ESPRIT PROJECT EP28425

BISANTE

Broadband Integrated Satellite Network Traffic Evaluations

Deliverable 2.3

Report on LEO Satellite Network Characteristics

Status / Version : **DELIVERABLE / FINAL**
Date : **August 4, 1999**
Distribution : **Public**
Code : **BISANTE/DEL23**
Author (s) : **H. Cruickshank (Ed), Z. Sun, L. Wood and T. Örs (University of Surrey)**
R. Dhaou and M. Becker (INT),
C. Xenakis (Solinet)

Abstract:
The goal of this deliverable is to provide the LEO satellite network characteristics which have an impact on the upper layers of applications. The LEO satellite network interface unit as well as uplink/downlink characteristics are described. Then the considerations which have to be taken into account when designing a LEO satellite constellation are explained. The emphasis is on topologies using a primarily ground- and space- based networks. Visibility, diversity and handover issues are also discussed. The impact of LEO satellite on the performance of TCP layer and applications is presented. LEO satellite networking scenarios using on-board switching and repeater satellites are defined
Keyword list: BISANTE, TCP/IP, Network, LEO Satellite, Traffic Engineering

The BISANTE Consortium consists of :

Thomson-CSF Communications	Partner	France
Netway	Partner	Austria
Solinet	Partner	Germany
University of Vienna	Associated Partner	Austria
University of Surrey	Associated Partner	United Kingdom
Institut National des Télécommunications (INT)	Associated Partner	France

TABLE OF CONTENTS

1. INTRODUCTION..1

2. LEO SATELLITE NETWORK INTERFACE UNIT CHARACTERISTICS........1

2.1. NETWORK LAYER..2
2.1.1. *Protocol Mapping/Tunnelling*..2
2.1.2. *Multiplexing/Demultiplexing*..2
2.1.3. *Shaping*..3
2.2. MEDIA ACCESS CONTROL LAYERS..3
2.3. CHANNEL CODING AND PHYSICAL LAYER..6
2.3.1. *Modulation techniques*..6
2.3.1.1. Impact of burst errors on the ATM layer..8
2.3.1.2. Impact of burst errors on AAL protocols..9
2.3.1.3. Solutions for Improving Error Characteristics..9

3. LEO UPLINK/DOWNLINK CHARACTERISTICS..10

3.1. UPLINK AND DOWNLINK PROPAGATION DELAY..10
3.2. UPLINK AND DOWNLINK PROPAGATION IMPAIRMENTS..12

4. LEO CONSTELLATION DESIGN..14

4.1. CONSTELLATION GEOMETRY AND PLACEMENT OF GROUND TERMINALS AND GATEWAYS TO THE TERRESTRIAL NETWORKS...14
4.1.1. *Orbit Parameters*..14
4.1.1.1. Semi-Major Axis..14
4.1.1.2. Eccentricity..14
4.1.1.3. Inclination..14
4.1.1.4. Longitude of the Ascending Node (Ω)..15
4.1.1.5. Argument of Perigee..16
4.1.2. *Orbits*..16
4.1.2.1. Circular orbits..16
4.1.2.2. Elliptical orbits..18
4.1.3. *An Overview of Constellation Geometry*..19
4.1.3.1. Polar and Near-Polar Star Constellations..19
4.1.3.2. Delta Constellations..21
4.1.3.3. Notation..22
4.1.4. *Network topology*..23
4.1.4.1. Primarily ground-based networks..23
4.1.4.2. Primarily space-based networks..23
4.1.4.3. The case for intersatellite links..23
4.1.4.4. Generalising to the rosette..26
4.1.5. *Design choices*..30
4.2. AVAILABILITY, VISIBILITY AND DIVERSITY..31
4.2.1. *Earth-to-space diversity*..31
4.2.2. *In-orbit network diversity*..31
4.2.3. *Diversity in Skybridge*..31
4.2.4. *Visibility Statistics for SkyBridge and Teledesic*..32
4.3. HANDOVER BETWEEN SATELLITES AND BETWEEN GATEWAYS...33
4.3.1. *Earth-Fixed Cells vs Satellite-Fixed Cells*..35
4.3.2. *Frequency sharing between LEO and GEO satellites*..37
4.3.2.1. SkyBridge Interference with GSO systems..37
4.3.2.2. SkyBridge Interference with terrestrial systems..38

4.3.3. Analysis of SkyBridge Frequency sharing with GSO Systems 38
4.3.4. Channel Allocation Techniques..39
4.3.4.1. Fixed Channel Allocation.. 39
4.3.4.2. Dynamic Channel Allocation 39
4.4. ROUTING METHODS WITHIN THE CONSTELLATION BETWEEN SATELLITES AND GATEWAYS.40
5. LEO SATELLITE CHARACTERISTICS...42
5.1. TRANSPARENT ‘BENT-PIPE’ SATELLITE CHARACTERISTICS 42
5.2. ON-BOARD-SWITCHING SATELLITE CHARACTERISTICS 42
5.2.1. Intersatellite link MAC and LLC configuration 44
5.2.2. Intersatellite Link Characteristics..44
5.2.3. Routing Characteristics..44
5.2.4. Intersatellite Link Delay...44
6. SATELLITE NETWORK SCENARIOS 45
6.1. LEO SATELLITE CONSTELLATION USING REPEATER SATELLITES.................................45
6.2. LEO SATELLITE CONSTELLATION USING OBS SATELLITES 46
7. PERFORMANCE ISSUES OF TCP/IP OVER LEO SATELLITES......................50
7.1. TCP THROUGHPUT ISSUES 50
7.1.1. Impact of satellite channel characteristics on TCP 52
7.1.2. Congestion Control 53
7.1.2.1. Large window sizes 53
7.1.2.2. Selective Acknowledgments.. 54
7.2. IP THROUGHPUT ISSUES..54
7.2.1. IP Time To Live 55
7.2.2. IP Fragmentation 56
7.3. PERFORMANCE ANALYSIS OF TCP/IP APPLICATIONS 57
7.3.1. First set...57
7.3.1.1. Bulk Transfer Protocols 57
7.3.1.2. Semi-interactive Protocols 57
7.3.1.3. Interactive Protocols.. 58
7.3.2. Second set 58
7.3.2.1. FTP Performance... 58
7.3.2.2. http Performance 58
7.3.2.3. Telnet Performance 59
7.3.3. Result..59
7.4. METHODS FOR REDUCING DATA FLOW 59
7.4.1. Satellite distribution of cache content 60
7.5. NEW AND EMERGING PROTOCOLS 61
8. CONCLUSIONS 62
9. REFERENCES..63
10. APPENDIX 1: TERMINOLOGY..67

TABLE OF FIGURES

Figure 2-1 Protocol stack for the SkyBridge satellite ATM network 2
Figure 2-2 Multiple Access for the air interface 3
Figure 4-1 Semi-major axis of orbit 14
Figure 4-2 Inclination of Orbit 15
Figure 4-3 RAAN of Orbit 15
Figure 4-4 Orbital altitudes for satellite constellations 18
Figure 4-5 Streets of coverage 19
Figure 4-6 Polar view of Walker star pattern 20
Figure 4-7 Map projection of constellation (stylised) 20
Figure 4-8 Ground Tracks of approximate *Teledesic* constellation (288-satellite design) 21
Figure 4-9 Simplest Walker delta constellation 22
Figure 4-10 Example Rosette Ground Track (Skybridge 80-satellite approximation) 22
Figure 4-11 Flattening the star network 25
Figure 4-12 Untwisting the network 25
Figure 4-13 How seam distance varies apparent network distance 26
Figure 4-14 How a toroidal mesh maps into the earth 27
Figure 4-15 Simulation of sixway connectivity for 8/18/2 constellation 28
Figure 4-16 Cross-section through equator of rosette with ISLs 28
Figure 4-17 Example network Topology 29
Figure 4-18 ISL routing approach a la Teledesic 30
Figure 4-19 Relay ‘bent-pipe’ satellite approach a la Skybridge 30
Figure 4-20 Part of Teledesic network mesh showing dual network diversity 31
Figure 4-21 Minimum and Mean Elevation Angle For SkyBridge and Teledesic 32
Figure 4-22 Satellite Visibility for SkyBridge and Teledesic 33
Figure 4-23 Relation between elevation angle and altitude 34
Figure 4-24 Maximum Inter-Handover Delay 35
Figure 4-25 Handover Procedure, difference in propagation delays 35
Figure 4-26 Problems that accurs on the access point/portable way 36
Figure 4-27 Problems that occur on the portable/access point way 36
Figure 4-28 System allocations of the proposed SkyBridge frequency bands 39
Figure 5-1 Transparent Bent-Pipe satellite Architecture 42
Figure 5-2 On-Board Switching Satellite Architecture 43
Figure 6-1 LEO Satellite Network using Repeater Satellites 45
Figure 6-2 LEO Satellite Network using On-Board Switching Satellites 46
Figure 7-1. TCP header 50
Figure 7-2 TCP Slow start algorithm 51
Figure 7-3. IP packet header 55
Figure 7-4: Simple satellite configuration with caches 61

1. INTRODUCTION

This document discusses low-earth-orbiting (LEO) satellite characteristics and the various networking options available for a LEO satellite network based around a constellation consisting of multiple satellites. The LEO satellite network will be presented using layered block diagrams to show the separate networking aspects that can be rendered as separate modules.

Section 2 describes the Satellite network Interface Unit (SIU) which converts/maps the terrestrial network protocol to the satellite protocol and vice versa. The SIU multiplexes/demultiplexes traffic and also does shaping. The multiple access functionality as well as channel coding/decoding is also implemented in the SIU.

The satellite uplink and downlink characteristics such as delay and propagation impairments are described in Section 3.

The design of a satellite constellation is discussed in Section 4. First, the orbital parameters which will also be used for simulation, are defined. Then the properties of circular and elliptical orbits are illustrated. This is followed by an overview to constellation geometry which shows the difference between near-polar and delta constellations. Primarily space- and ground- based network and the case for Inter-Satellite-Links (ISLs) are discussed. Design considerations such as diversity and multiple satellite visibility are also mentioned in Section 4, where elevation and satellite visibility statistics for SkyBridge and Teledesic are provided. Information on handover according to Earth Fixed Cells (EFC) and Satellite Fixed Cells (SFC) constellations is provided. Frequency sharing between LEO and GEO satellites as implemented by SkyBridge is explained. This section concludes with routing for satellite constellations.

Section 5 describes the repeater and on-board switching satellite characteristics. The ISL link and routing characteristics for the OBS satellite scenario are also discussed. This is followed by the definition of satellite network scenarios using repeater and OBS satellites in Section 6. The constellation characteristics for Teledesic and SkyBridge are provided.

Section 7 presents the upper protocol layers such as TCP/IP protocol suite with the focus on aspects that directly affect TCP/IP throughput over LEO satellites.

2. LEO SATELLITE NETWORK INTERFACE UNIT CHARACTERISTICS

The characteristics of the satellite network will differ considerably from terrestrial networks that interwork with it, most noticeably at the layers closest to the physical layer. The transport layer at the end host, specifying the end-to-end or point-to-multipoint communication paradigm in use, is of considerable importance in defining overall communication performance.

Simulation of specific implementations of network and transport layers may be required in order to support the satellite environment correctly. Interaction between layers (e.g. TCP congestion control interacting with physical errors that appear bursty due to coding scheme choices) will result in complex relationships between layers and modules. Network termination module characteristics to consider for the involved terrestrial and satellite network are described below.

2.1. NETWORK LAYER

LEO satellite constellations are based on 'ATM-like' network protocols or proprietary protocols. For instance, the Teledesic network [TELE99] uses fast packet switching technology based on their own proprietary protocol. All communication is treated identically within the network as multiplexed streams of short fixed-length packets. Each packet contains a header that can be expected to include address and sequence information, an error-control section used to verify the integrity of the header and possibly payload, and a payload section that carries the digitally-encoded voice or data. Conversion to and from the packet format takes place in the terminals.

The fast packet switched network combines the advantages of a circuit-switched network (low delay 'digital pipes'), and a packet-switched network (efficient handling of multi-rate and bursty data). Fast packet switching technology is ideally suited for the dynamic nature of a LEO network. Each satellite in the constellation is a node in the fast packet-switched network, and has intersatellite communication links with adjacent satellites.

Skybridge [SKYB99], is based on ATM protocol and does not use inter-satellite links but instead relies on switching in the ground stations. The protocol stack of SkyBridge is shown in Figure 2-1.

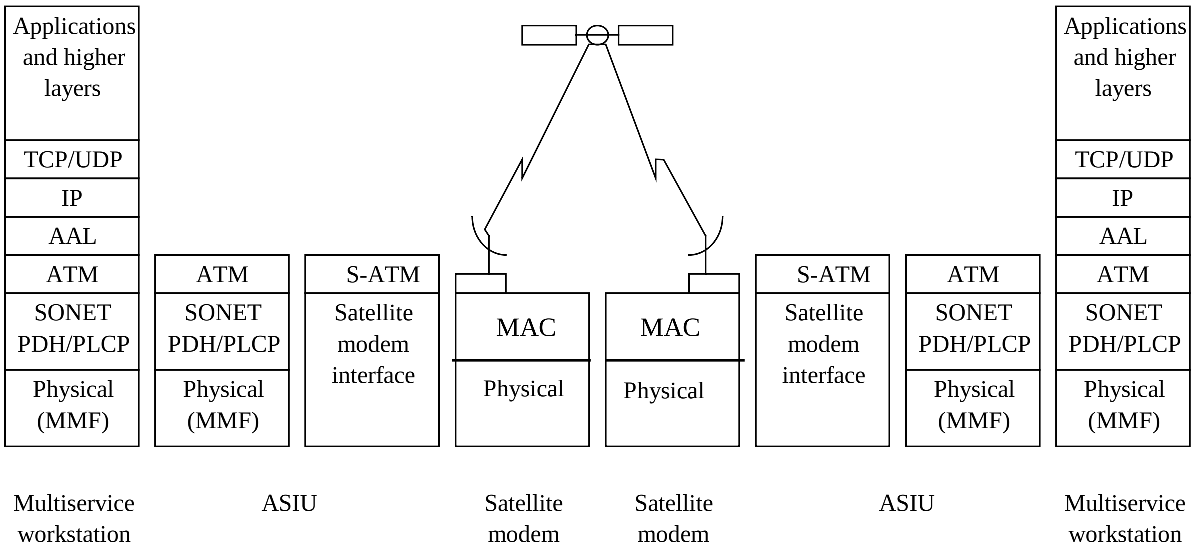


FIGURE 2-1 PROTOCOL STACK FOR THE SKYBRIDGE SATELLITE ATM NETWORK

2.1.1. Protocol Mapping/Tunnelling

Whether the satellite network uses ATM-like packets or proprietary packets, mapping/tunnelling of the terrestrial network protocol to the satellite protocol is necessary. The satellite interface unit may therefore need to carry out some buffering.

2.1.2. Multiplexing/Demultiplexing

The gateway (SIU) is also responsible for multiplexing/demultiplexing of streams from the uplink and downlink Again buffering may be required for this purpose.

2.1.3. Shaping

Shaping can be used in order to satisfy the delay variation requirements of various applications. A shaping buffer is used to space packets in order to give them the same characteristics as at the entrance to the satellite network.

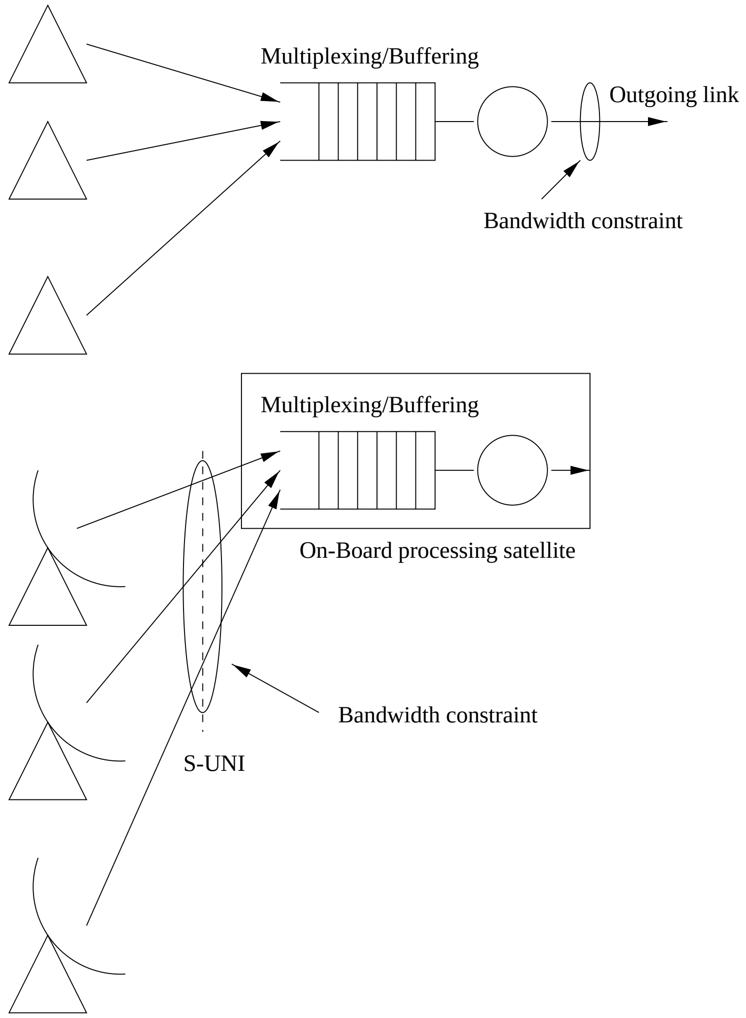


FIGURE 2-2 MULTIPLE ACCESS FOR THE AIR INTERFACE

2.2. MEDIA ACCESS CONTROL LAYERS

Considering that satellite communications uses multiple access on a shared medium, a MAC layer, which is left undefined by the IP and ATM protocols, is needed. The MAC protocol plays a central role as a means of accessing the Radio Physical Layer (RPL) from the ATM layer. The access scheme refers to the physical layer multiplexing technique to a share a common channel among multiple users of possibly multi-services The problem of statistical multiplexing at the satellite air interface is slightly different from that in the fixed network as illustrated in Figure 2-2. In the fixed network, the problem is associated with control of bandwidth on an outgoing link from some multiplexing point after buffering has occurred. It is generally assumed that the access links from the source are dimensioned in such a way that they do not impose any constraints on the traffic (e.g. sources can transmit at their peak bit rate). In the air interface the constraint is on the bandwidth available in total to all sources before the buffering/multiplexing point. MAC allocation methods can have considerable affect on delay and traffic shaping and policing policies.

This section examine the techniques which are available for allowing a number of users to communicate via a common satellite transponder. This is broadly termed ‘multiple access’, where the transponder’s available power and bandwidth are shared between a number of different channels and earth stations, which may themselves have quite different transmit powers and signal characteristics. Efficient use of these resources, while meeting the need of the user’s traffic demands, is important.

The choice of the multiple access scheme has a great impact on the performance of the satellite network for a scenario where a large number of terminals are interconnected. The primary goal in the assignment process is to satisfy the users' QoS and to maximise the utilisation of the uplink. In satellite networks, once a cell from a connection has been chosen to be assigned to an MF-TDMA frame, it is the responsibility of lower layers to choose a specific slot on the MF-TDMA frame for the cell. All these functions constitute the MAC implementation.

In this section, the term "connection" will be used to mean either a stream of satellite-specific signalling information that is exchanged between the satellite and a given terminal, or a stream of signalling cells/packets, or the data cells/packets of a particular connection.

MAC layer access schemes can be typically categorised into four classes: Fixed Access, Random Access, Demand Assignment Multiple Access (DAMA) and Adaptive Access. The first three techniques have evolved to meet the needs of constant high traffic with long durations, sporadic traffic with short to medium durations, and sporadic traffic with long durations, respectively [BOHM93]. Finally, adaptive access is used to meet the needs of multiple media which consists of traffic with all of the above characteristics.

With **random access**, connections from different terminals may broadcast cells simultaneously resulting in "collisions" (corrupted data), in which case retransmission is necessary. Random access schemes, such as Aloha, obtain reasonable throughput only at low loads and they offer little in terms of performance guarantees. In the satellite environment with long wireless delays, detecting collisions adds considerable overhead, so variants such as slotted aloha are much more likely to be implemented. Hence, the use of random access is not suitable to provide QoS. It can however be used for best-effort services.

With **fixed assignment**, a terminal's connection is permanently assigned a constant number of slots per frame (or some multiple number of frames) for the lifetime of the terminal. This means that when the connection is idle, the slots are not utilised (i.e. that they are wasted). The terminal-to-terminal delay when using this access scheme is the propagation delay plus the processing and queuing delays onboard.

Demand assignment (DA) allocates slots on an as-needed basis. There are two basic types of DA: fixed rate and variable-rate DA.

With the **fixed-rate** variety, a connection is assigned a fixed number of slots per frame (or some multiple number of frames) for the duration of the connection. Like the fixed assignment scheme, if the source is idle, the slots will be wasted. So, the fixed-rate DA and fixed assignment schemes are the same, except that one is for the duration of the connection while the other is for the lifetime of the terminal.

With the **variable-rate DA** scheme, slots are only assigned when it is known that there are cells awaiting service at the connection's terminal queue. This works as follows: when a cell arrives at the terminal queue, signalling messages are sent to the satellite notifying it of the arrival. When the satellite receives this information, it dynamically assigns slot(s) to the connection. Variable-rate DA is a guaranteed assignment scheme in that slots are assigned based on previously allocated resources (bandwidth), which is available for use whenever it is needed. It avoids collisions and efficiently uses the uplink capacity because the satellite is aware of the needs of the source and it responds to the need by assigning slots on a frame-by-frame basis. If the connection does not need a slot which has been allocated to it during connection establishment, the satellite may assign the slot to others. The drawback of this scheme is the propagation delay starts from the time signalling information is sent until the the satellite response (i.e., assignment) is received at the ground terminal.

With variable-rate DA, some slots would be needed for transmitting notifications to the satellite. One way to do this is to overlay this signalling information on the regular signalling and synchronisation slots in an "out-of-band" signalling scheme. A second way, adapting to the dynamic nature of the traffic, uses the in-band signalling opportunities in data slot headers.

Adaptive Assignment mechanisms utilise various MAC protocol variants depending upon the instantaneous traffic conditions. All protocols have their strength and weaknesses; no access scheme outperforms all others under every condition. To accommodate a combination of traffic types, channels can be partitioned into several sections, each operating under its own protocol. Adaptive protocols attempt to provide good performance over a large range of conditions; the access scheme itself changes, adapting to network traffic load fluctuations, yielding an access procedure appropriate for the actual traffic type mixture. Thus to meet the design objectives in a multi-service environment an Adaptive Access mechanism seems to be the best choice.

In order to optimise the performance of MAC schemes methods such as free assignment have been proposed [NGOC95, NGOC96]. Free assignment is concerned with the remaining slots in a frame which have not been assigned by the fixed or demand schemes. These remaining slots are the spare uplink capacity that the network can freely assign to connections in order to increase overall throughput, to relieve congestion at the ground terminal queues, or to reduce the terminal-to-terminal delay. Free assignment could use signalling information to determine the terminal queue's cell occupancy, like variable-rate DA.

An adaptive techniques called Random-Reservation Adaptive Assignment (RRAA) which combines schemes from the first three classes to meet the needs of multiple media which consists of all the above mentioned traffic profiles is studied in [ORS98, WERN98]. Here rt-VBR can be assigned peak rate only in order to avoid unacceptable delay. In a second step, an additional optimised access module for bandwidth-efficient support of rt-VBR traffic is considered in detail. This scheme is called Advanced Packet Reservation Multiple Access (A-PRMA). It realises variable-rate demand assignment by random access with implicit reservation, yielding significantly lower delays than a contention-free variable-rate DAMA scheme at the slight expense of cell loss.

RRAA is capable of working in both LEO and GEO environments, whereas due to the poor performance facing large round-trip delay (as in GEO) A-PRMA is considered for the LEO scenario only. In the latter, RRAA and A-PRMA could be combined in order to form an even more sophisticated hybrid protocol for highly complex traffic environments.

[ORS99b] discusses the main multiple access schemes:

- Frequency Division Multiple Access (FDMA),
- Time Division Multiple Access (TDMA) and,
- Code Division Multiple Access (CDMA).

A brief comparison of FDMA, TDMA and CDMA is provided in the following table. The properties in which we are interested concern the use and reuse of frequency spectrum and the ultimate capacity that the multiple access technique can deliver.

Characteristic	**FDMA**	**TDMA**	**CDMA**
Bandwidth utilisation	Single Channel Per Carrier (SCPC)	Multiple Channels Per Carrier - partial allocation	SCPC, partial or full allocation
Interference rejection	Limited	Limited with frequency hopping	Can suppress interference, up to noise limit
Intermodulation effects	Most sensitive (most backoff required)	Less sensitive (less backoff required)	Least sensitive (least backoff required)
Doppler frequency shift	Bandwidth limiting	Burst time limiting	Removed by receiver
Spectrum flexibility	Uses least bandwidth per carrier	Moderate bandwidth use per carrier	Largest demand for contiguous segment
Capacity	Basic capacity available	Can provide capacity improvement through hopping	Capacity indeterminate due to loading unknowns

TABLE 2-1 COMPARISON OF MAIN MULTIPLE ACCESS METHOD PROPERTIES

In LEO constellations, where satellites are more dynamic than in GEO ones, the use of synchronised access methods is more complex, especially on the uplink.

MAC options include TDMA, FDMA, and CDMA and can use contention-based, reservation-based, or fixed media access control. TDMA/FDMA combinations can be used to realize the advantages of both systems. Demand Assignment Multiple Access (DAMA) can be used with any of the MAC options. If on-board processing is not performed, DA must be done by the Network Control Centre (NCC). On-board DAMA decreases the response time of the media access policy by half because link access requests need not travel to the NCC on the ground.

2.3. CHANNEL CODING AND PHYSICAL LAYER

2.3.1. *Modulation techniques*

The different modulation methods for satellite channels were discussed in [ORS99b]. The relationship between C/N and the bit error rate of the channel is one measure of performance for a digital link. This is computed from the Carrier-to-Noise Density ratio., C/N_0 ratio, for a particular modulation scheme by:

$$\frac{E_b}{N_0} = \begin{cases} C/N_0 - 10\log_{10}(DataRate) \\ C/N - 10\log_{10}(DataRate/Bandwidth) \end{cases} \tag{2.5}$$

The data rate over bandwidth ratio, R/B, is called the spectrum or bandwidth efficiency of the modulation. Because of the limited bandwidth, the ideal is to have this value as large as possible. To improve the bandwidth efficiency, modulation has become more complex. The simplest binary modulation scheme, BPSK, is composed of a binary alphabet. The higher the modulation order the larger the alphabet order. QPSK, for example, uses a two-bit alphabet.

The error relationships allows to use error functions to compute the symbol error rates using the equations given below. Depending on the alphabet level, the symbol error rate is different from the bit error rate. To compare modulation methods, convert the symbol error rates computed to bit error rates. For a simple gray-coded M-level modulation, the bit error rate is related to the symbol error rate by :

$$p_e = \frac{P_s}{\log_2(M)} \tag{2.6}$$

It is possible to choose from several modulation methods and some of these methods have different names. [ORS99b] shows the common alternate names for each modulation.

Satellite Channel Error Characteristics

For a satellite channel, there are two general performance requirements, first low cost of ground station and, transmission power while maintaining required signal-to-noise ratio at the receiver. The other requirement is maximising the communication throughput by minimising data retransmission. This is especially important if Go-back-N sliding window based flow-control is used at the transport layer. When the propagation delay is large, data that is transmitted after the missing segment and before the retransmission request reaches the source, may be discarded. This is addressed e.g. in TCP with selective acknowledgements (SACK) in RFC2018

Inherently, satellite channels produce random single-bit errors in the data being transmitted. The **Bit Error Rate** (BER) depends on the Signal-to-Noise ratio at the receiver, or link margin. Thus for an acceptable level of error rate, a certain minimum signal-to-noise ratio must be ensured at the receiver and hence maintained at the transmitter.

Forward Error Correction (FEC) techniques provide a solution that satisfies both these requirements. These techniques introduce some redundancy in the transmitted data. When the receiver gets the corrupted data, it uses this redundancy to decide if received data is corrupted and find out what must have been the original data. FEC codes can broadly be classified as block codes and tree codes. Block codes are 'memory-less' codes that map ‘k’ input binary signals to ‘n’ output binary signals, where ‘n’ > ‘k’ for redundancy. Tree codes, on the other hand, use 'memory' by remembering ‘v’ input signals immediately preceding the target block of ‘k’ input signals. These ‘v’ + ‘k’ input binary signals are used in the generation of ‘n’ output binary signals corresponding to ‘k’ input signals.

Convolutional coding, a subset of tree codes, and Viterbi decoding are the most popular FEC techniques used on satellite channels. Thus, when the transmitted signal is FEC coded, the receiver during decoding is able to decide in most cases if the signal has been corrupted during transmission and in some cases the receiver is able to correct the corrupted signal. Thus, the receiver makes requests for data retransmission only when it detects loss of data or when data is so much corrupted that receiver cannot correct it. Since receiver can tolerate a certain level of errors in the received data, the required signal-to-noise ratio at the receiver reduces. Thus less power is required for transmission.

The reduction in required signal-to-noise ratio at the transmitter to maintain an acceptable BER can also be viewed as the reduction in satellite channel's BER for a given signal-to-noise ratio at the transmitter. Thus, use of **FEC coding reduces the BER** of the satellite channel for a given signal-to-noise ratio at the receiver.

However, whenever the receiver commits a mistake in detecting corrupted data or in deciding what must have been the original data before corruption, a whole bunch of successive bits are affected i.e. a 'burst' of errors occurs. Thus, **the original random error nature of satellite channels gets transformed to one with bursty errors**. This change from random error environment to bursty error environment for satellite channels profoundly affects the operation of ATM and AAL protocols and their transport over SDH/PDH frames as described in the next three subsections.

Forward Error Correction coding reduces the Bit Error Rate of Satellite links but makes the errors bursty.

2.3.1.1.Impact of burst errors on the ATM layer

ATM was designed for transmission on a physical medium with excellent error characteristics, such as optical fibre, which has improved dramatically in performance since the 1970s. Therefore, many of the features included in protocols that cope with an unreliable channel were removed from ATM. While this results in considerable protocol simplification in the optical fixed networks ATM was designed for, it also causes severe problems when ATM is transmitted over an error-prone channel, such as the satellite link.

The most important impact of burst errors on the functioning of ATM layer is the dramatic increase in the Cell Loss Ratio (CLR). The 8 bit ATM Header Error Control (HEC) field in the ATM cell header can correct only single bit errors in the header. However, in a bursty error environment, if a burst of errors hits a cell header, it is likely that it will corrupt more than a single bit. Thus the HEC field becomes ineffective for burst errors and the CLR rises dramatically.

It has been shown by a simplified analysis and confirmed by actual experiments [GOYA98] that for random errors, CLR is proportional to the square of the bit error rate (BER) and for bursty errors, CLR is linearly related to BER. Hence, for the same BER, in case of burst errors, the CLR value (proportional to BER) is orders of magnitude higher than the CLR value for random errors (proportional to square of BER). Also, since for bursty errors, CLR is linearly related to BER, the reduction in CLR with reduction in BER is not as steep as in the case of channels with random errors (where CLR is proportional to square of BER). Finally, for bursty errors, the CLR increases with decreasing average burst length. This is because for the same number of total bit errors, shorter error bursts mean that a larger number of cells are affected [GOYA98].

Another negligible but interesting problem is that of misinserted cells. Since 8 HEC bits in the ATM cell header are determined by 32 other bits in the header, there are only 2^{32} valid ATM header patterns out of 2^{40} possibilities (for 40 ATM header bits). Thus for a cell header, hit by a burst of errors, there is a $2^{32}/2^{40}$ chance that corrupted header is a valid one. Moreover, if the corrupted header differs from a valid header by only a single bit, HEC will 'correct' that bit and accept the header as a valid one. Thus for every valid header bit pattern (out of 2^{32} possibilities), there are 40 other patterns (obtained by inverting one bit out of 40) that can be 'corrected'. The possibility that our 'error burst' hits the header is one of these patterns is $40\times2^{32}/2^{40}$. Thus overall, there is a $41\times2^{32}/2^{40}$ (= 41/256 ≈ 1/6) chance that a random bit pattern, emerging after an ATM cell header is hit by a burst of errors, will be taken as a valid header. In that case a cell, that should have been discarded, is accepted as a valid cell. (Errors in the payload must be detected by the transport protocol at the endpoints.) Such a cell is called a 'misinserted' cell. Also, the probability P_{mi} that a cell will be misinserted in a channel with bursty errors is around 1/6th of the cell loss ratio on the channel, i.e.,

$$P_{mi} \approx (1/6) \times CLR \qquad (2.7)$$

Since CLR can be written as a constant times BER, the misinserted cell probability is also a constant times BER, i.e.,

$$P_{mi} = k \times BER \qquad (2.8)$$

The cell insertion rate, C_{ir}, the rate at which cells are inserted in a connection, is obtained by multiplying this probability by the number of ATM cells transmitted per second (r), divided by total possible number of ATM connections (2^{24}), i.e.,

$$C_{ir} = (k \times BER \times r) / 2^{24} \qquad (2.9)$$

Due to the very large number of total possible ATM connections, the cell insertion rate is negligible (about one inserted cell per month) even for high BER (≈10^{-4}) and data rates (≈34 Mbps) [GOYA98].

A transition from random errors to burst errors causes the ATM Cell Loss Ratio metric to rise significantly.

2.3.1.2.Impact of burst errors on AAL protocols

The cyclic error detection codes employed by AAL protocols type 1, 3/4 and 5 are susceptible to error bursts in the same way as the ATM HEC code. A burst of errors that passes undetected through these codes may cause failure of protocol's mechanism or corruption in data. AAL type 1's segmentation and reassembly (SAR) header consists of 4 bits of Sequence Number (SN) protected by a 3 bit CRC code and a single bit parity check. There is a 15/255=1/17 chance that an error burst on the header will not be detected by the CRC code and parity check. Such an undetected error at the SAR layer may lead to synchronisation failure at the receiver's convergence sublayer. AAL 3/4 uses a 10-bit CRC at the SAR level. Here, bursty errors and scrambling on the satellite channel increases the probability of undetected error. However, full byte interleaving of ATM cell payload can reduce undetected error rate by several orders of magnitude by distributing the burst error into two AAL 3/4 payloads. The price to be paid for distributing burst error into two AAL payloads is doubling of the detected error rate and AAL 3/4 payload discard rate. AAL type 5 uses a 32-bit CRC code that detects all burst errors of length 32 or less. For longer bursts, the error detection capability of this code is much stronger than that of AAL 3/4 CRC. Moreover, it uses a length check field, which finds out loss or gain of cells in an AAL 5 payload, even when CRC code fails to detect it. Hence it is unlikely that a burst error in AAL 5 payload would go undetected [GOYA98].

ATM AAL 1 and 3/4 are susceptible to bursty errors.

AAL 5 is more robust against bursty errors.

2.3.1.3.Solutions for Improving Error Characteristics

In the previous sections, it was seen that the bursty error characteristics of FEC-coded satellite channels adversely affect the performance of physical, ATM and AAL protocols. Currently, two popular methods used to get around the problem of bursty errors are:

- Use of an outer Reed-Solomon (RS) coding/decoding in concatenation with 'inner' convolutional coding/Viterbi decoding. Outer RS coding/decoding will perform the function of correcting error bursts resulting from inner coding/decoding. RS codes consume little extra bandwidth (e.g. 9% at 2Mbps). In December 1992, INTELSAT approved the use of RS codes as an optional feature. [GOYA98] discusses some of the tests and field trials conducted to test the performance of Reed Solomon codes.
- CRC codes used in ATM and AAL layer headers are able to correct single bit errors in the header. Thus, if the bits of N headers are interleaved before encoding and de-interleaved after decoding, the burst of errors will get spread over N headers such that two consecutive headers emerging after de-interleaving will most probably never have more than a single bit in error. Now the CRC code will be able to correct single bit errors and going by dual mode of operation, no cell/AAL PDU will be discarded. Interleaving involves re-shuffling

of bits on the channel and there is no overhead involved. However, process of interleaving and deinterleaving requires additional memory and introduces delay at both sender and receiver. [GOYA98] discusses the basic ideas behind interleaving scheme ALE (ATM Link Enhancement) developed by COMSAT and how it addresses the burst error problems of ATM, AAL and physical layer protocols.

Burst errors can be mitigated either by additional encoding (like Reed-Solomon) or by using "interleaving" techniques.

Different methods have been proposed which are applicable for two scenarios. The performance of these scheme is directly related to the code rate (bandwidth efficiency) and/or the coding gain (power efficiency), provided the delay involved is acceptable to any ATM-based application. These methods were summarised in [ORS99a].

3. LEO UPLINK/DOWNLINK CHARACTERISTICS

3.1. UPLINK AND DOWNLINK PROPAGATION DELAY

The propagation delay for the packets of a connection is the sum of the following three quantities : The source ground terminal to source satellite propagation delay (t_{up}), the Inter-satellite link propagation delays (t_i) (if ISLs are used) and the destination satellite to destination ground terminal propagation delay (t_{down}).

The uplink and downlink satellite-ground terminal propagation delays (t_{up} and t_{down} respectively) represent the time taken for the signal to travel from the source ground terminal to the first satellite in the network (t_{up}), and the time for the signal to reach the destination ground terminal from the last satellite in the network (t_{down}).

t_{up}= distance from terminal to satellite/speed of light (3.1)
t_{down}= distance from satellite to destination terminal/speed of light (3.2)

The Link delays depends on the constellation design. In contrast to GEO satellites, the LEO uplink and downlink propagation delay is variable over time, in particular for Earth Fixed Cells (EFC) constellations.

In this section, we develop a simple uplink and down link delay model of a satellite network. This model can be used to estimate the end-to-end delay of LEO satellite networks. The end-to-end delay (D) experienced by a data packet traversing the satellite network is the sum of the transmission delay (t_t), the uplink (t_{up}) and downlink (t_{down}) ground segment to satellite propagation delays, the inter-satellite link delay (t_i), the on-board switching and processing delay (t_s) and the buffering delay (t_q). The inter-satellite, on-board switching, processing and buffering delays are cumulative over the path traversed by a connection. In this model, we only consider the satellite component of the delay. The total delay experienced by a packet is the sum of the delays of the satellite and the terrestrial networks. This model does not incorporate the delay variation experienced by the cells of a connection. The delay variation is caused by orbital dynamics, buffering, adaptive routing (in LEOs) and on-board processing. Quantitative analysis of delay jitter in satellite systems is not considered here. The end-to-end delay (D) is given by:

$$D = t_t + t_{up} + t_i + t_{down} + t_s + t_q$$

Transmission delay: The transmission delay (t_t) is the time taken to transmit a single data packet at the network data rate.

$$t_t = \frac{packet\_size}{data\_rate}$$

For broadband networks with high data rates, the transmission delays are negligible in comparison to the satellite propagation delays. For example, a 9180-byte TCP packet is transmitted in about 734 microseconds at an information rate of 100 Mbit/s (for a 155Mbit/s uplink). This delay is much less than the propagation delays in satellites.

Propagation delay: The propagation delay for the cells of a connection is the sum of the following three quantities:

- The source ground terminal to source satellite propagation delay (t_{up})
- The inter-satellite link propagation delays (t_i)
- The destination satellite to destination ground terminal propagation delay (t_{down})

The *uplink and downlink satellite-ground terminal propagation delays* (t_{up} and t_{down} respectively) represent the time taken for the signal to travel from the source ground terminal to the first satellite in the network (t_{up}), and the time for the signal to reach the destination ground terminal from the last satellite in the network (t_{down}).

For an ATM cell with the length of 53 bytes = 424bits, and for the approximate data transmission speed of 150 000 km/s in fibre and 300 000 km/s on satellite link :

$$t_p = \frac{Distance}{PropagationSpeed} \qquad (3.3)$$

$$t_t = \frac{tp\ BitRate}{424} \qquad (3.4)$$

The next table shows the minimum propagation and transmission delays for different satellite systems (for an elevation angle of 90° with the satellite directly over the ground terminal) and the number of cells that can be emitted at a certain bit rate and propagation time.

Distance (Km)	**t_p (in fibre) (µs)**	t_p (on satellite) (µs)	**t_t (Sat.) 64 Kbps (cells)**	**t_t (Sat.) 2 Mbps (cells)**	**t_t (Sat) 155 Mbps (cells)**
Iridium : 780	5200	2600	0.39	12.26	950.47
Teledesic : 1375	9400	4700	0.71	22.17	1718.16
Skybridge : 1469	9793	4897	0.74	2.09	1790.05
LEO[1] : 1603	10687	5343	0.81	25.20	1953.34
MEO[2] : 20000	133333	66667	10	314	24371
GEO : 38000	253333	126667	19	597	46305

TABLE 3-1 PROPAGATION AND TRANSMISSION TIMES

The *inter-satellite link delay* (t_i) is the sum of the propagation delays of the inter-satellite links (ISLs) traversed by the connection. Inter-satellite links (crosslinks) may be *intraplane* or *inter-plane* links. Intraplane links connect satellites within the same orbit plane, while inter-plane links connect satellites in different orbit planes. In GEO systems, ISL delays can be assumed to be constant over a connection's lifetime because GEO satellites are almost stationary over a given point on the Equator, and with respect to one another. In LEO constellations, the ISL delays depend on the orbital radius, the number of satellites per orbit, and the inter-orbital distance (or the number of orbits). All intraplane links in circular orbits, including GEO rings,

[1] Distance for the LEO constellation MSS chosen by "Constellation for Multimedia" project.
[2] Distance for the MEO constellation FSS chosen by "Constellation for Multimedia" project.

are constant. Interplane ISL delays change over time, break at highest latitudes, and must be reformed. As a result, LEO systems can exhibit a high variation in ISL delay.

$$t_i = \frac{\sum ISL\_lengths}{speed\_of\_signal}$$

LEO satellites have lower propagation delays due to their lower altitudes, but many satellites are needed to provide global service. While LEO systems have lower propagation delay, they exhibit higher delay variation due to connection handovers and other factors related to orbital dynamics.

The effects of the propagation delays for LEO systems are further intensified by buffering delays that could be of the order of the propagation delays especially for best-effort IP traffic.

The large delays in GEOs and delay variations in LEOs affect both real time and non-real time applications. Many real time applications are sensitive to the large delay experienced in GEO systems, as well as to the delay variation experienced in LEO systems. In an acknowledgement and time-out based congestion control mechanism (like TCP), performance is inherently related to the delay-bandwidth product of the connection. Moreover, TCP Round Trip Time (RTT) measurements are sensitive to delay variations that may cause false time-outs and retransmissions. As a result, the congestion control issues for broadband satellite networks are somewhat different from those of low-latency terrestrial networks. Both interoperability, as well as performance issues between satellite and terrestrial networks must be addressed before data, voice and video services can be provided over a LEO satellite network.

3.2. UPLINK AND DOWNLINK PROPAGATION IMPAIRMENTS

For deriving statistics of coded bit error rate (BER) from propagation models, whereas the latter uses the results of the link-level simulation to calculate the service availability statistics for dedicated operational scenarios.

The major consideration in planning an overall satellite link is the quality required in the baseband. This is measured in terms of S/N for an analogue system and in terms of Bit Error Rate (BER) for a digital system. In both cases the quality of the link is proportional to the carrier to total noise (C/N_T) at the input of the receiver demodulator. The link budget is a calculation of the (C/N) power ratio at the receiving side of a transmission link, taking into consideration the transmission medium and the transmitter/receiver characteristics.

[ORS99b] discusses respectively uplink and downlink propagation impairments for satellite channels.

The atmospheric attenuation is computed from the water vapour density, ambient temperature, frequency, elevation angle and altitude of the ground station.

The rain loss is computed from frequency, rain rate, latitude and altitude of the ground station, and path elevation angle from the ground station to the satellite.

C/N and Eb/No are computed from atmospheric loss, rain loss, propagation loss, antenna efficiency, link margin, transmit antenna pointing loss and coding.

The parameters required for the uplink channel model and possible values are:

Uplink Parameters	Possible Values
distance [km] between ground station and satellite	up to 2000 km
Elevation angle [rad] ground station to satellite	$\pi/2$ to $\pi/9$
altitude [km] of ground station	-1 to 30 km
latitude [rad] of ground station	- $\pi/2$ to $\pi/2$ rad (note Skybridge tails off above 70 degrees)
Transmission frequency [GHz].	Ku band, Ka band
Polarization [deg].	0-360 deg
Bandwidth [MHz]	
Data rate [Mbps]	1-155 Mbps
Transmit antenna efficiency	0.4-0.8
Transmit antenna diameter [m]	0.6-20 m
Transmit antenna power [Watts]	1-100 Watts
Transmit antenna pointing loss [dB]	0.5-3 dB
Temperature [degrees C]	Night Time Winter: 0.0 degrees C Day Time Winter: 6.0 degrees C Night Time Summer: 15 degrees C Day Time Summer: 20 degrees C
Water vapour level [g/m^3]	Dry: 3.0 g/m^3 Medium: 5.0 g/m^3 High: 8.0 g/m^3 Very High: 12.0 g/m^3
Rain rate [mm/h]	see [ORS99b] Depending on region
Link Margin [dB]	1-3 dB
G/T	-5 to –15 dB
Multiple Access	TDMA, FDMA, CDMA, PRMA
Number of Users (simultaneous)	1-500
Modulation	BPSK, QPSK, DPSK, MQPSK
Coding Scheme	Convolutional, Block or Reed Solomon Codes

TABLE 3-2 POSSIBLE UPLINK PARAMETER VALUES

4. LEO CONSTELLATION DESIGN

The overall design of the LEO constellation as a discrete autonomous system or private network will have a considerable effect upon the design of its various components. As such, the constellation design should be considered as an early criterion in its own right. The choices made here affect the choices made in the individual modules; there are a large number of interrelated factors. Network design choices to consider are discussed in this section.

4.1. CONSTELLATION GEOMETRY AND PLACEMENT OF GROUND TERMINALS AND GATEWAYS TO THE TERRESTRIAL NETWORKS.

4.1.1. Orbit Parameters

In order to define the trajectory of a satellite in space, orbital parameters are required. The shape of an orbit is described by two parameters - the semi-major axis (a) and the eccentricity (e). The position of the orbital plane in space is specified by means of another two parameters - the inclination i and the longitude of the ascending node Ω. These parameters which will also be used for the simulation model are described in this section.

4.1.1.1.Semi-Major Axis

This element specifies the size of the orbit (in km). It is defined as one-half of the major axis, which is the length of the chord which passes through the two foci of the orbit's ellipse. For circular orbits, the semi-major axis (a) is simply the radius of the circle. Note that the earth is at one of the foci.

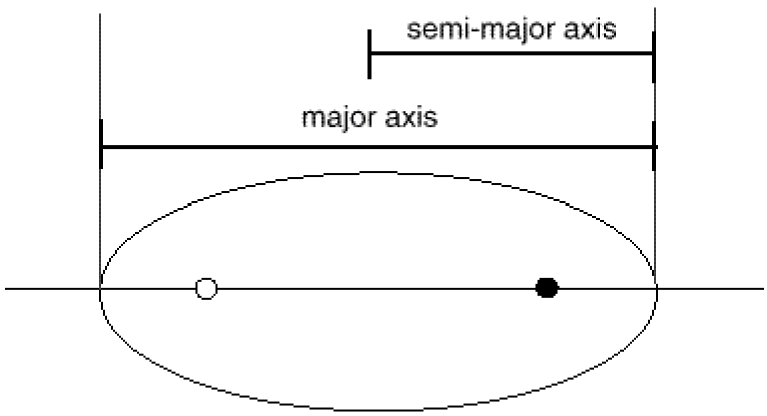


FIGURE 4-1 SEMI-MAJOR AXIS OF ORBIT

4.1.1.2.Eccentricity

Eccentricity (e) determines the shape of the orbit. It is a unitless geometric constant with a value between zero and one. A purely circular orbit has an eccentricity of zero.

4.1.1.3.Inclination

The Inclination (i) determines the tilt of the orbital plane with respect to the equatorial plane of the earth and is an angle measured in degrees. This element is defined as the angle between the two normal vectors K and W shown in figure 4-2. An orbit with an inclination of zero degrees is equatorial; an orbit with an inclination of 90 degrees is polar. Inclinations of less than 90 degrees correspond to direct orbits (i.e., the satellite is rotating around the north pole heading east) and inclinations between 90 and 180 degrees correspond to retrograde orbits (i.e., the satellite is rotating around the north pole heading west). Inclinations are limited to a maximum of 180 degrees.

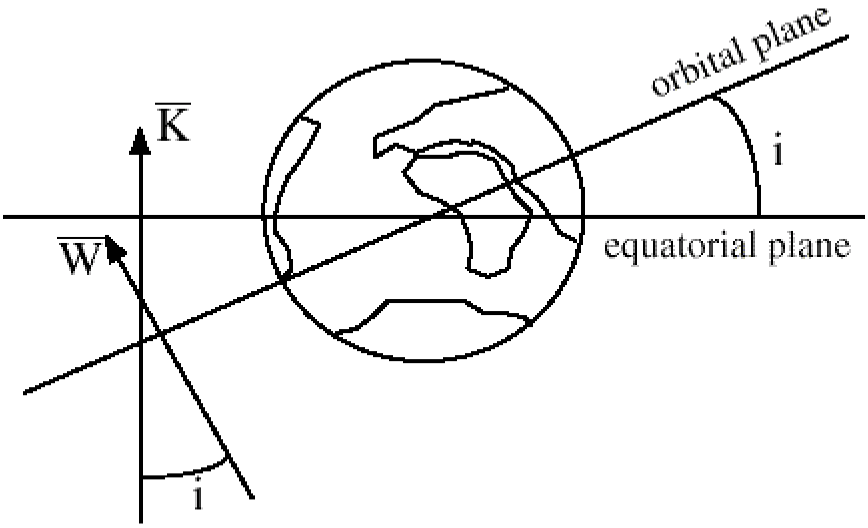


FIGURE 4-2 INCLINATION OF ORBIT

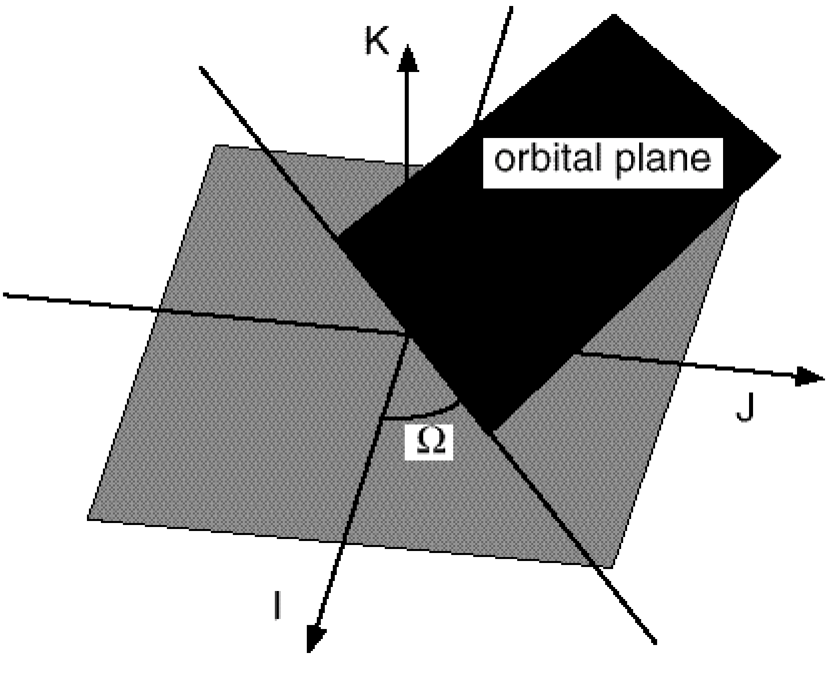


FIGURE 4-3 RAAN OF ORBIT

4.1.1.4. Longitude of the Ascending Node (Ω)

This element determines the rotation of the orbital plane around the normal vector K, and is an angle measured in degrees. It is defined as the angle in the equatorial plane between the line formed by the intersection of the orbital plane and the vector I as shown in Figure 4-3. Thus, this "longitude" is not a normal longitude tied to the earth's surface, but is an angle measured in the equatorial plane. It is therefore also called Right Ascension of the Ascending Node (RAAN).

4.1.1.5. Argument of Perigee

The argument of perigee (ω) determines the rotation of the orbit in its own plane. This parameter is used to describe thhe location of the orbit in its plane. The argument of perigee is undefined in the case of circular orbits, since the perigee is undefined.

Orbital Perturbations

There are many subtle effects that perturb earth satellite orbits, invalidating the simple orbits predicted by two-body gravity equations. Some of the factors that perturbate the orbit are:

- Earth's Oblateness: The earth bulges at the equator, which leads to a much more complex gravity field than the spherically symmetric field of a "point" gravity source.
- Solar and Lunar Effects: These effects of the sun and moon are the most influential gravitational forces on earth satellites besides the earth's own field.
- Atmospheric Drag: The friction that a satellite encounters as it passes through the diffuse upper layers of the earth's atmosphere.
- Solar Radiation Pressure: Solar radiation pressure is caused by collisions between the satellite and photons radiating from the sun, which are absorbed or reflected.

4.1.2. Orbits

Orbital geometry has a considerable effect on the design of satellite constellation network, and influences satellite coverage and diversity, physical propagation considerations such as power constraints and link budgets, and - particularly important from a networking viewpoint - the resulting dynamic network topology and round-trip latency and variation. As a result of this, the choice of orbits must be considered carefully in order to be able to characterise the resulting class of satellite network accurately.

We can classify the orbital choices for a constellation into two categories, based on whether the orbits are circular or elliptical:

4.1.2.1. Circular orbits

Satellites in circular orbits can provide continuous coverage of an area of ground beneath them (the ‘footprint’), but the area covered by the satellite moves as the satellite moves in its orbit. The altitude of these orbits is selected as a result of physical and geometric considerations, including signal power, time of satellite visibility, coverage area, and avoidance of the Van Allen radiation belts.

Low Earth Orbit (LEO)

At altitudes of typically between 500 and 2000km, beyond the upper atmosphere but below the inner Van Allen belt, a large number of satellites are required to provide simultaneous full-earth coverage. The actual number of satellites required depends upon the coverage required and upon the frequency bands used; large amounts of frequency reuse across the earth becomes possible to provide large system capacity. Propagation delay between earth station and satellite is under 15 milliseconds (ms), depending on their relative locations. LEO satellite constellations include:

- *voice:* Iridium, Globalstar
- messaging: Orbcomm
- proposed for broadband data: Teledesic, Skybridge

Medium Earth Orbit (MEO)

At altitudes of around 10,000km, between the inner and outer Van Allen belts, these orbits can permit full earth coverage with fewer, larger satellites, with increasing resulting delay. Propagation delay between earth station and satellite is under 40ms. MEO satellite constellations include:

- voice: ICO
- *proposed for broadband data:* Orblink, Hughes Spaceway NGSO, @contact.

Geostationary Earth Orbit (GEO)

- At an altitude of 32,768km above the equator, the angular velocity of a satellite in this orbit matches the daily rotation of the earth's surface, and this orbit has been widely used as a result. Providing coverage of high latitudes (above 75°) is generally not possible, so full earth coverage cannot be achieved with a geostationary constellation. Propagation delay between earth station is around 0.125ms; this leads to the widely-quoted half-second round-trip latency quoted for communications via geostationary satellite. GEO satellites and constellations include:
- *television broadcast*: Astra etc.
- *voice*: Inmarsat and proposed single satellites for targeted service areas such as Thuraya, ACeS, APMT.
- *proposed for broadband data:* Hughes Spaceway, Loral Astrolink

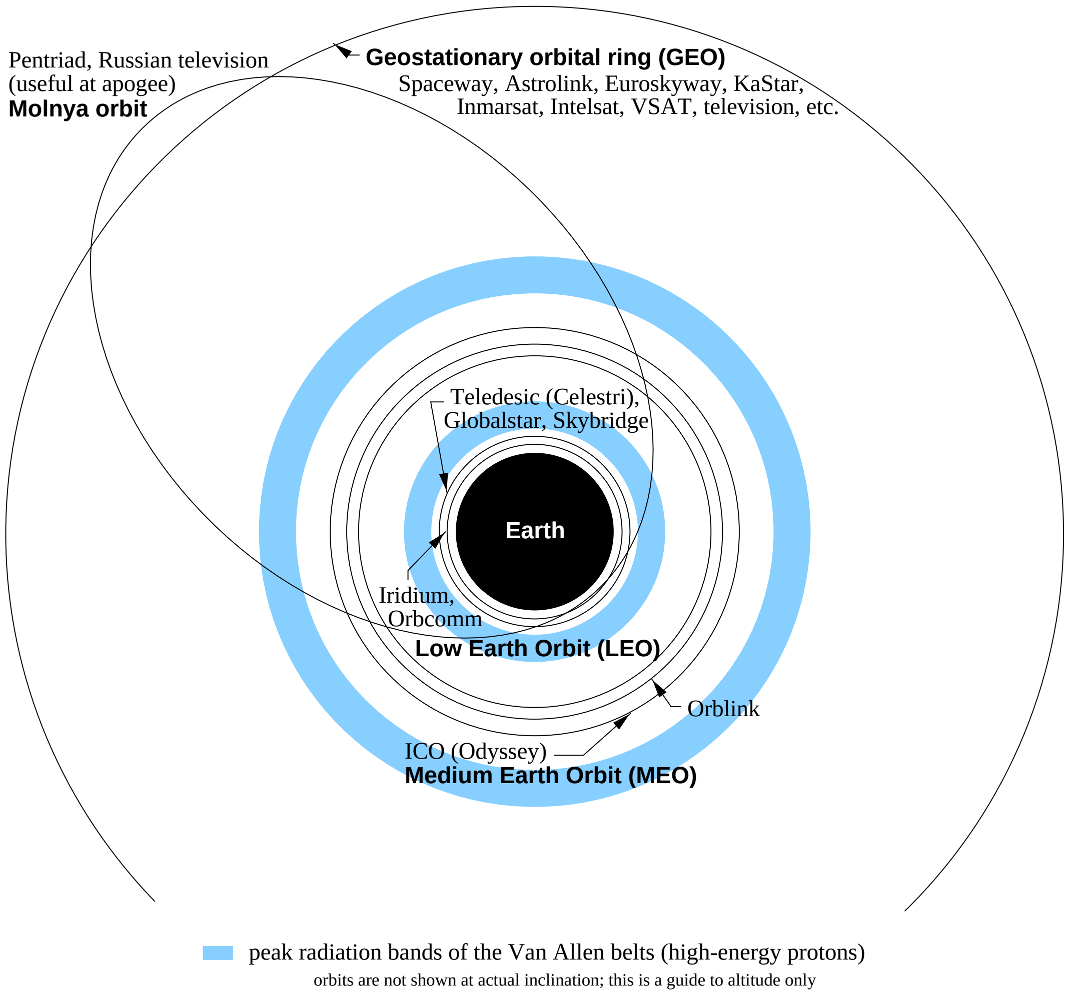


FIGURE 4-4 ORBITAL ALTITUDES FOR SATELLITE CONSTELLATIONS

4.1.2.2. Elliptical orbits

Elliptical orbits differ from the continuous-but-moving coverage of circular orbits, in that coverage is only provided when the satellite is moving very slowly relative to the ground while at apogee, furthest from the earth's surface, and power requirements in link budgets are dimensioned for this distance. When the satellite moves faster from high apogee to low perigee and back, and as its coverage area zooms in size (and other satellites are at apogee providing services in its place) it does not provide service coverage; in fact, its electronics may be shut down while it passes through the Van Allen radiation belts.

These orbits are generally at an inclination of 63.4 degrees so that the orbit is quasistationary with respect to the earth's surface. This high inclination enables coverage of high latitudes, and Russian use of Molnya and Tundra elliptical orbits for satellite television to the high-latitude Russian states is well-known. Elliptical constellations include:

- *proposed for broadband data*: Virtual GEO and Pentriad. These have apogees beyond geostationary orbit, with resulting larger propagation delays. Virtual GEO also intends to establish intersatellite links between satellites at the apogees of different elliptical orbits.

As elliptical orbits are the exception rather than the general rule, and generally provide carefully-targeted selected, rather than general worldwide, coverage, we will not consider them further here. The properties of elliptical orbits have been explored extensively by John Draim [DRAI98].

4.1.3. An Overview of Constellation Geometry

Simulating a constellation network requires an appreciation of how the satellites move over time, and when handover between satellites occurs. Handover is discussed in section 4.3; why and when it occurs is discussed here.

Many satellite constellations are based around the idea of corotating planes, slightly offset to provide full overlap, giving *streets of coverage*, as illustrated in Figure 4-5.

These corotating planes have the same inclination with respect to a reference plane. It would be possible to place these orbital planes at constant inclination to any reference plane, which does not have to be the equator. However, if we place the orbits with constant inclination to a plane other than the equator, we complicate the ground paths, and the varying action of precession due to the oblateness of the earth will act to distort the network, making controlling ground coverage and maintaining ISLs via pointing much more complex. As a result, the equator is the only reference plane we consider.

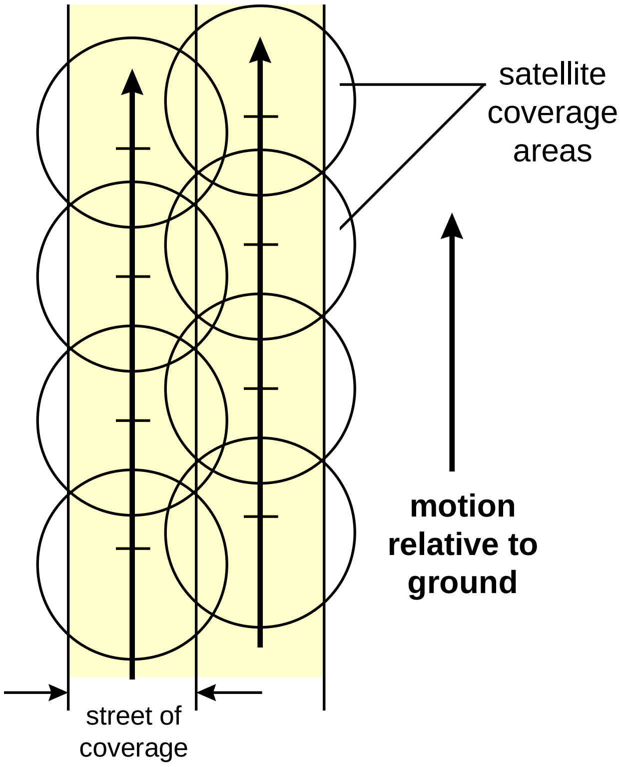


FIGURE 4-5 STREETS OF COVERAGE

4.1.3.1. Polar and Near-Polar Star Constellations

The polar orbit has an inclination of exactly 90 degrees and the near polar orbit constellation has an inclination angle near to that, but tailored to the particular requirements of the orbit.

Walker [WALK71, WALK84] explored different types of constellations, often using a streets approach to coverage. Because of his contribution to the field near-polar constellations with an orbital seam between ascending and descending planes are also named the *Walker star pattern*. This due to fact that all of the orbits cross near the poles, and if viewed from one of the poles the orbital planes intersect to make a star, as illustrated in Figure 4-6 below.

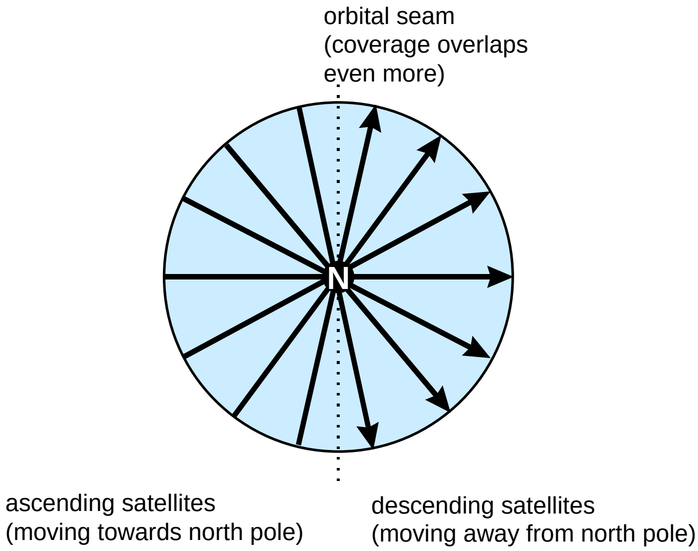


FIGURE 4-6 POLAR VIEW OF WALKER STAR PATTERN

The right ascensions of the ascending nodes of the p orbital planes Ω_1 ..Ω_p are such that they are approximately evenly spaced, with the exception of the two contra-rotating planes at the ring edges. Here the separation between the contra-rotating planes is slightly less than between other planes, to ensure full, overlapping, ground coverage as the streets of coverage move over each other [WALK71]. ISLs between the planes may not be supported due to the high relative velocities (twice the orbital velocity) of the satellites moving in opposite directions.

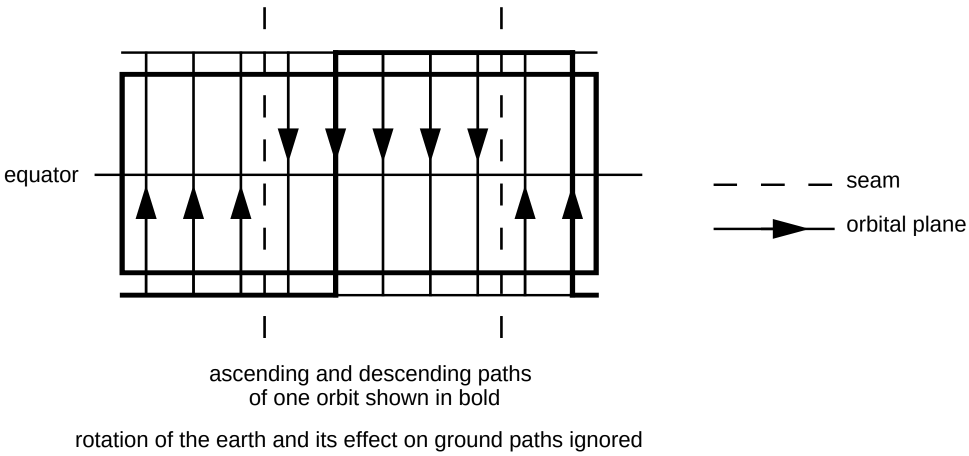


FIGURE 4-7 MAP PROJECTION OF CONSTELLATION (STYLISED)

Figure 4-7 ignores the effect on the orbital paths of the earth's rotation. This effect can be considerable, as illustrated in Figure 4-8 for Teledesic.

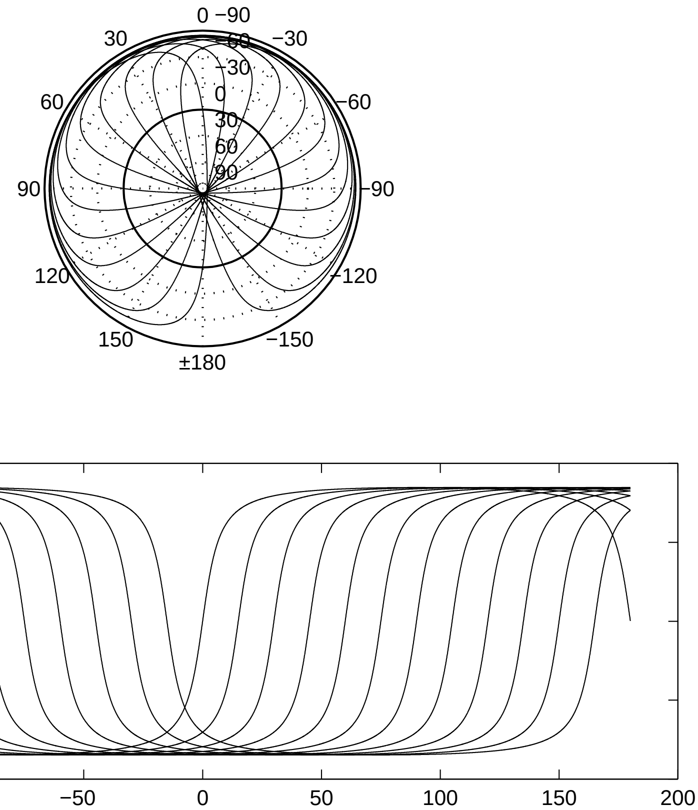


FIGURE 4-8 GROUND TRACKS OF APPROXIMATE *TELEDESIC* CONSTELLATION (288-SATELLITE DESIGN)

With the exception of near the poles, any point on the earth's surface will see overhead satellites moving at regular intervals from north to south or south to north with a star constellation. As satellites in neighbouring planes are closer to each other at the poles than at the equator, coverage of the polar constellation is not evenly spread with varying latitude. The equator is the largest separation that the coverage and distance between orbits must be defined for. At the poles, the overlapping of satellite footprints will cause interference and multiple coverage, requiring some footprints to be disabled, and the high relative velocities of satellites travelling in neighbouring planes will make maintaining ISLs very difficult due to Doppler shift, high tracking rate, and the need to swap neighbours and reestablish links as orbital planes cross.

4.1.3.2. Delta Constellations

The Delta constellation, also known as the Walker delta, or as a rosette, is discussed in detail in [WALK71]. This is a more general constellation case than the polar case. The orbital planes are inclined with constant inclination δ, and the even spacing of the right angles of the ascending nodes Ω_1 ..Ω_p across the full 360° of longitude means that ascending and descending planes of satellites continuously overlap (shown in Figure 4-9), rather than being completely separate as with the Walker star.

Ballard [BALL80] concentrates on the bounds of multiple satellite visibility by interleaving low-inclination multiple planes containing few satellites and using careful phasing to fill in the gaps between satellite footprints in the same plane. Ballard calls the delta constellation an inclined rosette constellation and does not use the 'street of coverage' approach (required for near-polar constellations with near-parallel planes) that is assumed by Rider and detailed in

[RIDE85]. Although interleaving and careful phase alignment of planes for coverage is adopted by *Skybridge* to decrease the number of satellites required, the inclined orbits place more severe constraints on inter-satellite networking with intersatellite links. The SkyBridge constellation satellite tracks are shown in Figure 4-10.

There is no coverage above a certain latitude depending upon the value of δ; inclined rosette constellations generally neglect polar coverage.

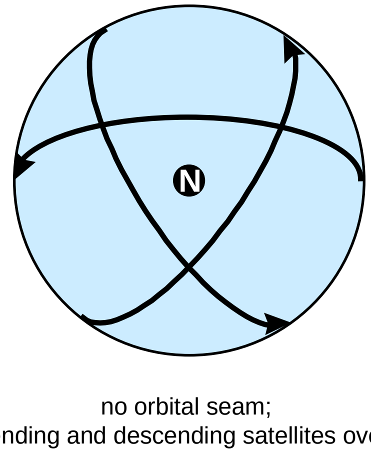


FIGURE 4-9 SIMPLEST WALKER DELTA CONSTELLATION

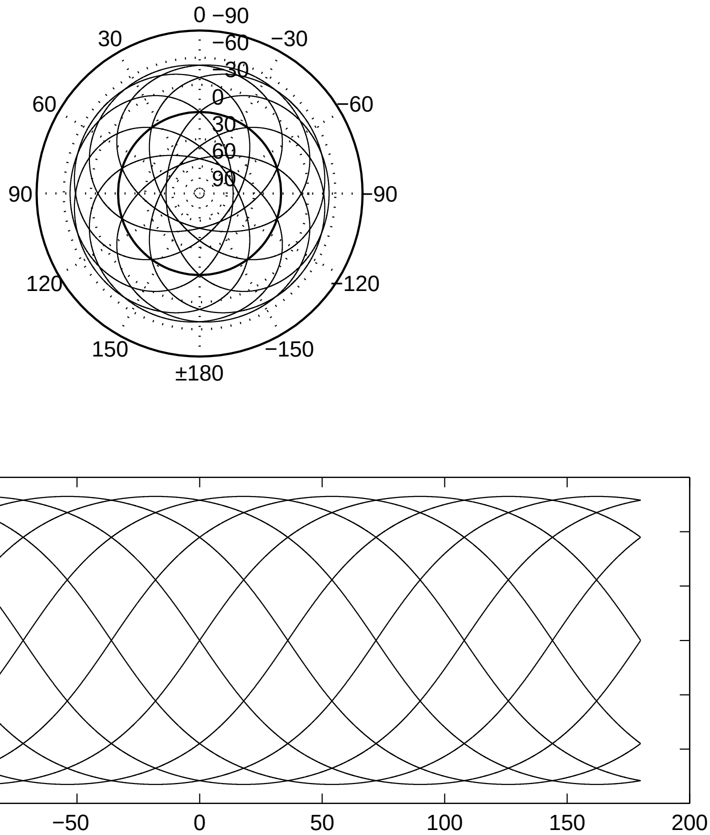


FIGURE 4-10 EXAMPLE ROSETTE GROUND TRACK (SKYBRIDGE 80-SATELLITE APPROXIMATION)

4.1.3.3. Notation

Constellations are usually described in one of two forms in the literature:

Walker notation:

N/P/p

number of satellites per plane N/number of planes P/number of distinct phases of planes to control spacing offsets in planes. Thus, a Teledesic Boeing design approximation could be described as 24/12/2.

Ballard notation:

(NP,P,m)

(total number of satellites NP, number of planes P, harmonic factor m describing phasing between planes). A Skybridge design approximation could be described as two subconstellations, both (40,8,2), offset from one another.

The Walker notation is more commonly seen, although the Ballard notation can more accurately describe possible offsets between planes, especially when m is a fractional. Note that the Ballard notation is assumed to describe a rosette where the ascending nodes are spaced over 360°, while in Walker notation this is unspecified and unclear.

4.1.4. Network topology

4.1.4.1. Primarily ground-based networks

The topology of a ground-based network, where the satellites are only used to provide last-hop connectivity, is entirely arbitrary, and governed primarily by financial considerations. It is likely that all satellite telemetry, tracking and control (TT&C) ground stations will be networked, to share information about the state of the constellation, but beyond that there are a large number of networking possibilities and a number of ways the constellation can be integrated with existing terrestrial networks.

As a result of this, the design of the terrestrial network component of a ground-based constellation like Skybridge must be explicitly determined.

4.1.4.2. Primarily space-based networks

Primarily space-based networks have a slightly more predictable topology, as a result of their use of ISLs and the constraints placed on ISL use by orbital geometry. (Of course, these space-based networks can be complemented by ground networks as briefly discussed above, opening up the design space.).

To get an idea of the topology of the space component of a satellite network with intersatellite links, we can take the network off the earth, breaking the ring formed by the orbital planes, and lay it flat. We will do this with a Walker star geometry, simply because most constellations with ISLs have adopted Walker star geometries (Iridium, Teledesic), avoiding complex diversity questions. (Motorola’s Celestri design was proposed as a full rosette with ISLs, however.)

4.1.4.3. The case for intersatellite links

It is possible to design a satellite constellation network as a primarily ground-based network where an overhead satellite simply provides the last hop, as a primarily space-based network where radio or laser intersatellite links (ISLs) provide direct connectivity between satellites, or as a combination of the two approaches providing redundancy.

The intersatellite-link (ISL) approach, where satellites communicate directly with each other by line of sight, can decrease earth-space traffic across the limited air frequencies by removing the need for multiple earth-space hops, but requires more sophisticated and complex processing/switching/routing onboard satellite to support the ISLs. This allows completion of communications in regions where the locally-overhead satellite cannot see a ground gateway station, unlike simpler ‘bent-pipe’ frequency amplifying/shifting satellites which act as simple transponders.

For circular orbits, fixed fore and aft intersatellite link equipment to communicate with satellites in the same plane is possible. This is not possible for the interplane case between satellites in different orbits, as the line-of-sight paths between these satellites will change angle and length as the orbits separate and converge between orbit crossings, giving rise to:

- high relative velocities between the satellites,
- tracking control problems as antennas must slew around
- Doppler shift

In elliptical orbits, a satellite would see the relative positions of satellites 'ahead' and 'behind' appear to rise or fall considerably throughout the orbit, and controlled pointing of the fore and aft intraplane link antennas would be required to compensate for this, whereas interplane crosslinks between quasi-stationary apogees can be easier to maintain (the Virtual GEO design).

Of the proposed broadband data constellations, Teledesic is an example of a space-based network using ISLs, while Skybridge is an example of a ground-based network without ISLs. The two approaches can be contrasted by comparing constellations that are similar to Teledesic and to Skybridge, especially since they are planned for similar orbital altitudes at 1400km.

4.1.4.3.1.The topology of star networks

Join the highest latitudes (**SS** to **SS**) and stretch the network around (Figure 4-11) so that the longitudinal edges join to form the seam, and the sphere returns.

We can simplify this diagram still further by saying that it is topologically equivalent to the same diagram with the half-twists, caused by orbits crossing each other at the poles, removed, giving us an easier-to-visualise rectangle (Figure 4-12). This assumes that there are no ISLs near the poles when we do this.

Our constellation network has the topology of a cylinder, whose edges are formed by the seam.

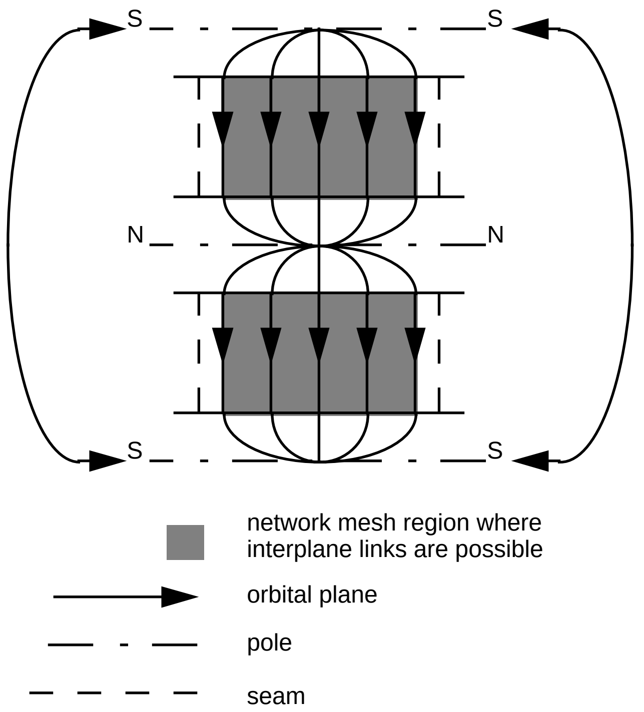


FIGURE 4-11 FLATTENING THE STAR NETWORK

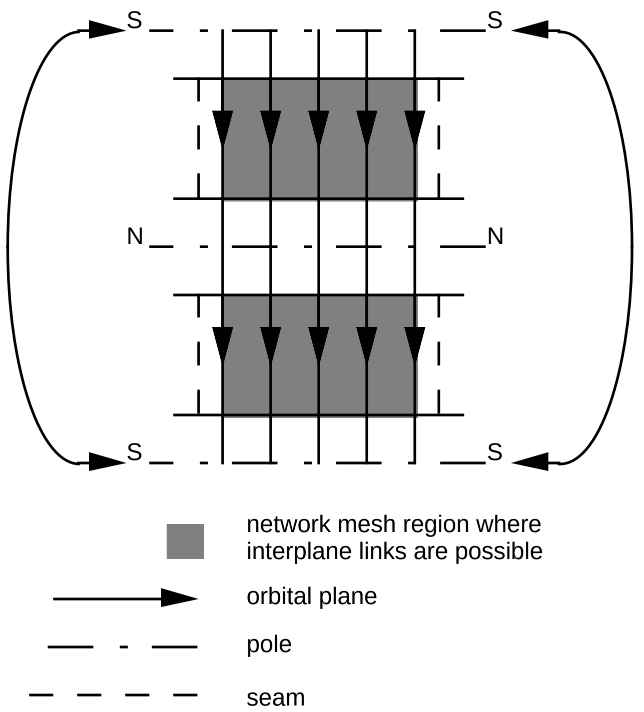


FIGURE 4-12 UNTWISTING THE NETWORK

The seam imposed by the contra-rotating planes, where use of ISLs is unlikely, can make routing in the mesh longer because traffic between opposite parts of the earth must be passed via intra-plane links over the poles. The distance in the network between two ground stations on the earth will vary, depending on where the seam is, and the communication time between them will vary as a result of this. Although the ground stations are always the same physical distance apart on the earth's surface, the apparent network distance between them changes over time in a seamed network , and failures in the ISL network mesh make these changes more visible.

The case of single network coverage with seam, giving the network topology of a twisted cylinder whose axis lies on the equator, is demonstrated in *Iridium* and in *Teledesic*.

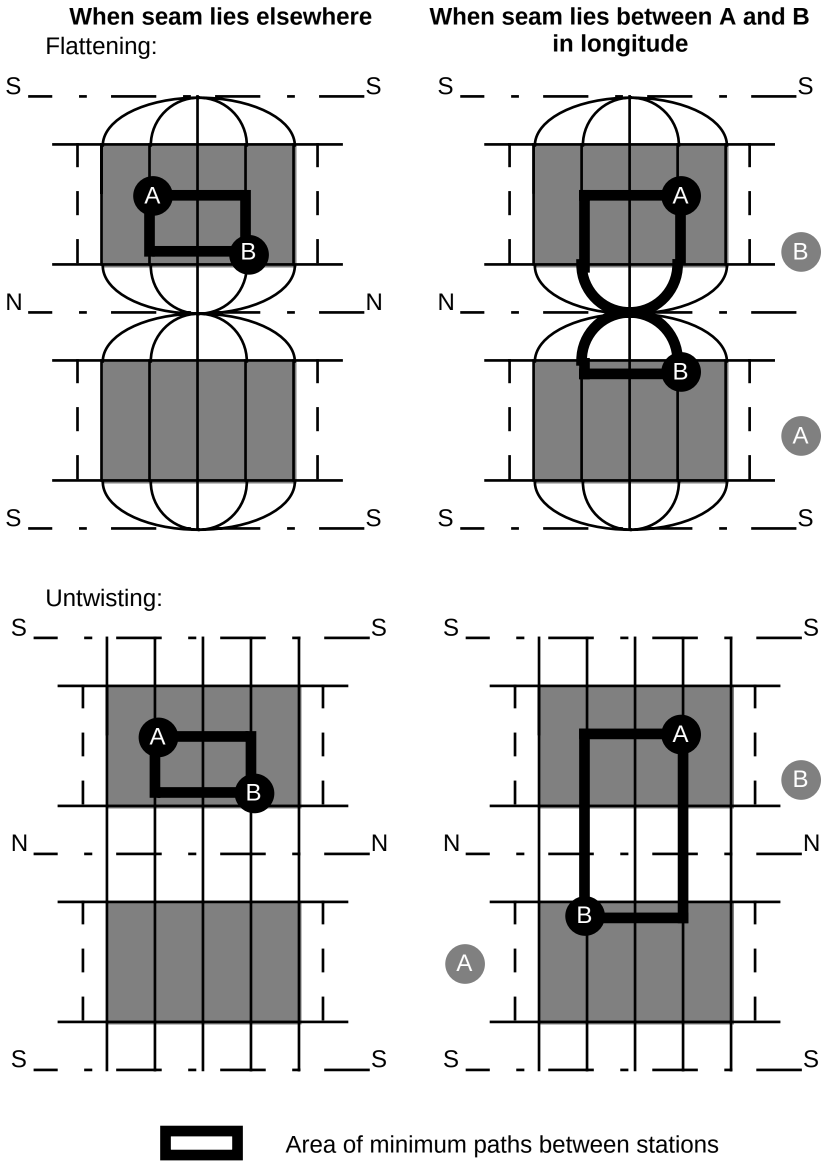


FIGURE 4-13 HOW SEAM DISTANCE VARIES APPARENT NETWORK DISTANCE

4.1.4.4. Generalising to the rosette

The star network is simply a special case of the full rosette topology, discussed here.

If we double the number of orbital planes, so that each plane overlaps entirely with its retrograde, we can eliminate this seam by moving from a star to a delta constellation. However, we then also double the number of satellites, and we have two planes of satellites at any point, both ascending and descending. This is effectively dual network diversity and double network coverage, so that any point on the ground has not just one, but two distinct and widely separated areas of the network that it can see and communicate with. (This network coverage is a somewhat different and separate concept to the *double global coverage* detailed in [ADAM87], which is concerned with dual satellite diversity - seeing two satellites at a time from any point within given latitudes on the earth .)

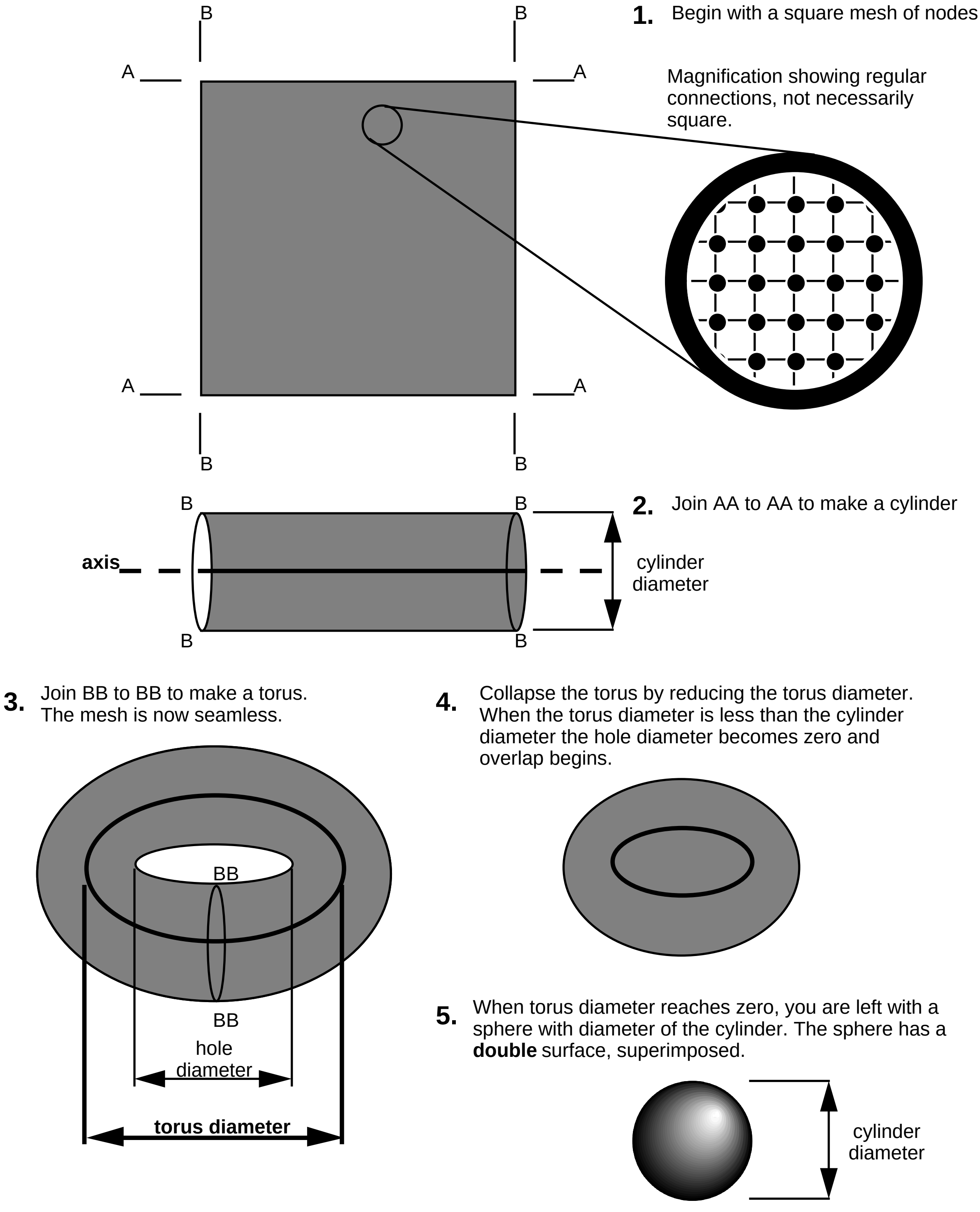


FIGURE 4-14 HOW A TOROIDAL MESH MAPS INTO THE EARTH

The cylinder has been widened so that its longitudinal seam edges now wrap over and touch themselves. With all edges touching, you have a torus. This torus maps onto a sphere that is the earth the satellites orbit around. To do that, the diameter of the axis of the torus is reduced to zero, so that the two 'sides' of the torus coincide. We can obtain the same double-network-coverage sphere from a square mesh. This double-surface sphere gives us double network coverage, where each point on the ground sees two planes of network and two widely-separated points in the network, one in each plane. If a station on the ground can see two points, it can communicate with those points. If those points are in different planes, the point on the ground is communicating with two distinct and separate parts of the network.

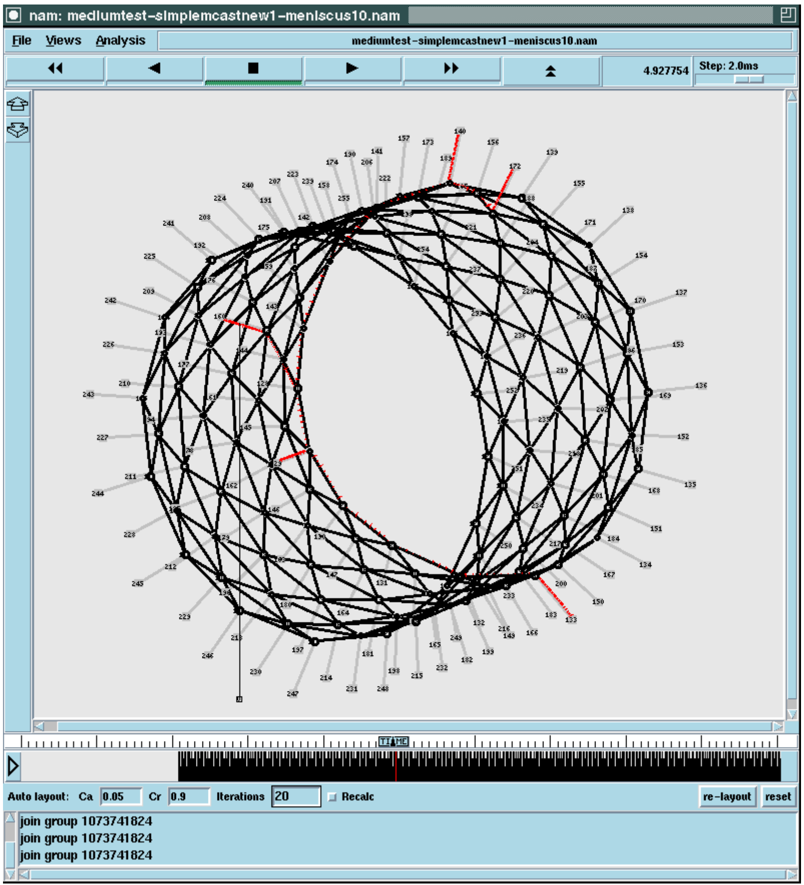
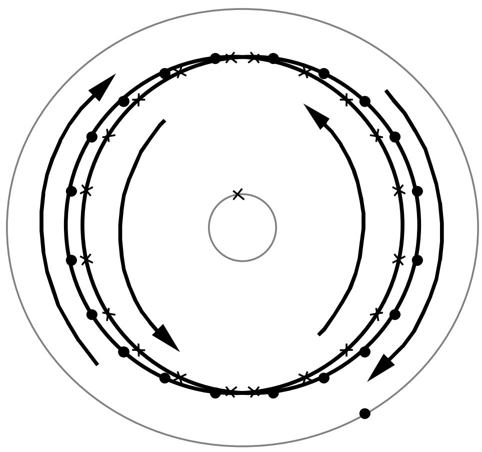


FIGURE 4-15 SIMULATION OF SIXWAY CONNECTIVITY FOR 8/18/2 CONSTELLATION



ascending satellites (dots indicated by clockwise arrows) form one surface
descending satellites (crosses indicated by anticlockwise arrows) form the other

FIGURE 4-16 CROSS-SECTION THROUGH EQUATOR OF ROSETTE WITH ISLS

Such a double-plane constellation cannot be conveniently described by the geometrical notations used by Walker or Ballard. This notation does not fully indicate the topology or size of the network, and so is not really suitable for satellite networks. The ISLs in the constellation do not have to be fourway, as illustrated in Figure 4-15.

In a rosette constellation the seam vanishes, since communications pass over it due to the nature of double coverage. In fact, the seam is now simply a reference plane that the orbits are placed relative to, and this reference plane can be at any longitude.

If we look at the equator and plot both normal and retrograde orbits' hemisphere entry/exit points (Figure 4-16) , we can see the wrapped torus clearly

Seamless networks allow fixed-delay communication between two ground stations, as there is no seam changing position relative to the ground stations to affect in-mesh routing. The bi-directional Manhattan network, or double-network-coverage constellation (shown in Figure 4-17), offers the ground station a choice of satellites for paths of the same delay length.

Manhattan networks are extensively discussed in computing literature [e.g. GOOD86, MAXE87], and recognising that the satellite network constellation can be a bi-directional or unidirectional Manhattan network allows this literature to be applied in modelling the network performance. Results from this literature can be modified to deal with the seamed star constellation.

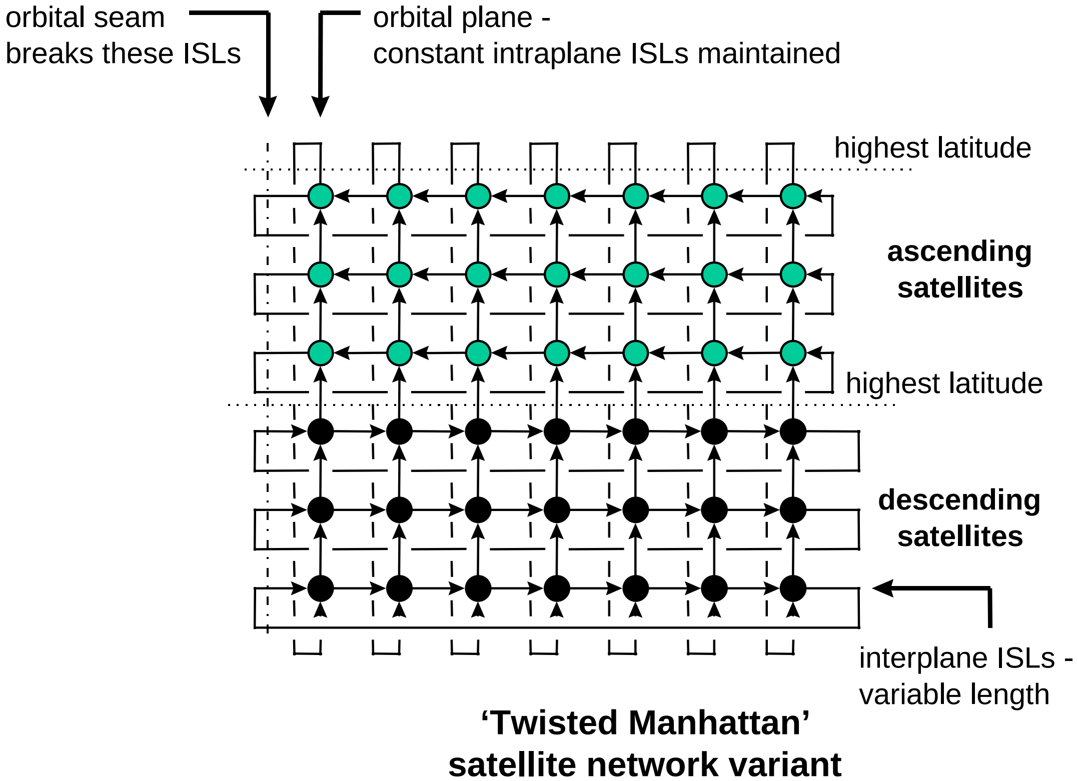


FIGURE 4-17 EXAMPLE NETWORK TOPOLOGY

The number of satellites that must be launched for a seamless network, and the problem of double network coverage, make building any form of seamless constellation network of little benefit unless fixed delay between ground stations via ISLs is essential. However, this does not invalidate the seamless network as a starting point for analysis, provided that we are aware of the limitations of the assumptions we make.

4.1.5. Design choices

The number of possible physical design choices for a constellation are considerable, and not easily summarised. From a networking perspective, we can raise the following choices at a macroscopic level:

space-based network with ISLs	**vs**	ground-based network sans ISLs
star constellation without overlapping ascending and descending planes	**vs**	seamless delta constellation with overlapping planes

These choices can be expressed by considering the commercially-proposed Teledesic and Skybridge constellations. Teledesic exemplifies the former; Skybridge the latter. In both cases the constellation's ground network is undefined; we can compare the two by considering identical users with identical requirements in identical locations.

We can represent the different approaches from a network viewpoint by the following stack diagrams shown in Figure 4-18 and Figure 4-19.

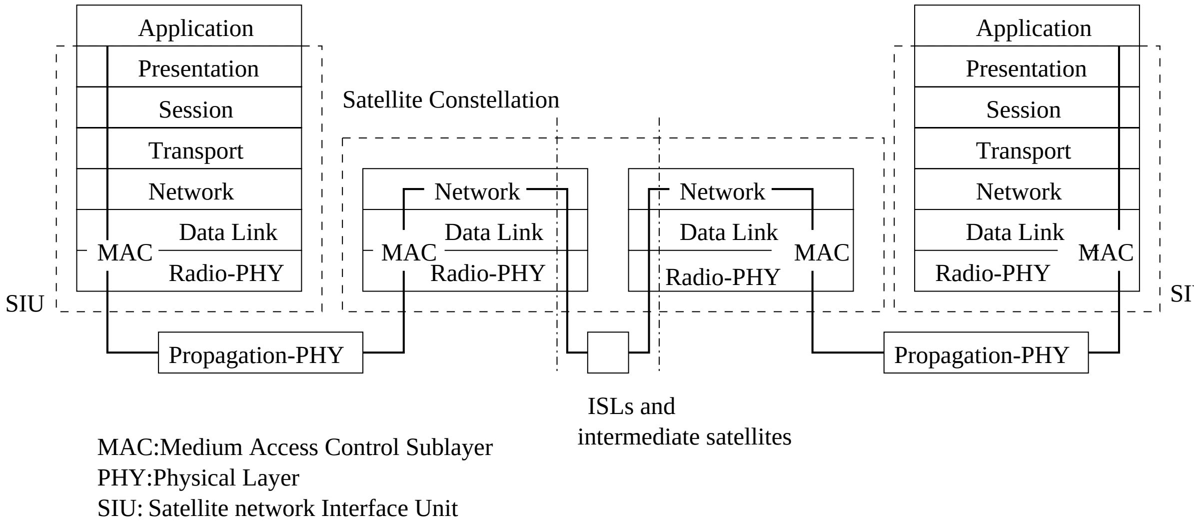


FIGURE 4-18 ISL ROUTING APPROACH A LA TELEDESIC

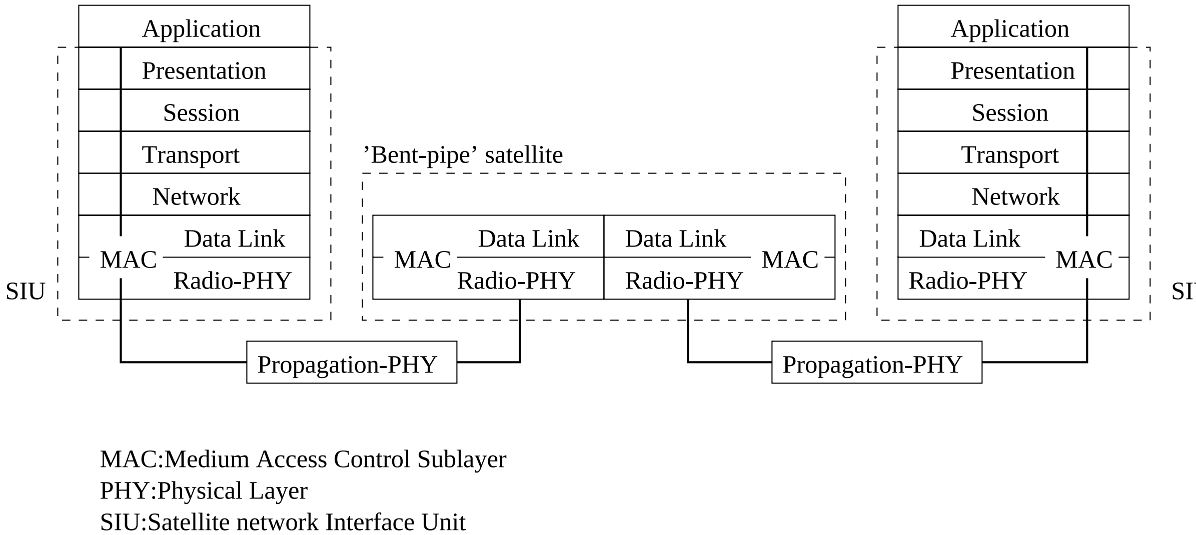


FIGURE 4-19 RELAY 'BENT-PIPE' SATELLITE APPROACH A LA SKYBRIDGE

4.2. AVAILABILITY, VISIBILITY AND DIVERSITY

4.2.1. Earth-to-space diversity

Earth-to-space diversity is the use of more than one satellite at once for communication. This allows an improvement in physical availability, by decreasing the impact of shadowing (buildings obstructing the path between the ground terminal and satellite) and providing redundancy at the physical or data-link level. Diversity is also exploited for soft handovers. An overview of earth-to-space diversity can be found in [WISL96].

This earth-to-space diversity can be exploited at various layers of the network stack. Physical diversity can be exploited in rosette constellations, e.g. Globalstar's use of CDMA and recombination of signals across multiple satellite transponders. It can be exploited at the data-link layer, via TDMA management. Beyond that is research work; coding diversity and network-layer diversity are not currently proposed for exploitation in commercial constellations.

4.2.2. In-orbit network diversity

In-orbit diversity is exploited in the original 840-active-satellite design of Teledesic as shown in Figure 4-20 to provide redundancy for failures in links and satellites. It is only possible due to the large number of satellites in the Teledesic constellation and their close spacing. [LAW94].

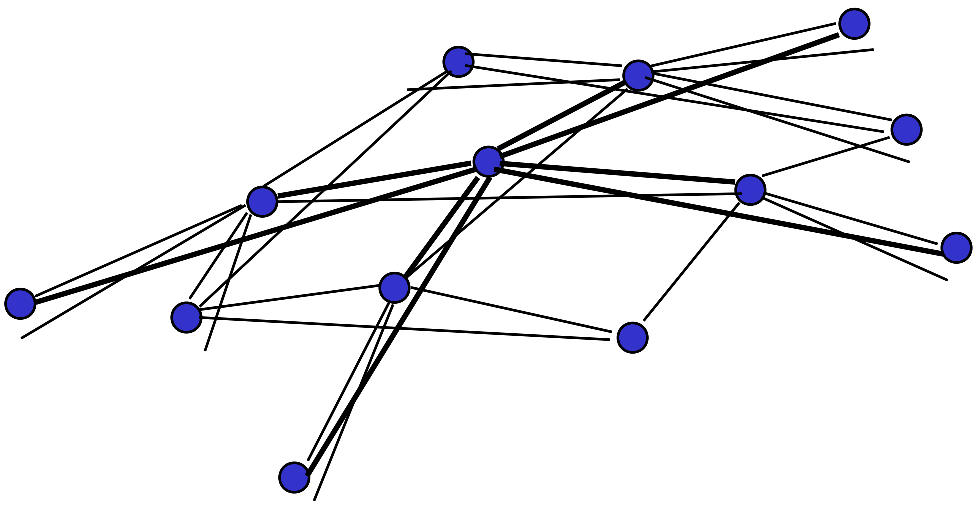


FIGURE 4-20 PART OF TELEDESIC NETWORK MESH SHOWING DUAL NETWORK DIVERSITY

As this can affect routing across the ISL mesh, it can have a considerable effect on end-to-end delivery.

4.2.3. Diversity in Skybridge

Skybridge utilises dual diversity between neighbouring satellites in the two subconstellations to avoid sending transmissions from the part of the sky inhabited by the geostationary arc. This is judged necessary because Skybridge is based around reusing Ku-band frequencies already used by geostationary satellites [SKYB97].

From a networking viewpoint, diversity should only mean a small change in end-to-end delay, with no other visible effects.

4.2.4. *Visibility Statistics for SkyBridge and Teledesic**

The minimum and mean elevation angle statistics for Teledesic and SkyBridge are shown in Figure 4-21. These statistics influence satellite link availability since usually line-of-sight (LOS) for a user to the a satellite is required. A high mean elevation angle means that the probability of having a satellite in LOS is higher. Note that SkyBridge's mean elevation angle for high latitudes is low (as it does not plan to provide service in these areas).

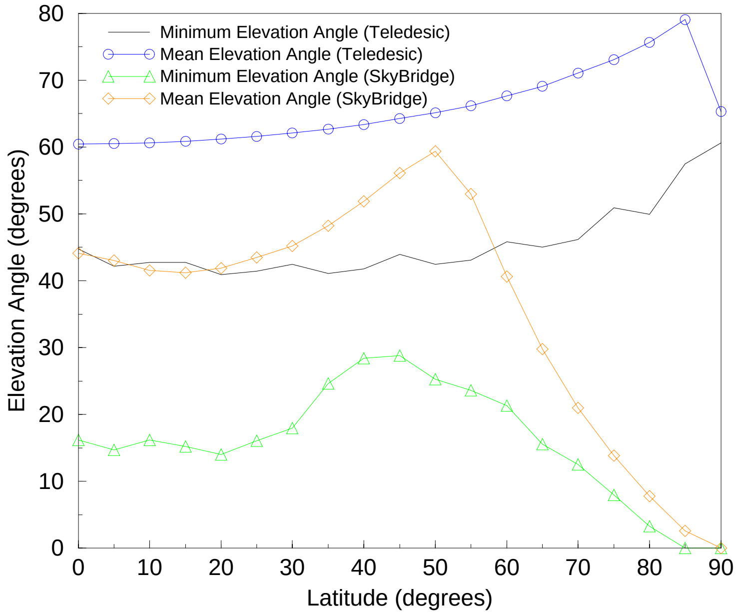
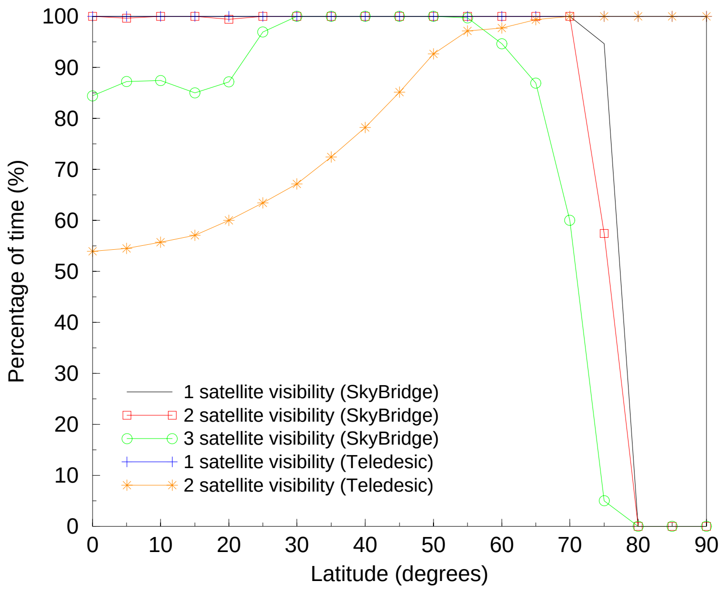


FIGURE 4-21 MINIMUM AND MEAN ELEVATION ANGLE FOR SKYBRIDGE AND TELEDESIC

As SkyBridge is planning to have at least dual satellite diversity (in order not to interfere with GEO satellites at ku-band) the visibility of 2 satellites in percentage of time is high as shown in Figure 4-22. Delta constellations such as SkyBridge also achieve triple satellite visibility. On the other hand Teledesic (near-polar constellation) was not designed for Earth-to-space diversity.



* The authors would like to thank A. Sammut for the production of satellite elevation angle and diversity statistics for SkyBridge and Teledesic.

FIGURE 4-22 SATELLITE VISIBILITY FOR SKYBRIDGE AND TELEDESIC

4.3. HANDOVER BETWEEN SATELLITES AND BETWEEN GATEWAYS.

Whereas the handovers (handoffs) of communications are rather well understood in the terrestrial wireless networks, the handovers in non-geostationary satellite networks supporting multimedia traffic remain a vast field of investigation.

Handover is needed to avoid ongoing call drops. Satellite cells move along with the satellite and calls must be handed over from one beam spot to the next (beam handover), and eventually to the next satellite (satellite handover). In the event the next beam or satellite has no idle circuit to take over the handed-over call, the call is terminated (forced termination), and this event is referred to as a handover failure. Premature handover generally results in unnecessary handover and delayed handover results in increased probability of forced call termination. Handover can be initiated based on the signal level and/or distance measurements.

Beam handover has two scenarios:

- **Intra-beam Handover Scenario** Assume that the subscriber is in beam A using frequency 1 and is associated with satellite S. As the beam approaches another geographic region, frequency 1 may no longer be available. There are two possible reasons for this. The first is governmental regulatory, meaning the particular set of frequencies are not available in the approaching region. Another reason is interference, which may be caused when satellite S moves too close to another satellite using the same frequency. In this case, even though the subscriber is still within beam A (satellite S), the satellite will send a message to the portable unit to change to frequency 2 in order to maintain the communication link. The satellite is the intelligent entity in this handover case.
- **Inter-beam Handover Scenario** The portable unit (PU) continually monitors the radio frequency (RF) power of frequency 1 used in beam A. The PU also monitors the RF power of two adjacent candidate handover beam, B and C, via general broadcast channel (information channel). The PU determines when to handover based on RF signal strength. If the beam B signal becomes stronger than the signal used in beam A, the PU will initiate a handover request to the satellite to switch the user to beam B. The satellite assigns a new frequency (3) to the PU because two adjacent beams cannot use the same frequency (Iridium for Instance use a 12-beam reuse pattern). Inter-beam handover can be extremely frequent (every 2 minute or even less). The PU is the intelligent entity in this handover case.

Two handover scenarios for satellite handovers are possible in situations where diversity is not exploited:

- **Intra-plane Satellite Handover Scenario** Assume that the subscriber moves from beam to beam within satellites S's coverage area. The gateway knows the subscriber is approaching the boundary between satellite S and satellite T because it knows the subscriber's location area code and the satellite's locations. The gateway will send a message to the trailing satellite S to prepare to handover the subscriber, and another message to the leading satellite T in the same plane to prepare to accept the subscriber. The gateway will then send a message to the portable unit via satellite S to resynchronize the Doppler shift and timing arrival of the signal. Finally, the handover is completed when the satellite sends a message to PU informing it of which new frequency to use. The gateway is the intelligent entity in this handover case.

- **Inter-plane Satellite Handover Scenario** Inter-plane satellite handover is the same as intraplane satellite handover except that instead to handing over the connection to a satellite in the same plane, it is handed over to a satellite in a different plane. The reason of doing a handover to a satellite in another plane can be that no satellite in the same plane is able to cover the subscriber or there are no available channels in the satellite of the same plane to do a handover. Another reason can be that the satellite in a different plane is able to provide better service (lower plane satellites have more problems with shadowing than higher level satellites).

Time necessary for launching and executing the handover must be very short. In addition, the handovers should not harm the negotiated connections quality of service.

With the satellites' orbital velocity , and the dimension of the cells, the time to cross the overlap area of two cells is relatively short (a few seconds). However, due to the characteristics of the satellite constellation, a terminal can be covered by at least two satellites. That offers the possibility to optimise the handover respecting each connection quality of service of each connection and serving a greater number of connections.

Figure 4-23 shows the relation between elevation angle and altitude :

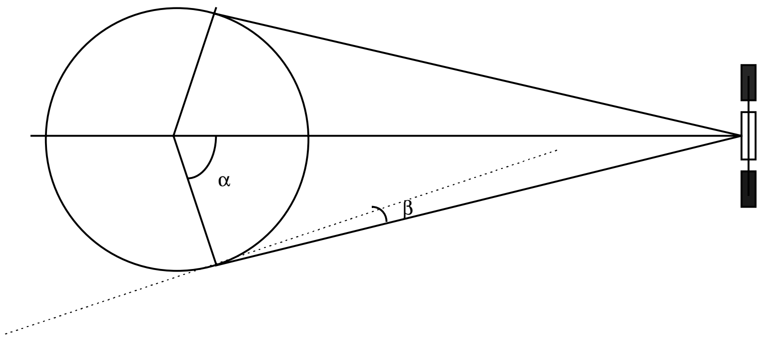


FIGURE 4-23 RELATION BETWEEN ELEVATION ANGLE AND ALTITUDE

While, α is the Coverage Angle, β the Elevation Angle and h the Altitude. The relation is given in [GALT98] by :

$$h = R_T\left(\frac{1}{\cos(\alpha)(1+\tan(\alpha)\tan(\alpha+\beta))}-1\right) \qquad (4.1)$$

The evolution of the Inter-Handover delay according to altitude and minimum elevation angle is shown in Figure 4-24.

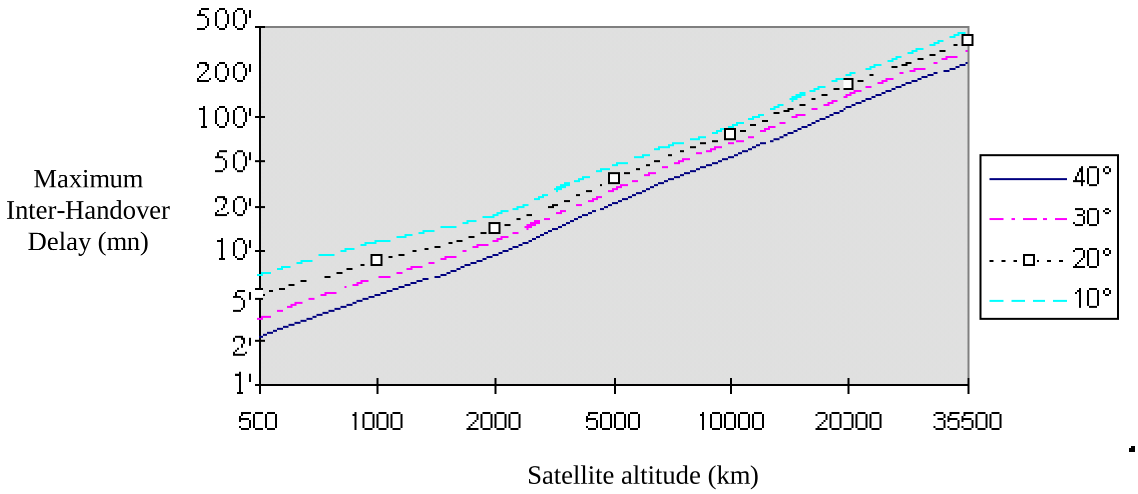


FIGURE 4-24 MAXIMUM INTER-HANDOVER DELAY

4.3.1. Earth-Fixed Cells vs Satellite-Fixed Cells

The handover problem is considered according to the constellation. In fact, EFC (Earth-Fixed Cells) constellations allow the transmission of all users in a geographical cell in the same time, but introduce delay variations. In SFC (Satellite Fixed Cell) constellations the handover rate is more important, so handover time is more critical.

The problems that occur in EFC constellations are due to the exaggerated difference in propagation delays in the radio signal of each satellite. The difference, due to different satellite localisation, results in the loss of sequence, loss or duplication of cells according to the position of satellites relative to earth units.

Figure 4-25 shows the actions the handover of the portable from Satellite 1 to Satellite 2. The distances between the access point and each satellite are different.

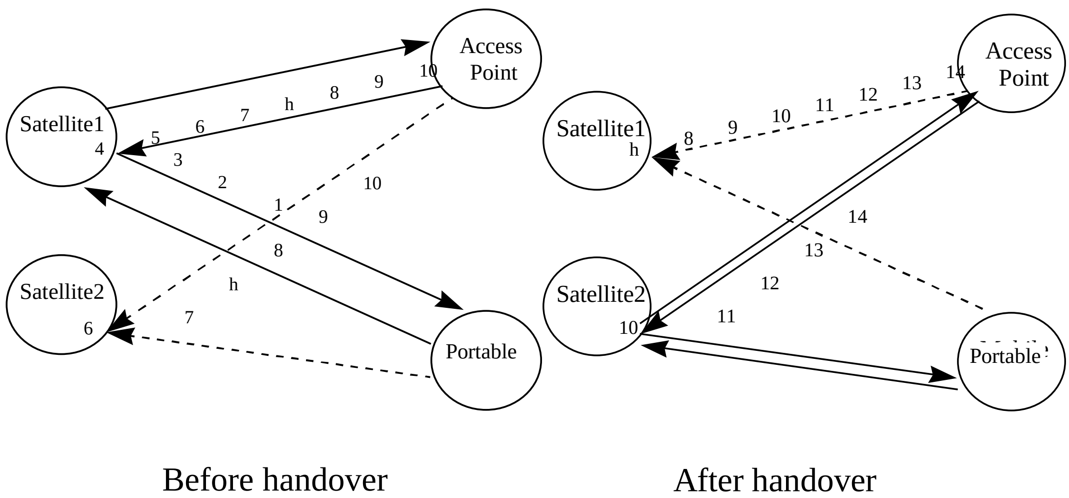
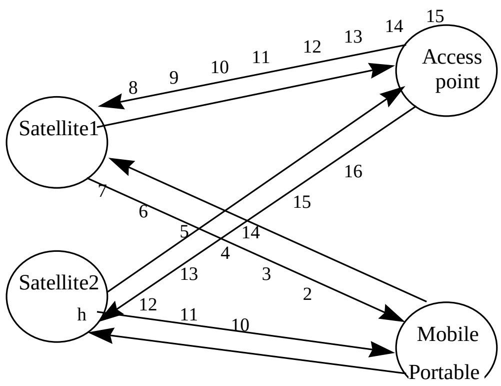


FIGURE 4-25 HANDOVER PROCEDURE, DIFFERENCE IN PROPAGATION DELAYS

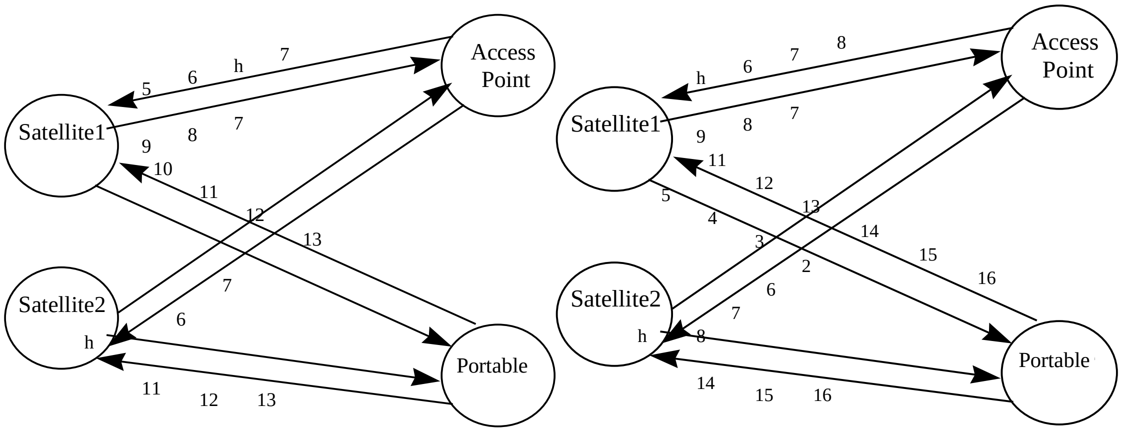
The first problem is the non respect of sequencing of segments. Let us consider the case when Satellite2 receives the handover command before Satellite1 (see Figure 4-26). During the critical period, both satellites will transmit data to destination to the terminal. If the interference problems are resolved then data will be received out of order.



Desequencement of Segments

FIGURE 4-26 PROBLEMS THAT ACCURS ON THE ACCESS POINT/PORTABLE WAY

From the portable to the access point, there is never a handover command. However, if the distance between access point and satellite is lower than the mobile-satellites distance then loss of segments is possible. Otherwise duplication of segments can occur as shown in Figure 4-27. Loss and duplication of segments can be taken in consideration in upper layers of the protocol. It is important to take into account the segments sequence and the respective jitter constraints.



Segments Duplication Segments Loss

FIGURE 4-27 PROBLEMS THAT OCCUR ON THE PORTABLE/ACCESS POINT WAY

One benefit of a narrow satellite footprint is that each satellite can serve its entire coverage area with a number of high-gain scanning beams, each illuminating a single small cell at a time. Small cells allow efficient reuse of spectrum and resulting high system capacity, high channel density, and low transmitter power. However, if this small cell pattern swept the Earth’s surface at the velocity of the satellite, a terminal would be served by the same cell for only a few seconds before a channel reassignment or "hand-off" to the next cell would be necessary (as for Iridium).

As in the case of terrestrial cellular systems, frequent hand-offs result in inefficient channel utilisation, high processing costs, and lower system capacity. The Teledesic, and Skybridge Networks use an Earth-fixed cell design to minimise the hand-off problem.

For Teledesic, channel resources (frequencies and time slots) are associated with each cell and are managed by the current "serving" satellite. As long as a terminal remains within the same

Earth-fixed cell, it maintains the same channel assignment for the duration of a call, regardless of how many satellites and beams are involved. Channel reassignments become the exception rather than the normal case, thus eliminating much of the frequency management and hand-off overhead.

A database contained in each satellite defines the type of service allowed within each Earth-fixed cell. Small fixed cells allow Teledesic to avoid interference to or from specific geographic areas and to contour service areas to national boundaries. This would be difficult to accomplish with large cells or cells that move with the satellite.

The Skybridge system uses up to 18 simultaneous spot beams per satellite. Using active antennas, each spot is fixed on earth and illuminate 350 Km radius cells. 200 gateways are used to provide global coverage. Every satellite covers a 3000 km radius area.

4.3.2. Frequency sharing between LEO and GEO satellites

A key function is to manage the frequency sharing with other current (or planned) users of the frequency bands. SkyBridge shares the Ku-band. It re-utilises the frequency assignments for Fixed Satellite Services, Broadcasting Satellite Services, and Fixed Services between 10 and 18 GHz. It does this while protecting the geostationary satellite system's use of the spectrum through a specific technique called geostationary arc avoidance. Traffic from a gateway cell is transparently handed over to another satellite whenever harmful interference could potentially be generated on a GSO receiver. SkyBridge is also designed to avoid interference with terrestrial wireless systems operating in the same frequency bands.

As the satellites move across the sky, the earth stations track one satellite and when it moves into the avoidance arc, or falls below the horizon, an automatic hand-over of traffic occurs from one satellite to another facilitating:

- a continuous user link when a SkyBridge satellite reaches the minimum elevation angle and the terminal loses sight of the satellite,
- GEO arc avoidance hand-over, which occurs when a SkyBridge satellite is within +/-10° of the protected GEO arc,
- optimisation of resource hand-over resources.

4.3.2.1. SkyBridge Interference with GSO systems

4.3.2.1.1. Service and Infrastructure Links

As the path distance between non-geostationary orbit (NGSO) satellites to geostationary orbit (GSO) ground stations is small when compared to NGSO earth stations and GSO satellites, it is concluded that the major contribution of interference towards the GSO system is in the downlink, from SkyBridge satellites to GSO ground stations. SkyBridge has accepted that interference power into the GSO systems must be limited in order to facilitate band sharing with such systems, power into GSO systems can originate from both main beam and sidelobes of SkyBridge antennas and be received in either main beam or sidelobes of GSO ground station receivers. Sidelobe to sidelobe interference is discounted.

In potential interference conditions, a SkyBridge satellite will cease transmission in the direction of the GSO ground station. A ‘non-operating zone’ is defined around a SkyBridge Gateway Station (GS) such that it is an area around a Gateway Cell (GC) where a potential interference exists between a GSO satellite and associated earth stations. A GC is one of the

SkyBridge stationary cells in which a GSO gateway station exists. A SkyBridge satellite can only illuminate/serve this cell if it is outside the non-operating zone which is described around the boresight line from the GSO ground station to the GSO satellite.

The effectiveness of the non-operating zone's definition depends upon the assumption of directivity of the GSO and SkyBridge's ground station antennas. When the satellite enters the non-operating zone it will cease transmission to the cell in which the gateway stands. SkyBridge earth stations (including user terminals) will also cease transmitting to the satellite. The satellite will also cease transmission to adjacent cells to reduce the possibility of interference through sidelobes. NB, the satellite will only cease transmission to certain cells, it will not shut down totally.

When transmission ceases due to a satellite entering the non-operating zone its traffic is handed over to an adjacent satellite. The constellation has been designed such that there will always be another satellite which can serve the cells affected by the non-operating system. The total period of time spent in the non-operating zone by any single satellite is a very small proportion of the total operating time (<1% is quoted but not demonstrated).

The sequence of spotbeams which will be shut down is determined prior to the actual event and is transmitted to the satellite by a centralised controlling station but the actual handover os coordinated by the GS which is in view of the satellite at he time of handover.

4.3.2.1.2. Telemetry, Tracking & Control (TT&C) Links

TT&C signals will be broadcast continuously via a wide beam antenna, interference with other systems will be mitigated by the use of spread spectrum techniques, which will reduce the power flux density (pfd) of the signals on the ground.

4.3.2.2. SkyBridge Interference with terrestrial systems

SkyBridge states that the probability Density Function (PDF) of its downlink transmissions will always be within the FCC's rules protecting terrestrial systems. It also states that because the satellites are moving this will reduce the power received by a terrestrial station. It must be assumed that this is referring to *mean* power levels received on the ground.

The minimum elevation angle of the ground stations will ensure that terrestrial stations are not affected directly by SkyBridge ground stations' transmissions in the uplink. Coordination with systems which are in the direction of transmission is adopted, with SkyBridge ground stations always deferring to the terrestrial stations through non-use of the same frequency bands.

4.3.3. Analysis of SkyBridge Frequency sharing with GSO Systems

The SkyBridge FCC proposal presents the limiting criteria for interference for each of the service types involved in the desired frequency bands shown in Figure 4-28, planned FSS and BSS and unplanned FSS. The ITU definition of aggregate eirp and effective PDF, i.e. the accumulation of all power from the SkyBridge network into the GSO systems, is used to demonstrate that the SkyBridge system will not breach the criteria in either the long or short term scenarios. This is only the case if the switch-off mechanism is applied, as the isolation of the GSO ground station antennas is taken into account. The techniques which SkyBridge use are relevant in any future NGSO system as it is likely that there may be some band sharing with future MSS geostationary systems and other NGSO systems [SAMM97].

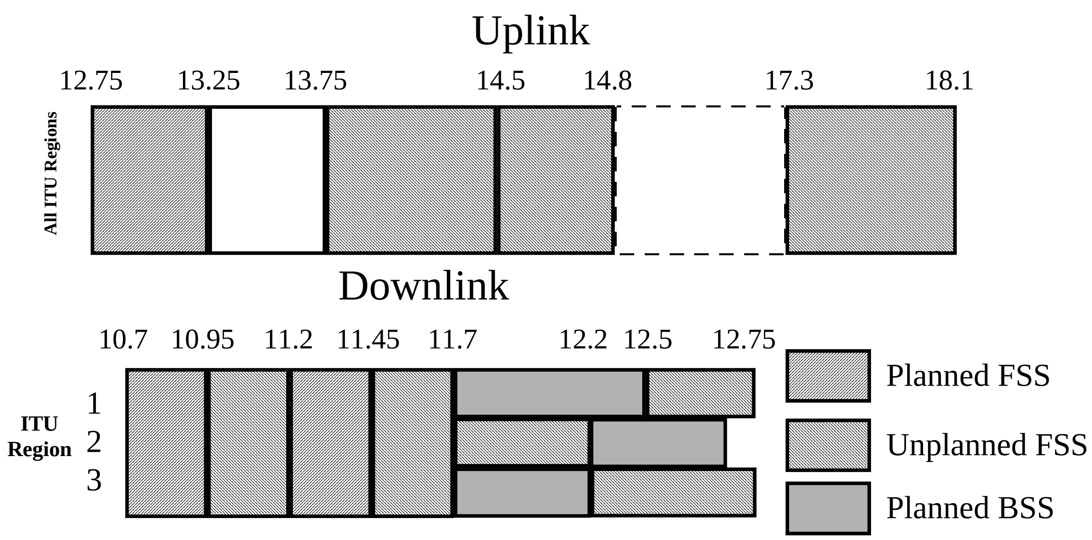


FIGURE 4-28 SYSTEM ALLOCATIONS OF THE PROPOSED SKYBRIDGE FREQUENCY BANDS

4.3.4. Channel Allocation Techniques

Channel allocation techniques have to fullfill the following constraint : two different cells on the earth may reuse the same channel provided that they are at a suitable distance, called reuse distance D which allows tolerable levels for the cochannel interference. As stated before, we consider FCA and DCA techniques.

4.3.4.1.Fixed Channel Allocation

With the FCA technique, a set of channels is permanently assigned to each cell, according to the allowed reuse distance D. A call can only be served by an available channel (if any) belonging to the set of the cell. If an arriving call does not find any free nominal channel in its cell, the call is blocked and lost.

For uniform traffic conditions, the full set of M channels is divided into equal groups each composed by S channels

$$S=M/K \text{ where } K=D^2/3R^2. \quad (4.2)$$

K is the number of cells that form the FCA cluster. The spatial repetition of this cluster assures the territory coverage as in a mosaic.

The use of FCA in a situation with nonuniform traffic requires complex network planning in order to assign more capacity in the cells where a higher traffic is expected. In the LEO context under consideration, such planning is difficult because the traffic offered to a given cell is unpredictable due to the fast satellite motion with respect to the earth. This is the reason why a more suitable solution for LEO-MSS's is given by DCA approach, as described below.

4.3.4.2.Dynamic Channel Allocation

A DCA strategy allows that any system channel can be temporarily assigned to any cell, provided that the constraint on the reuse distance D is fulfilled. Let x be the cell of the arrival, I(x) the set of interfering cells with x (i.e. those cells that lie at a distance less than D from x), and Λ(x) the set of available channels in x (i. e., those channels that are not used either in x or in cells belonging to I(x)).

The DCA technique considered here selects the channel i^* to be allocated in the cell of the call arrival x according to the following minimum cost criterion :

$$C_x(i^*) = \min_{i \in \Lambda(x)} \{ C_x(i) \}. \qquad (4.3)$$

The cost function $C_x(i)$ has been defined as follows :

$$C_x(i) = q_x(i) + \sum_{k \in I(x)} \{C_x(k, i)\}. \qquad \forall i \in \Lambda(x) \qquad (4.4)$$

where the cost contribution for channel $i \in \Lambda(x)$, due to the interfering cell $k \in I(x)$, $C_x(k, i)$, is given by

$$C_x(k, i) = v_k(i) + 2\,(1 - q_k(i)\}. \ \forall k \in I(x) \qquad (4.5) \text{ and}$$

$v_k(i) = \{1$, if $i \in \Lambda(k)$, 0 otherwise$\}$

$q_k(i) = \{0$, if $i \in F_D(k)$, 1 otherwise.

This DCA technique selects (whenever possible) channels belonging to $F_D(x)$, i. e., the set of channels that are assigned to x by FCA. Otherwise, the DCA strategy selects in $\Lambda(x)$ the channel that becomes locked in the minimum number of interfering cells of x. finally, if $\Lambda(x)=0$, the call is blocked.

In order to enhance the DCA performance, whenever a call termination occurs in a cell x (due to either the physical end of call or a handover), a channel is released in x according to a deallocation criterion with a cost-function complementary to that used in the allocation phase. The deallocation cost-function selects (to be freed in x) the channel j* that becomes available in the greatest number of interfering cells and, possibly, a channel that

4.4. ROUTING METHODS WITHIN THE CONSTELLATION BETWEEN SATELLITES AND GATEWAYS.

The routing methods within constellation depends on the constellation design. The topology of a LEO-based network is dynamic. The network connectivity between any two points is dynamic. The satellite nodes move with time above a rotating earth. Each satellite keeps the same position relative to other satellites in its orbital plane. Its position and propagation delay relative to earth terminals and to satellites in other planes change continuously and predictably. In addition to changes in network topology, as traffic flows through the network, queues of packets accumulate in the satellites, changing the waiting time before transmission to the next satellite. All of these factors affect the packet routing choice made by the fast packet switch in each satellite.

The maximum delay between two terrestrial end points, including the hops across satellite is constrained by real-time echo delays. This constraints limit the hop count in systems utilising inter-satellite links.

Satellite failure can create islands of communication within the LEO network. The network routing algorithm must accommodate these failure modes.

Due to the satellite orbital dynamics and the changing delays, most LEO systems are expected to use some form of adaptive routing to provide end-to-end connectivity. Adaptive routing inherently introduces complexity and delay variation. In addition, adaptive routing may result in packet reordering. These out of order packets will have to be buffered at the edge of the network resulting in further delay and jitter.

Teledesic's network architecture is built around a custom combined routing and MAC protocol, intended to be transparent to the outside world and protocols travelling over it. Routing is performed by all satellites reporting information to constellation command and control, which then passes information back to the satellites. Routing information distributed focuses primarily on delay metrics, where even expected queuing delay within a satellite is expressed as a delay metric, in addition to the intersatellite link delay states. This allows satellites to forward packets based upon minimum-delay computations. This is described in [PAT98]. Although this patent was filed in 1994 when Teledesic was planning a large 840-active-satellite constellation, before Boeing and Motorola involvement, it is still reasonable to assume that this work still applies to the current Teledesic design.

Teledesic's custom packets are described as 512-bit, with 80-bit headers and a 16-bit cyclic redundancy check (CRC) on both header and data. This leaves a payload of 416 bits - just large enough to contain a single ATM cell with redundant header protection removed. Teledesic's approach differs from the ATM approach of validating the header, but not the payload.

Teledesic have indicated that they have performed simulations demonstrating that TCP/IP performance is not adversely impacted by the delay variations inherent in switching across an orbital mesh. This appears due to the simulated delay variations presented being less than the granularity of TCP's retransmission timeout (RTO), so that retransmissions are not triggered [CHAN99].

The Teledesic network uses a "connectionless" protocol. Packets of the same connection may follow different paths through the network. Each node independently routes the packet along the path that currently offers the least expected delay to its destination. The required packets are buffered, and if necessary put in sequence, at the destination terminal to eliminate the effect of timing variations.

Recently, intense research work has been undertaken on the definition of adaptive routing in many constellations. [WERN97] propose an ATM-based concept for routing of information in (LEO/MEO) satellite system including inter-satellite links based on use of a modified Disjkstra Shortest Path Algorithm. The focus of the research results presented is on traffic routing aspects.

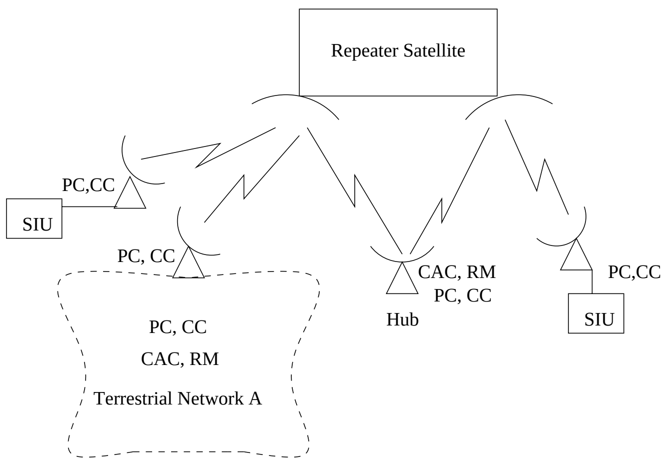
Routing deals with providing paths for the exchange of information between two users on earth. This entails identifying a start and end satellite, and connecting them via a time-variant ISL infrastructure. Due to the motion of the satellite both the start and end ones may change with time and also the ISL path. Indeed, since ATM switching implies low delay at each satellite node on the ISL route, the advantage gained from low propagation delay on the LEO/MEO up- and downlink can be retained.

5. LEO SATELLITE CHARACTERISTICS

Here we can consider several different types of satellites that lead to considerably different satellite networks. There are traditional ‘repeater’ satellites, where signals are amplified and frequency-shifted. There are satellites where on-board switching is done, with varying degrees of complexity, before sending to ground terminals. There are satellites where on-board routing is performed to route to other satellites via inter-satellite links, or to ground terminals.

5.1. TRANSPARENT ‘BENT-PIPE’ SATELLITE CHARACTERISTICS

In this simple satellite model without on-board processing or switching, shown in Figure 5-1, no on-board buffering is required. The repeater satellite is only introducing a small processing delay and can be modelled by a delay module.



PC:Priority Control CC:Congestion Control
CAC:Connection Admission Control SIU:Satellite network Interface Unit
RM:Resource Management

FIGURE 5-1 TRANSPARENT BENT-PIPE SATELLITE ARCHITECTURE

Note that the complexity is put on the ground segment (hub and gateway) if transparent ‘bent pipe’ satellites are used for the constellation.

5.2. ON-BOARD-SWITCHING SATELLITE CHARACTERISTICS

This constellation include cell/packet switching capabilities in order for fixed/portable terminals to access ATM B-ISDN directly.

This architecture is able to multiplex and switch all the traffic on-board the satellite reducing the terminal complexity and cost.

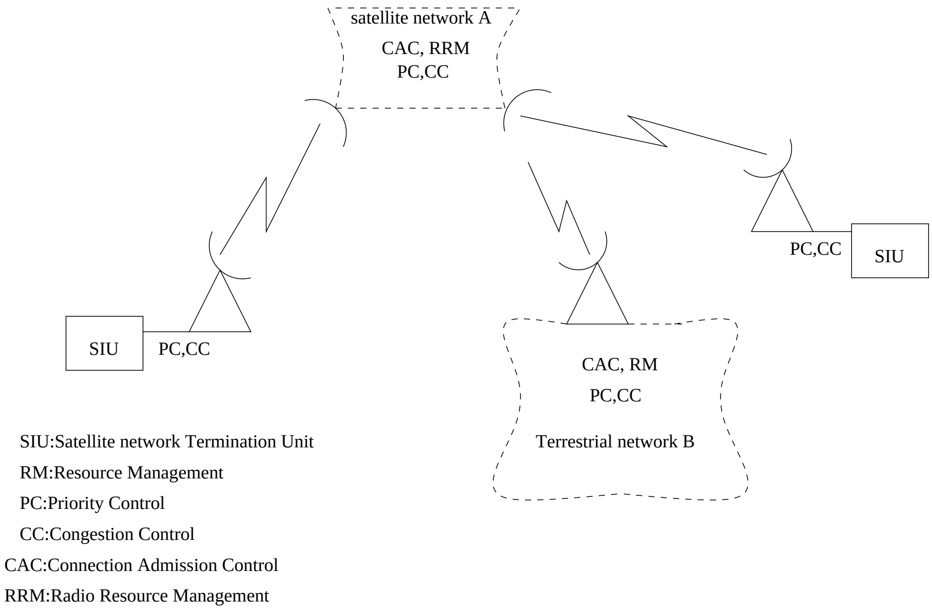


FIGURE 5-2 ON-BOARD SWITCHING SATELLITE ARCHITECTURE

On-board switching on a satellite is needed to ensure that an incoming message on an input channel is switched to an appropriate output channel. There are three different traffic sources which compete for resources on the satellite. Traffic from a mobile or stationary user. Input from a gateway containing multiple message which may be multiplexed into a single carrier and input from another satellite through an inter-satellite link. The signal received has to be demultiplexed to the individual packets/cells and switched to the appropriate output queues. Increasing the area covered by a single satellite increase the probability that the origin and destination points are covered by the same satellite, thus reducing the need for intersatellite communications. Due to the added complexity and power requirements of on-board switching, some of the proposed LEO and MEO systems don’t use on-board switching altogether and all the switching is done on the ground (such as SkyBridge). The speed of on-board switching is of critical importance in LEO systems that use intersatellite links or multibeam antennas..

About 50% of the planned Ka-band satellite networks propose to use on-board ATM-like fast packet switching. In a simple satellite model without on-board processing or switching, no on-board buffering is required. However, if on-board processing is performed, then on-board buffering is needed. On-board processing can be used for resource allocation and media access control (MAC). Demand Assignment Multiple Access (DAMA) can be used with any of the access scheme options (TDMA, FDMA, and CDMA). If on-board processing is not performed, DAMA must be done by the NCC. On-board DAMA decreases the response time of the media access policy by half because link access requests need not travel to the NCC on the ground any more. In addition to media access control, ABR explicit rate allocation or EFCI control, and UBR/GFR buffer management can also be performed on-board the satellite. On-board switching may be used for efficient use of the network by implementing adaptive routing/switching algorithms. Trade-offs must be made with respect to the complexity, power and weight requirements for providing on-board buffering, switching and processing features to the satellite network. In addition, on-board buffering and switching will introduce some additional delays within the space segment of the network. For fast packet or cell switched satellite networks, the switching delay is negligible compared to the propagation delay, but the buffering delay can be significant. Buffering also results in delay variations due to the bursty nature of traffic.

5.2.1. Intersatellite link MAC and LLC configuration

Future broadband LEO satellite communication systems will increasingly rely on an intersatellite link (ISL) trunk network with time-variant topology. Cross-link capacity of the orbiting satellites are a function of their position with respect to one another. For example, satellites in polar orbits must reduce their output power as they become closer nearer the northern and southern latitudes. Thus the satellite network cannot be considered uniformly equal in potential capacity flow. Routing algorithms, failure recovery reconfigurations, and flow control are all affected by this constraint.

5.2.2. Intersatellite Link Characteristics

Earlier studies on the routing in ISL networks of polar or near-polar constellations have clearly identified the seam between contra-rotating orbits and the on/off-switching of inter-plane ISLs as two fundamental drawbacks for connection-oriented operation. An inclined delta or rosette constellation avoides the orbital seam, but still has inter-plane intersatellite links breaking and reforming at highest latitudes.

5.2.3. Routing Characteristics

In LEO constellations context, the routing is usually split into Up-Down-Link (UDL) routing and ISL routing. Uplink (UL) routing is the process by which the source ground station selects the source satellite used to forward the packets of the connection, while Downlink (DL) routing is the process by which the destination ground station selects the destination satellite from which the packets of the connection will arrive.

The criteria used for UDL routing is the availability of resources in the satellite and in the ground station, the minimisation of handover rate on the UDL, and the quality of the communication between the ground station and the satellite.

Given a source satellite and a destination satellite, as provided by UDL routing, ISL routing computes the (or at least one) optimal path between these two satellites. The criteria used are : ressource availability in the satellites and ISLs, minimisation of the handover rate, quality of the communication among satellites, and length of the path.

5.2.4. Intersatellite Link Delay

The inter-satellite link delay (ti) is the sum of the propagation delays of the inter-satellite links traversed by the connection. Inter-satellite links (crosslinks) may be in-plane or cross-plane links. In In-plane links connect satellites within the same orbit planes. In GEO systems, ISL delays can be assumed to be constant over a connection's lifetime because GEO satellites are almost stationary over a given point on the earth, and with respect to an other. In contrary, in LEO constellations, the ISL delays depend on the orbital radius, the number of satellites-per-orbit, and the inter-orbital distance (or the number of orbits). Also, the interplane ISL delays change over the life of a connection due to satellite movement and the overall delay changes due to adaptive routing techniques in LEOs. As a result, LEO systems can exhibit a high variation in ISL delay.

$$t_i = \frac{\sum ISL\_Length}{Speed\_of\_Signal} \quad (5.1)$$

6. SATELLITE NETWORK SCENARIOS

Two main LEO constellations can be identified for broadband network access:

- Intelligence and routing in the ground segment [Skybridge].
- Intelligence and routing on the space segment [Teledesic].
- Both of these scenarios will be described in this section.

6.1. LEO SATELLITE CONSTELLATION USING REPEATER SATELLITES

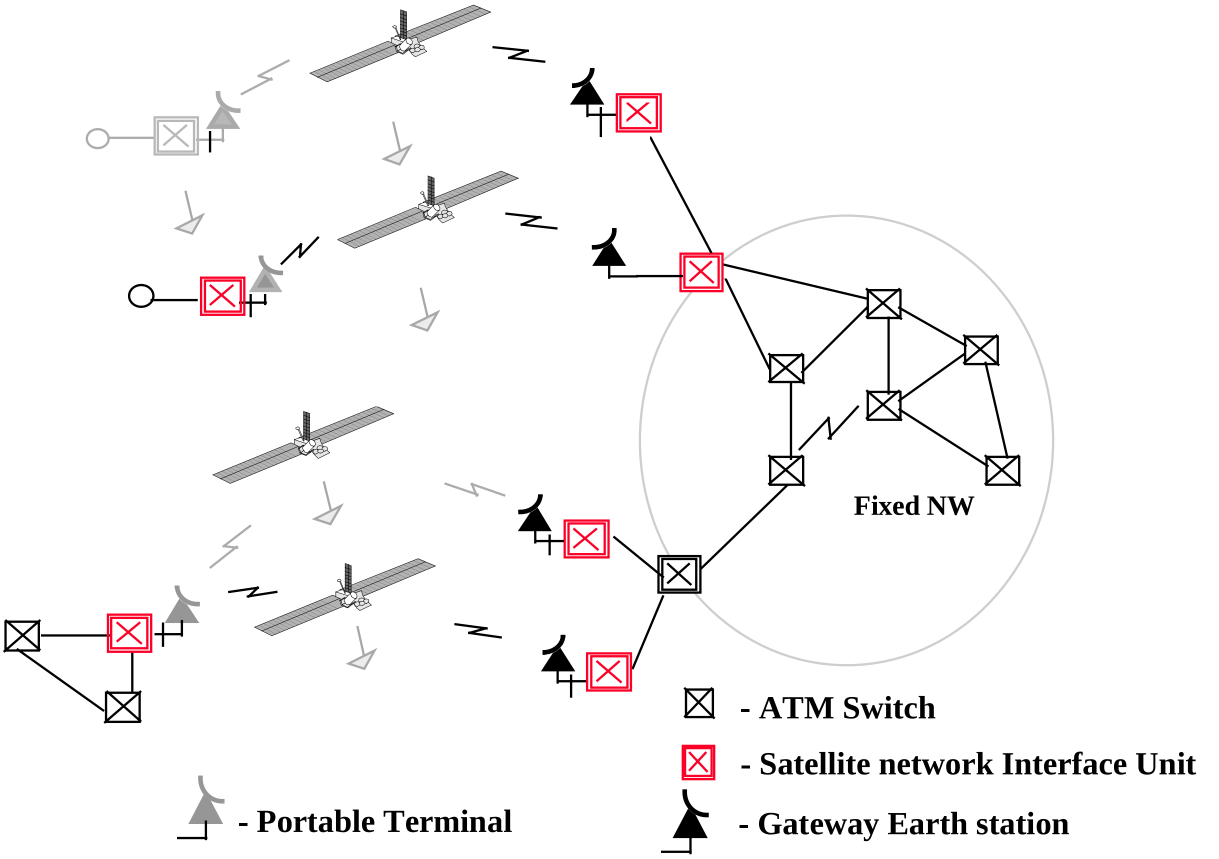


FIGURE 6-1 LEO SATELLITE NETWORK USING REPEATER SATELLITES

In this scenario terminals can only communicate if there is a gateway earth station in the area. The following two cases are shown in Figure 6-1: (i) case one where the user terminal is portable requiring mobility management functions to support its migration from one satellite beam coverage to another and (ii) case two where the user terminal can be fixed, however the satellite is mobile requiring mobility management functions to support handover of the connection from one satellite or beam to another satellite or beam.

6.2. LEO SATELLITE CONSTELLATION USING OBS SATELLITES

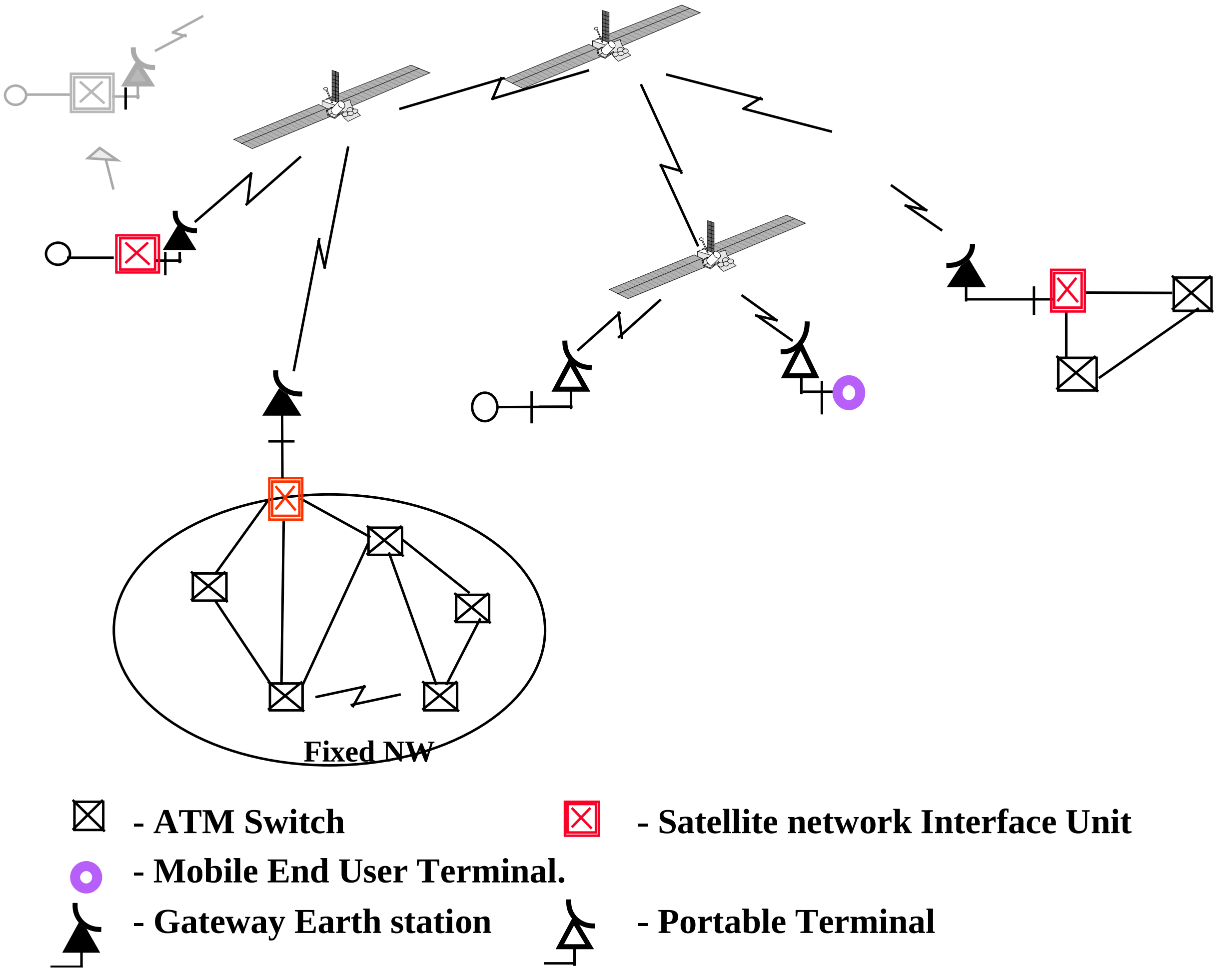


FIGURE 6-2 LEO SATELLITE NETWORK USING ON-BOARD SWITCHING SATELLITES

In this scenario, several satellites form network in space and terminals can communicate as long as they are in the coverage area of the satellite (there is no dependency on the gateway). The satellite networks can be used to interconnect several terrestrial networks and also to allow remote terminals to establish connections to remote stations and hubs. Inter-satellite links are necessary here in order to support connections that span beyond a single satellite hop. Such communications demand routing capability. In LEO constellations the satellite position changes over time and hence rerouting of ongoing connections is necessary.

Table 6-1 summarises some characteristics of the two constellation targets of this paper.

TABLE 6-1 CONSTELLATIONS CHARACTERISTICS

	Skybridge	Teledesic
Services and cost		
Service scheduled to	2002	2003
Backers	Alcatel Space, Aerospatiale, CNES (FR), Loral Space & (US) Communication, Toshiba Mitsubishi Electronic, Sharp (JP) , SPAR Aerospace (CA), Société Régionale d'Investissements de Wallonie (BE)	Bill Gates, Craig Mc Caw Joined by Boeing, Motorola, Matra Marconi Space and Saudi prince Alwaleed Bin Talal (in 1998)
Modulation	QPSK	--
FCC licenced	yes(1997)	yes (Nov 1997)
Data Throughput	(Residential) Uplink : 16 Kbps- 2 Mbps Downlink : 16 Kbps - 20 Mbps (Professional) Uplink : 10 Mbps Downlink : 100 Mbps	Portable (E1) 16 kbps + 2 kbps signalling 2048 Mbps (128 basic channels) Fixed (OC) OC-3 155 Mbps OC-24 1.2 Gbps
Supported service (Protocol / Routing / Packet format)	ATM Switching on the ground/ transparent satellite	Proprietary packet/Connectionless/ adaptive/ 512 bits packet
Downlink Throughput	20 Mbps-nx20 Mbps	16 Kbps-64 Mbps
Down link Granularity	16 Kbps	16 Kbps
Uplink Throughput	2 Mbps-nx2 Mbps	16 Kbps-2 Mbps
Uplink Granularity	16 Kbps	16 Kbps
Orbits and Geometry		
Orbit Class	LEO	LEO
Altitude (km)	1469	1375
Number of satellites	80 active	288 active 3 in orbit spares
Constellation type	Delta	Near-polar Star
Number of planes	10x2	12

Satellite/ Plane	(4/10)x2	24/12
Inclination (°)	53	84.7
Inter-satellite links	0	8
Minimum Terminal elevation angle (°)	10	40
Min. link one-way propagation delay (ms)	4.9	4.6
Number of earth stations	200	
Coverage (%)	± 68°	100 % Earth surface 100% population
Coverage concept	EFC	EFC
Availability (%)	99.9	99.9
Beam and re-use characteristics		
Multiple access method	TDMA/ FDMA/ CDMA	TDMA / SDMA / FDMA / ATDMA
Beams per satellite	18 Spot beams	64 beams (supercells) 576 cells (beam data for 840-sat design)
Total number of beams	1440	18 432
Beam diameter (km)	700	200 (cell diameter)
Footprint diameter (km)	6000	700
Satellite antenna	steerable, earth-fixed cells	steerable, earth-fixed cells
Dual satellite visibility	equal to or more than two satellites most of the time	equal to or more than two satellites most of the time
Dual or higher satellite path diversity exploited ?	-	no (gateways only)
Frequencies and miscellaneous		
Downlink frequencies (MHz)	Ku-band [3] 10-18 GHz	Ka-band[4]
Uplink frequencies (MHz)	Ku-band [5]	Ka-band[6]
Feeder downlink frequencies (GHz)	Ku-band	Ka-band
Feeder uplink frequencies (GHz)	Ku-band	Ka-band
On-board switching	no	yes

[3] downlink 10.7 – 12.75 GHz (Skybridge)
[4] downlink 28.6 – 29.1 GHz (Teledesic)
[5] uplink 12.75 – 14.5 GHz (Skybridge)
[6] uplink 18.8 – 19.3 GHz (Teledesic)

Handover performed ?	yes	yes
Antenna size (cm)	50 (Residential) 80 (Professional)	8 up to 180

L-band is 0.5 to 1.5GHz, C-band is 4 to 8 GHz, Ku-band is 10.9 to 17 GHz, Ka-band is 18 to 31 GHz

7. PERFORMANCE ISSUES OF TCP/IP OVER LEO SATELLITES

Over the past decade there have been widely divergent reports (both formal and informal) about how well TCP/IP performs over satellite links. Some reports indicate TCP/IP throughput is poor, and others report that TCP/IP throughput is quite good [PART97]. This section discusses TCP/IP performance analytically, indicating what LEO satellite features impact TCP/IP performance.

7.1. TCP THROUGHPUT ISSUES

The Transmission Control Protocol (TCP) was designed to provide a reliable end-to-end byte stream over an unreliable internetwork. An internetwork differ from a single network because different parts may have different underlying protocols, topologies, bandwidth, delays, packet sizes. TCP is initially formally defined in RFC 793 and updated in RFC1122 and extensions are given in RFC 1323.

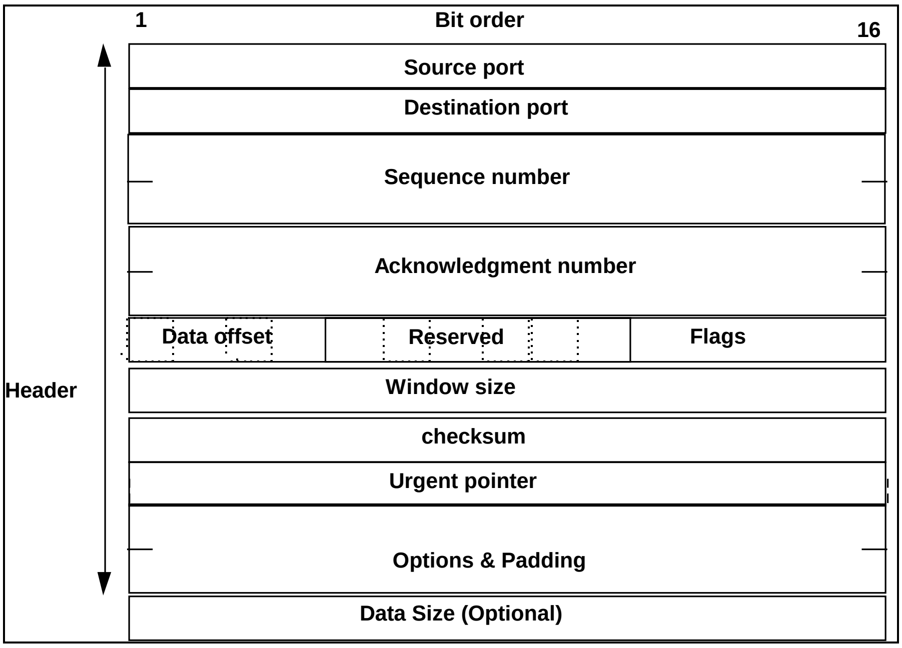


FIGURE 7-1. TCP HEADER

TCP is a byte stream, not a message stream, and message boundaries are not preserved end-to-end. All TCP connections are duplex connections and point-to-point. As such TCP does not support multicasting or broadcasting.

The sending and receiving TCP entities exchange data in the form of segments. A segment consists of a fixed 20-byte header (plus an optional part) followed by zero or more data bytes. Two limits restrict the TCP segment size:

- Each segment must fit into 65,535 byte IP payload (RFC 2147 describes adapting TCP and UDP to use IP6 that supports datagrams larger than 65,535 bytes long).
- Each subnetwork has a Maximum Transfer Unit (MTU). The segment should fit into the minimum path MTU to avoid local explicit fragmentation.

In practice, the MTU is a few thousands of bytes and thus defines the upper boundary of the segment size. The TCP segment is shown in Figure 7-1.

The TCP header fields are as follows:

- Source and destination ports: These define the local end points of the connection. They are used in conjunction with the IP source and destination addresses (in the IP packet header) to provide an identification of the connection.
- Sequence and acknowledgement numbers perform the usual functions and they relate to data bytes (not messages) in the TCP stream. The acknowledgements follow the sliding-window protocol principles (where the receiver indicates the next expected sequence number).
- Header length (Data offset) and Options: The header length indicates the length of the TCP header, where options field can be of varying length, and used to specify parameters such as the maximum segment size that a host can accept.
- Flag bits: These contain control functions such as :
    1. URG - urgent pointer field valid
    2. ACK - acknowledgment field valid
    3. PSH - deliver data at receipt of this segment
    4. RST - reset sequence and acknowledgment numbers
    5. SYN - sequence number valid
    6. FIN - end of byte stream from sender
- Checksum: This provides a degree of reliability and detects misdelivered packets. The checksum scope includes the TCP segment length and IP source and destination addresses. (This violates strict protocol hierarchy since the IP addresses belong to the IP layer)
- Window size: This fields provide flow control, where slow receivers can control the transmission rate of the senders. It follows the sliding-window principles. The window size indicates the amount of data that the sending host can transmit without waiting for acknowledgements from the receiver.

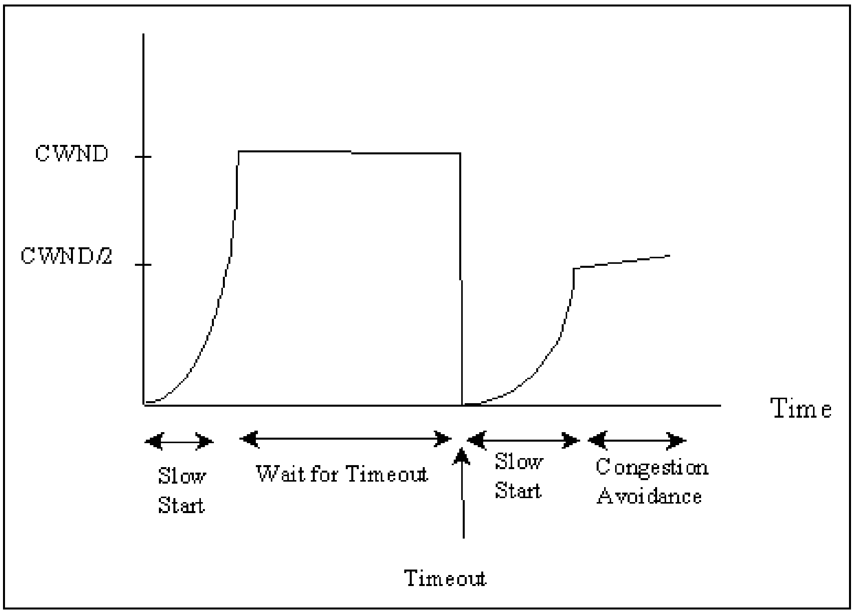


FIGURE 7-2 TCP SLOW START ALGORITHM

As part of implementing the reliable service, TCP is responsible for flow and congestion control: ensuring that data is transmitted at a rate consistent with the capacities of both the receiver and the intermediate links in the network path. Figure 7-2 shows the slow start algorithm with the slow start (with exponential increase in congestion window size) and congestion avoidance (with linear increase in congestion window size) phases.

Since there may be multiple TCP connections active in a link, TCP's congestion control is also responsible for ensuring that a link's capacity is responsibly shared among the connections using it. As a result, most throughput issues are rooted in TCP.

7.1.1. Impact of satellite channel characteristics on TCP

Satellite channels have several characteristics that differ from most terrestrial channels. These characteristics may degrade the performance of TCP. These characteristics include:

- Long feedback loop: Due to the propagation delay of some satellite channels it may take a long time for a TCP sender to determine whether or not a packet has been successfully received at the final destination. This delay affects interactive applications such as telnet, as well as some of the TCP congestion control algorithms.
- Large bandwidth-delay product: The bandwidth-delay product (BDP) defines the amount of data that a protocol should have "in flight" (data that has been transmitted, but not yet acknowledged) at any one time to fully utilize the available channel capacity. The delay used in this equation is the RTT (end-to-end) and the bandwidth is the capacity of the bottleneck link in the network path. Because the delay in some satellite environments is large, TCP will need to keep a large number of packets "in flight" (that is, sent but not yet acknowledged) .
- Transmission errors: Satellite channels exhibit a higher bit-error rate (BER) than typical terrestrial networks. TCP uses all packet drops as signals of network congestion and reduces its window size in an attempt to alleviate the congestion. In the absence of knowledge about why a packet was dropped (congestion at the network or corruption at the receiver), TCP must assume the drop was due to network congestion to avoid congestion collapse. Therefore, packets dropped due to corruption cause TCP to reduce the size of its sliding window, even though these packet drops do not signal congestion in the network.
- Asymmetric use: Due to the expense of the equipment used to send data to satellites, asymmetric satellite networks are often constructed. A common situation is that uplink has less available capacity than the downlink due to the expense of the transmitter required to provide a high bandwidth back channel. This asymmetry may have an impact on TCP performance.
- Variable Round Trip Times: In LEO constellations, the propagation delay to and from the satellite varies over time.
- Intermittent connectivity: In non-GSO satellite orbit configurations, TCP connections must be transferred from one satellite to another or from one ground station to another from time to time. This handoff may cause packet loss if not properly performed.

7.1.2. Congestion Control

To avoid generating an inappropriate amount of network traffic for the current network conditions, during a connection TCP employs four congestion control mechanisms . These algorithms are:

- Slow start.
- Congestion avoidance.
- Fast retransmit.
- Fast recovery.

These algorithms are described in detail in RFC 2581. They are used to adjust the amount of unacknowledged data that can be injected into the network and to retransmit segments dropped by the network.

TCP senders use two state variables to accomplish congestion control. The first variable is the congestion window (cwnd). This is an upper bound on the amount of data the sender can inject into the network before receiving an acknowledgment (ACK). The value of cwnd is limited to the receiver's advertised window. The congestion window is increased or decreased during the transfer based on the inferred amount of congestion present in the network. The second variable is the slow start threshold (ssthresh). This variable determines which algorithm is used to increase the value of cwnd. If cwnd is less than ssthresh the slow start algorithm is used to increase the value of cwnd. However, if cwnd is greater than or equal to (or just greater than in some TCP implementations) ssthresh the congestion avoidance algorithm is used. The initial value of ssthresh is the receiver's advertised window size. Further more, the value of ssthresh is set when congestion is detected.

The above algorithms have a negative impact on the performance of individual TCP connection's performance because the algorithms slowly probe the network for additional capacity, which in turn wastes bandwidth. This is especially true over long-delay satellite channels because of the large amount of time required for the sender to obtain feedback from the receiver. However, the algorithms are necessary to prevent congestive collapse in a shared network. Therefore, the negative impact on a given connection is more than offset by the benefit to the entire network.

7.1.2.1.Large window sizes

The standard maximum TCP window size (65,535 bytes) is not adequate to allow a single TCP connection to utilize the entire bandwidth available on some satellite channels. TCP throughput is limited by the following formula:

throughput = window size / RTT

Therefore, using the maximum window size of 65,535 bytes and a geosynchronous satellite channel RTT of 560 ms [Kru95] the maximum throughput is limited to:

throughput = 65,535 bytes / 560 ms = 117,027 bytes/second

However, TCP has been extended to support larger windows [RFC1323]. The window scaling options can be used in satellite environments, as well as the companion algorithms PAWS (Protection Against Wrapped Sequence space) and RTTM (Round-Trip Time Measurements).

	1.5 Mb/s			45 Mb/s			155 Mb/s		
	LAN	LEO	GEO	LAN	LEO	GEO	LAN	LEO	GEO
Requires PAWS	No	No	No	No	No	No	Yes	Yes	Yes
Requires large windows	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Slow start yime	0.01s	1.8s	5.6s	0.2s	3.5s	9.8s	1.9s	4.1s	11.3s
Slow start data (in bytes)	1,760	76,600	197,870	115,900	2,405,000	6,003,000	4,123,814	8,292,000	20,650,000

TABLE 7-1 COMPARISON OF TCP OVER LAN, GEO AND LEO SATELLITE INTERACTIONS

Table 7-1 illustrates for a range of common bandwidths, when the TCP enhancements of PAWS and large windows are required to fully utilize the bandwidth on a LEO link (100 ms one-way), for a range of link speeds. The table indicates how long slow start takes to get to full link speed, assuming 1 KB datagrams (a typical size) are transmitted .

7.1.2.2.Selective Acknowledgments

Selective acknowledgments (SACKs) [RFC 2018] allow TCP receivers to inform TCP senders exactly which packets have arrived. SACKs allow TCP to recover more quickly from lost segments, as well as avoiding needless retransmissions. When SACK is employed, the sender is generally able to determine which segments need to be retransmitted in the first RTT following loss detection. This allows the sender to continue to transmit segments (retransmissions and new segments, if appropriate) at an appropriate rate and therefore sustain the ACK clock.

7.2. IP THROUGHPUT ISSUES

IP (the Internet Protocol) is the network layer protocol in the TCP/IP protocol suite. An IP datagram consists of a header part and payload part. The header has a 20-byte fixed part and a variable length optional part. Figure 7-3 shows the IP datagram header format, which consists of the following fields:

- Version: Protocol version such as IPv4 or IPv6.
- Header Length: The length of IP header.
- Type of service: Various combinations of speed and reliability are possible. In IPv4, this field is rarely used in practice except in some intranets. RFC 2475 defines the use of this field to provide Internet differentiated services (diffserv)
- Total Length: Header plus user data (up to 65kB maximum).
- Identification: To determine the datagram that fragments belong to.
- DF: Don’t fragment, to tell routers to not fragment datagram.
- MF: More fragments, all fragments except the last one have this bit set to 1.
- Fragment offset: Where the fragment belongs in the current datagram.
- Time to Live (TTL): A counter used to limit the packet lifetime.
- Protocol: Transport protocol type such as TCP or UDP.
- Header checksum: Verifies the header only.

- Source and destination addresses: IP addresses (see section 4.2 for more information).
- Options: Additional services such as security, source routing, recording routes and time stamping.

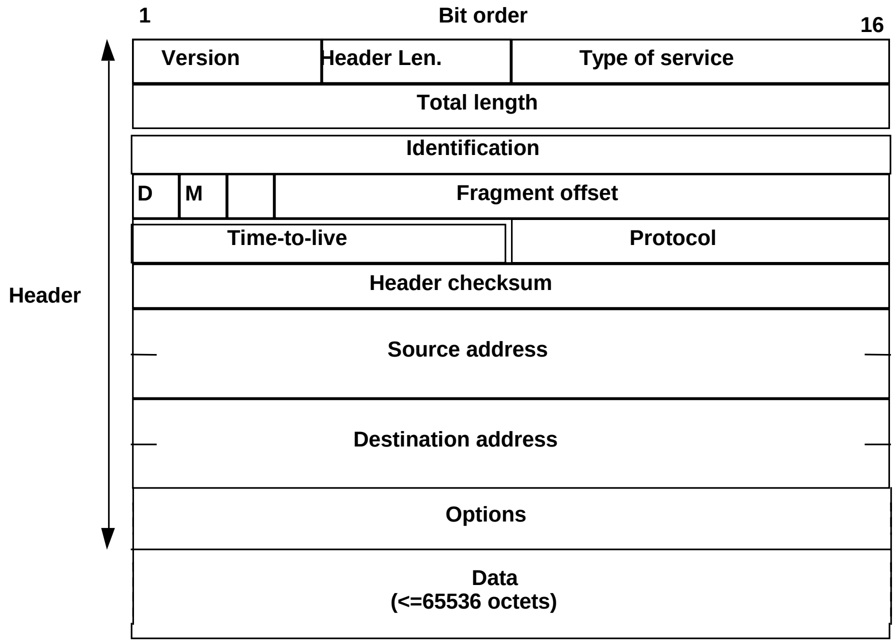


FIGURE 7-3. IP PACKET HEADER

IP's function is to provide a protocol to integrate heterogeneous networks together. In brief, a medium-specific way to encapsulate IP datagrams is defined for each media (e.g., satellite, Ethernet, or Asynchronous Transfer Mode). Devices called routers move IP datagrams between the different media and their encapsulations. Routers pass IP datagrams between different media according to routing information in the IP datagram. This mesh of different media interconnected by routers forms an IP internet, in which all hosts on the integrated mesh can communicate with each other using IP.

The actual service IP implements is unreliable datagram delivery. IP simply promises to make a reasonable effort to deliver every datagram to its destination. However IP is free to occasionally lose datagrams, deliver datagrams with errors in them, and duplicate and reorder datagrams. Because IP provides such a simple service, one might assume that IP places no limits on throughput. Broadly speaking, this assumption is correct. IP places no constraints on how fast a system can generate or receive datagrams. A system transmits IP datagrams as fast as it can generate them.

However, IP does have two features that can affect throughput: the IP Time to Live and IP Fragmentation.

7.2.1. IP Time To Live

In certain situations, IP datagrams may loop among a set of routers. These loops are sometimes transient (a datagram may loop for a while and then proceed to its destination) or long-lived. To protect against datagrams circulating semipermanently, IP places a limit on how long a datagram may live in the network. The limit is imposed by a Time To Live (TTL) field in the IP datagram. The field is decremented at least once per hop. When the TTL reaches zero, the datagram is discarded.

Originally, the IP specification also required that the TTL also be decremented at least once per second. Since the TTL field is 8-bits wide, this means a datagram could live for approximately 4.25 minutes. In practice, the injunction to decrement the TTL once a second is ignored, but, perversely, specifications for higher layer protocols like TCP usually assume that the maximum time a datagram can live in the network is only a couple of minutes.

The significance of the maximum datagram lifetime is that it means higher layer protocols must be careful not to send two similar datagrams (in particular, two datagrams which could be confused for each other) within a few minutes of each other. This limitation is particularly important for sequence numbers (hence PAWS). If a higher layer protocol numbers its datagrams, it must ensure that it does not generate two datagrams with the same sequence number within a few minutes of each other, lest IP deliver the second datagram first and confuse the receiver. We discuss this issue more in the next section when we discuss TCP sequence space issues.

7.2.2. IP Fragmentation

Different network media have different limits on the maximum datagram size. This limit is typically referred to as the Maximum Transmission Unit (MTU). When a router is moving a datagram from one media to another, it may discover that the datagram, which was of legal size on the inbound media, is too big for the outbound media. To get around this problem, IP supports fragmentation and reassembly, in which a router can break the datagram up into smaller datagrams to fit on the outbound media. The smaller datagrams are reassembled into the original larger datagram at the destination (not the intermediate hops).

Fragments are identified using a fragment offset field (which indicates the offset of the fragment from the start of the original datagram). Datagrams are uniquely identified by their source, destination, higher layer protocol type, and a 16-bit IP identifier (which must be unique when combined with the source, destination and protocol type). Observe that there's a clear link between the TTL field and the IP identifier. An IP source must ensure that it does not send two datagrams with the same IP identifier to the same destination, using the same protocol within a maximum datagram lifetime, or fragments of two different datagrams may be incorrectly combined.

Since the IP identifier is only 16 bits, if the maximum datagram lifetime is two minutes, we are limited to a transmission rate of only 546 datagrams per second. That's clearly not fast enough. The maximum IP datagram size is 64 KB, so 546 datagrams is, at best, a bit less than 300 Mb/s. The problem of worrying about IP identifier consumption has largely been solved by the development of Path MTU Discovery: a technique for IP sources to discover the MTU of the path to a destination .

Path MTU Discovery is a mechanism that allows hosts to determine the MTU of a path reliably. The existence of MTU discovery allows hosts to set the Don't Fragment (DF) bit in the IP header, to prohibit fragmentation, because the hosts will first learn through MTU discovery whether their datagrams are too big. Routers use an Internet Control Message Protocol (ICMP, RFC 792) stating ‘unreachable error’ if these routers are asked to forward IP datagrams with the DF bit set when the MTU size is less than the datagram size. Therefore path MTU discovery can be a painful recursive process.

7.3. PERFORMANCE ANALYSIS OF TCP/IP APPLICATIONS

A series of experiments were performed to analyse the performance of various satellite network configurations. The most experiments fell into two categories. The first set of experiments looked at performance of protocols using TCP over satellite links with bandwidths of 9.6 kbps, 64 kbps and 1Mbps. The second set concentrated on the performance of limited subset of application using a reduced bandwidth return link.

TCP applications can be categorized into three main classes:

- Bulk transfer applications such as FTP, RPC, and SMTP.
- Semi interactive applications such as http, rsh and NNTP
- Interactive applications such as telnet, rlogin and talk.

In order to fully evaluate the performance of TCP applications over satellite links it was necessary to consider the performance of protocols in each of these classes.

7.3.1. First set

The first set of experiments examines the performance of FTP, http, and Telnet protocols using TCP over satellite links with bandwidths of 9.6, 64 kbps and 1 Mbps.

7.3.1.1.Bulk Transfer Protocols

The File Transfer Protocol (FTP) can be found on many TCP/IP installed systems and provides an example of the most commonly executed bulk transfer protocol. FTP allows the user to log onto a remote machine and download files either from, or to the machine.

At bandwidths of 64 kbps and 9.6 kbps, throughput was proportional to the bandwidth available and delay had little effect on the performance. This was due to the 24KB window size, which was large enough to prevent any window exhaustion. However, at a bandwidth of 1Mbps, window exhaustion occurred and the delay had a detrimental effect on the throughput of the system. Link utilisation dropped from 98% at 64 kbps and 9.6% kbps to only 30% for 1Mbps. The throughput was however still higher for the 1Mbps case (due to reduced serialisation delay of the data). All transfers were conducted with 1-MB file, which was large enough to negate the effect of the slow start algorithm. Other bulk transfer protocols, e.g. SMTP and RCP recorded similar performance using typical application file size.

7.3.1.2.Semi-interactive Protocols

Netscape is a web browser that uses the http protocol to view graphical pages downloaded from remote machines. The perceived performance of the http protocol is largely dependent on the structure and content of the html files are being downloaded.

At bandwidths of 1Mbps and 64 kbps the throughput was largely governed by the delay, due to the majority of the session being spent in the open/close and slow start stages of transfer, which are affected by the RTT of the system. At 9.6 kbps this effect was overshadowed by the serialisation delay caused by the limited bandwidth on the outbound link. With bandwidths of 1 Mbps and 64 Kbps the performance was found as expected. At 9.6 kbps the users tended to get frustrated when downloading large files and abandon the session.

7.3.1.3. Interactive Protocols

A telnet session allows the user to log onto a remote system, using his computer as a simple terminal. This provides the way for a poor-performance computer to make use of the resources of a higher-power CPU at a remote site or to access resources not available at a local site.

The telnet sessions were judged subjectively by the user. At 1 Mbps and 64 Kbps bandwidth, users noticed the changes in delay more than the bandwidth, but at 9.6 Kbps the delay due to serialisation was the more noticeable effect and became annoying to the user. The performance of interactive session was greatly dependent on the type of session. Telnet sessions used to view and move directories/files were performed satisfactorily down to 9.6 kbp. Similar performance was observed for other interactive protocols (e.g. rlogin, SNMP, etc).

7.3.2. Second set

The second set of experiments concentrated on the performance of limited subset of application using a reduced bandwidth return link.

7.3.2.1. FTP Performance

At 64 Kbps link capacity the return link could be reduced to 4.8 Kbps with no effect on the throughput of the system. This was due to the limited bandwidth availability for the outbound connection, which experienced congestion. At 2.4 kbps on return link bandwidth, transfer showed a 25% decrease in throughput, resulting from congestion of ACKs in the return link.

At a 1 Mbps outbound link speed, the performance of FTP was affected more by the TCP window size (24-KB) than any variation in the bandwidth of the return link. It was not affected until the return link dropped to 9.6 kbps and started to show congestion. A 15% drop in performance was recorded for the returned of 9.6 kbps. Delay again had a significant effect on the performance at 1 Mbps due to the window exhaustion.

The high ratio of outbound to inbound traffic experienced in the FTP session means that it is well suited for delivery of files across asymmetric links with limited return bandwidth. For a 64 kbps outbound link, FTP delivery will perform well with return links down to 4.8 kbps.

7.3.2.2. http Performance

At 1 Mbps and 64 Kbps, the speed of the return link had a far greater effect than any variation in delay. This was due to congestion in the return link, arising due to the low server/client traffic ratio. The lower ratio was a result of the increased number of TCP connections required to download each object. At 9.6 kbps the return link was close to congestion, but still offered throughputs comparable to that at 64kbps. At 4.8 kbps the return link became congested and the outbound throughput showed a 50% drop off. A further 50% reduction in the outbound throughput occurred when the return link dropped to 2.4 kbps.

For both the 1 Mbps and 64 kbps inbound, the user found that return link speeds down to 19.2 kbps were acceptable. Bellow this rate, users started to become frustrated by the time taken to request a web page. A return bandwidth of at least 19.2 kbps is therefore recommended for web applications.

7.3.2.3. Telnet Performance

During interactive sessions, reducing the bandwidth of the return link increased the serialisation delay of the routers. This was counterbalanced by the fact that most of the datagrams were sent from the remote side consisted of only one or two bytes of the TCP payload and therefore could be serialised relatively quickly. Reducing the bandwidth was noticeable only to the competent typist where the increased data flow from the remote network resulted in increased serialisation and round trip times.

7.3.3. Result

Bulk transfer protocols showed the highest tolerance to any reduction in bandwidth, working acceptably down to bandwidths of 9.6 kbps for 1Mbps outbound and 4.8 kbps for 64 kbps. Return link delay also had little effect on the overall throughput. Most interactive sessions also functioned acceptably and showed a greater sensitivity to reduction in link delay than bandwidth. Semi-interactive sessions, like the web, required a bandwidth of at least 19.2 kbps to provide adequate performance.

Application	Link settings Inbound (outbound)	Delay ms Inbound (outbound)	Comments
FTP	1 Mbps (64, 38.4, 19.2, 9.6 kbps)	280 (0, 100, 280)	OK down to 9.6 kbps
Telnet	1 Mbps (64, 38.4, 19.2, 9.6 kbps)	280 (0, 100, 280)	OK for simple screen edit sessions
Http	1 Mbps (64, 38.4, 19.2, 9.6 kbps)	280 (0, 100, 280)	OK down to 19.2 kbps

TABLE 7-2 SUMMARY OF PERFORMANCE OF TCP/IP PROTOCOLS OVER SATELLITE NETWORKS WITH 1MBPS OUTBOUND.

Application	Link settings Inbound (outbound)	Delay ms Inbound (outbound)	Comments
FTP	64 (9.6, 4.8, 2.4) kbps	280 (0, 100, 280)	OK down to 4.8 kbps
Telnet	64 (9.6, 4.8, 2.4) kbps	280 (0, 100, 280)	OK for simple screen edit sessions
HTTP	64 (9.6, 4.8, 2.4) kbps	280 (0, 100, 280)	OK down to 9.6 kbps

TABLE 7-3 SUMMARY OF PERFORMANCE OF TCP/IP PROTOCOLS OVER SATELLITE NETWORKS WITH 64 KBPS OUTBOUND

7.4. METHODS FOR REDUCING DATA FLOW

User requests on the Internet are often served by a single machine. Very often, and especially when this server exists in a rather distant location, the user experiences reduced throughput and network performance. This low throughput is caused due to bottlenecks that can be either the server itself or one or more congested Internet routing hops. Furthermore, that server represents a single point of failure - if it is down, access to the information is lost.

To preserve the usability of the information distributed in Internet, the following issues need to be addressed at the server level:

- Document retrieval latency times must be decreased.
- Document availability must be increased, perhaps by distributing documents among several servers.
- The amount of data transferred must be reduced - certainly an important issue for anyone paying for network usage.
- Network access must be redistributed to avoid peak hours.
- Improvements in general user-perceived performance.

Of course, these goals must be implemented so that we retain transparency for the user as well as backward compatibility with existing standards. A popular and widely accepted approach to address at least some of these problems is the use of caching proxies.

A user may experience high latency when accessing a server that is attached to the network with limited bandwidth. Caching is a standard solution for this type of problem, and it was applied to Internet (mainly to the web) early for this reason. Caching has been a well known solution to increase the computer performance since the 1960s. The technique is now applied in nearly every computer's architecture. Caching relies on the principle of locality of reference which assumes that the most recently accessed data has the highest probability of being accessed again in the near future. The idea of Internet caching relies on the same principle.

ICP (Internet Caching Protocol) is a well organised, universit- based effort that deals with these issues. ICP is currently implemented into the public domain Squid proxy server. ICP is the protocol used for communication among squid caches. ICP is primarily used within a cache hierarchy to locate specific objects in sibling caches. If a squid cache does not have a requested document, it sends an ICP query to its siblings, and the siblings respond with ICP replies indicating a ‘‘HIT’’ or a ‘‘MISS.’’ The cache then uses the replies to choose from which cache to resolve its own MISS. ICP also supports multiplexed transmission of multiple object streams over a single TCP connection. ICP is currently implemented on top of UDP. Current versions of Squid also support ICP via multicast.

Another way of reducing the overall bandwidth and the latency, thus increasing the user-perceived throughpu, is by using replication. This solution can also provide a more fault-tolerant and evenly-balanced system. Replication offers promise towards solving some of the deficiencies of the proxy caching method.

A recent example of replication was the web information on NASA's latest mission to the Mars planet. In that case the information about the mission was replicated in several sites in US, Europe, Japan and Australia in order to be able to satisfy the million of user requests.

7.4.1. Satellite distribution of cache content

The concept of web caching is quite popular since many Internet Service Providers (ISPs) already use central servers to hold popular web pages, thus avoiding the increased traffic and delays created when thousands of subscribers request and download the same page across the network. Caches can be quite efficient but they have several weak points as they are limited by the number of people who are using each cache.

A solution can be provided by using a satellite system to distribute caches among ISPs. This concept can boost Internet performance, since many already fill multiple T1 (1.5-million-bit-per-second) lines primarily with web traffic. The broadcast satellite link could avoid much of that backhaul, but research is needed for delivering proof of this.

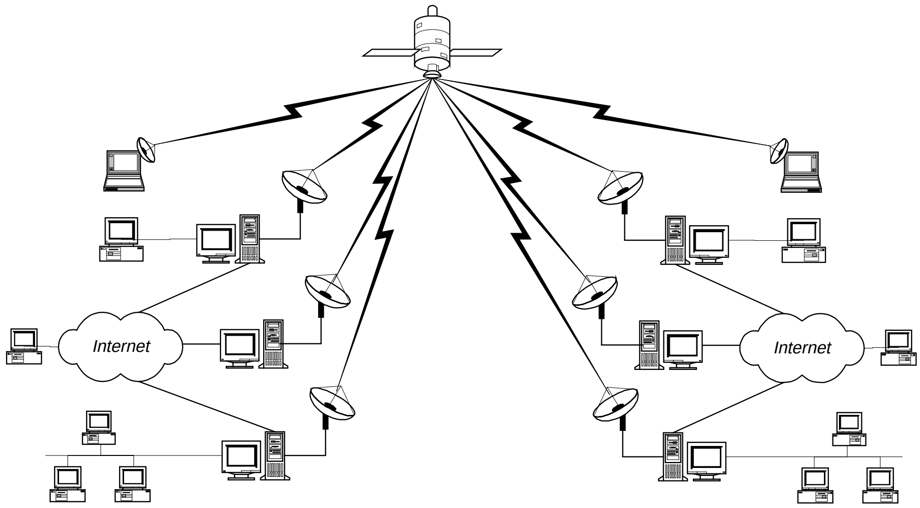


FIGURE 7-4: SIMPLE SATELLITE CONFIGURATION WITH CACHES

Such a satellite system can be useful and become significantly exploited in circumstances where bandwidth is expensive and traffic jams and delays are significant i.e. trans-Atlantic accesses. For example, a large amount of web content resides in the U.S., and European ISPs face a heavy bandwidth crunch to move data their way. A complete satellite system where caching can be introduced in most of its points (i.e. ISP, Internet, LAN, etc.) is being presented in figure 7-4.

7.5. NEW AND EMERGING PROTOCOLS

Originally the Internet protocols (e.g. TCP/IP) where primarily specified for the transmission of raw data between computer systems. For a long time the TCP/IP protocol suite was adequate for the transmission of still pictures and other row-data based documents. However, the emergence of modern applications, and mainly of those based on real-time video presented new requirements to the IP protocol suite. Though the former IP is not the ideal protocol for these suite of services, many applications appeared which present real-time (or near real-time) characteristics using IP. Products are available today supporting streaming audio, streaming video and audio-video conferencing.

On the other hand the emergence of new services caused the development of new Internet protocols and mainly the Internet Stream Protocol, Version 2 (ST 2). ST 2 is an experimental connection oriented internetworking protocol that operates at the same layer as the connectionless IP. ST 2 has been developed to support the efficient delivery of data streams to single or multiple destinations in applications that require guaranteed quality of service. ST2 is part of the IP protocol family and serves as an adjunct to, not a replacement for IP. The main application of ST 2 is the real time transport of multimedia data.

In addition to ST 2, considerable effort has been devoted for the definition of the new version of IP (IP version 6 - IPv6 or IPng). The new characteristics of IPv6 fall primarily into the following categories: expanded addressing capabilities, a simplified header format, improved support for extensions and options, flow labelling capabilities and authentication and privacy capabilities.

Among the above features of IPv6, the major one is the flow labelling capability. A flow is a sequence of packets sent from a particular source to a particular (unicast or multicast) destination for which the source desires special handling by the intervening routers. The nature of that special handling might be conveyed to the routers by a control protocol, such as a resource reservation protocol, or by information within the flow's packets themselves, e.g., in a hop-by-hop option. There may be multiple active flows from a source to a destination, as well as traffic that is not associated with any flow. A flow is uniquely identified by the combination of a source address and a non-zero flow label. Packets that do not belong to a flow carry a flow label of zero. In general in order to support various classes of service a special scheme to map the IPv6 flows to different classes of service is required. Another feature that is supported using the flows is the scaleable routing.

There is a clear need for IP networks to provide QoS. A relatively simple and coarse method, providing differentiated classes of service for Internet traffic to support various types of applications, has been standardized using the 'DS' field (RFC 2474 and RFC 2475). This differentiated services approach employ a small, well-defined, set of building blocks from which a variety of services may be built, providing QoS. RSVP (Resource Reservation Protocol) the reservation protocol for the integrated services architecture (RFC 2205) was designed to enable senders, receivers and routers of communication sessions (either multicast or not), to reserve resources in order to support services that require a QoS agreement.

A defining property of real time applications is the ability of one party to signal to one or more others parties and initiate a call. **SIP** is a client server protocol that enables peer users to establish a virtual connection (association) between them and then refers to a **RTP** (Real Time Protocol) (RFC 1889) session carrying a single media type. **RTP** provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services. RTP does not address resource reservation and does not guarantee QoS for real-time services.

8. CONCLUSIONS

This deliverable examines the LEO satellite network characteristics and its impact on the upper layers and applications. The characteristics of the LEO satellite network differ considerably from GEO satellites and terrestrial networks.

The LEO satellite network interface unit and uplink/downlink characteristics are described in this document. Various multiple access techniques, allowing the efficient use of satellite resources to meet the user's traffic demands, are examined. In this environment, a number of users communicate via a common satellite transponder. The transponder's available power and bandwidth are shared between a number of different channels and earth stations, which may themselves have quite different transmit powers and signal characteristics. The choice of the multiple access scheme has a great impact on the performance of the satellite network in a scenario where a large number of terminals are interconnected.

One of the characteristics of satellite channels is the bursty bit errors in the data being received. The Bit Error Rate (BER) depends on the Signal-to-Noise ratio at the receiver. Thus for an acceptable level of error rate, a certain minimum signal-to-noise ratio must be ensured at the receiver and hence maintained at the transmitter. The Forward Error Correction (FEC) techniques provide a solution that satisfies both these requirements. These techniques introduce some redundancy in the transmitted data. When the receiver gets the corrupted data, it uses this redundancy to decide if received data is corrupted and find out what must have been the original data. FEC codes can broadly be classified as block codes and tree codes.

The delays introduced by LEO satellites are smaller than for GEO satellites. The LEO satellite transmission delay is the sum of the three quantities: the uplink and downlink satellite-ground terminal transmission delays and the sum of the Inter-Satellite Link (ISL) transmission delays. The link delays depend upon the constellation design. In contrast with GEO satellites, the LEO transmission delays are variable over time.

There are two types of satellites: The traditional ‘repeater’ satellites (bent pipe), where signals are amplified and frequency-shifted. The other type involves satellites with on-board processing. In LEO constellations with on-board routing, the routing is usually split into Up-Down-Link (UDL) routing and ISL routing. Uplink (UL) routing is the process by which the source ground station selects the source satellite used to forward the packets of the connection, while Downlink (DL) routing is the process by which the destination ground station selects the destination satellite from which the packets of the connection will arrive. Given a source satellite and a destination satellite, as provided by UDL routing, ISL routing computes an (or at least one) optimal path between these two satellites.

The transport layer at the end hosts is of considerable importance in defining overall end-to-end communication performance. TCP is one example of a transport-layer protocol. Satellite channels have several characteristics that differ from most terrestrial channels and degrade the performance of TCP. Examples of these characteristics include higher bit-error rate, variable round trip delay, asymmetric design and use and intermittent connectivity, where ground terminals must be transferred from one satellite to another from time to time. This handover may cause packet loss if not properly performed.

9. REFERENCES

[ADAM87] Adams W. S. and Rider L., "Circular polar constellations providing continuous single or multiple coverage above a specified latitude", Journal of the Astronautical Sciences, Vol. 35, No. 2, April-June 1987, pp. 155-192.

[AKYI97] Akyilidiz I.F. and Jeong S.H., "Satellite ATM Networks: A Survey", IEEE Communications Magazine, July 1997

[ALTM98A] Altman E. and Barakat Ch., TCP over Satellite, INRIA- Sophia-Antipolis, Monday, December 7, 1998.

[ALTM98B] Altman E., Communication Satellite, INRIA-Sophia-Antipolis, DEA RSD ESSI, 1998.

[BALL80] Ballard A. H., "Rosette constellations of earth satellites", IEEE Transactions on Aerospace and Electronic Systems, vol. 16, no. 5, September 1980.

[BARA97a] Baras J.S., Corson S., Papademetriou S., Secka I.; Fast asymmetric Internet over wireless satellite-terrestrial networks, MILCOM 97 Proceedings Volume: 1 , Page(s): 372 -377 vol.1

[BARA97b] Baras J.S.; Ball, M. Roussopoulos, N. Jang, K.; Stathatos K.; Valluri J. "Integrated management of large satellite-terrestrial networks", MILCOM Proceedings Volume: 1 , Page(s): 383 -387, 1997.

[BOHM93] Bohm S. and Elhakeem A.K. Murthy V.K.M., Analysis of a movable boundary random/DAMA accessing technique for future integrated services satellites, In IEEE Globecom, pages 1283--1289, 1993.

[CHAN99] Chan D., Assesing the Impact of LEO Satellite Networks on TCP Performance, Satellite Communications Architectures and Network Workshop, IEEE ICC, 1999.
[CLAU97] Clausen H.D., Nocker B., Internet services via direct broadcast satellites, IEEE IPCCC, Page(s): 468 -475, 1997.
[CONG97] Cong X., Wireless ATM - An overview. [On-line]. ohio-state (USA): CIS. Available from internet: <URL: http://www.cis.ohio-state.edu/~jain/cis788-97/satellite_nets/>
[DABB98] Dabbous W., Integrating Satellite links within the Internet, INRIA-Sophia-Antipolis, 26 February 1998.
[DELR99] Del Re E., Fantacci R., Giambene G., "Handover Queuing Strategies with Dynamic and Fixed Channel Allocation Techniques in Low Earth Orbit Mobile Satellite Systems", IEEE VTC, Volume: 1 , Page(s): 89 -101, 1999.
[DRAI98] Drain, “Satellite continuous coverage constellations”, US patent 4809935, “Elliptical satellite system which emulates the characteristics of geosynchronous satellites”, US patent 58745206, and others.
[ELBE97] Elbert Bruce R., The satellite communication application handbook. Norwood, MA: Artech House, ISBN 0-89006-781-3, 1997.
[GALT98] Galtier Jerome, "Réseaux satellitaires", CENT/INRIA Sophia Antipolis, Ecole d’hiver des Télécommunications de Sophia Antipolis, 1998.
[GOOD86] Goodman J. and Greenberg A.G., "Sharp approximate models of adaptive routing in mesh networks", preliminary report, Teletraffic analysis and computer performance evaluation, Boxma O. J., Cohen J. W. and Tüms H. C. (ed), Elsevier Science Publishers B.V. (North-Holland), pp. 255-270, 1986.
[GOYA97] Goyal R., Jain R., Kalyanaraman S., Fahmy S., Vandalore B., Kota S.; Makki K., Ni L.M., Singhal M.,Pissinou N., TCP selective acknowledgments and UBR drop policies to improve ATM-UBR performance over terrestrial and satellite networks, Computer Communications and Networks, Page(s): 17 -25, 1997
[GOYA98] Goyal R., Jain R., Goyal M., Fahmi S., Vandalor B., VonDeak T., "Traffic Management in ATM Networks Over Satellite Links", Department of Computer Information Science and NASA Lewis Research Center, University of Ohio State and NASA, 1998.
[GRAZ97] Graziosi F., Ruggieri M., Santucci F. "Traffic performance of integrated satellite-terrestrial mobile network", Personal, Indoor and Mobile Radio Communications, 1997. Waves of the Year 2000. PIMRC '97., The 8th IEEE International Symposium on Volume: 3 , Page(s): 1029 -1033, 1997.
[HARR97] Harrington E., Josephson B., Paclik J., Modeling 21st century tactical communications networks, MILCOM 97 Proceedings Volume: 2 , Page(s): 671 -675 vol.2
[HEND97a] Hendrickson R., Globalstar for the military, MILCOM 97 Proceedings Volume: 3 , Page(s): 1173 -1178, 1997.
[HEND97b] Henderson T.R., Katz R.H., "Satellite Transport Protocol (STP) : An SSCOP-based Transport Protocol for Datagram Satellite Networks", University of California at Berkley. 1997.
[HEND99] Henderson T.R., Katz R.H., "Transport Protocols for Internet-Compatible Satellite Networks", Electrical Engineering and Computer Science Department at the University of California, Berkeley, 1999.
[PART97] Partridge C., and Shepard T.,"TCP/IP Performance over ", October 1997.

[ISNA98] Isnard, Olivier. "Etude de la qualité de service dans les réseaux ATM satellitaires en présence de Handovers". Masters Thesis Report, INT, IIE, 1998.
[JAIN97] Jain Raj ; Hart David, Satellite Communications. [On-line]. ohio-state (USA): Available from <URL: http://www.cis.ohio-state.edu/~jain/cis788-97/satellite_nets/>.

[LIRO98] Liron Moshe Lerner, US patent 5,740,164, “Traffic routing for satellite communication system”, Teledesic LLC.
[MARA98] Maral G., Restrepo J., Del Re E., Fantacci R., Giambene G., "Performance Analysis for a Guaranteed Handover Service in an LEO Constellation with a Satellite-Fixed Cell System", IEEE Vehicular Technology Conference, Volume: 4 , Page(s): 1200 -1214, 1998.
[MARG98] Marchent B.G., "Development of ATM and IP Networks for Future Mobile Multimedia Systems", Fujitsu Europe Telecom R&D Centre, 22 July 1998
[MAUG97] Mauger R., Catherine Rosenberg, "QoS Guarantees for Multimedia Services on TDMA-Based Satellite Network", Nortel Technology, July 1997
[MAXE87] Maxemchuk N. F., "Routing in the Manhattan street network", IEEE Transactions on Communications, vol. 35 no. 5, pp. 503-512, May 1987.
[NGOC95] Ngoc T.L. and Krishnamurthy S.V., Performance of Combined Free/Demand Assignment Multi-Access (CFDAMA) Protocol with Pre-Assigned Request Slots in Integrated Voice/Data Satellite Communications, In IEEE ICC, pages 1572--1576, 1995.
[NGOC96] Ngoc T.L. and Krishnamurthy S.V., Performance of combined free/demand assignment multiple-access schemes in satellite communications, International Journal of Satellite Communications, 14:11--21, 1996.
[ORS96] Örs T., Hadjitheodosiou M. H., Sun Z. and Evans B.G., "ATM over VSAT Networks using traditional bent-pipe or on-board processing satellites", ATM Networks: Performance modelling and analysis ,Volume III, ISBN 0-412-809702, pp 321-341, Chapman and Hall, London, 1996.
[ORS98] Ors T., Sun Z., and Evans B.G., An Adaptive Random-Reservation MAC Protocol to Guarantee QoS for ATM over Satellite, In Broadband Communications: The future of telecommunications (IFIP), pages 107—119, Chapman and Hall, 1998.
[ORS99a] Örs T., Chotikapong Y. and Sun Z., “GEO Satellite Network Characteristics”, BISANTE Deliverable D2.2, March 1999.
[ORS99b] Örs Tolga, Traffic and Congestion Control for ATM over Satellite to provide QoS, Ph. D. Thesis, Jan 99. Surrey University, United Kingdom
[PATT98] Patterson D. P. and Liron, M. L., "Non-blocking dynamic fast packet switch for satellite communication system", US Patent 5,796,715, granted 18 August 1998.
[RAIN97] Raines R.A., Janoso R.F., Gallagher D.M., Coulliette, D.L., Simulation of two routing protocols operating in a low Earth orbit satellite network environment, MILCOM 97 Proceedings, Volume: 1 , Page(s): 429 -433, 1997.
[REST97] Restrepo J.M., Analyse comparative entre constellations de satellites en orbite basse (cellule rattachee au satellite et cellule fixes au sol) pour un service simultane a des usagers fixes et mobiles, Ph. D. Thesis, Jul 97. ENST-Toulouse, France
[RFC1323] Jacobson, V., Braden, R. and D. Borman, "TCP Extensions for High Performance", May. 1992
[RFC2018] Mathis, M., Mahdavi, J., Floyd, S. and Romanow A.,"TCP Selective Acknowledgement Options", October 1996.
[RFC2488] Allman M., Glover D. and Sanchez L., "Enhancing TCP over satellite channels using Standard Mechanisms", January 1999.
[RFC2186] Wessels D.,Claffy K. “Internet Cache Protocol (ICP)”, version 2. September 1997.
[RIDE85] Rider L., "Optimised polar orbit constellations for redundant earth coverage", Journal of the Astronautical Sciences, Vol. 33, pp. 147-161, April-June 1985.
[SAMM97] Sammut A., Coordination with equatorial systems at ku-band, GIPSE project contribution, 1997.

[STUA97] Stuart J.R.; Stuart J.G., Revolutionary next generation satellite communications architectures and systems, Aerospace Conference, 1997. Proceedings., IEEE Volume: 3 , Page(s): 535 -545, 1997.
[STUR92] Sturza Mark A., LEOs- The communication satellites of the 21st Century, Teledesic Corporation, 1992.
[SKYB97] Spector et al, Application of Skybridge LLC for authority to launch and operate the Skybridge system, FCC submission, 28 February 1997.
[SKYB99] Skybridge, Available from internet: <URL: http://www.skybridgesatellite.com/>
[TIA98] TIA/EIA Telecommunication Systems Bulletin 91 (TSB91), "Satellite ATM Networks: Architectures and Guidelines," April 1998.
[TELE99] The Teledesic Network, Available from internet: <URL: http://www.teledesic.com/>
[UZUN97] Uzunalioglu, H.; Wei Yen , "Managing connection handover in satellite networks", IEEE GLOBECOM '97., Volume: 3 , Page(s): 1606 -1610, 1997.
[VAUG98] Victor G. Vaughn ; Fred J., "Residing Time Distribution and Call Handover Trafic Performance of Cells in LEO Mobile Satellite Systems", Department of Electrical Engineering, Virginia Tech, 1993.
[WALK71] Walker J. G., "Some circular orbit patterns providing continuous whole earth coverage", Journal of the British Interplanetary Society, vol. 24, pp. 369-384, 1971.
[WALK84] Walker J. G., “Satellite constellations”, Journal of the British Interplanetary Society, Vo. 37, pp. 559-571, 1984.
[WERN97] Werner M., Delucchi C., Vogel H.-J., Maral G., De Ridder J.-J., "ATM-based routing in LEO/MEO satellite networks with intersatellite links", IEEE Journal on Selected Areas in Communications, Volume: 15 1 , Page(s): 69 -82, 1997.
[WERN98] Werner M., Bostic J., Örs T., Bischl H. and Evans B. G., Multiple Access for ATM-based Multiservice Satellite Networks, Broadband Access and Network Management, pp 290-297, ISBN 9-05-199400-1, I O S Press, Proceedings of European Conference on Networks and Optical Communications, Manchester, England, June 23-25 1998.
[WERN99] Werner M., Wauquiez F.; Frings J. "Capacity Dimensioning of ISL Networks in Broadband LEO Satellite Systems", IMSC’99, Ottawa, 1999.
[WISL96] Wisløff Tor E., Dual satellite path diversity in non-geostationary satellite systems, Dr. Ing. Thesis 1996:84 at the Norwegian University of Science and Technology, Department of Telematics.
[WOOD95] Wood Lloyd, Network perfomance of non-geostationary constallations equipped with intersatellite links, Surrey University, United Kingdom , November 1995.
[WOOD99] Wood Lloyd, Lloyd’s satellite constellations, Online at: www.ee.surrey.ac.uk/Personal/L.Wood/constellations/
[YOUN98] Young-Ki Hwang; Hariri S., Improving TCP performance over wireless ATM networks, IEEE Information Technology Conference, Page(s): 157 -160, 1998.
[ZHAN97] Zhang B. and Leung V.C.M. "Performance evaluation of rate-based congestion control schemes for ABR service over broadband satellite networks interworking with ATM/B-ISDNs", IEEE ICC ,Volume: 2 , Page(s): 679 -683, 1997.

10. APPENDIX 1: TERMINOLOGY

- **A cell** is an area of the earth illuminated by one satellite spot beam.
- **A footprint** is the whole coverage area of a satellite, i.e. it is the sum of the areas covered by its spot beams.
- **An overlap area** is the area in which a ground station (i.e. a single subscriber or a concentrator) can receive a signal with an acceptable power level from more than one adjacent spot beam.
- **A UDL** is the aggregation of all spot beams pertaining to the same footprint; it has a fixed capacity, and is uni-directional.
- **A beam** is the communication medium between a satellite and a spot on the ground. A beam has a variable capacity which must not exceed the capacity of the UDL the beam belongs to.
- **A node** of the network is any station or any satellite. Satellites have multibeam antennas for up-link reception and down-link transmission, and are connected to neighbouring satellites by means of inter-satellite links (ISL) which are uni-directional.
- **A handover (or hand-off)** occurs when either a UDL connecting a satellite to a ground station is cut off, or when a beam change occurs (inside the same UDL), or when an ISL is cut off. All connections passing through that link must be re-routed.
- **Connections** are assumed to be full-duplex, with forward and return channels, where forward channels are intended to be from source to destination, and return channels from destination to source.
- **A call connection drop** occurs when an existing connection has to be dropped. It may happen either when there is a handover and the connection cannot be re-routed, or when high priority traffic preempts all the resources used by a connection.
- **A call block** occurs when a new connection cannot be established. It may happen when there are no resources available in the network in order to support the new connection.
- **A satellite** is associated with a space position which varies deterministically over time. Satellite movements are described through orbital mechanics; a ground station is associated with a ground position which may vary randomly over time if the station is mobile.