m_2 + m_4\right) & \text{if N:odd} \\ m_1 \dfrac{r_{\max} N}{4} + 2m_3 r_{\max} R_{ST-GT} + 2N\left(m_2 + m_4\right) & \text{if N:even} \end{cases} \tag{5.47}$$

Observations:

1. As N increases, the effective system cost increases linearly. At first glance, under this traffic pattern, it is always best to use the fewest number of satellites. However, as shown in Chapter 3, all of the cost parameters are functions of N. The lower the N, the longer the distance between the links and the higher the link costs. Hence, depending on the cost parameters, there exists an optimal N that may be higher than the minimum of three satellites for a GEO network.
2. The effective system cost depends on the relative fractions of traffic classes. For higher fractions of ST-GT and GT-ST traffic, the effective system cost for this topology is higher since $\frac{dJ}{dR_{ST-GT}} = 2m_3 r_{\max}$ is positive. This relationship makes sense, since ST-GT and GT-ST traffic generate extra gateway link costs that are not generated by ST-ST traffic.
3. Since all traffic are routed on the ISLs, ISL link capacity is high and is independent of the relative fractions of different traffic classes. On the other hand, the gateway link capacities are sensitive to the relative fractions of traffic classes. If the actual R_{ST-GT} is higher than predicted, then the gateway links will become major bottleneck links in the network.

2. All Traffic Carried by the Terrestrial Network

For the worst case traffic realization, r_{max}, let all traffic be routed using only gateway and terrestrial links. In this case, $C_{ISL} = 0$. The maximum gateway link dimension is:

$$C_{up/dn} = \frac{r_{\max} R_{ST-GT}}{N} + r_{\max} R_{ST-ST}\left(\frac{1}{N} - \frac{1}{N^2}\right) \tag{5.48}$$

Compared to Equation 5.46, it can be observed that ST-GT and GT-ST traffic generate equal amount of traffic on the gateway links in both cases. Most ST-ST traffic must traverse two gateway links if routed using the terrestrial network. The ST-ST traffic that originate and terminate in the same satellite footprint do not need to be routed on gateway links.

Routing cost is generated by using the terrestrial links. The amount of traffic flow on each terrestrial link interconnecting the gateways is:

$$y = \begin{cases} \dfrac{r_{\max}\left(N^2-1\right)}{8N^2} & \text{if N:odd} \\ \dfrac{r_{\max}}{8} & \text{if N:even} \end{cases}$$

The effective system cost is:

$$J = \begin{cases} 2Nm_3 r_{\max}\left(\dfrac{R_{ST-GT}}{N} + R_{ST-GT}\left(\dfrac{1}{N} - \dfrac{1}{N^2}\right)\right) + 2Nm_4 + \zeta\left[\dfrac{r_{\max}\left(N^2-1\right)}{4N}\right] & \text{if N:odd} \\ 2Nm_3 r_{\max}\left(\dfrac{R_{ST-GT}}{N} + R_{ST-GT}\left(\dfrac{1}{N} - \dfrac{1}{N^2}\right)\right) + 2Nm_4 + N\zeta\left[\dfrac{r_{\max}}{4}\right] & \text{if N:even} \end{cases} \quad (5.49)$$

Observations:

i. The effective system cost depends on the relative fractions of traffic classes. The higher the fractions of ST-GT and GT-ST traffic, the lower the effective system cost for this topology. This is intuitive since ST-ST traffic must traverse two gateway links whereas ST-GT and GT-ST traffic only traverse one gateway link. Substituting in $R_{ST-ST} = 1 - 2R_{ST-GT}$, Equation 5.49 can be rewritten as:

$$J = \begin{cases} 2Nm_3 r_{\max}\left(\dfrac{(N-1)-(N-2)R_{ST-GT}}{N^2}\right) + 2Nm_4 + \zeta\left[\dfrac{r_{\max}\left(N^2-1\right)}{4N}\right] & \text{if N:odd} \\ 2Nm_3 r_{\max}\left(\dfrac{(N-1)-(N-2)R_{ST-GT}}{N^2}\right) + 2Nm_4 + N\zeta\left[\dfrac{r_{\max}}{4}\right] & \text{if N:even} \end{cases}$$

$$\frac{dJ}{dR_{ST-GT}} = -2m_3 r_{\max}\left(\frac{N-2}{N}\right) \quad (5.50)$$

From Equation 5.50, it can be observed that compared to routing on ISLs only, the dependence of effective system cost on R_{ST-GT} is not as high for small N.

ii. For this routing method, the ISLs are not required. The gateway links must be dimensioned to have high capacity.

3. Hybrid Routing

If $C_{ISL} < r_{\max}$, then some traffic must be routed on the terrestrial links. We will show that ST-GT and GT-ST traffic should be routed to the terrestrial links first before ST-ST traffic. If ST-ST traffic must be routed to the ground, it is optimal to route higher hop traffic to ground first. First we show that, in a ring network, each ISL carry M-hop traffic from M different OD pairs.

Lemma 5.3.1 *(M-hop Traffic on Each ISL):*

In a N-node ring network, under min-hop routing, each ISL carry M-hop traffic from M different OD pairs, where:

$$1 \le M \le \begin{cases} \dfrac{N-1}{2} & \text{N:odd} \\ \dfrac{N}{2} & \text{N:even} \end{cases} \tag{5.51}$$

Two examples are shown in Figure 5-12.

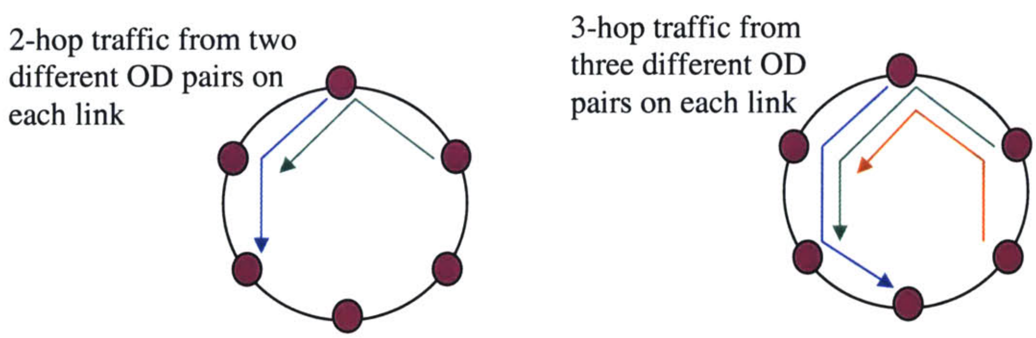


Figure 5-12 Illustration of M-hop Traffic on a Link

Proof:

Take an arbitrary node and label it node 0. Label all other nodes as shown in Figure 5-13.

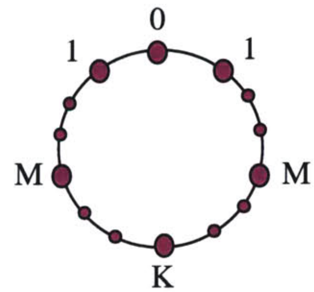


Figure 5-13 M-hop Traffic in a Ring Network

For N:even, $K = \frac{N}{2}$ and there is only one node labeled with K as illustrated in Figure 5-13. For N:odd, $K = \frac{N-1}{2}$ and there are two nodes labeled with K. Hence $1 \le M \le K$. Consider the link $(0,1)$. For every node labeled 1 to *M*, one *M*-hop traffic terminates at the node. All of these traffic must traverse the link $(0,1)$. Hence on link $(0,1)$, there are *M*-hop traffic from *M* different OD pairs. Since a ring network is symmetrical, any node can be considered to be node 0. Therefore, each ISL carries *M*-hop traffic from *M* different OD pairs.

□

Now we present the optimal routing scheme for a hybrid network.

Theorem 5.3.2 *(Optimal Routing for Hybrid Network):*

For the GEO satellite network shown in Figure 5-10, and under the uniform all-to-all traffic assumption, if the ISL links cannot accommodate all traffic by min-hop routing, then ST-GT and GT-ST traffic must be routed to the terrestrial links first before ST-ST traffic to minimize the effective system cost. If ST-ST traffic must be routed to ground, routing longer hop traffic to the ground first minimizes the effective system cost.

Proof:

Assume C_{ISL} is dimensioned to accommodate all traffic by min-hop routing. The gateway links must be dimensioned to accommodate all ST-GT and GT-ST traffic, since these traffic must traverse downlink or uplink at least once. Now consider decreasing the capacity on each ISL by one unit. The excess traffic that cannot be carried by ISLs must be carried by the terrestrial network. Hence, each terrestrial link flow will increase by one unit regardless of the class of traffic routed to the ground. If the excess traffic routed to the ground is ST-GT or GT-ST traffic, the gateway link flows remain the same; however, if the excess traffic routed to the ground is ST-ST traffic, gateway link flow must increase. Hence, if ST-ST traffic is routed to the ground, it will incur higher cost on the gateway links compared to routing ST-GT and GT-ST traffic to the ground. Therefore, the optimal routing strategy when ISLs are congested is to route ST-GT and GT-ST traffic to the ground before ST-ST traffic.

Now we show that if ST-ST traffic must be routed to the ground, it is optimal to route higher hop traffic to ground first. Assume the ISLs can support all ST-ST traffic using min-hop routing. Decreasing the capacity of each ISL by one unit results in one unit of flow on each terrestrial link. From Lemma 5.3.1, each ISL carry M-hop traffic from M different OD pairs. By decreasing each ISL capacity by one unit, the amount of M-hop traffic that needs to be routed to the ground is $\frac{1}{M}$ from each M-hop OD pair. Hence, routing higher hop traffic to the ground first results in less traffic being routed to the ground per M-hop OD pair. This implies lower cost on the gateway links. Therefore, if ST-ST traffic must be routed to the ground, it is optimal to route the higher hop traffic to ground first.

□

From Theorem 5.3.2, if C_{ISL} cannot accommodate all of $r_{\max}$, but can accommodate all ST-ST traffic, some GT-ST and ST-GT traffic must be routed to the ground. This occurs when:

$$\begin{aligned} &\frac{r_{\max}\left(N^2-1\right)}{8N^2} > C_{ISL} \geq \frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2} \quad \text{if N:odd} \\ &\frac{r_{\max}}{8} > C_{ISL} \geq \frac{r_{\max}R_{ST-ST}}{8} \quad \text{if N:even} \end{aligned} \tag{5.52}$$

In the subsequent analysis, we assume N:odd. On each ISL, we try to route all ST-ST traffic first. After routing all ST-ST traffic by min-hop routing, each ISL must carry $\frac{r_{\max} R_{ST-ST}\left(N^2-1\right)}{8N^2}$ amount of traffic. The amount of capacity left over on each ISL after routing ST-ST traffic is $C_{ISL}-\frac{r_{\max} R_{ST-ST}\left(N^2-1\right)}{8N^2}\geq 0$. Therefore, on each ISL link, $\left(C_{ISL}-\frac{r_{\max} R_{ST-ST}\left(N^2-1\right)}{8N^2}\right)$ amount of ST-GT and GT-ST traffic may be carried, giving a total of C_{ISL} amount of traffic on each ISL link (full utilization of ISL links). The un-routed ST-GT and GT-ST traffic on each link, $\frac{r_{\max}\left(N^2-1\right)}{8N^2}-C_{ISL}$, must be routed on the ground. The gateway links still have $\frac{r_{\max} R_{ST-GT}}{N}$ traffic since only ST-GT and GT-ST traffic needs to be routed on the ground. The ground links will each have $\frac{r_{\max}\left(N^2-1\right)}{8N^2}-C_{ISL}$ amount of traffic.

The effective system cost is:

$$J=\begin{cases} 2Nm_1C_{ISL}+2m_3r_{\max}R_{ST-GT}+2N\left(m_2+m_4\right)+2N\zeta\left(\frac{r_{\max}\left(N^2-1\right)}{8N^2}-C_{ISL}\right) & \text{if N:odd} \\ 2Nm_1C_{ISL}+2m_3r_{\max}R_{ST-GT}+2N\left(m_2+m_4\right)+2N\zeta\left(\frac{r_{\max}}{8}-C_{ISL}\right) & \text{if N:even} \end{cases} \qquad (5.53)$$

Observations:

1.The effective system cost is a linear function of C_{ISL} in this region.

$$\frac{dJ}{dC_{ISL}}=2N\left(m_1-\zeta\right) \qquad (5.54)$$

if $m_1 - \zeta > 0$, then J is an increasing function of C_{ISL} ; otherwise, J is a decreasing function C_{ISL} .

2.In this scenario, C_{ISL} is lower compared to using only ISL, while $C_{up/dn}$ is unchanged in this region. Because some of the GT-ST and ST-GT traffic must be carried by terrestrial links, routing costs are generated.

For $C_{ISL} < \frac{r_{\max} R_{ST-ST}\left(N^2-1\right)}{8N^2}$, the ISL links can no longer support any ST-GT and GT-ST traffic. Furthermore, some ST-ST traffic also has to be routed using ground links. On each ISL link, the amount of overflow ST-ST traffic is: $\frac{r_{\max} R_{ST-ST}\left(N^2-1\right)}{8N^2} - C_{ISL} \geq 0$. The optimal routing strategy is to route the longer hop traffic to ground first to reduce the amount of flow on the gateway links as shown in Theorem 5.3.2. Next, we will show that the optimal routing strategy results in a piecewise-linear, convex function of J with respect to C_{ISL} .

***Theorem 5.3.3** (Piecewise-Linear Convex Effective System Cost):*

For the GEO satellite network shown in Figure 5-10, and under the uniform all-to-all traffic assumption, if traffic are routed according to the optimal routing strategy shown in Theorem 5.3.2, the effective system cost is a piecewise linear, convex function with respect to C_{ISL} .

Proof:

Observe that by routing ST-GT and GT-ST traffic to ground first, the effective system cost, obtained by rearranging Equation 5.52, is:

$$J = \begin{cases} 2NC_{ISL}\left(m_1-\zeta\right)+2m_3 r_{\max} R_{ST-GT}+2N\left(m_2+m_4\right)+\dfrac{\zeta r_{\max}\left(N^2-1\right)}{4N} & \text{if N:odd} \\ 2NC_{ISL}\left(m_1-\zeta\right)+2m_3 r_{\max} R_{ST-GT}+2N\left(m_2+m_4\right)+\dfrac{N\zeta r_{\max}}{4} & \text{if N:even} \end{cases} \quad (5.55)$$

This is a linear function of C_{ISL}, valid over the range:

$$\frac{r_{\max}\left(N^2-1\right)}{8N^2} > C_{ISL} \geq \frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2} \quad \text{if N:odd}$$

$$\frac{r_{\max}}{8} > C_{ISL} \geq \frac{r_{\max}R_{ST-ST}}{8} \quad \text{if N:even}$$

According to Theorem 5.3.2, if ST-ST traffic must be routed to the ground, the longer hop traffic must be routed to the ground before shorter hop traffic. For subsequent analysis, we analyze for the case of N:odd only.

For each node in a ring, there are two nodes that are $\frac{N-1}{2}$ hops away. This is the longest distance between any two nodes on a ring. Each ISL carries $\frac{N-1}{2}$ longest distance traffic under min-hop routing as shown in Lemma 5.3.1. Consider decreasing the capacity on each ISL by one unit, each node must drop $2\left(\frac{2}{N-1}\right)$ amount of traffic. Hence, each gateway link must increase capacity by $\frac{4}{N-1}$ since all of the traffic dropped by ISLs must be routed to the ground network. The effective system cost can then be expressed as:

$$\begin{aligned} J &= 2Nm_1C_{ISL} + 2Nm_3\left(\frac{r_{\max}R_{ST-GT}}{N} + \left(\frac{4}{N-1}\right)\left(\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2} - C_{ISL}\right)\right) \\ &\quad + 2N\left(m_2+m_4\right) + 2N\zeta\left(\frac{r_{\max}\left(N^2-1\right)}{8N^2} - C_{ISL}\right) \\ &= 2N\left(m_1 - \left(\frac{4}{N-1}\right)m_3 - \zeta\right)C_{ISL} + 2Nm_3\left(\frac{r_{\max}R_{ST-GT}}{N} + \left(\frac{4}{N-1}\right)\left(\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}\right)\right) \\ &\quad + 2N\left(m_2+m_4\right) + 2N\zeta\left(\frac{r_{\max}\left(N^2-1\right)}{8N^2}\right) \end{aligned} \tag{5.56}$$

This is a linear function with respect to C_{ISL}, valid for:

$$\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}-\frac{r_{\max}R_{ST-ST}}{N^2}\left(\frac{N-1}{2}\right)\le C_{ISL}<\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}.$$

In general, for each node in a ring, there are two nodes that are $\frac{N-1}{2}-i=\frac{N-1-2i}{2}$ hops away, where i is an integer, $i\in\left[0,\frac{N-3}{2}\right]$. When these traffic are dropped, the effective system cost can be written as:

$$\begin{aligned}J&=2Nm_1C_{ISL}+2N\left(m_2+m_4\right)+2N\zeta\left(\frac{r_{\max}\left(N^2-1\right)}{8N^2}-C_{ISL}\right)\\&+2Nm_3\left(\frac{r_{\max}R_{ST-GT}}{N}+\sum_{j=1}^{i}\left(\frac{2r_{\max}R_{ST-ST}}{N^2}\right)+\left(\frac{4}{N-1-2i}\right)\left(\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}-\sum_{k=0}^{i-1}\frac{r_{\max}R_{ST-ST}}{N^2}\left(\frac{N-1-2k}{2}\right)-C_{ISL}\right)\right)\end{aligned}\tag{5.57}$$

Rearranging Equation 5.57, we obtain:

$$\begin{aligned}J&=2N\left(m_1-\left(\frac{4}{N-1-2i}\right)m_3-\zeta\right)C_{ISL}\\&+2Nm_3\left(\frac{r_{\max}R_{ST-GT}}{N}+\sum_{j=1}^{i}\left(\frac{2r_{\max}R_{ST-ST}}{N^2}\right)+\left(\frac{4}{N-1-2i}\right)\left(\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}-\sum_{k=0}^{i-1}\frac{r_{\max}R_{ST-ST}}{N^2}\left(\frac{N-1-2k}{2}\right)\right)\right)\\&+2N\left(m_2+m_4\right)+2N\zeta\left(\frac{r_{\max}\left(N^2-1\right)}{8N^2}\right)\end{aligned}\tag{5.58}$$

valid for

$$\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}-\sum_{k=0}^{i}\frac{r_{\max}R_{ST-ST}}{N^2}\left(\frac{N-1-2k}{2}\right)\le C_{ISL}<\frac{r_{\max}R_{ST-ST}\left(N^2-1\right)}{8N^2}-\sum_{k=0}^{i-1}\frac{r_{\max}R_{ST-ST}}{N^2}\left(\frac{N-1-2k}{2}\right)\tag{5.59}$$

Equation 5.58 is a linear function with respect to C_{ISL}, in the form of:

$$J=a_iC_{ISL}+b_i\tag{5.60}$$

where a_i decreases and b_i increases with increase in i. Increasing in i corresponds to lower C_{ISL} regions. Hence, as C_{ISL} increases from one region to the next, a_i increases and b_i decreases for discrete regions of C_{ISL}. Within each region, the effective system cost is a linear function. Since b_i decreases monotonically from one region to another as C_{ISL} increases, the effective system cost is a piece-wise linear convex function. Hence, an optimal effective system cost exists over the closed interval of C_{ISL} and the optimal is unique.

□

To clearly illustrate the results in Theorem 5.3.3, we derive analytical expressions for optimal link capacities and optimal effective system cost for three and four satellites. The results are shown in Table 5-2.

Table 5-2 Optimal Effective System Cost for Three and Four GEO Satellite Systems

N	Link Capacities	Effective System Cost	Break points
3	$C_{ISL}=0$ $C_{up/dn}=\frac{(2-R_{ST-GT})r_{\max}}{9}$	$J=2m_3\left(\frac{r_{\max}(2-R_{ST-GT})}{3}\right)+6(m_4)+\frac{2\zeta r_{\max}}{3}$	One $C_{ISL}=\frac{r_{\max}R_{ST-ST}}{9}$
	$C_{ISL}\le\frac{r_{\max}R_{ST-ST}}{9}$ $C_{up/dn}=\frac{r_{\max}R_{ST-GT}}{3}+2\left(\frac{r_{\max}R_{ST-ST}}{9}-C_{ISL}\right)$	$J=6(m_1-2m_3-\zeta)C_{ISL}+2m_3\left(\frac{r_{\max}(2-R_{ST-GT})}{3}\right)+6(m_2+m_4)+\frac{2\zeta r_{\max}}{3}$	
	$\frac{r_{\max}R_{ST-ST}}{9}\le C_{ISL}\le\frac{r_{\max}}{9}$ $C_{up/dn}=\frac{r_{\max}R_{ST-GT}}{3}$	$J=6(m_1-\zeta)C_{ISL}+2m_3r_{\max}R_{ST-GT}+6(m_2+m_4)+\frac{2\zeta r_{\max}}{3}$	
	$C_{ISL}=\frac{r_{\max}}{9}$ $C_{up/dn}=\frac{r_{\max}R_{ST-GT}}{3}$	$J=2m_3(r_{\max}R_{ST-GT})+6(m_2+m_4)+2m_1\left(\frac{r_{\max}}{3}\right)$	
4	$C_{ISL}=0$	$J=2m_3\left(r_{\max}R_{ST-GT}+\frac{3r_{\max}R_{ST-ST}}{4}\right)+8(m_4)+\zeta r_{\max}$	Two $\frac{r_{\max}R_{ST-ST}}{16}=C_{ISL}$ $\frac{r_{\max}R_{ST-ST}}{8}=C_{ISL}$
	$C_{ISL}\le\frac{r_{\max}R_{ST-ST}}{16}$	$J=8(m_1-2m_3-\zeta)C_{ISL}+2m_3\left(r_{\max}R_{ST-GT}+\frac{3r_{\max}R_{ST-ST}}{4}\right)+8(m_2+m_4)+\zeta r_{\max}$	
	$\frac{r_{\max}R_{ST-ST}}{16}\le C_{ISL}\le\frac{r_{\max}R_{ST-ST}}{8}$	$J=8(m_1-m_3-\zeta)C_{ISL}+2m_3\left(r_{\max}R_{ST-GT}+\frac{r_{\max}R_{ST-ST}}{2}\right)+8(m_2+m_4)+\zeta r_{\max}$	
	$\frac{r_{\max}R_{ST-ST}}{8}\le C_{ISL}\le\frac{r_{\max}}{8}$	$J=8(m_1-\zeta)C_{ISL}+2m_3r_{\max}R_{ST-GT}+8(m_2+m_4)+\zeta r_{\max}$	
	$C_{ISL}=\frac{r_{\max}}{8}$	$J=m_1r_{\max}+2m_3r_{\max}R_{ST-GT}+8(m_2+m_4)$	

From these expressions, we can plot effective system cost with respect to ISL link capacity for different cost parameter values. For $N = 3$, there are three scenarios:

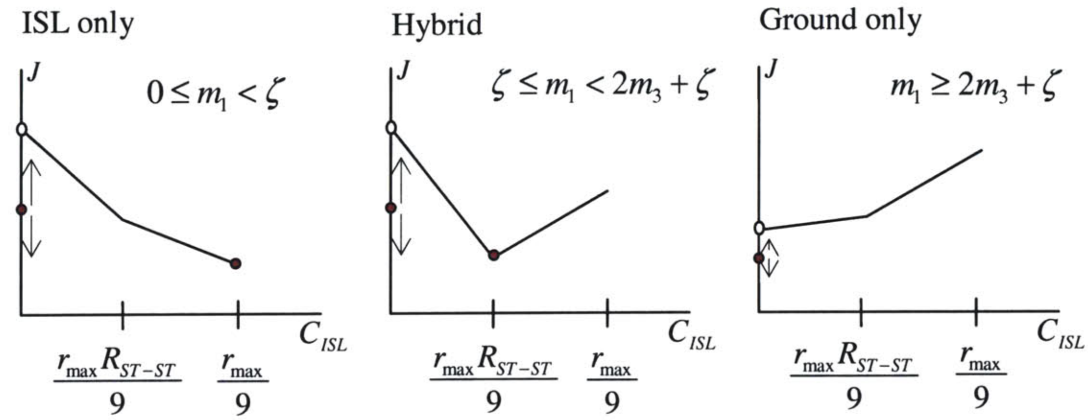


Figure 5-14 Optimal Effective System Cost for Three GEO Satellites

Note that depending on the value of fixed cost for ISL, m_2, there exists a possible optimal effective system cost when $C_{ISL} = 0$ for all three scenarios. When m_2 is high, the cost value at $C_{ISL} = 0$ decreases, and may fall below the predicted optimal points in Figures a and b. Thus if m_2 is very high, the ground only topology is optimal regardless of the value of the other cost parameters. Comparing the optimal effective system cost of the ground only and ISL only topologies for $0 \le m_1 < \zeta$, observe that the ISL only topology should be chosen if:

$$2m_3\left(\frac{r_{max}\left(2-R_{ST-GT}\right)}{3}\right)+6\left(m_4\right)+\frac{2\zeta r_{max}}{3} > 2m_3\left(r_{max}R_{ST-GT}\right)+6\left(m_2+m_4\right)+2m_1\left(\frac{r_{max}}{3}\right)$$

$$\frac{r_{max}}{9}\left(2m_3R_{ST-ST}+\zeta-m_1\right) > m_2$$

Comparing the optimal effective system cost of the hybrid and ground only topologies for $0 \le m_1 < 2m_3 + \zeta$, observe that the hybrid topology should be chosen if:

$$2m_3\left(\frac{r_{max}\left(2-R_{ST-GT}\right)}{3}\right)+6\left(m_4\right)+\frac{2\zeta r_{max}}{3} > 6\left(m_1-\zeta\right)\frac{r_{max}R_{ST-ST}}{9}+2m_3r_{max}R_{ST-GT}+6\left(m_2+m_4\right)+\frac{2\zeta r_{max}}{3}$$

$$\frac{\left(2m_3+\zeta-m_1\right)r_{max}R_{ST-ST}}{9} > m_2$$

If the above conditions are not satisfied, then it is always optimal to use the ground only topology for a three GEO satellite network. Note that for both conditions the fixed cost of an ISL has an upper bound that depends on R_{ST-ST}. If ST-ST traffic is a small percentage of the total traffic, it is more likely that a ground only topology is more economical. In general, if the fixed cost of an ISL is too high, it is never optimal to use ISLs.

For $N = 4$, there are four scenarios:

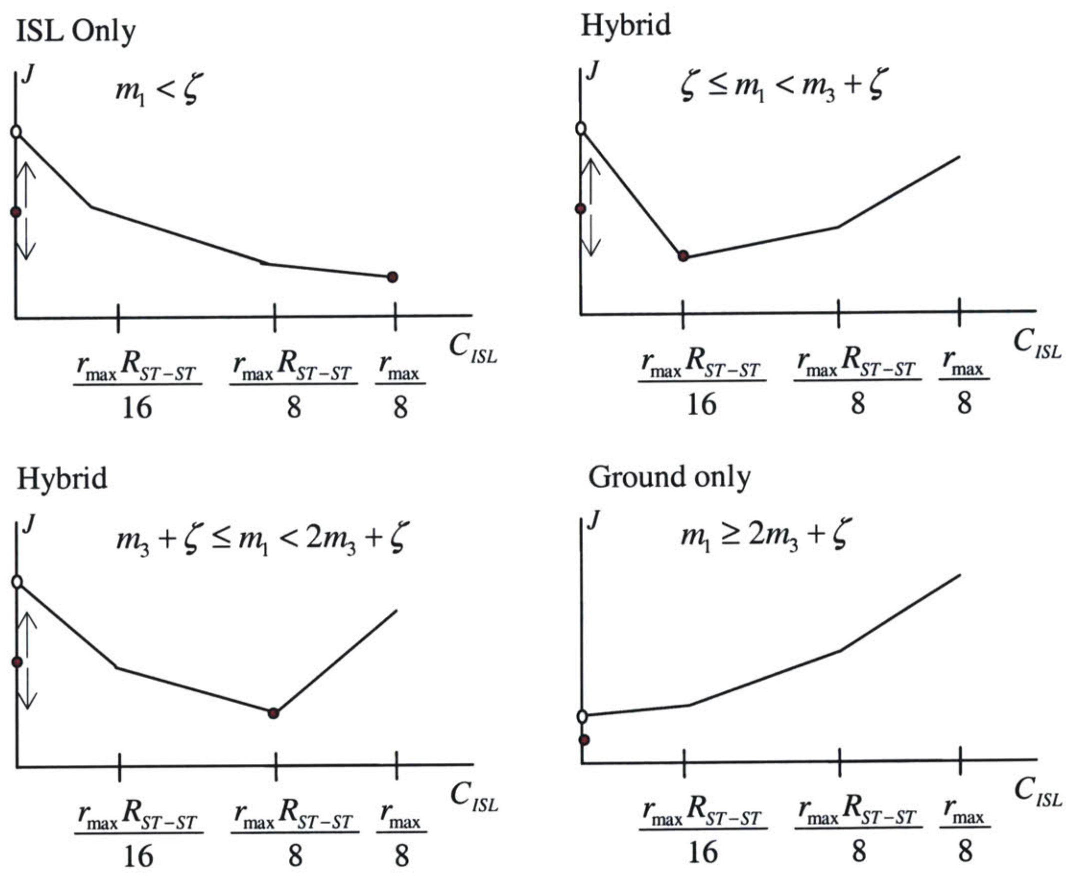


Figure 5-15 Optimal Effective System Cost for Four GEO Satellites

As with the three satellite case, there exist cost parameters such that one of the three topologies is more cost effective. For the hybrid topology, there are two different optimal effective system cost values depending on the relative link marginal costs. Again, if the fixed cost of an ISL is too high, it is always optimal to use the ground only topology.

We plot the feasible regions for a general GEO satellite network in Figure 5-16. Note that the feasible region for routing on the ground is the entire positive quadrant. If m_2 is high, then routing all of the traffic on the terrestrial network results in the most economical network. If m_2 is sufficiently low, then three regions may be identified for optimal network topology selection based on marginal link costs. If $m_1 < \zeta$, it is most economical to route all the traffic using ISLs. Hence, the ISLs should be dimensioned to support as much traffic as possible. If $m_1 > 2m_3 + \zeta$, it is most economical to route all the traffic using the terrestrial network. In this case, ISLs should not be used. Otherwise, a hybrid satellite-terrestrial network is the most economical. It can be observed that there exist more cost pairs (m_1, m_3) that make the hybrid topology optimal.

For the hybrid topology, the optimal routing strategy uses both ISLs and the terrestrial network. In particular, when ISL links are congested, it is most favorable to route ST-GT and GT-ST traffic to the terrestrial links before ST-ST traffic. This routing strategy ensures high utilization of the satellite segment and low satellite system cost.

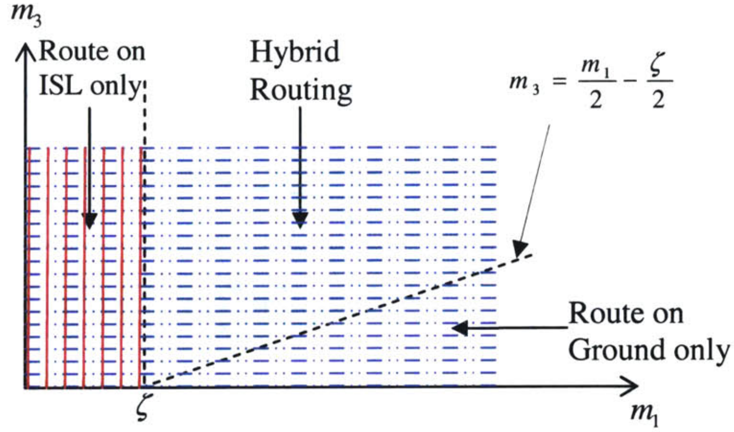


Figure 5-16 Feasible Regions for GEO Satellite Topologies

5.4 Constellation Size Optimization

For each satellite topology, we would like to find the optimal number of satellites that minimizes the effective system cost. From the link cost models developed in Chapter 3, is can be observed that link cost models are also functions of link distance. Since we assume that GEO satellites are positioned equidistant on a ring, increasing the number of satellites in the constellation decreases individual ISL and terrestrial link distances, which results in lower link costs. Hence there is a tradeoff between the incremental cost of adding more satellites and the decrease in cost due to shorter link distance. Ignoring access links in our analysis, we find that the optimal number of satellites using the ground only topology is three. For the ISL only and Hybrid topologies, there exist cost parameters such that a three satellite topology is not optimal. The effective system cost equations we use here are those obtained in Section 5.3.

From Chapter 3, as the number of satellites increase, the ISL and terrestrial routing cost decrease due to shorter link distance. ISL link distance is defined to be the line-of-sight distance between two adjacent satellites as shown in Figure 5-17 for a three satellite network.

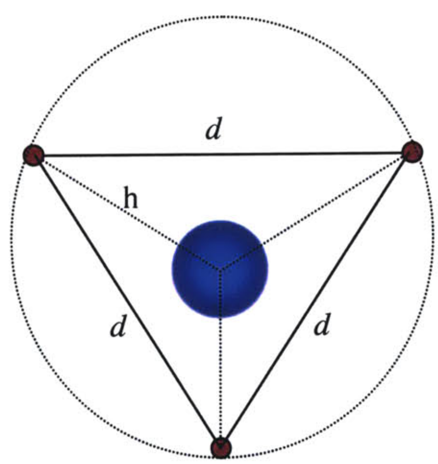


Figure 5-17 GEO Satellite ISL Link Distance

For a N-satellite network, ISL link distance is:

$$d = \sqrt{2h^2\left(1-\cos\frac{2\pi}{N}\right)} \tag{5.61}$$

For linear cost function with respect to C_{ISL}, Equation 3.7 can be rewritten as

$$\$_{ISL} = \left(k_1'\xi\right)C_{ISL}d^2 + k_2' = \left(2h^2k_1'\xi\right)\left(1-\cos\frac{2\pi}{N}\right)C_{ISL} + k_2' \tag{5.62}$$

The marginal cost of an ISL link is then:

$$m_1(N) = K_1\left(1-\cos\frac{2\pi}{N}\right), \text{ where } K_1 = 2h^2k_1'\xi \tag{5.63}$$

The marginal cost of routing on the ground is assumed to be:

$$\zeta(N) = \frac{K_2}{N} \tag{5.64}$$

where $K_2 = k2\pi r_e$, k is a cost coefficient and r_e is the radius of Earth.

The marginal cost of gateway links is assumed to be independent of the number of satellites since link distance is approximately constant. We also assume that the fixed costs are independent of number of satellites.

1. If all traffic is routed using ISL:

ISLs must be dimensioned to carry worst-case traffic. This topology is used when $0 \le m_1 < \zeta$. Rewriting Equation 5.47, we obtain:

$$J(N) = \begin{cases} K_1\left(1-\cos\frac{2\pi}{N}\right)\left(\frac{r_{max}\left(N^2-1\right)}{4N}\right) + 2m_3 r_{max} R_{ST-GT} + 2N\left(m_2+m_4\right) & \text{if N:odd} \\ K_1\left(1-\cos\frac{2\pi}{N}\right)\left(\frac{N r_{max}}{4}\right) + 2m_3 r_{max} R_{ST-GT} + 2N\left(m_2+m_4\right) & \text{if N:even} \end{cases} \tag{5.65}$$

Let $A(N)=\left(1-\cos\frac{2\pi}{N}\right)\left(\frac{(N^2-1)}{N}\right)$ for N:odd, and $B(N)=\left(1-\cos\frac{2\pi}{N}\right)N$ for N:even.

These are both decreasing functions of N. A plot of these two functions is shown in Figure 5-18.

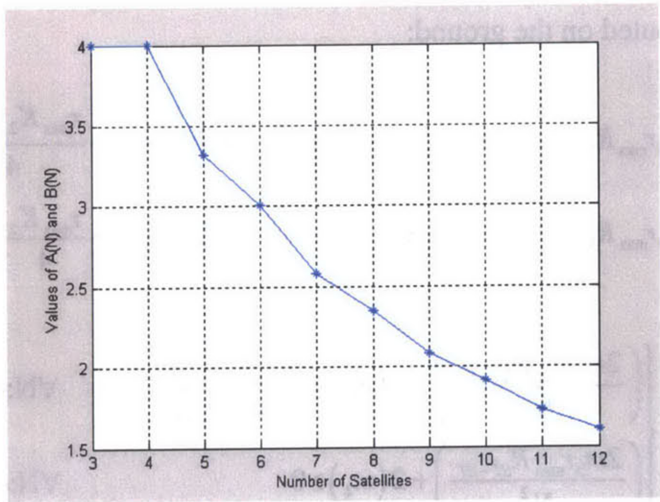


Figure 5-18 Cost Parameters vs. Number of Satellites

To find the number of satellites that minimizes the effective system cost, we approximate the effective system cost function by a continuous function:

$$\frac{dJ(N)}{dN}=K_1\left(\frac{r_{\max}}{4}\right)\left(\left(1-\cos\frac{2\pi}{N}\right)-\frac{2\pi}{N}\left(\sin\frac{2\pi}{N}\right)\right)+2(m_2+m_4) \qquad \text{if N:even}$$

$$\text{where } \frac{2-\pi}{2}\leq\left(1-\cos\frac{2\pi}{N}\right)-\frac{2\pi}{N}\left(\sin\frac{2\pi}{N}\right)<0,\ \forall N\text{:even}\geq 4$$

The left hand equality is obtained when $N=4$ for even number of satellites.

Although variable cost decreases with increase in *N*, fixed cost increases linearly with increase in *N*. In general, if $\frac{dJ(N)}{dN}>0,\ \forall N\geq 3$, then $J(N)$ is a strictly increasing function of *N*. In this case the optimal number of satellites is 3. This is the minimum number of GEO satellites needed to provide global coverage (excluding the polar regions). If $\frac{dJ(N)}{dN}<0,\ \forall N\geq 3$, then every increment in the number of satellites results in a lower cost. This case is not very likely in reality. Another scenario is $\frac{dJ(N)}{dN}=0$, for an optimal *N*. Hence, if all routing is done using ISLs, there exist some cost parameters such that a three satellite system is not optimal.

2. If all traffic is routed on the ground:

$$J(N)=\begin{cases}\left(2m_3r_{\max}R_{ST-GT}+2m_3r_{\max}R_{ST-ST}\left(1-\frac{1}{N}\right)\right)+2N(m_4)+\frac{r_{\max}K_2(N^2-1)}{4N^2} & \text{if N:odd}\\ \left(2m_3r_{\max}R_{ST-GT}+2m_3r_{\max}R_{ST-ST}\left(1-\frac{1}{N}\right)\right)+2N(m_4)+\frac{r_{\max}K_2}{4} & \text{if N:even}\end{cases} \tag{5.66}$$

$$\frac{dJ(N)}{dN}=\begin{cases}\left(\frac{2m_3r_{\max}R_{ST-ST}}{N^2}\right)+2(m_4)+\frac{r_{\max}K_2}{4}\left(\frac{1}{N^3}\right)>0 & \forall\text{N:odd}\\ \left(\frac{2m_3r_{\max}R_{ST-ST}}{N^2}\right)+2(m_4)>0 & \forall\text{N:even}\end{cases} \tag{5.67}$$

In this case, $J(N)$ is a strictly increasing function of *N*. Hence, the lowest number of satellites, $N=3$, minimizes the effective system cost function.

3. Hybrid Routing

In this case, the effective system cost function contains cost from ISL, gateway and ground. As in the ISL only case, there exist some cost parameters such that a three satellite network is not optimal. The optimal number of satellites depends on the actual cost parameters.

5.5 Summary

In this chapter, we have applied the formulation presented in Chapter 4 on GEO satellite network design problems. For a single OD pair and single traffic class applications, we have derived feasible regions of three different satellite topologies: ISL only, Ground only, and Hybrid. These feasible regions are described by marginal link costs. An optimal topology may be selected based on marginal and fixed link costs.

For a general N-node GEO satellite network with multiple classes of traffic, we assumed that rejection cost of traffic is extremely high such that all traffic demands must be satisfied by the satellite network. This reduces the problem to a static, worst-case dimensioning problem. For this problem, analytical solutions are obtained for three different routing strategies: ISL only, Ground only and Hybrid. Feasible regions for the three different routing strategies are shown. We observe that the effective system cost function is a piece-wise linear, convex function with respect to C_{ISL}.

Based on the link cost functions derived in Chapter 3, we attempt to optimize the size of a GEO satellite network to minimize the effective system cost. For the Ground only topology, a three satellite network is optimal. For the ISL only and Hybrid topologies, there exist cost parameters such that a three satellite network is not optimal.

The feasible regions derived in this chapter may be used as a first order decision making tool for optimal satellite network topology selection. Note that throughout the analysis, linear cost functions are used. For general cost functions, if the resulting effective system cost is convex, the problem can be solved computationally; however, if the resulting effective system is non-convex, the problem becomes extremely difficult to solve as there may exist many minima in the cost function. Another issue with the analytical solutions is that they are specific to the topologies we analyzed. If the satellite network topology does not exhibit symmetry or if the traffic pattern is not uniform, then the problem must be solved by computation.

Chapter 6

LEO and MEO Satellite-Terrestrial Networks

LEO and MEO satellite systems have gained much attention over the past ten years. Both Iridium and Globalstar systems use low earth orbits for their satellite constellations. A proposed ICO satellite system for data communications is to be placed in medium earth orbit. In general, altitudes of 500 to 1500km are considered to be low earth orbit while altitudes of 5000 to 10000km are considered to be medium earth orbit. Compared to a GEO system at altitude of about 35786km, the propagation delay between an earth terminal and a satellite is much lower for LEO and MEO systems. Although LEO and MEO satellite systems may offer advantages in terms of lower propagation delay, lower satellite bus complexity, and possible diversity gain, there are several disadvantages associated with LEO and MEO systems:

1. Due to lower altitude, many satellites are required to provide global service. Furthermore, many ISLs or gateways are needed to provide connectivity between regions on Earth.
2. Unlike the GEO system where only a single satellite is needed to start generating revenue, the entire constellation of LEO and MEO satellites must be in orbit prior to revenue generation. Launching an entire constellation of satellites may take a long time as many different launches must be scheduled.

3. Since satellites move relative to earth, to ensure connectivity for a communication pair, traffic must be handed over between antenna beams and between satellites. To add to the complexity, it may not be feasible to maintain ISL connectivity at all times. Changes in the network topology make routing and traffic monitoring difficult.
4. Satellite systems in lower orbits experience more degradation effects from the Earth's atmosphere. This reduces satellite system lifetime and hence the amount of time the system can be used to generate revenue.
5. It is more difficult to direct satellite resources to regions of high demand using LEO and MEO satellite systems. The satellites in the constellation are all identical and cover the Earth approximately uniformly. Hence most of the satellite resources are under-utilized most of the times.

Due to these disadvantages, it is not apparent that LEO and MEO systems are favorable architectures for future commercial broad-band data satellite networks. Nevertheless, we will analyze optimal capacity and routing issues for LEO and MEO systems that have regular topology.

The polar constellation is a regular graph that is amenable to analysis. This constellation is used by the Iridium satellite system as depicted in Figure 6-1. In this constellation, satellites are equally positioned in one plane (intra-plane). Several such planes are needed to provide complete coverage of the Earth.

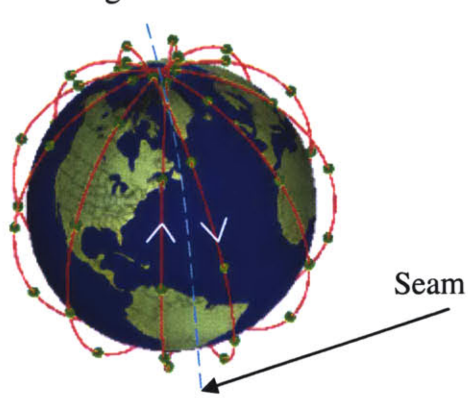


Figure 6-1 Iridium Satellite Constellation

Adapted from: www.geom.umn.edu/~worfolk/ SaVi/constellations.html

The intra-plane satellites are each interconnected by ISLs to their two nearest neighbors. These ISLs may be maintained at all times since the distance between any two intra-plane satellites is fixed regardless of the relative motion between the satellites and the Earth. Each satellite can also be interconnected to satellites in the two neighboring planes (inter-plane). As satellites orbit Earth, inter-plane satellites move closer to each other at the polar regions and move apart near the equator. For this reason, each satellite must track its two inter-plane neighbors. At the polar regions, inter-plane ISLs may need to be shut off due to difficulty in tracking satellites. As shown in Figure 6-1, this constellation also results in a seam where satellites are moving in counter-rotating planes, making satellite tracking across the seam a difficult task. We compare the polar constellation shown in Figure 6-1 with and without the seam.

6.1 Polar Constellation without Seam

We assume that all ISLs can be maintained at all times in this section. This results in a regular 2-D torus graph for the satellite segment with M rows and N columns. For the ground segment, the number and location of gateways influence routing of ST-GT and GT-ST traffic, which in turn influence the cost of the system. In order to derive analytical solutions for effective system cost and optimal link capacity, we assume that each satellite has access to one gateway at any time instant. Clearly, if ISLs are used, the number of gateways can be greatly reduced while maintaining connectivity between users. However, if ISLs are not used, this assumption is a necessary condition to maintain connectivity between all users. We provide analysis for routing using only ISLs and routing using only the ground network for uniform, all-to-all traffic. Hybrid routing is considerably more difficult to analyze for LEO and MEO constellations and the results would not offer much more insight than what we have shown in Chapter 5. Hence this routing scheme is not analyzed here.

6.1.1 Routing Using ISLs Only

In [4], the author has computed the capacities required on each link for a LEO satellite network with uniform all-to-all traffic. Detailed derivations can be found in Appendix B. We summarize the results in Table 6-1.

Table 6-1 Capacity Required on each ISL Link

	C_x	C_y
M and N are odd integers >= 3	$\frac{r_{max}}{8M}\left(1-\frac{1}{N^2}\right)$	$\frac{r_{max}}{8N}\left(1-\frac{1}{M^2}\right)$
M and N are even integers >= 2	$\frac{r_{max}}{8M}$	$\frac{r_{max}}{8N}$
M is odd integer >=3 and N is even integer >= 2	$\frac{r_{max}}{8M}$	$\frac{r_{max}}{8N}\left(1-\frac{1}{M^2}\right)$
M is even integer >=2 and N is odd integer >= 3	$\frac{r_{max}}{8M}\left(1-\frac{1}{N^2}\right)$	$\frac{r_{max}}{8N}$

To satisfy all the input traffic demand, each gateway link must carry

$$C_{up/dn} = \frac{r_{max} R_{ST-GT}}{MN} \tag{6.1}$$

For the subsequent analysis, we only show the case where M and N are both even integers. Since all traffic is carried by the ISLs, there is no traffic on the ground. Hence the effective system cost is:

$$\begin{aligned} J &= 2MN\left(m_1\left(\frac{r_{max}}{8N}+\frac{r_{max}}{8M}\right)+2m_2\right)+2MN\left(m_3\left(\frac{r_{max}R_{ST-GT}}{MN}\right)+m_4\right) \\ &= \frac{(M+N)r_{max}m_1}{4}+\left(2R_{ST-GT}r_{max}m_3\right)+2MN\left(2m_2+m_4\right) \end{aligned} \tag{6.2}$$

If $M = N$, then all ISLs have the same capacity. In this case, the effective system cost becomes:

$$J = \frac{Nr_{max}m_1}{2}+\left(2R_{ST-GT}r_{max}m_3\right)+2N^2\left(2m_2+m_4\right) \tag{6.3}$$

6.1.2 Routing Using Terrestrial Network Only

We assume that the gateways are also interconnected in a Torus network. On the gateway links, each gateway link must carry

$$C_{up/dn} = \frac{r_{\max} R_{ST-GT}}{MN} + r_{\max} R_{ST-ST}\left(\frac{1}{MN} - \frac{1}{(MN)^2}\right) \tag{6.4}$$

For the subsequent analysis, we only show the case where M and N are both even integers. The effective system cost is:

$$\begin{aligned} J &= 2MN\left(m_3\left(\frac{r_{\max} R_{ST-GT}}{MN} + r_{\max} R_{ST-ST}\left(\frac{1}{MN} - \frac{1}{(MN)^2}\right)\right) + m_4\right) + 2MN\zeta\left(\frac{r_{\max}}{8N} + \frac{r_{\max}}{8M}\right) \\ &= 2r_{\max} m_3\left(R_{ST-GT} + R_{ST-ST}\left(1 - \frac{1}{MN}\right)\right) + 2MNm_4 + \frac{\zeta r_{\max}(M+N)}{4} \end{aligned} \tag{6.5}$$

For $M = N$,

$$J = 2r_{\max} m_3\left(R_{ST-GT} + R_{ST-ST}\left(1 - \frac{1}{N^2}\right)\right) + 2N^2 m_4 + \frac{\zeta r_{\max} N}{2} \tag{6.6}$$

For a polar constellation without seam and under uniform all-to-all traffic, we can see that the input traffic can be distributed evenly across network links through min-hop routing. This will be contrasted with the case where a seam exists in the constellation. From an architectural standpoint, if ISLs are used, then a hybrid routing scheme is more sensible since the terrestrial network can be used to alleviate congestion on the satellite links and allow the ISLs to achieve high utilization. The routing strategy for hybrid routing is identical to the routing rules we have obtained in Theorem 5.3.2. ST-GT and GT-ST with longer hops between source and destination should be routed to the ground first.

6.2 Polar Constellation with Seam

Now we analyze a general LEO or MEO system in polar constellation with seam. We still assume that each satellite has one gateway in its footprint.

6.2.1 Routing Using ISLs Only

We show the detailed derivation in Appendix B. Here we summarize the main results.

In the y direction, all ISLs have equal capacity:

$$\text{If M = odd, } C_y = \frac{r_{\max}}{8N}\left(1-\frac{1}{M^2}\right) \tag{6.7}$$

$$\text{If M = even, } C_y = \frac{r_{\max}}{8N} \tag{6.8}$$

In the x direction, the center links are required to carry more capacity. The highest capacity required on a link is:

$$\text{If N = odd, } C_{x,\max} = \left(\frac{r_{\max}}{4M}\right)\left(1-\frac{1}{N^2}\right) \tag{6.9}$$

$$\text{If N = even, } C_{x,\max} = \frac{r_{\max}}{4M} \tag{6.10}$$

Since all ISLs must be identical in the x-direction, the links must be dimensioned to the capacity above. This leads to underutilization of the links on the edge of the constellation. We may derive the amount of flow on the least loaded link:

$$C_{x,\min} = \frac{r_{\max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right) \tag{6.11}$$

We define a link utilization factor, ρ_x, to be the ratio between the flow on the least loaded link to the maximum link capacity in the x-direction:

$$\text{If N = odd, } \rho_x = \frac{\frac{r_{max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right)}{\left(\frac{r_{max}}{4M}\right)\left(1-\frac{1}{N^2}\right)} = \frac{4}{N+1} \tag{6.12}$$

$$\text{If N = even, } \rho_x = \frac{\frac{r_{max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right)}{\frac{r_{max}}{4M}} = \frac{4(N-1)}{N^2} \tag{6.13}$$

ρ_x is plotted in Figure 6-2. As the number of satellites increase, the worst-case link utilization decreases tremendously. This illustrates one of the inefficiencies of the polar constellation with seam.

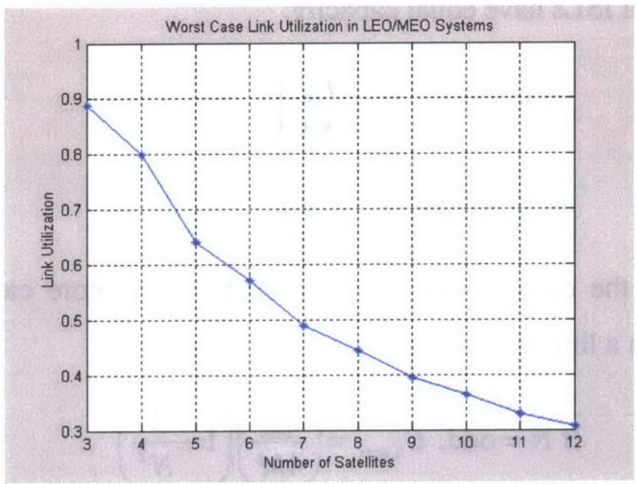


Figure 6-2 Worst Case Link Utilization for LEO/MEO Systems

On the gateway links, each gateway link must carry

$$C_{up/dn} = \frac{r_{max} R_{ST-GT}}{MN} \tag{6.14}$$

For the subsequent analysis, we only show the case where M and N are both even integers. Since all traffic is carried by the ISLs, there is no traffic on the ground. Hence the effective system cost is:

$$J = MN\left(m_1\left(\frac{r_{max}}{8N}+\frac{r_{max}}{4M}\right)+2m_2\right)+2MN\left(m_3\left(\frac{r_{max}R_{ST-GT}}{MN}\right)+m_4\right) \tag{6.15}$$

Here we used the assumption that all satellites are identical, even for the ones on the seam. In order for all ISLs to have the same capacity, we need to choose $M = 2N$.

For this topology, it can be observed that the links closer to the seam will be underutilized. In this case, hybrid routing can balance out the loads on the links and greatly improve the utilization of the inter-satellite links. Hence for a polar constellation with seam, if ISLs are used, it is best to dimension the links using a hybrid routing scheme.

6.2.2 Routing Using Terrestrial Network Only

We show the detailed derivation in Appendix B.

In the y direction, all ground links carry equal flow:

$$\text{If M = odd,} \quad f_y = \frac{r_{\max}}{8N}\left(1-\frac{1}{M^2}\right) \tag{6.16}$$

$$\text{If M = even,} \quad f_y = \frac{r_{\max}}{8N} \tag{6.17}$$

In the x direction, the center links are required to carry more flow. The sum of all the flows in the x direction is:

$$f_{x,total} = \frac{r_{\max}}{4} + \frac{r_{\max}}{2}\left(\frac{N}{2}-1\right)\left[1-\frac{(N-1)}{3N}\right] \tag{6.18}$$

On the gateway links, each gateway link must carry

$$C_{up/dn} = \frac{r_{\max}R_{ST-GT}}{MN} + r_{\max}R_{ST-ST}\left(\frac{1}{MN}-\frac{1}{(MN)^2}\right) \tag{6.19}$$

The effective system cost for even M and N is:

$$\begin{aligned} J = {} & 2MN\left(m_3\left(\frac{r_{\max}R_{ST-GT}}{MN} + r_{\max}R_{ST-ST}\left(\frac{1}{MN}-\frac{1}{(MN)^2}\right)\right)+m_4\right) \\ & +\frac{\zeta r_{\max}}{4}\left(1+\frac{1}{3}(N-2)\left(2+\frac{1}{N}\right)\right)+\frac{\zeta M r_{\max}}{8} \end{aligned} \tag{6.20}$$

6.3 Summary

In this section, we have analyzed two routing schemes using ISLs only and Ground only for LEO and MEO satellites in polar constellation. It has been shown that for a polar constellation without seam and for uniform, all-to-all traffic, traffic can be uniformly loaded on all satellite links through min-hop routing. For a polar constellation with seam, ISLs closer to the seam will be highly underutilized. Therefore, if ISLs are used, it is best to dimension the network using a hybrid routing scheme.

For LEO/MEO satellite systems, the analytical solutions provided do not capture the tradeoffs between different satellite topologies. The position and number of gateways play a more important role for LEO/MEO systems than for a GEO system. Furthermore, using ISLs tend to constrain the satellite network topology to a polar constellation. Without ISLs, the satellite segment has much more freedom in constellation design, at the expense of more gateways. We have not addressed these issues in this thesis.

Chapter 7

Conclusions and Discussions

In this thesis, we studied the problem of identifying the most cost effective satellite network topology for future broad-band data applications. In particular, we have presented the following analyses and results:

1. From fundamental relationships among power, antenna aperture, distance of the link, and link capacity, we have developed simple *link cost equations* for inter-satellite links (ISLs) and gateway links as a function of link capacity and distance. Constructing link cost equations this way allows us to study the tradeoffs between providing capacity on different links and the resulting effective system cost.

2. We have formulated the satellite network design problem using *two-stage stochastic programming*. This formulation can be used to obtain *optimal link capacity and routing strategy*, given a satellite network topology, link cost functions and input traffic demand under uncertainty. Incorporating uncertainty in input traffic demand prediction in the optimization is critical for satellite network design as it is nearly impossible to upgrade satellites in response to changes in traffic. The network obtained by solving the stochastic programming problem has the lowest effective system cost for a given set of inputs. Instead of giving a single-point solution to the optimization problem, the stochastic programming formulation allows design tradeoffs to be easily captured, which can give network designers some guidelines on the relationships between design parameters.

3. *Analytical solutions* to the optimization problem have been derived for a few classes of satellite topologies of interest. For GEO satellite networks interconnected with the ground network, we have shown that different traffic classes favor different network topologies. By favor, we mean that there exist more cost parameters leading to a particular network topology. Satellite Terminal to Satellite Terminal (ST-ST) traffic tends to favor a topology with only ISLs whereas Satellite Terminal to Ground Terminal (ST-GT) and Ground Terminal to Satellite Terminal (GT-ST) traffic tend to favor a topology that uses only the ground network. A mixture of the three traffic classes tends to favor a hybrid topology if the fixed costs of the ISL and gateway links do not dominate the link costs. The optimal topology for different traffic classes can be obtained if link cost parameters are known.

   If ISLs are used, a hybrid routing scheme improves utilization of the ISLs compared to routing all of the traffic on the ISLs. This is because the ISLs must be dimensioned to carry worst case traffic for the latter case, whereas ISL capacity can be much lower for the former case. For LEO and MEO satellite networks using a polar constellation with a seam, routing some traffic to the ground can help to balance the load on the satellite links and reduce the capacity requirement on those links.

   With hybrid routing, we have shown that it is optimal to route ST-GT and GT-ST traffic with higher number of hops between the source and destination satellite to the ground first. ST-GT and GT-ST traffic should always be routed to the ground before ST-ST traffic. This routing strategy minimizes the cost of routing traffic after the network is deployed.

   For a GEO satellite network, we have shown that if ISLs are not used, then a three-GEO satellite network is optimal. When ISLs are used, there exist link cost parameters such that a three-GEO satellite network is not optimal.

The mathematical formulation we have presented is useful for satellite network design and the analytical solutions we have obtained offer some insights into the design problem. In our analysis, we have made some stringent assumptions on the inputs to the optimization problem in order to obtain analytical solutions shown in Chapter 5 and 6. We will highlight the implications of our results and present some discussions on possible areas for improvement and directions for future research.

Link Cost Functions

In Chapter 3, we have shown simple cost models for ISL and gateway links. These cost functions are continuous functions with respect to capacity on the links. In real systems, however, link costs may take on discrete values due to physical component constraints. We would still expect the link cost to increase with increase in link capacity; however, the exact relationship between link cost and capacity needs a more in-depth study.

An ideal link cost function needs to fully capture the incremental cost in building and deploying the satellite network to incremental cost in link capacities. This entails a detailed system level study on the interactions among design parameters of different subsystems in the network. In reality, perhaps only a few link capacities and their associated system designs are feasible. This simplifies the optimization problem down to identifying the optimal link capacities that minimizes the effective system cost from the small set of feasible link capacities.

For ground links, we made the assumption that a per-flow cost is incurred whenever satellite traffic must be carried by the ground network. This cost structure may not be valid if the satellite system operator also operates the ground network, or if fiber is leased. In any case, since the satellite links costs are sunk once the system is deployed, these links should be maximally utilized. If routing traffic on the ground links incurs extra charges, then these links should be avoided, whenever possible. In general, it is more advantage to route ST-ST traffic on ISLs. When the ISLs are congested, ST-GT and GT-

ST traffic should be routed to ground before ST-ST traffic. Although we did not analyze the case where traffic is between two Ground Terminals, it is clear that if this class of traffic needs to be supported, then it should be routed to the ground first when ISLs are congested.

We did not consider the cost of providing different quality of service to different types of users in our analysis. Hence we assumed that the per-flow cost of rejecting traffic is a constant. Future satellite networks may need to support real-time and non-real-time traffic with a wide range of service requirements. The satellite system operator is likely to charge a different fee for different types of traffic in order to maximize the utilization of the satellite resources and to maximize the profit. In this case then, rejecting different types of traffic when the network becomes congested incurs different rejection costs.

All though our derivation in Chapter 5 and 6, we have assumed linear cost functions for ease of analysis. Although these solutions give us some insight into the design problem, for general link cost functions, analytical solutions cannot be obtained easily and computational analysis must be performed.

As a final observation on link costs, we note that power loss in an optical fiber is an exponential function of distance while power loss in free space is a quadratic function of distance. This difference between power loss in fiber links and ISL links may mean that ISLs are more cost effective for long distance communication compared to ground infrastructure. Detailed analysis on the cost of the various links in a hybrid satellite-terrestrial network is needed to give more indication on the optimal network topology.

Input Traffic Demand Matrix

We adopted a standard flow model for data traffic in our problem formulation which does not capture the dynamics of real traffic patterns. Real traffic varies on many time scales. Over the course of the satellite life time, business cycle will dictate aggregate traffic statistic. On a smaller time-scale, for a network of satellites serving the entire globe, time-of-day effects will be prominent. Traffic statistic may also change on minute, second, or even smaller time scale. Detailed traffic modeling taking into consideration temporal and distance effects can be useful in identifying locations of high traffic demand; thus providing indication on where the satellite network resources should be concentrated.

Network Topologies

In our analysis, we have considered several specific classes of satellite network topologies for comparison. For the GEO satellite network, the ring topology is a valid assumption; however, the number and locations of gateways as well as how gateways are interconnected with satellites have much more flexibility. In particular, we note that although the satellite segment cannot be upgraded once the system is launched into orbit, the ground facilities can be upgraded to reflect changes in traffic pattern. The ground segment can also be deployed in phases – more gateways may be added when traffic reach critical thresholds. Some flexibility should be build into the satellites to support potential gateway upgrades and modifications.

For LEO and MEO satellite networks, satellites can be positioned in many different constellations. Some constellations such as the polar constellation make ISL tracking easier to implement. Due to antenna tracking limitations, some constellations in which the relative position between satellites changes rapidly may not be feasible for ISL use. If the ISLs are not used, the set of possible constellations we can use is larger, but the number of gateways and the location of gateways are very much restricted. These design tradeoffs have significant impacts on the satellite system cost and require further study.

In our formulation, we did not include user terminal access links. To compare GEO, LEO, and MEO satellite systems, these links must be included. In order to compare these systems fairly, detailed cost studies must be performed on the various systems. As mentioned in Chapter 6, LEO and MEO have many disadvantages over the GEO satellite network. It seems that a GEO satellite network is more suitable for delay-insensitive, non-real-time data applications. A detailed comparison between the different systems in terms of overall system cost will be useful.

Conclusion

As terrestrial network infrastructures become more developed and widely spread, it will be increasingly more challenging for satellite networks to be competitive with the ground networks. Satellite network providers must find niche applications inherently suitable for satellite networks and design the satellite network which best meet those demands. Network optimization is a critical procedure for satellite network design that can help designers to identify the different tradeoffs in the design process in order to identify the most cost-effective system for future applications. The stochastic programming formulation and the results we have shown in this thesis are first steps toward designing an efficient hybrid satellite-terrestrial network.

References

[1] J. V. Evans, "Satellite Systems for Personal Communications," Proceedings of the IEEE, Vol. 86, No. 7, July 1998.

[2] E. Lutz, M. Werner, & A. Jahn, Satellite Systems for Personal and Broadband Communications, Springer-Verlag, Berlin, 2000.

[3] Eylem Ekici, Ian Akyildiz, & Michael Bender, "A Distributed Routing Algorithm for Datagram Traffic in LEO Satellite Networks," IEEE/ACM Transactions on Networking, Vol. 9, No. 2, April 2001.

[4] Kenneth Kwok, Cost Optimization and Routing for Satellite Network Constellations, M.S. dissertation, MIT, LIDS, Feb. 2001.

[5] A. Lisser, et al, "Capacity Planning Under Uncertain Demand in Telecommunications Networks," Hautes Etudes Commerciales – Genève, October 1999.

[6] Joel Schindall, MIT Satellite System and Architecture Class guest lecture notes, 2000.

[7] Martin Lo, "Satellite-Constellation Design," Computing in Science and Engineering, Volume: 1 Issue: 1, Jan.-Feb. 1999.

[8] Parametric Estimating Handbook, 2nd ed. International Society of Parametric Analysts. Spring 1999. http://www.ispa-cost.org/PEIWeb/cover.htm.

[9] Amitava Dutta & Dasaratha Rama. "An Optimization Model of Communications Satellite Planning," IEEE Transactions on Communications. Vol. 40. No. 9, Sept. 1992.

[10] Frank Kelly, "Loss Networks," The Annals of Applied Probability, Vol. 1, Issue 3, Aug. 1991.

[11] Dimitri Bertsekas & Robert Gallager, Data Networks, 2nd ed. Prentice Hall, 1992.

[12] Peter Kall, & Stein Wallace, Stochastic Programming, John Wiley & Sons Inc., 1994.

[13] Vincent Chan, "Optical Space Communications," IEEE Journal on Selected Topics in Quantum Electronics, Vol. 6, No. 6, Nov/Dec. 2000.

[14] Morris Katzman, et al. Laser satellite communications, Prentice-Hall Inc., 1987.

[15] R. Ahuja, T. Magnanti, & J. Orlin, Network Flows, Prentice-Hall Inc., 1993.

Appendix A

Derivation for ST-GT and GT-ST Traffic

A.1 ST-GT and GT-ST Terminal Traffic Carried by ISL Only

The network analyzed here is depicted in Figure 5-6.

Lemma:

For the topology shown in Figure 5-6 and for equal amount of ST-GT and GT-ST traffic, the optimal ISL link capacity equal to twice the optimal gateway link capacity, ie. $C^*_{ISL} = 2C^*_{up/dn}$.

Proof:

Suppose $C^*_{ISL} > 2C^*_{up/dn}$, then the maximum flow on the paths are limited by $C^*_{up/dn}$. Hence, even if C^*_{ISL} is large, the excess capacity on the ISL is simply wasted. Since a lower $C'_{ISL} < C^*_{ISL}$ decreases network investment cost while keeping routing cost the same, we can always get a lower effective system cost by using C'_{ISL}. This contradicts with the optimality assumption of C^*_{ISL}. On the other hand, if $C^*_{ISL} < 2C^*_{up/dn}$, the ISL link becomes the bottleneck and a lower effective system cost can be obtained by using $C'_{up/dn} < C^*_{up/dn}$, this also contradicts with the optimality assumption of $C^*_{up/dn}$. Thus, to achieve optimal effect system cost, $C^*_{ISL} = 2C^*_{up/dn}$.

□

Formulation

$$\Im = \min\left\{m_1 C_{ISL} + m_2 + 2\left(m_3 C_{up/dn} + m_4\right) + \underset{\tilde{r}}{E}\left[D\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)\right]\right\}$$

$$D\left(C_{ISL}, C_{up/dn}, r\right) = \min_s\left(\kappa s\right)$$

s.t.

$$2y \le C_{ISL} \qquad \text{capacity constraint}$$

$$2y + s = r \qquad \text{flow constraint}$$

$$\frac{2y}{C_{ISL} - 2y} + \frac{y}{C_{up/dn} - y} + T_{ISL} + T_{up/dn} \le T_{\max} \qquad \text{delay constraint}$$

$$y \ge 0$$

$$s \ge 0$$

Solution

Solving the second stage problem, the largest flow that can be carried by the link is:

$$y = C_{up/dn}\left(\frac{1}{1 + 2/\left(T_{\max} - T_{ISL} - T_{up/dn}\right)}\right) = \tau_3 C_{up/dn} \qquad (A.1)$$

$$\text{where } \tau_3 = \frac{1}{1 + 2/\left(T_{\max} - T_{ISL} - T_{up/dn}\right)} \qquad (A.2)$$

Thus, the amount of overflow on the link given a realization of input traffic demand, r, is:

$$s = \begin{cases} r - \tau_3 C_{ISL} & \text{if } r \ge \tau_3 C_{ISL} \\ 0 & \text{otherwise} \end{cases} \qquad (A.3)$$

Then the second stage cost is:

$$D\left(C, r\right) = \begin{cases} \kappa\left(r - \tau_3 C_{ISL}\right) & \text{if } r \ge \tau_3 C_{ISL} \\ 0 & \text{otherwise} \end{cases} \qquad (A.4)$$

Having obtained an expression for the second stage problem as a function of C, the first stage problem may be solved.

$$\Im = \min_{C_{ISL}} \left\{ (m_1 + m_3) C_{ISL} + m_2 + 2m_4 + \int_{\tau_3 C_{ISL}}^{r_{max}} \kappa (r - \tau_3 C_{ISL}) p_{\tilde{r}}(r) dr \right\} \quad (A.5)$$

$$J(C_{ISL}) = (m_1 + m_3) C_{ISL} + m_2 + 2m_4 + \int_{\tau_3 C_{ISL}}^{r_{max}} \kappa (r - \tau_3 C_{ISL}) p_{\tilde{r}}(r) dr \quad (A.6)$$

$$\frac{dJ(C_{ISL})}{dC_{ISL}} = m_1 + m_3 - \kappa\tau_3 \int_{\tau_3 C_{ISL}}^{r_{max}} p_{\tilde{r}}(r) dr \quad (A.7)$$

$$\frac{d^2 J(C_{ISL})}{dC_{ISL}^{\ 2}} = \kappa\tau_3^{\ 2} p_{\tilde{r}}(\tau_3 C_{ISL}) \geq 0 \quad (A.8)$$

The optimal link dimension can be found by setting Equation A.7 to zero and solving for C_{ISL}. In particular, the following condition must be satisfied: $\int_{\tau_3 C_{ISL}}^{r_{max}} p_{\tilde{r}}(r) dr = \frac{m_1 + m_3}{\kappa\tau_3} \leq 1$, with strict inequality when $C_{ISL} > 0$. Hence, this topology is feasible only if

$$\frac{m_1 + m_3}{\tau_3} < \kappa \quad (A.9)$$

A.2 ST-GT and GT-ST Terminal Traffic Carried by Ground Only

The network analyzed here is depicted in Figure 5-7.

Formulation:

$$\Im = \min_{C_{up/dn}} \left\{ 2(m_3 C_{up/dn} + m_4) + E\left[D(C_{up/dn}, \tilde{r}) + G(C_{up/dn}, \tilde{r}) \right] \right\}$$

$$D(C_{up/dn}, r) + G(C_{up/dn}, r) = \min(2\zeta y + \kappa s)$$

s.t.

$$y \le C_{up/dn} \quad \text{capacity constraint}$$

$$2y + s = r \quad \text{flow constraint}$$

$$\frac{y}{C_{up/dn} - y} + T_{up/dn} + T_{gnd} \le T_{\max} \quad \text{delay constraint}$$

$$y \ge 0$$

$$s \ge 0$$

In this problem, there are two scenarios.

3. If $\zeta > \kappa$, then traffic should always be rejected and optimal capacity is just zero. This case is not very interesting as the satellite system should not be built.
4. If $\zeta < \kappa$, then traffic should always be routed if possible. In this case, the maximum flow is

$$y = C_{up/dn}\left(\frac{1}{1 + 1/\left(T_{\max} - T_{up/dn} - T_{gnd}\right)}\right) = \tau_4 C_{up/dn} \tag{A.10}$$

$$\text{where } \tau_4 = \frac{1}{1 + 1/\left(T_{\max} - T_{up/dn} - T_{gnd}\right)} \tag{A.11}$$

$$\text{Thus, } s = \begin{cases} r - 2\tau_4 C_{up/dn} & \text{if } r \ge 2\tau_4 C_{up/dn} \\ 0 & \text{otherwise} \end{cases} \tag{A.12}$$

Then the second stage cost is:

$$D\left(C_{up/dn}, r\right) + G\left(C_{up/dn}, r\right) \begin{cases} = \zeta r & \text{if } r < 2\tau_4 C_{up/dn} \\ = 2\zeta\tau_4 C_{up/dn} + \kappa\left(r - 2\tau_4 C_{up/dn}\right) & \text{otherwise} \end{cases} \tag{A.13}$$

$$\Im = \min_{C_{up/dn}} \left\{ 2\left(m_3 C_{up/dn} + m_4\right) + \int_0^{2\tau_4 C_{up/dn}} \zeta r p_{\tilde{r}}(r)\,dr + \int_{2\tau_4 C_{up/dn}}^{r_{\max}} \left(2\zeta\tau_4 C_{up/dn} + \kappa\left(r - 2\tau_4 C_{up/dn}\right)\right) p_{\tilde{r}}(r)\,dr \right\}$$

$$J\left(C_{up/dn}\right) = 2\left(m_3 C_{up/dn} + m_4\right) + \int_0^{2\tau_4 C_{up/dn}} \zeta r p_{\tilde{r}}(r)\,dr + \int_{2\tau_4 C_{up/dn}}^{r_{\max}} \left(2\zeta\tau_4 C_{up/dn} + \kappa\left(r - 2\tau_4 C_{up/dn}\right)\right) p_{\tilde{r}}(r)\,dr$$

$$\frac{dJ\left(C_{up/dn}\right)}{dC_{up/dn}} = 2m_3 - 2(\kappa - \zeta)\tau_4 \int_{2\tau_4 C_{up/dn}}^{r_{\max}} p_{\tilde{r}}(r)\,dr \tag{A.14}$$

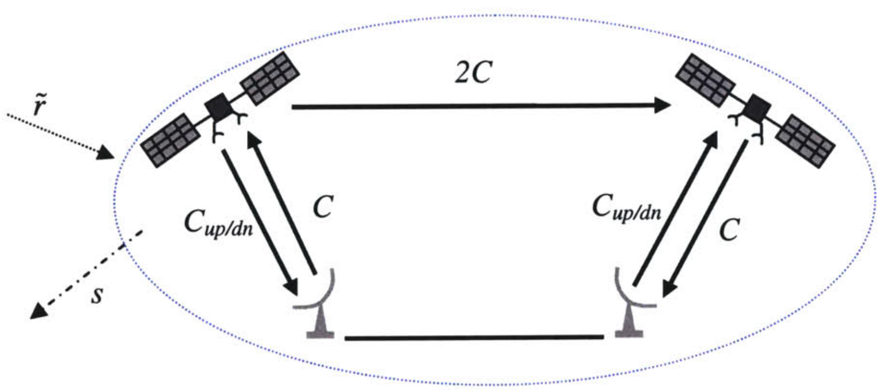
$$\frac{d^2 J\left(C_{up/dn}\right)}{dC_{up/dn}^{\;\;2}} = 4\left(\kappa - \zeta\right)\tau_4^{\;2} p_{\tilde{r}}\left(2\tau_4 C_{up/dn}\right) \geq 0 \tag{A.15}$$

The optimal gateway link dimension can be found by solving Equation A.18. In particular, the following condition must be satisfied: $\int_{2\tau_4 C^*_{up/dn}}^{r_{max}} p_{\tilde{r}}\left(r\right)dr = \frac{m_3}{\left(\kappa - \zeta\right)\tau_4} \leq 1$, with strict inequality when $C^*_{up/dn} > 0$. Hence, this topology is feasible only if

$$\frac{m_3}{\left(\kappa - \zeta\right)} < \tau_4,\ \zeta < \kappa \tag{A.16}$$

A.3 Hybrid Satellite-Terrestrial Link Dimensioning for ST-GT and GT-ST Traffic

The network analyzed here is depicted in Figure 5-8. As in the ISL only case, the optimal ISL link capacity equal to twice the optimal gateway link capacity, ie. $C^*_{ISL} = 2C^*_{up/dn}$. To facilitate computation of the feasible regions, we consider labeling the links as shown in Figure. Note that $C_{ISL} = 2C$ and $C_{up/dn} = C$.



Parameters:

C_{ISL} : Capacity of ISL and gateway link (design variable)

$C_{up/dn}$: Capacity of gateway link (design variable)

$\tilde{r}$: Random demand (with known probability density function, $p_{\tilde{r}}$)

r: A realization of the random demand,

$D\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)$: Cost function for routing on the ground link

$G\left(C_{ISL}, C_{up/dn}, \tilde{r}\right)$: Cost function for overflow traffic

m_1, m_2, m_3, m_4: Cost coefficients (assuming linear cost functions)

ζ: Cost per traffic unit routed on the ground link

κ: Cost per overflow traffic unit

y_1: Flow on ISL

y_2: Flow on second path

s: overflow traffic

Formulation:

$$\Im = \min\left\{m_1 2C + m_2 + 2\left(m_3 C + m_4\right) + 2\left(m_3 C_{up/dn} + m_4\right) + E\left[D\left(C, C_{up/dn}, \tilde{r}\right) + G\left(C, C_{up/dn}, \tilde{r}\right)\right]\right\}$$

$$D\left(C, C_{up/dn}, r\right) + G\left(C, C_{up/dn}, r\right) = \min\left(2\zeta y_2 + \kappa s\right)$$

s.t.

$$\begin{aligned}
2y_1 &\le 2C \\
y_2 &\le C_{up/dn} \\
2y_1 + 2y_2 + s &= r \\
\frac{2y_1}{2C - 2y_1} + \frac{y_1}{C - y_1} + T_{ISL} + T_{up/dn} &\le T_{\max} \\
\frac{y_2}{C_{up/dn} - y_2} + T_{up/dn} + T_{gnd} &\le T_{\max} \\
y_1, y_2 &\ge 0 \\
s &\ge 0
\end{aligned}$$

Solution:

As with ST-ST traffic analysis, there are two scenarios for the second stage cost:

1. If $\zeta > \kappa$, then traffic should not be routed to the terrestrial links and the optimal gateway link capacity, $C_{up/dn}$, is just zero. This case is identical to the ISL only case analyzed in Section 5.2.1.
2. If $\zeta < \kappa$, then traffic should always be routed if possible. In this case, ISL links will always be used first and then the terrestrial links. Excess input demand will be rejected.

The maximum flow on the ISL link is:

$$y_1 = C\left(\frac{1}{1+2/\left(T_{\max}-T_{up/dn}-T_{ISL}\right)}\right) = \tau_3 C \tag{A.17}$$

The maximum flow on the ground path is:

$$y_2 = C_{up/dn}\left(\frac{1}{1+1/\left(T_{\max}-T_{up/dn}-T_{gnd}\right)}\right) = \tau_4 C_{up/dn} \tag{A.18}$$

For the second stage problem,

$$s = \begin{cases} r-2\tau_3 C-2\tau_4 C_{up/dn} & \text{if } r \geq 2\tau_3 C+2\tau_4 C_{up/dn} \\ 0 & \text{otherwise} \end{cases} \tag{A.19}$$

$$\begin{aligned} & D\left(C, C_{up/dn}, r\right)+G\left(C, C_{up/dn}, r\right) \\ & = \begin{cases} 0 & \text{if } r \leq 2\tau_3 C \\ \zeta\left(r-2\tau_3 C\right) & \text{if } 2\tau_3 C < r \leq 2\tau_3 C+2\tau_4 C_{up/dn} \\ 2\zeta\tau_4 C_{up/dn}+\kappa\left(r-2\tau_3 C-2\tau_4 C_{up/dn}\right) & \text{otherwise} \end{cases} \end{aligned} \tag{A.20}$$

Substituting the second stage cost equation into the first stage equation, we obtain:

$$J\left(C_{ISL}\right) = 2\left(m_1+m_3\right)C+2m_3 C_{up/dn}+m_2+4m_4+\int_{2\tau_3 C}^{2\tau_3 C+2\tau_4 C_{up/dn}} \zeta\left(r-2\tau_3 C\right)P_{\tilde{r}}(r)\,dr$$

$$+\int_{2\tau_3 C+2\tau_4 C_{up/dn}}^{r_{\max}} \left(2\zeta\tau_4 C_{up/dn}+\kappa\left(r-2\tau_3 C-2\tau_4 C_{up/dn}\right)\right)P_{\tilde{r}}(r)\,dr$$

$$\frac{\partial J\left(C,C_{up/dn}\right)}{\partial C}=2\left(m_1+m_3\right)-2\tau_3\left(\zeta\int\limits_{2\tau_3C}^{2\tau_3C+2\tau_4C_{up/dn}}P_{\tilde{r}}\left(r\right)dr+\kappa\int\limits_{2\tau_3C+2\tau_4C_{up/dn}}^{r_{\max}}P_{\tilde{r}}\left(r\right)dr\right)\qquad\text{(A.21)}$$

$$\frac{\partial J\left(C,C_{up/dn}\right)}{\partial C_{up/dn}}=2m_3-2\left(\kappa-\zeta\right)\tau_4\int\limits_{2\tau_3C+2\tau_4C_{up/dn}}^{r_{\max}}P_{\tilde{r}}\left(r\right)dr\qquad\text{(A.22)}$$

For positive link capacities, the following conditions must be satisfied:

$$m_3\left(\frac{\kappa}{\tau_4\left(\kappa-\zeta\right)}-\frac{1}{\tau_3}\right)<\frac{m_1}{\tau_3}<m_3\left(\frac{1}{\tau_4}-\frac{1}{\tau_3}\right)+\zeta\qquad\text{(A.23)}$$

$$0<\frac{m_3}{\tau_4}<\left(\kappa-\zeta\right),\ \zeta<\kappa$$

If these conditions are not satisfied, then this topology should not be used.

Appendix B

Derivation for LEO/MEO Systems

<u>2-D *M* by *N* Torus</u>

A 2-D torus graph models a LEO or MEO satellite system in polar constellation without seam. The graph we analyze is depicted in Figure B-1.

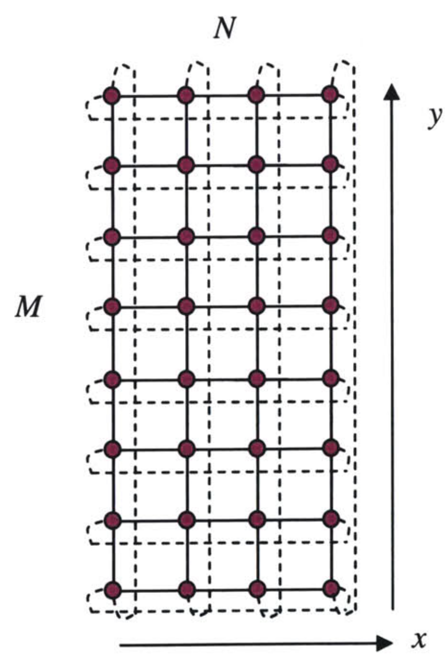


Figure B-1 2-D Torus Graph

In this graph, there are M rows and N columns of satellites. The total number of satellites in the constellation is MN and the total number of unidirectional ISLs is $4MN$. Under the uniform all to all traffic, and assuming high rejection cost, the maximum amount of input traffic demand is $r_{\max} = (MN)^2$.

The min-hop routing strategy is optimal in this case and the link capacities are summarized in Table 6-1. Here we show the detailed derivation used to obtain the results in Table 6-1.

1. *M* and *N* are odd integers >= 3.

If one node sends, the total number of minimum hops is:

$$H = \sum_{i=\left(\frac{-(N-1)}{2}\right)}^{\frac{N-1}{2}} \left[\sum_{j=\left(\frac{-(M-1)}{2}\right)}^{\frac{M-1}{2}} \left(|i| + |j|\right) \right] \tag{B.1}$$

$$H_x = M \sum_{i=\left(\frac{-(N-1)}{2}\right)}^{\frac{N-1}{2}} |i| = 2M \sum_{i=1}^{\frac{N-1}{2}} i = \frac{M(N-1)(N+1)}{4} \tag{B.2}$$

$$H_y = N \sum_{j=\left(\frac{-(M-1)}{2}\right)}^{\frac{M-1}{2}} |j| = 2N \sum_{j=1}^{\frac{M-1}{2}} j = \frac{N(M-1)(M+1)}{4} \tag{B.3}$$

The overall numbers of hops in each direction is:

$$H_{total,x} = \frac{M^2 N(N-1)(N+1)}{4} \tag{B.4}$$

$$H_{total,y} = \frac{MN^2(M-1)(M+1)}{4} \tag{B.5}$$

The unidirectional link capacities are:

$$C_x = \frac{M^2 N(N-1)(N+1)}{4(2MN)} = \frac{M(N^2-1)}{8} \tag{B.6}$$

$$C_y = \frac{MN^2(M-1)(M+1)}{4(2MN)} = \frac{N(M^2-1)}{8} \tag{B.7}$$

2. *M* and *N* are even integers >= 2.

If one node sends, the total number of minimum hops is:

$$H = \sum_{i=\left(\frac{-N}{2}\right)}^{\frac{N}{2}-1} \left[\sum_{j=\left(\frac{-M}{2}\right)}^{\frac{M}{2}-1} \left(|i|+|j|\right) \right] \tag{B.8}$$

$$H_x = M \sum_{i=\left(\frac{-N}{2}\right)}^{\frac{N}{2}-1} |i| = M\left[2\left(\sum_{i=1}^{\frac{N}{2}-1} i\right) + \frac{N}{2}\right] = \frac{MN^2}{4} \tag{B.9}$$

$$H_y = N \sum_{j=\left(\frac{-M}{2}\right)}^{\frac{M}{2}-1} |j| = N\left[2\left(\sum_{j=1}^{\frac{M}{2}-1} j\right) + \frac{M}{2}\right] = \frac{NM^2}{4} \tag{B.10}$$

Overall numbers of hops in each direction:

$$H_{total,x} = \frac{M^2 N^3}{4} \tag{B.11}$$

$$H_{total,y} = \frac{M^3 N^2}{4} \tag{B.12}$$

Per unidirectional link capacity:

$$C_x = \frac{M^2 N^3}{4(2MN)} = \frac{MN^2}{8} \tag{B.13}$$

$$C_y = \frac{M^3N^2}{4(2MN)} = \frac{M^2N}{8} \tag{B.14}$$

3. *M* is odd integer >=3 and *N* is even integer >= 2.

If one node sends, the total number of minimum hops is:

$$H = \sum_{i=\left(\frac{-N}{2}\right)}^{\frac{N}{2}-1} \left[\sum_{j=\left(\frac{-(M-1)}{2}\right)}^{\frac{M-1}{2}} \left(|i|+|j|\right) \right] \tag{B.15}$$

$$H_x = M \sum_{i=\left(\frac{-N}{2}\right)}^{\frac{N}{2}-1} |i| = M \left[2\left(\sum_{i=1}^{\frac{N}{2}-1} i \right) + \frac{N}{2} \right] = \frac{MN^2}{4} \tag{B.16}$$

$$H_y = N \sum_{j=\left(\frac{-(M-1)}{2}\right)}^{\frac{M-1}{2}} |j| = 2N \sum_{j=1}^{\frac{M-1}{2}} j = \frac{N(M-1)(M+1)}{4} \tag{B.17}$$

4. *M* is even integer >=2 and *N* is odd integer >= 3.

If one node sends, the total number of minimum hops is:

$$H = \sum_{i=\left(\frac{-(N-1)}{2}\right)}^{\frac{N-1}{2}} \left[\sum_{j=\left(\frac{-M}{2}\right)}^{\frac{M}{2}-1} \left(|i|+|j|\right) \right] \tag{B.18}$$

$$H_x = M \sum_{i=\left(\frac{-(N-1)}{2}\right)}^{\frac{N-1}{2}} |i| = 2M \sum_{i=1}^{\frac{N-1}{2}} i = \frac{M(N-1)(N+1)}{4} \tag{B.19}$$

$$H_y = N \sum_{j=\left(\frac{-M}{2}\right)}^{\frac{M}{2}-1} |j| = N \left[2\left(\sum_{j=1}^{\frac{M}{2}-1} j \right) + \frac{M}{2} \right] = \frac{NM^2}{4} \tag{B.20}$$

Since each node sends $\frac{r_{\max}}{(MN)^2}$ to every node including itself, this value can be multiplied to all the link capacity values to obtain the entries in Table 6-1.

2-D *M* by *N* mesh

A *M* by *N* mesh graph can be used to model half of a LEO or MEO satellite constellation. In this graph, there are *M* rows and *N* columns of satellites. The total number of satellites in the constellation is *MN* and the total number of unidirectional ISLs is $2(M(N-1)+N(M-1))=2(2MN-M-N)$. Under uniform all to all traffic, and assuming high rejection cost, the maximum amount of input traffic demand is $r_{\max}=(MN)^2$.

We can find the minimum hop in the graph using cut set arguments. The overall number of minimum hops is:

$$H=2\left[\sum_{a=1}^{N-1}\left(aM\left(MN-aM\right)\right)+\sum_{b=1}^{M-1}\left(bN\left(MN-bN\right)\right)\right] \tag{B.21}$$

$$\begin{aligned} H_{total,x} &= 2\sum_{a=1}^{N-1}\left(aM\left(MN-aM\right)\right) \\ &= 2\left(\frac{M^2N^2(N-1)}{2}-\frac{M^2N(N-1)(2N-1)}{6}\right)=\frac{M^2N\left(N^2-1\right)}{3} \end{aligned} \tag{B.22}$$

$$\begin{aligned} H_{total,y} &= 2\sum_{b=1}^{M-1}\left(bN\left(MN-bN\right)\right) \\ &= 2\left(\frac{M^2N^2(M-1)}{2}-\frac{N^2M(M-1)(2M-1)}{6}\right)=\frac{N^2M\left(M^2-1\right)}{3} \end{aligned} \tag{B.23}$$

We can derive an average per unidirectional link capacity. This is a lower bound for the worst case link capacity. This bound is achieved with equality when *M=N=3*.

$$C_x = \frac{H_{total,x}}{2M(N-1)} = \frac{MN(N+1)}{6} \tag{B.24}$$

$$C_y = \frac{H_{total,y}}{2N(M-1)} = \frac{MN(M+1)}{6} \tag{B.25}$$

In summary, if we fix the total amount of traffic sent to be $r_{\max}$, then each node sends $\frac{r_{\max}}{(MN)^2}$ amount of traffic to every node including itself. The average per unidirectional link capacity is:

$$C_x = \frac{r_{\max}(N+1)}{6MN} \approx \frac{r_{\max}}{6M} \quad \text{compare with } \frac{r_{\max}}{8M}\frac{(N-1)(N+1)}{N^2} \approx \frac{r_{\max}}{8M} \text{ for a torus graph}$$

$$C_y = \frac{r_{\max}(M+1)}{6MN} \approx \frac{r_{\max}}{6N} \quad \text{compare with } \frac{r_{\max}}{8N}\frac{(M-1)(M+1)}{M^2} \approx \frac{r_{\max}}{8N} \text{ for a torus graph}$$

The actual link capacities can be obtained simply by making horizontal and vertical cuts in the graph. We can compute the link capacity required for the link with the highest load.

In the x direction:

$$\text{If N = odd, } C_x = \frac{1}{M}\frac{r_{\max}}{(MN)^2}\left(M\left(\frac{(N+1)}{2}\right)\right)\left(M\left(\frac{(N-1)}{2}\right)\right) = \left(\frac{r_{\max}}{4M}\right)\left(1-\frac{1}{N^2}\right) \tag{B.26}$$

$$\text{If N = even, } C_x = \frac{1}{M}\frac{r_{\max}}{(MN)^2}\left(\frac{MN}{2}\right)^2 = \frac{r_{\max}}{4M} \tag{B.27}$$

In the y direction:

$$\text{If M = odd, } C_y = \frac{1}{N}\frac{r_{\max}}{(MN)^2}\left(N\left(\frac{(M+1)}{2}\right)\right)\left(N\left(\frac{(M-1)}{2}\right)\right) = \left(\frac{r_{\max}}{4N}\right)\left(1-\frac{1}{M^2}\right) \tag{B.28}$$

$$\text{If M = even, } C_y = \frac{1}{N}\frac{r_{\max}}{(MN)^2}\left(\frac{MN}{2}\right)^2 = \frac{r_{\max}}{4N} \tag{B.29}$$

We can also compute the link capacity required for the link with the lowest load.

In the x direction: $$C_x = \frac{1}{M}\frac{r_{\max}}{(MN)^2}(M)(M(N-1)) = \frac{r_{\max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right) \quad \text{(B.30)}$$

In the y direction: $$C_y = \frac{1}{N}\frac{r_{\max}}{(MN)^2}(N)(N(M-1)) = \frac{r_{\max}}{N}\left(\frac{1}{M}-\frac{1}{M^2}\right) \quad \text{(B.31)}$$

The ratio between the lowest link capacity and the highest link capacity is:

In the x direction:

If N = odd, $$C_x = \frac{\frac{r_{\max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right)}{\left(\frac{r_{\max}}{4M}\right)\left(1-\frac{1}{N^2}\right)} = \frac{4}{N+1} \quad \text{(B.32)}$$

If N = even, $$C_x = \frac{\frac{r_{\max}}{M}\left(\frac{1}{N}-\frac{1}{N^2}\right)}{\frac{r_{\max}}{4M}} = \frac{4(N-1)}{N^2} \quad \text{(B.33)}$$

In the y direction:

If M = odd, $$C_y = \frac{\frac{r_{\max}}{N}\left(\frac{1}{M}-\frac{1}{M^2}\right)}{\left(\frac{r_{\max}}{4N}\right)\left(1-\frac{1}{M^2}\right)} = \frac{4}{M+1} \quad \text{(B.34)}$$

If M = even, $$C_y = \frac{\frac{r_{\max}}{N}\left(\frac{1}{M}-\frac{1}{M^2}\right)}{\frac{r_{\max}}{4N}} = \frac{4(M-1)}{M^2} \quad \text{(B.35)}$$