Sponsored by: Hughes Network Systems, NASA and ATIRP

ABSTRACT

Title of Thesis: RESOURCE ALLOCATION FOR KA-BAND BROADBAND SATELLITE SYSTEMS

Degree candidate: Vineet Birmani

Degree and year: Master of Science, 1999

Thesis directed by: Professor John S. Baras
Department of Electrical and Computer Engineering

Forthcoming high data rate geostationary satellite systems, which are planned to work in rain affected Ka-band, would need effective resource management from a network management perspective. In this thesis we have looked at satellite power and bursting time-slots needed for the downlink channel, as satellite resources that need to be managed. Some solutions are proposed in order to tackle the above-mentioned problem.

RESOURCE ALLOCATION FOR KA-BAND BROADBAND SATELLITE SYSTEMS

by

Vineet Birmani

Thesis submitted to the Faculty of the Graduate School of
the University of Maryland, College Park in partial
fulfillment of the requirements for the degree of
Master of Science
1999

Advisory Committee:

Professor John S. Baras, Chairman/Advisor
Professor Nicholas Roussopoulos,
Dr. Subramanian Raghavan

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I am greatly indebted to my advisor, Dr. John S. Baras, for his guidance, support and encouragement through the course of this work. I would also like to thank Dr. N. Roussopoulos and Dr. S. Raghavan for their valuable comments on the thesis. A special word of thanks to Mr. David Whitefield of the Hughes Network Systems who introduced me to many new ideas through several illuminating discussions pertaining to this thesis.

This work was supported by Hughes Network Systems through the HNS-ISR Industrial Fellowship Program, by NASA through NASA Cooperative Agreement - NCC3-528, and by Advanced Telecommunications and Information Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program through cooperative agreement DAAL01-96-2-0002.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Brief History of the *Ka*-Band

In the late 1970's, some space agencies, including NASA in the United States, NASDA in Japan, ESA in Europe, and ASI in Italy, identified the need to develop new space technologies to extend the role of satellite communications. It was also recognized, at the same time, that the frequencies used by satellites up to that point would not provide sufficient bandwidth to meet the anticipated demand for services. Therefore it was decided to pioneer the use of higher frequencies, especially the Ka-band (20 / 30 GHz band). This frequency band had been considered useless until then because of the excessive signal attenuation due to rain, and other forms of moisture in the atmosphere. It was agreed that this band could provide the needed bandwidth and, through higher gain antennas, allow larger on-board effective isotropic radiated power (EIRP) levels to ensure reliable communications during rain periods. The *Ka*-band was thus selected as the frequency band for these developments.

1.2. Some Ka Band Success stories.

Over the years some Ka-band experimental missions have been launched. ITALSAT, the Italian Ka Band program was launched in late 1970's for the development and deployment of a regenerative multi-beam Ka band payload, with Italian coverage, obtained by six very narrow spot-beams. A gross total of about 0.9 Gbit/s was achieved with a 147 Mbit/s time division multiple access (TDMA) in the uplink. A synchronous baseband space-switch matrix provided the inter-spot connectivity, with TDM in the downlink. ITALSAT-F1 was successfully launched in 1991. The system became the first operational regenerative Ka band system integrated with the terrestrial networks. It provided operational experience for reallocation of capacity in a fast and flexible way.

Another, more recent mission was the NASA’s ACTS program [7], whose satellite was launched in 1993. It demonstrated the technologies that have become the foundation of the current interest in the use of Ka band by a number of global interactive multimedia systems. ACTS introduced very small hopping spot-beams (0.3° beamwidth) to concentrate the available satellite signal power over a small area and thus penetrate through the rain, and it introduced the use of coding to enhance transmission. The ACTS program also developed an infrastructure for On-Board Processing (OBP) equipped with memory and switching to relay messages from users in one beam to users in another beam, so that the available satellite transponders could be time-shared to activate the large number of beams needed for complete US coverage.

Soon this turn of events stimulated a strong industrial interest. Pushed by the spread of this interest, the Federal Communications Commission (FCC) set a September 1995 deadline for filing for the use of Ka-band frequencies to provide interactive multimedia and other services, and the FCC awarded 13 licenses in 1997 to build them. Applications were also filed with the ITU for many other systems, for a worldwide total of well over one thousand satellites. This happy end for a frequency band long maligned as being useless has been referred to as the “Ka-band Rush”.

1.3. Future Ka-band systems and their new set of requirements

Some of the proposed systems, like Spaceway (see [23] for details) and Astrolink are being designed to support millions (see Appendix A for details about the Astrolink constellation) of users. To offer an attractive solution to the communication needs of millions of users, both the size and cost of the earth stations or user terminals have to be reduced by using high gain multiple spot-beam satellites with on-board processing and switching capability.

In order to be able to achieve this futuristic design plan of supporting millions of users, it is going to be imperative that these systems do an efficient management of the various satellite and spectrum resources. Some of these resources, like the frequency spectrum, have been a limiting factor in most of the old and present day systems, so a lot of work has already been done in designing good resource management algorithms. On the other hand the satellite power, although always a limited resource has gained much greater importance in these Ka band systems because of the effect of rain attenuation.

Another limiting resource is the small number of satellite antennas. With these next-generation broadband systems being planned to support hundreds to thousands of spot-beams, a much smaller number of antennas would have to implement a scheduling discipline so that they are able to serve the thousands of downlink cells in a fair manner. The fairness being referred to in these systems is with respect to the Quality of Service (QoS) parameters that are going to form the basis of the charge, which a subscriber would pay for using the system resource to make a call. This would require the scheduling algorithms to manage the downlink TDM bursts in a manner such that under conditions of no congestion all cells are served, while in conditions of congestion calls with higher QoS are given preference. Hence the primary objective becomes to maximize profit while ensuring fairness.

Congestion, it has been agreed, would be a frequently occurring phenomenon in the downlink channel of these satellite systems. Further, when rain attenuation occurs during these periods of congestion, rain margins would need to be allowed for while serving areas with rain. This would result in greater congestion. As a result of these factors it would become necessary that the antenna allocation (for downlink bursts) and power management algorithms for these systems be so designed such that they work well when working alone; and also result in maximization of desired-profit when the need for them to work together arises.

1.4. Contributions of this work

The considered network scenario is mesh configured, comprising a satellite operating in the Ka frequency band with on-board packet-switching and control functionalities. In this scenario, we particularly address the resource allocation on the forward channel (satellite to earth terminals) of a geostationary satellite link. We have looked at algorithms for burst scheduling and power allocation (for rain-fade compensation) for the satellite systems providing QoS guarantees.

These problems have been studied from a network management perspective. This implies that the various algorithms are being proposed under the assumption that they would be implemented *off-line* at *a Network Operation and Control Center (NOCC)*, and then the plans would be transmitted up to the satellite in an asynchronous manner, i.e. whenever the NOCC feels the need to change the existing plan of service. This would typically be when there is:

- A significant variation in traffic volume due to the hour of the day.
- Considerable change in the buffer fill-factor. These buffers are at the output of the packet-switch on-board the satellite, and contain traffic meant for various downlink cells. Each buffer serves one downlink cell.
- Weather change, resulting in a need for change in power allocation to the antennas while they are bursting down to certain specific ground cells to compensate for the rain-fade.

Through this work we hope to suggest the use of QoS and priority information to schedule bursts in the satellite for the downlink. This is based on the assumption that not all traffic that makes it to the downlink buffers after the cell-switch would be successfully beamed down.

Also if information about the rain intensity in various downlink cells is available at the NOCC on a near-real-time basis it can be used to generate power allocation plans so as to ensure lower percentage of outages due to rain fades which can plague the communication in Ka band.

Similar algorithms could be implemented on the satellite payload on a real time basis. They might give better results if implemented there although it would result in much greater complexity and cost.

In the next chapter the problem has been described in detail. Some details of past work done on these issues and some related problems have also been given. Chapter three gives details of the solutions proposed and analyzed in this thesis. In the fourth chapter some simulation results are described, which we obtained when we implemented some of the suggested algorithms on a simulated model of the Ka-band satellite system. These results have been compared, in terms of implementation complexity and maximization of ‘desired-profit’ variable, with some other approaches. Chapter five provides conclusions and suggestions for further research, based on the study reported in this thesis.

2. NEED FOR RESOURCE MANAGEMENT

2.1. NEXT GENERATION BROADBAND COMMUNICATION SATELLITES: TYPICAL SYSTEM DESCRIPTION

The satellite system studied here supports on-board ATM-like switching and a spot-beam intensive architecture. The requirement for satellite switching results from the need for small inexpensive user terminals. This is for the following reason. One of the ways to achieve greater system capacity and inexpensive user terminals is by using multiple beams. However, a multi-beam configuration means some form of switching between beams. The on-board switch thus employed is also expected to provide a statistical multiplexing gain in a bursty multimedia environment.

Figure 1 below illustrates a typical configuration of a broadband satellite system operating in Ka band to provide multimedia services. Each spot-beam serves a number of earth stations and the on-board packet-switch provides the required inter-connectivity between spot-beams. Multimedia (voice/video/data) traffic is assumed to be segmented into *ATM-like* equal size packets. Variable Bit Rate (VBR) and Available Bit Rate (ABR) are used for real-time (voice, video) and jitter tolerant traffic (data), respectively.

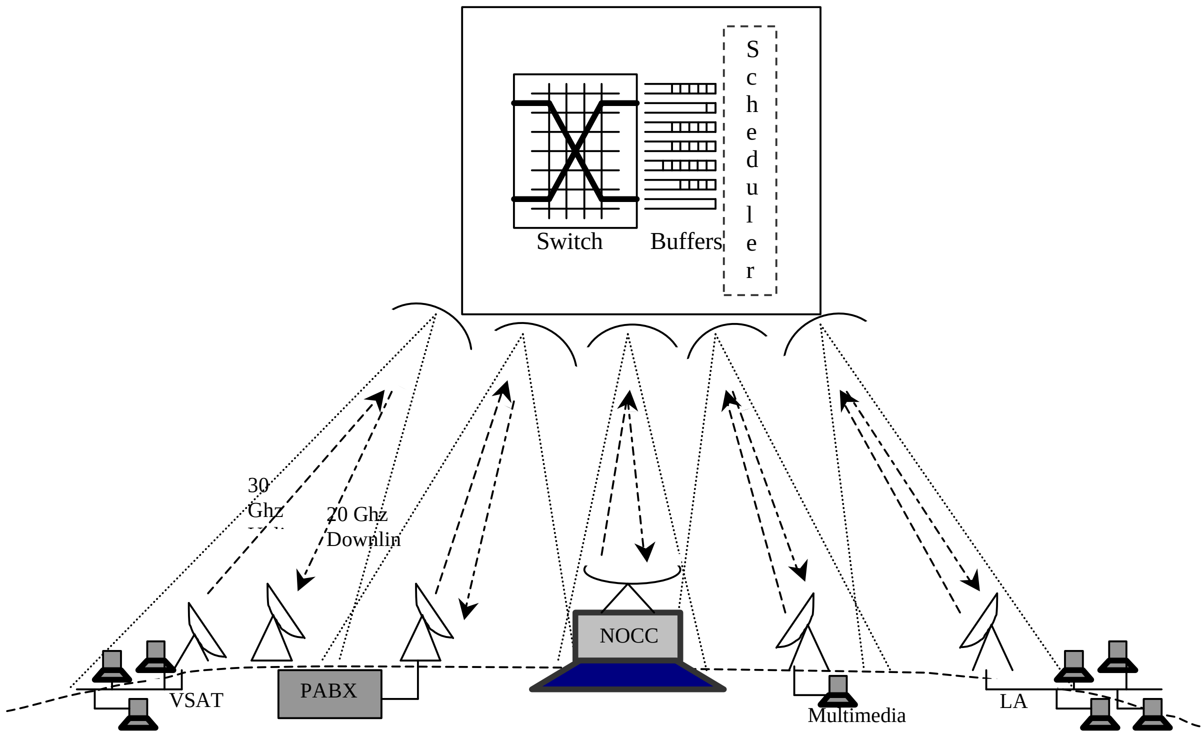


Figure 1: Illustration of a typical configuration of forthcoming high data rate satellites

In most proposed systems, the uplink channel is shared by earth stations using a combined Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) scheme, called the Multiple Frequency-TDMA (MF-TDMA). In the MF-TDMA scheme the total bandwidth allocated to each spot-beam is first divided into a number of non-overlapping carriers [11]. It is desired to keep a low transmission rate in order to reduce the size and cost of the earth stations transmitter power amplifier. However the transmission rate should be high enough to accommodate peak traffic rate required by the user services. For example a 2.048 Mbps is selected to include multimedia services including MPEG-1 video traffic. Each channel is further divided into non-overlapping time slots, each of which can accommodate one fixed-size packet. (TDMA frames considered here can be subdivided into a number of frame units, or packets, each of them having the same length and format.) The earth station is equipped with a frequency-hopping synthesizer so that the transmitter can access any time-frequency slot in a two-dimensional MF-TDMA frame. This allows for flexibility and bandwidth efficiency provided by TDMA without preamble, while reducing the size of the ground station due to reduced burst rate.

On the downlink, the access mechanism is assumed to be some type of time-division muliplexing (TDM), [11]. Here again the downlink transmission to the various ground cells is organized into bursts, made up of packets of length and format similar to those on the uplink. The burst is assumed to be consisting of a fixed number of time-slots, with each time-slot long enough to transmit one fixed size packet. The combination of MF-TDMA and TDM allows for the use of powerful bandwidth-on-demand capacity allocation.

Scheduling is used to assign slots on the airlink. Because of the onboard regenerative packet switching, *it is possible to dissociate the uplink from the downlink scheduling and resolve some contention on-board.*

Services supported by the system include: audio, videoconference, interactive on-demand services and database consultation, in addition to the traditional Internet applications. A user terminal - (UT) involved in a multimedia connection or serving as a gateway for a local network can also simultaneously support a combination of these applications.

User terminals are *independently and spatially distributed across the footprint of the satellite*. But at the same time it is assumed that the number of UT's in a certain spot-beam is closely related to the population of the area covered by the spot-beam. With this assumption it is reasonable to model the traffic demands of the various spot-beams in the footprint of the satellite in a manner such that it follows the population density variation in the footprint.

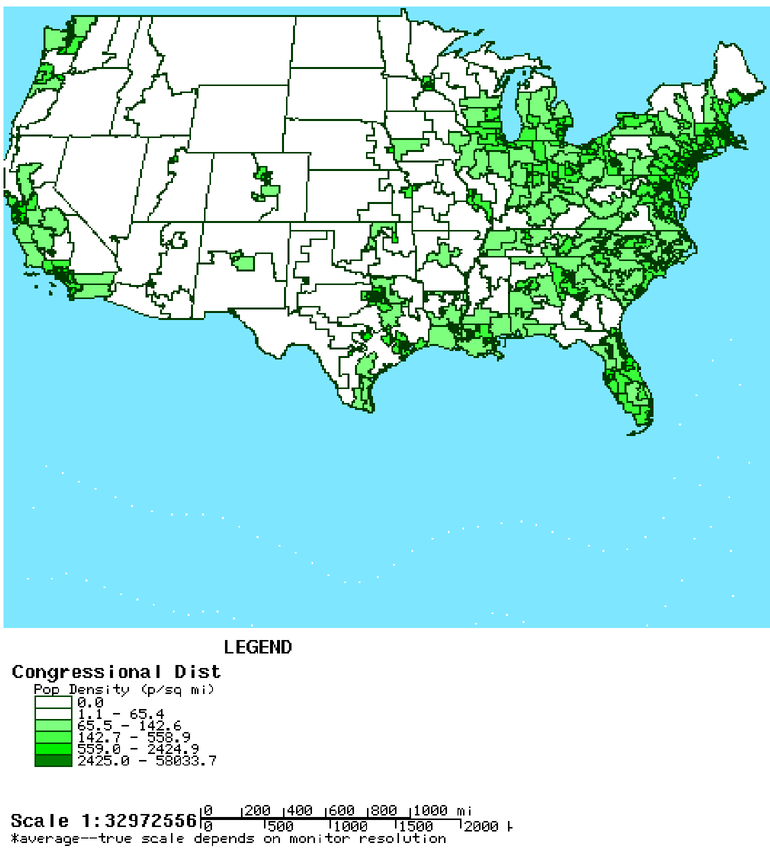


Figure 2: Population density distribution over USA (Diagram obtained from the Dept. of Commerce, Census Bureau, http://www.census.gov)

The spot-beam architecture of the satellite systems that are being emulated here is such that the number of uplink cells is less than the number of downlink cells by at least a factor of four [see appendix A for details of a typical system]. As a result the cell size of the downlink is much smaller than that of the uplink. This is essentially so that the satellite power can be concentrated on a smaller area on the ground. This also means that the packet-switch, which switches packets between the uplink and downlink beams, has unequal number of input and output ports. (The input ports correspond to the uplink beams while the output ports correspond to the downlink beams.)

2.2. THE ANTENNA SUBSYSTEM

Phased array antennas are attractive for geostationary satellites. The primary reason being that the satellite systems that are being proposed plan to achieve global communication coverage by using high gain multibeam satellite antennas. These antennas are sorted into two types: direct radiating phased arrays and large reflector antennas fed by cluster feeds.

Phased arrays offer the capability of power allocation between beams according to traffic demand, reconfigurability of beams, maintainability of beam pointing and tolerance of amplifier malfunction. These are the kind of antennas which the systems studied here are assumed to possess [22].

A brief description of various kinds of phased array antennas is provided below. Three types of phased array antennas are briefly discussed here: the direct radiating array, the dual reflector fed by planar array, and a single reflector fed by a radially aligned array. Configuration 1 (Figure 3a), is a planar direct radiating array whose elements are arranged in a triangular grid. The following parameters are common to all antenna types: N_{fi} is the number of elements, D_{fi} is the element diameter and D_{ai} is the diagonal length of the hexagonal array. Here subscript *i* refers to configuration number shown in (Figure 3). Configuration 2 (Figure 3b), is a single reflector antenna with an array feed. Its additional parameters are reflector diameter D_{m2}, focal length f_{m2} and feed displacement length L_{f2}. Configuration 3 (Figure 3c), is a dual reflector antenna composed of two paraboloidal reflectors and a planar array. Additional parameters are main reflector diameter D_{m3}, main reflector focal length, F_{m3}, subreflector diameter D_{s2}, subreflector focal length F_{s2} and feed displacement from the subreflector L_{f2}. The number of elements determine the complexity of the antenna and of the beam-forming network.

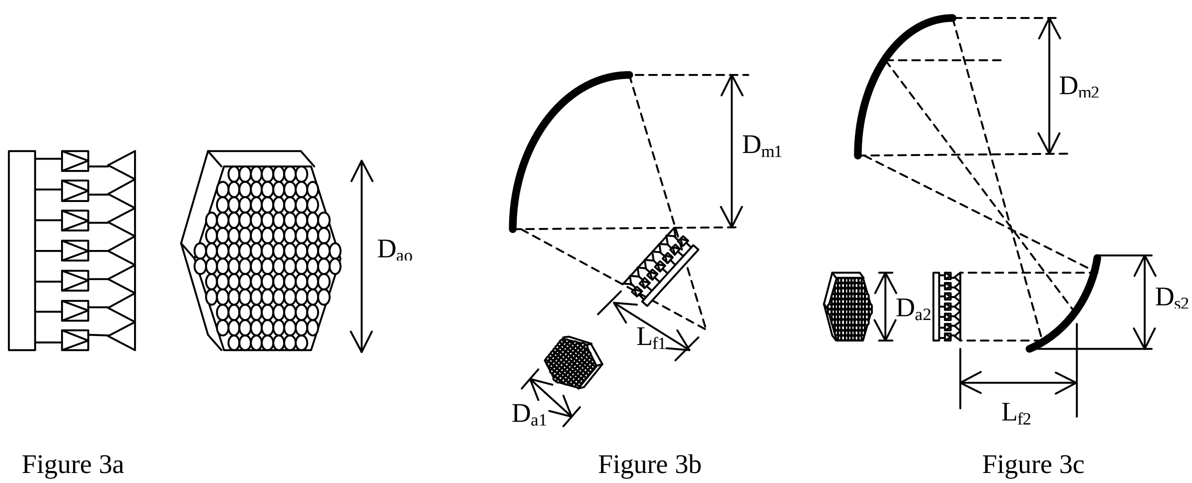


Figure 3: Three types (configurations) of phased array antennas

2.3. NEED FOR RESOURCE MANAGEMENT

Many next-generation satellite systems using GEO, LEO and MEO or their hybrids are being proposed for both residential and business applications to support multimedia services. Multiplexing of the traffic bound for various user terminals is done on the satellite on a beam-per-beam basis. This is done at the output (downlink) buffers of the packet-switch. The antennas on-board the satellite would serve these buffers by transmitting the packets in the form of bursts. *The number of antennas serving the buffers (one buffer per downlink cell) is much smaller than the number of buffers.* This is generally because of the cost factor. But at the same time, these antennas burst down to various ground cells, so that under normal load conditions the buffers are served at rates that prevent buffer-overflow.

The scheduling algorithm, which is described in this work, is meant to work at this point of the system, under conditions of congestion. Congestion is considered to occur in this system when packet inflow into the buffers at the output of the packet-switch is at rates higher than the aggregate of the antenna service rate. This could occur under three circumstances:

(i) when the rate of packet inflow increases (described in section 2.3.1); or

(ii) when the service rate drops, (described in section 2.3.2); or

(iii) when both the above conditions hold true.

Under these conditions of congestion it would be desirable to ensure that the system resources are allocated in an efficient manner. Further, given the large variety of traffic mix which makes it to these buffers at the output of the packet-switch, it would be desirable that the packets belonging to traffic with higher quality of service guarantees are not dropped at these queues due to non-availability of servers (in this case, antennas), while the lower priority packets are being able to make it. This is equivalent to saying that the satellite resources (bursting time-slots at the antennas and the satellite power) should be allocated such that the 'aggregate-priority' of packets which make it to the ground is maximized.

2.3.1. Rate of packet inflow into buffers greater than provisioned

Multimedia traffic is bursty with high peak to average transmission rate. If capacity is allocated on a per-call basis, as in a conventional circuit-switched scheme, then it must be equivalent to peak rate, even if the transmission is not so. This approach leads to under-utilization of the system capacity. As satellite communication systems are known to be power and bandwidth limited, it is important, from the network management standpoint, to maximize the utility of the space-segment resources by using an efficient dynamic resource allocation scheme at every stage in conjunction with an on-board fast ATM-like switch. In each MF-TDMA frame, the uplink scheduler dynamically allocates the uplink time-frequency slots to the earth stations according to their instantaneous capacity requirements. An example of a scheduling strategy based on a tunable Leaky Bucket scheme is as follows. At a given time voice, video and data traffic are admitted on the basis that the average traffic rates don't exceed the uplink capacity. The uplink scheduler then sets the token rates and the bucket depths corresponding to the average and the peak rates of the traffic type (e.g. voice/video/data) or the QoS guarantees for each active earth station. In each MF-TDMA frame, the uplink scheduler generates tokens into the buckets according to the set rates, and allocates to each active earth station a number of time frequency slots equivalent to the instantaneous capacity requested by the earth station, or the number of tokens remaining in the bucket, whichever is smaller. See [6] for greater details.

Each earth station sends its traffic packets into the assigned time-frequency slots. Based on the destination address contained in their headers, the on-board switch transfers received traffic cells to the corresponding downlink beams. It can be seen that the on-board switch provides a statistical multiplexing mechanism, and the number of packets received in a frame from various uplink beams to a downlink beam may exceed the downlink capacity. Therefore an on-board downlink queue (or buffer) is provided to absorb the fluctuations of traffic flow, see Figure 1. These fluctuations are a cause of congestion, which must be handled by the scheduler.

2.3.2. ANTENNAS ARE UNABLE TO SERVE AT FULL CAPACITY DUE TO NEED FOR RAIN-FADE COMPENSATION

The Ka-band spectrum, while providing wider bandwidth and permitting small antennas, is at the same time susceptible to rain-induced fades that cause system outages [1]. These impairments include signal absorption by rain, clouds, and gases, and amplitude scintillation arising from refractive index irregularities. Some examples demonstrating the extent of the problem are:

(i) Rain can result in attenuation in excess of 20 dB in many areas of the world;

(ii) Rain-attenuation at 20 GHz is almost three times that at 11 GHz (Ku-band).

The resulting outages can become a limiting factor in guarantees about the network availability that a service provider can promise the end user. This is a direct impact of the rain fades. Another relatively indirect effect is on the Quality of Service (QoS) guarantees that the network service providers' promise. If satellite communication is continually subjected to rain fades, and hence transmission errors, it would be tough to maintain the promised QoS.

Techniques have been developed to provide compensation for rain attenuation, usually at the expense of system capacity. In the case of the ACTS system, 10 dB of dynamic fade compensation for the affected users can be added by means of data rate reduction and convolutional coding.

But not every user needs this expensive compensation. As a result most satellite systems being planned for the future will opt for a large volume of low cost uncompensated availability to the users, with special accommodations, for select customers who must have higher availability. *This transforms directly to providing higher QoS in terms of availability to users at higher cost*.

If satellite power is readjusted between various downlink antennas to provide for rain margin, in order to achieve the above requirement of selective provision of higher availability, it would mean that some antennas will not carry traffic at each bursting instant. (This is assuming that all downlink phased array antennas burst in a synchronized manner with packets destined for various downlink cells.) Hence the antennas will not be able to serve at full capacity, resulting in queue build-up at the buffers.

2.4. The Network Operations and Control Center (NOCC)

The implementation of various off-line resource allocation algorithms is assumed to be done at the NOCC. The NOCC is assumed to have a high priority control channel access to the satellite through which it dictates among other things, power reallocation plans for the antennas, and high and low traffic spot-beam bursting parameters.

Satellite channels are being increasingly allocated on a demand assigned basis so that the limited capacity can be efficiently managed. But this generates a traffic variation between beams, which cannot be handled by a purely preprogrammed plan for the satellite hardware as has been done in the past. Further, the satellite systems normally have a lifetime of 15 years or more. With the Internet traffic growing at an enormous rate, it would seem wasteful to provide for capacities for that far into the future right now. To overcome this problem of not wanting to put excessive capacity on-board the satellite, while at the same time wanting to be able to support the growth of traffic in the future, it would be desirable to control the satellite capacity from the ground. These satellite systems are, as mentioned before spot-beam intensive, and are planning to offer wide variety of QoS guarantees. With the footprint of a geostationary satellite covering a vast area on the ground, a significant improvement in efficient allocation of capacity can be obtained if the spatial and temporal variation of the traffic in the footprint of the satellite is adequately taken into account. The network control centers are being designed to handle just that, using good resource allocation algorithms.

It is assumed that the NOCC shall have access to information about:

(i) on-board buffer size and fill-factor which will help generate queue dynamics,

(ii) historical and near-real-time traffic volume variations in various uplink and downlink cells,

(iii) weather forecast on a near-real time basis,

(iv) information about signal attenuation due to rain, at user terminals (such information can be made available by selective polling),

(v) rain models [9] which provide coarse information about rain and its movement.

3. SOLUTIONS TO THE RESOURCE ALLOCATION PROBLEM

3.1. SCHEDULING AND FAIRNESS

Computer networks that allow users to share (or multiplex) resources must also incorporate some mechanism of solving the problem of contention for the shared resource. The server implements a scheduling discipline to decide which request to serve next. Different types of schedulers quite often differ from each other based on what the scheduling discipline they are implementing is trying to achieve. Intuitively they all want to be fair to the various input streams that enter the multiplexer. Hence the difference lies in what is considered as fair. Given the Asynchronous Transfer Mode (ATM) QoS definitions, fairness currently is interpreted differently by schedulers at different levels in the network. Quite often fairness is applied among traffic streams within each class of service, while a different fairness criterion may be applied to traffic from different classes.

In providing a solution to the burst scheduling problem we also try to achieve fairness among bursts destined for various downlink beams. The scheduler proposed here is implemented after the ATM-like-switch, when packets for the various cells have been queued in their respective buffers (with each buffer for a specific downlink). The downlink beams are dynamically allocated to the antennas. This is done in such a way that the antennas can burst down packets from various beams in a manner which, under stable load conditions (that is when the rate of entry of packets into the queues is less than or equal to the aggregate rate of service), transmits all packets. On the other hand, under unbalanced load conditions the aggregate-priority of packets, which reach the earth terminals, is maximized. *Aggregate-priority is defined as the summation over all packets received at the destination of a quantity which is arrived at by calculating the product of weight attached to each packet and the number of packets of that weight received.* We associate different degree of fairness to the packets of different priorities. It is furthermore required that the packets with highest priority i.e. priority 4 are never dropped. *(In our work we have assumed that priority 4 is the highest priority while priority 1 corresponds to the lowest priority.)*

3.2. Generalized Processor Sharing and Weighted Round Robin Scheduling

Generalized Processor Sharing (GPS) is an ideal, unimplementable work-conserving scheduling discipline, which most scheduling disciplines try to emulate when they want to provide fairness.

Weighted Round Robin (WRR) is one such scheme, and is the simplest emulation of the GPS. This scheme has a disadvantage of being unfair over periods smaller that one round time, but with present day packet switches operating at gigabits per second, the round times are not very large, so the unfairness is not noticed over most time scales. It is a variant of this scheme, which is proposed in this thesis for the purpose of burst scheduling. As will be explained in section 3.3, the weights attached to the input stream are not constant but change with time-of-day and day-of-week.

This is how the scheme developed here is different from the generally understood WRR. The algorithm would be simulated at the NOCC but executed in real time on the satellite.

3.3. THE KNAPSACK PROBLEM

Suppose a hitch-hiker has to fill up his knapsack by selecting from among various possible objects, which will give him maximum comfort. This knapsack problem [15] can be mathematically formulated by numbering the objects from *1 to n* and introducing a vector of binary variables x_j $(j = 1, 2, \ldots n)$ having the following meaning:

$$x_j = \begin{cases} 1, & \text{if item j is assigned to knapsack;} \\ 0, & \text{otherwise.} \end{cases}$$

Then if p_j is a measure of the comfort given by the object j, w_j its size and c the capacity of the knapsack, the problem is to select, from among all binary vectors x satisfying the constraint

$$\sum_{j=1}^{n} w_j x_j \le c$$

the one which maximizes the objective function

$$\sum_{j=1}^{n} p_j x_{i\,j}\ .$$

A trivial approach would be to examine all possible binary vectors x, selecting the best of those that satisfy the constraint. Unfortunately the number of such vectors is 2^n.

If the number of knapsacks is 1 it is referred to as the *0-1 knapsack problem*. On the other hand for more than one knapsack this is called a *Multiple Knapsack Problem (MKP)*. It will be shown in section 3.6 that the rain-fade compensation by power reallocation bears a close resemblance to the (MKP).

The Knapsack problem is known to be an NP complete problem. However it is not strongly NP complete. As such there are a reasonably large number of exact algorithms which have been developed, each of which try to reduce the running time required to get to the final solution. Besides that, some good approximate schemes [15] have also been developed for cases where the problem's input size is relatively large.

3.4. SOLUTION TO THE BURST SCHEDULING PROBLEM

The system as described would require adaptive resource allocation algorithms so that it can modify the system's availability to meet specific QoS requirements. As mentioned in section 2.2.1 the buffers at the output of the switch are there to absorb the fluctuations in the traffic. The buffers are assumed to be relatively large and can carry many bursts of packets. (The downlink antenna bursts down a burst-length worth of packets at a time.)

Packet loss occurs when the downlink queue overflows. In addition, a packet can be considered lost if it is not-jitter tolerant and it has stayed in the queue longer than permitted. Therefore a priority packet needs to be burst down to the ground before it becomes stale. The challenge to this approach is to maintain the cell loss probability below an acceptable threshold while guaranteeing the required quality of service. To this end, it is assumed that the NOCC adjusts the (uplink) token rates and bucket depths adaptively so that the flow of aggregate traffic to a downlink beam is temporarily reduced when the queue level becomes high according to predictions made there. At the same time, the network controller also needs to ensure that the high priority packets (for example the packets belonging to rt-VBR flows) are scheduled to burst down while the token rate variations are yet to take effect. This problem can be restated as follows:

In the problem considered here, we are given a set of queues, each queue representing the traffic destined for a downlink cell, and a set of servers which are the downlink antennas serving the bursts arriving from those queues. We want to allocate the downlink queues, or buffers (whose number is much larger than the number of antennas), to the antennas, with the primary objective of maximizing the aggregate priority of packets transmitted by the antennas in a given time, while ensuring fairness in the allocation policy at the burst level. Bursts are formed by a fixed number of packets, while packets belong to various flows. As every queue carries packets from different flows, each flow having one of four different priorities, our scheduler aims at providing fairness (of resources allocated) among queues with varying priorities or *weights (or profits)*.

The queues are known to be FIFO and are filled with packets. Each packet has a priority, which is the priority of the flow to which the packet belongs. There are four priority levels, as described in the following section.

3.5. ATM TO MAC LAYER SERVICE MAPPING

At the ATM layer, several ATM transfer capabilities (ATC) have been defined by the ITU-T (defined as service categories by the ATM forum). Service categories are classified to be either real time or non-real time. According to the ATM forum definition, CBR (constant-bit rate) and RT-VBR (real time variable bit rate) belong to the real-time traffic, while NRT-VBR (non-real time variable bit rate), UBR (unspecified bit rate), and the ABR (available bit rate) belong to the non-real time category. In this thesis we have proposed a certain mapping from the ATM transfer categories to the corresponding service classes at the MAC layer, in order to facilitate their integration at the buffers in the satellite. We refer to these classes as the *MAC Buffer Handling Capabilities or MBHC.*

In this thesis, we assumed (based on ATM priority levels) four distinct MBHC's: CBR with drop bit unset, CBR cells with drop bit set, bursty data, and best effort. These categories differ in the guarantees given about their delivery to the end system. CBR with drop bit unset is the highest priority. The scheduling discipline would ensure that packets in this class would be delivered to the ground cells more than 99% of the time under conditions of congestion. The other three classes would be given similar (but lesser percentage) guarantees of delivery after they've made it to the output buffers. (Although this is not the primary objective of the scheduling discipline, through simulation we hope to show that similar guarantees can be given and be met in most circumstances.)

Therefore the end (ideal) performance desired from the scheduling discipline is:

(i) ensure that the CBR packet with drop bit unset is minimally dropped;

(ii) maximize the aggregate priority of packets transmitted by the antennas in a given time, and

(iii) implement fairness in the allocation policy at the burst level.

The variant of the *Weighed Round Robin* (WRR) scheduling strategy, which we implement to ensure the above objectives, has the following features.

Ideally the weights attached to each competing queue would be equal to the aggregate of the priorities of the cell in the queue up to *a search depth.* The search depth should be equal to the length of the burst that an antenna can burst at a time. Further, for CBR packets with drop bit unset it is required that the weights be so assigned that the packet is never dropped.

As mentioned before this is an ideal situation. This is because the packets are flowing into the queue at aggregate speeds of a few gigabits per seconds (See Appendix A); the antennas are bursting down again at about the same speed. It would be impractical to assume that the scheduler would be able to look at each packet's header and check out the priority and compute aggregate weights at that speed.

For that reason the scheduling mechanism we propose should be statistically implemented (more economically) at the NOCC. At the same time it should emulate the process of weight calculation and schedule generation on a non-real time basis based on statistical values of the following information:

(i) historical and near-real-time diurnal variation of *traffic volume and typ*e for the various kinds of spot-beams,

(ii) queue length variations at output buffers of the on-board switch,

(iii) spatial variation of traffic volume and type. (This can be safely assumed to follow the population density variations on the footprint of the satellite.)

It is required that the NOCC maintains the above information in the form of a database for easy retrieval. Further, it is also required that the NOCC maintains information about trends observed in these data.

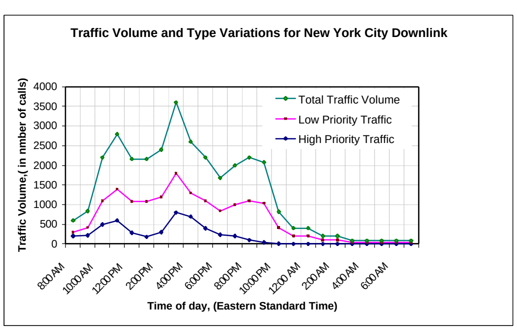
In this thesis it is assumed that the footprint of the GEO satellite is the mainland USA. Further, the numbers used below assume that there are *n* disjoint downlink ground cells covering the whole footprint of the satellite. These *n* downlink cells are numbered from 1 to *n.*

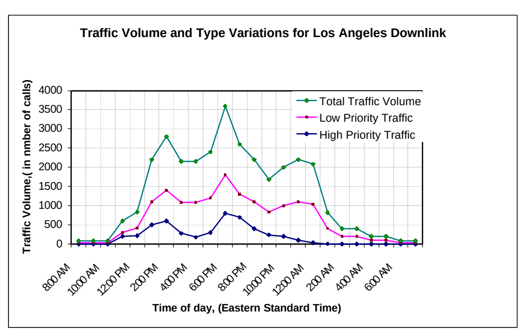
The kind of historical data trends that should be maintained are:

(i) The traffic volume variations for downlink *i* Î *{1, 2, . . . n}*, with time-of-the-day and day-of-the-week. *The granularity of time would depend on how much fluctuation the on-board buffers can absorb and how fast the NOCC can compute and upload.* As an example the downlink corresponding to hot-spots like New York City, and suburban Maryland (where population density is smaller by a factor of 100 than that in NYC) could show trends as shown in the example graphs below in Figure 4, and also in Table 1. These graphs have been drawn using synthetic data (justified by intuition) and are not based on any actual traffic demand traces. Further, the x-axis of the graph is the time-of-day in terms of the Eastern Standard Time (EST); This shows clearly how the three hour difference between EST and Pacific Time (PT) can be used to dynamically allocate system capacity (a well known and widely practical technique).

(ii) The composition and mixture of packet types that make up the traffic volume for downlinks *i Î {1, 2, . . . n}*. Such information would keep track, for example, of what fraction of traffic that makes a fixed length burst of *x* packets is high priority traffic. It is suggested that *x* be the *search depth* mentioned before. This information is not-real-time, but statistical and is obtained based on queue dynamics information, which the NOCC can get periodically from the satellite.

(iii) The fraction of traffic on the downlinks *i Î {1, 2, . . . n}*, shown falls distinctly into the various priority categories, whose graphs are shown below, based on the stored traffic demand trends for each of the downlinks.





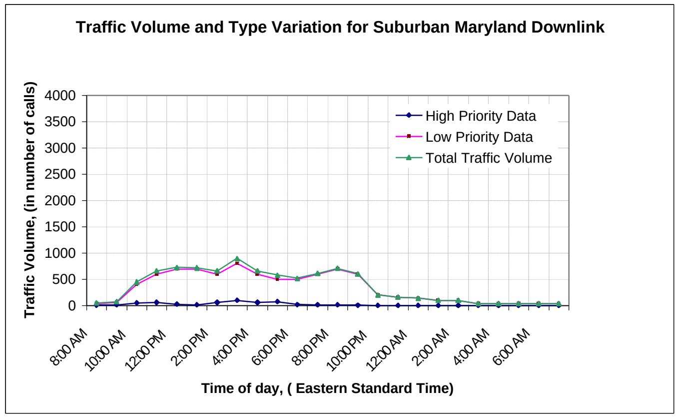


Figure 4: Graphs illustrating a synthetic (based on common sense experiences) mix of traffic for the beams to 3 specific destinations

TRAFFIC CHARACTERISTIC	NEW YORK CITY	SUBURB - MARYLAND
Traffic Volume between 8:00 AM and 7:00 PM on a weekday		
1. High priority, important industry traffic – e.g. Video Conferencing, commercial documents, stock-market traffic etc.	Very High Volume	Very Low Volume
2. Lower priority residential traffic – e.g. web surfing, multimedia file downloads	Low Volume	Low Volume
Traffic Volume on weekends and between 7:00 PM and 8:00 AM on weekdays		
1. High priority, important industry traffic – e.g. Video Conferencing, commercial documents, stock-market traffic etc.	Very Low Volume	Very Low Volume
2. Lower priority residential traffic – e.g. web surfing, multimedia file downloads	High Volume	Very high Volume

Table 1: More details of synthetic traffic trends example.

This kind of storage of traffic trends is needed in order to ensure that the NOCC can make the following types of queries on the database and get fast response:

(i) What is the most likely queue length variation for the downlinks for the next *x minutes*? The value of *x* would again be determined by the amount of fluctuations that the on-board buffers are able to absorb. It would also depend on how frequently the schedule plan is to be updated.

(ii) What is the (statistically obtained) aggregate priority of packets which are expected to fill up the downlink buffers of the *n* beams for the next *x minutes*. This value can be obtained from the traffic trends collected at the NOCC.

Based on replies to these types of queries and other information the downlink queues can be ordered according to weights (these weights would indicate the type of traffic filling up the queues, and the buffer overflow probability) which will in turn reflect the relative importance of those downlinks. Once the order of importance is generated, it should be transmitted from the NOCC up to the satellite, which will implement this ranking priority for serving the queues (i.e. Burst Scheduling) in a round robin manner (one burst at a time). Hence the Weighted Round Robin scheme results.

The frequency with which the order for serving the queues is updated, should be higher at times when the traffic volumes are large and show significant variations; for example during the day. During the night a default plan can be implemented.

3.6. THE RAIN FADE PROBLEM AND VARIOUS SOLUTIONS FOR COMPENSATION

As mentioned before, future satellite systems operating in the Ka-band are going to be subject to degradation produced by the troposphere, which is much more severe than those found at lower frequency bands. As a consequence, in order to meet the Bit Error Rate (BER) guarantees, most Ka-band systems are expected to employ different forms of fade mitigation, which can be implemented relatively easily and at a modest cost.

One way of achieving that is to do resource sharing on-board the satellite. Some systems set aside a portion of system capacity to provide for what is referred to as the *rain margin*. In Time Division Multiplex (TDM) systems, this capacity could be time-slots. Hence in the event of rain in some portion of the footprint which causes BER to exceed the permitted limit, these extra time slots are assigned for transmission to rain areas in order to build-in adequate redundancy. For example, in the case of the ACTS system, 10 dB of dynamic fade compensation for the affected users can be added by means of data rate reduction and convolutional coding.

Another method of compensation is to provide for extra transmission power while transmitting in areas affected by rain. The phased array antennas that are assumed to be used in these Ka-band systems may support such dynamic allocation of power.

A variant of this latter approach forms of the basis the research reported in this thesis.

Rather than setting aside some system-wide resources to compensate for rain-fade, the approach taken in this thesis, uses information made available at the NOCC through various channels to reallocate power in a systematic manner, with the objective of maximizing the aggregate priority of the traffic supported by the satellite system.

3.6.1. AVERAGE RAIN TIME

Various weather charts have shown that the average rain time for which the compensation must be applied is usually very small [9], [21]. Typically, this may be less than 5% to 10% of time during an average year. As such, fade compensation through dynamic resource allocation is more cost effective than setting aside constant resources, which will be unnecessary for over 90% of the time.

3.6.2. SIMULTANEOUS RAIN FADE OVER EXTENDED AREAS

It has been observed that rain events over areas like the mainland USA (the footprint area of the typical satellite of interest in our study) are widely distributed in time. It is unlikely that significant rain intensity would be observed in more than a handful of areas simultaneously.

For this reason it is suggested that the NOCC maintains a database of the level of signal attenuation in the different downlinks $i \in \{1, 2, . . . n\}$ for different rain intensities. Then on the basis of information gathered,

(i) by polling various user terminals (UT) for attenuation, and

(ii) radar weather data about rain intensity in different downlinks,

the satellite power can be dynamically allocated to the various affected downlinks whenever significant variation in rain intensity is observed.

3.6.3. FADE DURATION

If the rain intensity information is available on a near-real-time (NRT) basis the power allocation plan can be modified just as quickly. This would ensure that satellite power is not wasted by providing for excessive rain-margin where it is not required. As fade duration could last from a few seconds to a few hours, the amount of rain margin that is applied would also depend on how often the NOCC can get the information about the level of attenuation suffered by the signal at various downlinks.

3.6.4. THE GRANULARITY OF RAIN-FADE COMPENSATION LEVELS

It is assumed the systems provide for up to three levels of power for each downlink cell. These levels differ from one cell to another and depend on factors like the elevation angle etc. If the above mentioned attenuation information is available periodically, on the basis of intervals with few minutes duration, from various UT's, then it would be to the advantage of the system to implement a larger number of levels for power allocation. This would, as mentioned before, result in power allocation to the beams in a manner that further reduces waste. It would also result in a system that would be able to support a greater number of simultaneous transmissions for the same amount of satellite power.

Although it is evident that such a scheme would put additional pressure on the communication infrastructure due to NOCC initiated communications and signaling to get the required information on a NRT basis, its effect can be minimized by:

(i) Polling only some of the User Terminals (UTs) in downlinks, which are experiencing, rain. All UTs need not be polled.

(ii) Using a beacon signal from the satellite for gauging the signal degradation experienced by downlink transmissions. It is assumed that the uplink transmission uses a closed-loop power control mechanism [1]. If the beacon signal strength is such that it indicates rain-induced signal degradation, an asynchronous message should be sent by a selected set of UTs from each downlink cell to the NOCC. It is in response to such messages that the NOCC would initiate a re-run of the power allocation algorithm.

3.7. Power Allocation Methodology

Quality of Service guarantees also include guarantees of system availability, besides guarantees on delivery. In rain events covering wide areas of the footprint, satellite power would have to be taken from some beams and allocated to other beams (this is a result of not having any spare capacity). Many optimization related decisions would need to be made, to implement such reallocations.

Some of these decisions are:

(i) To which low priority cells transmission would have to be sacrificed temporarily in order to compensate for rain affected areas in high priority cells.

(ii) Whether traffic demand in the rain affected areas can justify allocation of excessive power at the expense of some other area, which would have to miss a transmission burst.

(iii) Assess the need for more than one downlink cells to miss a burst in order to provide for sufficient power for the downlink in the rain affected area.

We have used the aggregate priority of the packets, which make up the burst for various downlink cells. As it was mentioned before this aggregate priority cannot be calculated in the satellite on a real-time basis because of the gigabit per second rates which are operating on these broadband satellites. Instead, the NOCC makes use of the statistical priority information available to it, which it gets from the various traffic volume and type databases it maintains (similar to the WRR information it maintains and/or receives).

Thus under no rain conditions the system will operate normally, while in conditions of severe wide-spread rain the same serving capacity of the system would be used to serve higher priority users before serving lower priority users. Unlike the method employed in the ACTS where the system operated under reduced capacity so as to compensate for rain, this method results in the system operating at full capacity at all times, but providing better service to users who pay more for higher quality service.

It should be noted that because traffic segregation of traffic at the granularity of downlink cells, some low priority users in the ‘hot-spot’ cells might end up benefiting, while some higher priority users of suburban downlinks might experience some disadvantage. However, the weighted-round-robin manner of serving the downlinks, along with the manner suggestion for implementing the ‘CBR with drop bit unset’ option, would ensure that the high priority users are sufficiently compensated for. These implementation details were implemented in simulations we performed and are explained in chapter 4 of this thesis, ‘Simulation model and results’. The overall flow of information, data and decisions, for the power allocation scheme developed and investigated in this thesis is shown in Figure 5.

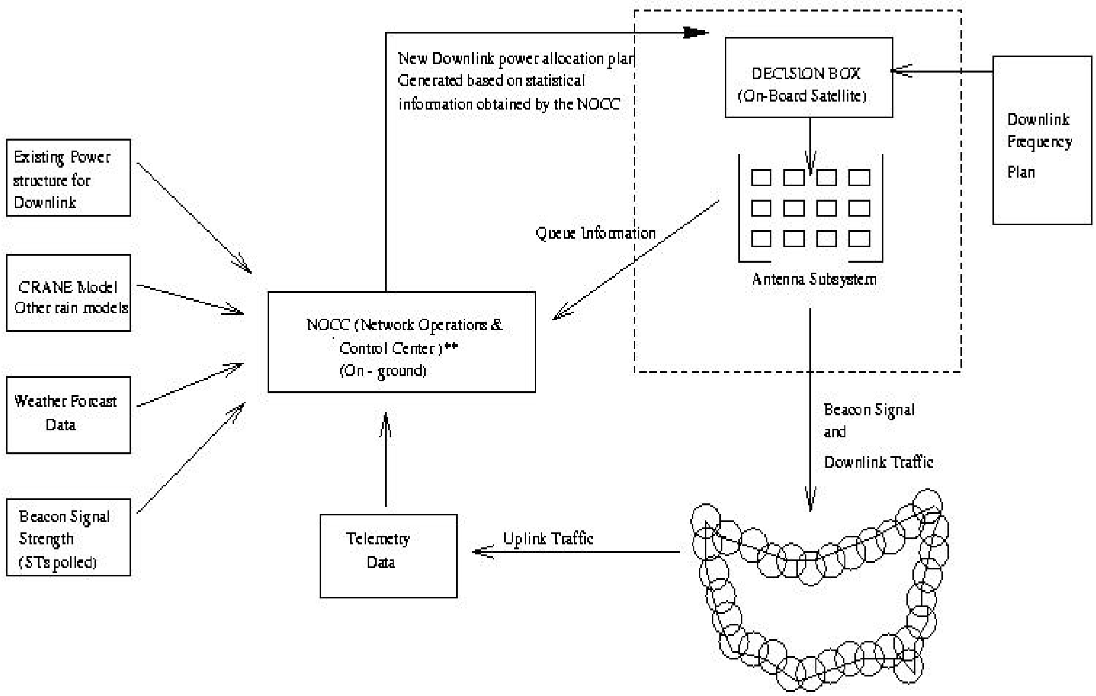


Figure 5: Power Allocation on the Downlink for Ka Band Geo Satellite - System Description

** This facility maintains a historical and spatial database (per downlink) of rain areas and the corresponding attenuation statistics.

3.8. LINEAR PROGRAMMING FORMULATION OF THE RAIN-FADE COMPENSATION PROBLEM FOR KA-BAND SATELLITE SYSTEMS

In this section we develop a Linear Programming formulation of the rain-fade compensation problem. In this formulation, the aggregate priority of the downlinks is calculated based on the actual priority of packets making up the burst (as was described in the ideal case before). The definition of various variables used in this formulation is provided in Table 2, while that for constants in Table 3 below.

Variables	Definition
p_{ti}	The power assigned to the downlink beam being transmitted at time *t* from antenna *i*.
x_i	A measure of the granularity in power levels which can be assigned to the downlink beams. x_i can take values from a set of values as indicated in the formulation that follows. The set of values consists of *0* and positive values less than *1*.
d_{ti}	The aggregate priority of downlink transmitted in time-slot *t* from antenna *i*.
q_{ti}	The priority of the packets being transmitted in time-slot *t* from antenna *i* This would be used to calculate the aggregate priority of the downlink d_{ti}, by adding priorities over all the packets in the downlink.
z_i	z_i takes the value *0* or *1* depending on whether the corresponding x_i is *0* or greater that *0*. i.e. $z_i = 0 \text{ if } x_i = 0$ $z_i = 1 \text{ if } x_i > 0$
f	describes a function which describes how the required power is dependent on rain intensity and rain type.

Table 2: Variables used in LP formulation

Constants	Definition
P_{tot}	Total Power of the satellite system being studied.
p_{min}	Minimum power that can be assigned to the downlink beam. This is the power required to keep the system components warm. It is assumed that this is also the power that would be required to transmit to a spot beam under clear sky conditions, i.e. when the attenuation due to rain is 0 dB.
A	Constant which along with the value of x_i would be used in defining the maximum power, which can be assigned to the beam.
I_{max}	Maximum number of beams that can be transmitting at one time. It is equal to the number of antennas on the system.
M	A very large number which ensures that Mx_i is greater than 1 for all $x_i > 0$.
J_{max}	Maximum number of fixed size packets which can be transmitted during one time slot on the downlink.
q_k , k Î $\{1,2,3,4\}$	Priorities assigned to the packets. PVC packets have highest priority. Best effort traffic has lowest priority and can be dropped in case the resources become limited.

Table 3: Constants used in LP formulation.

Based on the description of the problem given so far, and the terminology and nomenclature introduced, the LP formulation that results is provided below.

LP Formulation

$$\textit{Maximize} \qquad Y = \sum_{i=1}^{I_{\max}} d_{ti} z_i \quad , \qquad \textit{for all } t \tag{1}$$

subject to:

$$p_{ti} = Ax_i + p_{\min} \tag{2}$$

$$p_{ti}^{\min} = f(rain) \tag{3}$$

$$p_{ti} \geq p_{ti}^{\min} \tag{4}$$

$$x_i \in \{x_1, x_2, \ldots. x_n\} \tag{5}$$

$$z_i \leq Mx_i \tag{6}$$

$$0 \leq z_i \leq 1 \tag{7}$$

$$\sum_{i=1}^{I_{\max}} p_{ti} \leq P_{tot} \quad , \qquad \textit{for all } t \tag{8}$$

$$\sum_{j=1}^{J_{\max}} q_{tj} = d_{ti} \quad , \qquad \textit{for all } t \textit{ and } I \tag{9}$$

$$q_{ti} \in \{q_1, q_2, q_3, q_4\} \tag{10}$$

3.9. RAIN-FADE COMPENSATION AND THE KNAPSACK PROBLEM

Based on the explanation of the knapsack problem and the LP formulation of the rain-fade compensation problem just given, it is evident that the two problems are similar. To delineate clearly the relationship between the two problems, we provide below a detailed description of the knapsack problem followed by an example case of the rain-fade compensation problem, which we have used in our simulation experiments.

3.10. THE 0-1 MULTIPLE KNAPSACK PROBLEM

The *0-1 Multiple Knapsack Problem (MKP)* is defined as follows.

Given is a set of n items and a set of m knapsacks $(m \leq n)$, with

p_j = profit of item j,

w_j = weight of item j,

c_i = capacity of the i^{th} knapsack.

We want to select m disjoint subsets of items so that the total profit of the selected items is a maximum, and each subset can be assigned to a different knapsack whose capacity is no less than the total weight of the items in the subset. Formally this is described as:

Maximize $$z = \sum_{i=1}^{m} \sum_{j=1}^{n} p_j x_{ij} \qquad (11)$$

Subject to $$\sum_{j=1}^{n} w_j x_{ij} \leq c_i \qquad i \in M = \{1, \ldots, m\} \qquad (12)$$

$$\sum_{i=1}^{m} x_{ij} \leq 1 \qquad j \in N = \{1, 2, \ldots, n\} \qquad (13)$$

$$x_{ij} = 0 \text{ or } 1, \qquad i \in M,\ j \in N \qquad (14)$$

where $$x_{ij} = \begin{cases} 1, & \text{if item } j \text{ is assigned to knapsack } i, \\ 0, & \text{otherwise.} \end{cases}$$

When m = 1, the MKP reduces to the 0-1 (single) knapsack problem, (KP).

It can be assumed without loss of generality, that

$$w_j,\ p_j, \text{ and } c_i \quad \text{are positive integers,} \tag{15}$$

$$w_j \leq \max_{i \in M} \{c_i\}, \quad \text{for} \quad j \in N, \tag{16}$$

$$c_i \geq \min_{j \in N} \{w_j\}, \quad \text{for} \quad i \in M, \tag{17}$$

$$\sum_{j=1}^{n} w_j > c_i \quad , \quad \text{for} \quad i \in M. \tag{18}$$

If assumption *(5)* is violated, fractions can be handled by multiplying through by a proper factor. Items j violating assumption *(6)*, as well as knapsacks i violating assumption *(7)* can be eliminated. If the knapsack i^* violates assumption *(8)*, then the problem has the trivial solution $x_{i^*j} = 1$ for $j \in N$. $x_{ij} = 0$ for $i \in M \backslash \{i^*\}$ and $j \in N$. Finally, if $m > n$ then the $(m-n)$ knapsacks of the smallest capacity can be eliminated. (In this problem all knapsacks are of equal capacity, therefore any $(m-n)$ knapsacks can be eliminated.)

When applied to the rain-compensation problem, c_i refers to the maximum satellite power available for allocation to the downlinks; p_j is the aggregate priority of the packets which will be burst down to cell *j* in the next bursting period, while w_j is the power required to burst down in to the downlink cell *j*.

A complete description of the rain-fade compensation problem is given in the following section in the form of an example that has been used to simulate a typical operational scenario.

3.11. AN EXAMPLE PROBLEM

Parameters, variables and configuration considered in this example:

- There are *n* disjoint downlink cells that serve as the destination for the downlink beams.
- There are *k* antennas that can burst down to these cells on the ground.
- Under no-rain condition the *k* antennas can serve all *n* cells before the next batch of packets or frames arrives at the input queues of the downlink beams (on-board the satellite). This simply implies that under normal ideal load conditions the input rate into the buffers is equal to the aggregate of the *k* servers' (antennas) service rate. This can further be assumed to imply that all *k* antennas are able to burst for *n/k* times each, (*let this value equal m, i.e. n/k = m*) before the next batch of packets fills up the buffers. That means that each antenna is able to serve *m* downlink-buffers before the next batch of packet arrives. Let the servers' aggregate service rate under this condition of stability be μ_{stable} and the arrival rate of packets into the queues be λ_{stable}.

- Although this is the case for situations where no congestion builds up on the downlink, it has been agreed that congestion could build up in the downlink buffers, and could be further aggravated under conditions of rain in some of the downlink cells.
- Under the following conditions of congestion build-up
  - no-rain in any of the downlink cells, i.e. $m = m_{stable,}$
  - packet inflow into the buffers is at a rate higher than the packet serving rate of the antennas, i.e. $l > l_{stable}$,

  to achieve fairness in serving *n* queues with *k* antennas we have used the Weighted Round Robin scheduling scheme. It is assumed that the weights would simply reflect the trends in the priorities of the packets filling up the queues, and in the ‘fill-factor’ of each of the *n* queues.

There is another way for congestion to build up besides the one described above. This could happen if the antennas serving the queues slow down due to power adjustments that needs to be made in order to compensate for rain-fade in some of the downlink cells. This condition of congestion would arise irrespective of whether $l > l_{stable,}$ *or* $l = l_{stable}$. That is:

- Rain in *x (x ³ 1)* (number of) downlink cells, which implies $m < m_{stable}$.
- Packet inflow into the corresponding downlink buffers is at a rate l, where l ³ l_{stable} .

This case can be modeled as a Multiple Knapsack Problem (MKP) which was described before. How the two problems are related is described below. To accomplish this linkage in a straight forward manner we use the terminology employed in the description of the MKP (see equations 11 to 14) and provide an explanation as to what each term means in terms of the rain-fade compensation problem.

The *0-1 Multiple Knapsack Problem (MKP)* explained as follows.

Given:

n items: Here the downlink cells which need to be served.

m knapsacks: Here the number of times the antennas can burst before the next batch of packets arrives into the buffers. Further, m £ n, is required by the problem definition which is evidently satisfied here. Furthermore m is here equal to the minimum number of bursts required to serve the n downlink cells, each of them at least once if all k antennas are able to burst down at each bursting instant. This would happen only if there is a 'no-rain-in-any-cell' condition. The minimum number can be obtained if the antennas serve in a round robin manner, and this would equal n/k.

With

p_j = profit of item j. This is the weight (weight, as in the WRR scheduling scheme discussed previously) attached to each of the *n downlinks, (j = 1, 2, 3, …,n).* These profits are obtained from the trends available at the NOCC about the type and volume of traffic segregated on the basis of downlink cells.

w_j = weight of item *j*. This is the power required by the antennas when they burst down to the j^{th} downlink. The power can be allocated in discrete steps, and is greater than or equal to the minimum power required to maintain acceptable C/N for the downlink. w_j corresponds to the variable p_{ti}, used in the equations of the LP formulation, which are reproduced below for the convenience of the reader.

$$p_{ti} = Ax_i + p_{\min}$$

$$p_{ti}^{\min} = f(rain)$$

$$p_{ti} \geq p_{ti}^{\min}.$$

c_i = capacity of the i^{th} knapsack. Here it is the maximum power available to the antenna subsystem in the satellite that can be used to burst down at any time instant *t*, i.e. it is equal to $P_{tot.}$

The problem hence is to select *m* disjoint subsets of beams so that the total profit of the selected beams is a maximum, and each subset can be assigned a different bursting time (knapsack) whose capacity is no less than the total weight of the items in the subset. Formally,

maximize $$z = \sum_{i=1}^{m} \sum_{j=1}^{n} p_j x_{ij} \qquad (19)$$

subject to $$\sum_{j=1}^{n} w_j x_{ij} \le c_i, \qquad i \text{ Î } M = \{1, . . . m\} \qquad (20)$$

$$\sum_{i=1}^{m} x_{ij} \le 1, \qquad j \text{ Î } N = \{ 1, 2,. . . . n\} \qquad (21)$$

$$x_{ij} = 0 \text{ or } 1, \qquad i \text{ Î } M,\ j \text{ Î } N \qquad (22)$$

where $$x_{ij} = \begin{cases} 1, & \text{if item j is assigned to knapsack i,} \\ 0, & \text{otherwise.} \end{cases}$$

The assumptions and constraints described above transform to the assumptions and constraints of the rain-fade compensation problem as follows.

We assumed that,

w_j, p_j, *and* c_i *are positive integers,* (23)

$$w_j \le \max_{i \in M} \{c_i\}, \qquad \text{for} \quad j \text{ Î } N \qquad (24)$$

$$c_i \ge \min_{j \in N} \{w_j\}, \qquad \text{for} \quad i \text{ Î } M \qquad (25)$$

$$\sum_{j=1}^{n} w_j \; > c_i, \qquad \textit{for} \quad i \in M \qquad (26)$$

These correspond to the following interpretations:

(23) Implies that the power required for transmission to each downlink, the weight attached to each of the downlink buffers and the total power available for allocation at each time instant are positive integer respectively. Further, in the example provided above the value of c_i *for all* $i \in \{1, 2, . . ., k\}$ is a constant.

(24) Implies that the power required for transmission to any of the n downlinks never exceeds the maximum power available for beaming down to the earth. If it is higher then that beam can never be scheduled.

(25) Implies that the system power available for distribution at any instant of time has to be greater than at least the minimum of the downlink power requirement. As c_i is a constant in the above example, assumption *(25)* turns out to be the same as *(24)*.

(26) Is required to be true or else there is a trivial solution to the problem as described earlier.

3.12. THE RELATION AND LINKAGE BETWEEN THE SOLUTIONS OF THE TWO KEY PROBLEMS

The two key problems in our formulation of resource allocation are burst scheduling and power allocation on the downlink. These are related and linked together as shown in Figure 6. The burst scheduling module would provide the set of queues that should be served in the next round along with the order in which they should be served.

Following this, if there is rain in large parts of the footprint, it would become imperative to boost power in some of those beams at the expense of some other beams in order to maximize priority throughput. So if the knapsacks represent the overall satellite power we would need to fill them up so as to maximize the aggregate-priority among the selected queues. This would result in a reduced service rate for the next bursting period, resulting in sacrificing some downlink cells.

A couple of such options, as to how the two solutions should be combined and interact so as to maximize the priority throughput have been tried using simulations. The details of the simulation model and the results obtained are discussed in the next chapter.

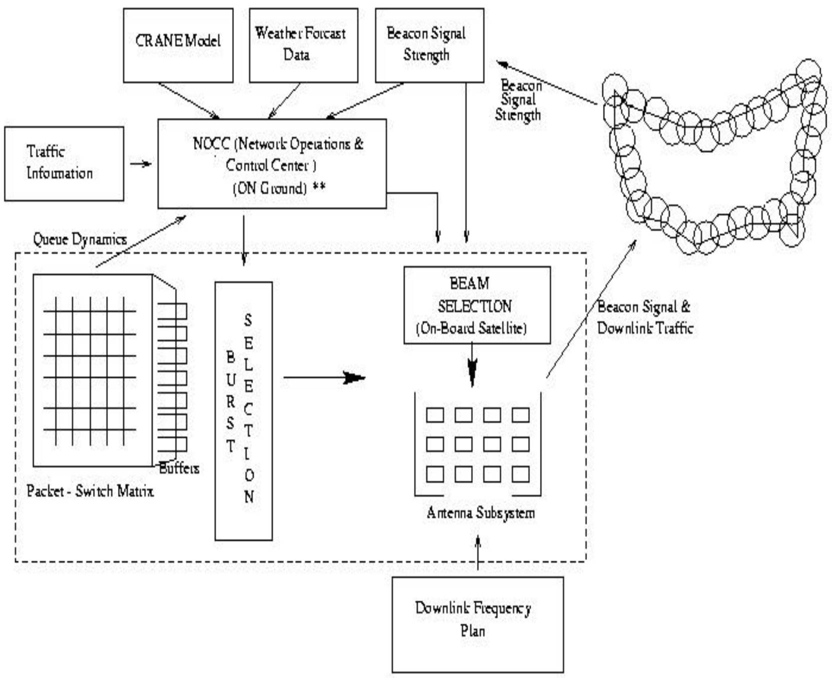


Figure 6: Resource Allocation on the Downlink for Ka Band GEO Satellite - System Description

** This facility maintains a historical and spatial database (per downlink) of rain areas and the corresponding attenuation statistics.

4. SIMULATION MODEL AND RESULTS

4.1. THE SIMULATION MODEL

Figure 7 shows a high level architecture (configuration) of the simulation model used for simulating the system and the resource allocation strategies we have designed in *OPNET-5.0.* The modules shown in the figure are described below.

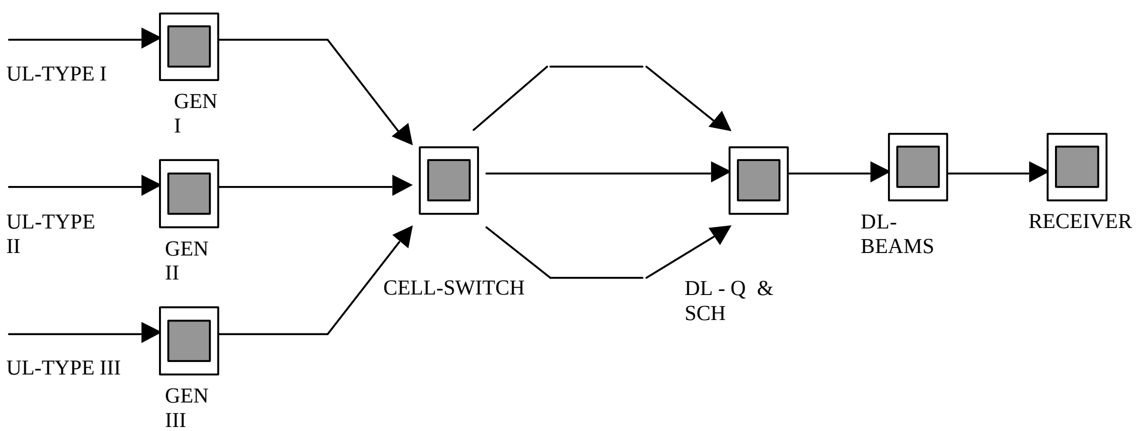


Figure 5: Simulation Model

4.1.1. UL-TYPE I, II & III AND GEN I, II & III

These modules are responsible for generation of traffic. The traffic model being emulated here is a superposition of batch-Poisson process. The superposition is achieved by the three types of uplinks (UL), each of which is responsible for traffic being fed to the satellite from different kinds of cells on the ground.

The UL-TYPE I along with GEN I generates traffic which is expected from all the suburban uplinks, i.e. uplinks carrying traffic from cells which have population density less than 100 persons per square mile. We refer to the map of the USA shown below in Figure 8. This area accounts for nearly 70 percent of the footprint (assuming the system being studied has the continental USA as the footprint). In case there are 100 uplink cells it would mean that 70 of them are expected to generate traffic similar to the one modeled here. The packets being generated here are at a rate that would be expected to be the rate of generation of traffic from the combined 70 uplinks. It consists of a majority of low priority traffic (i.e. priority 1 and priority 2).

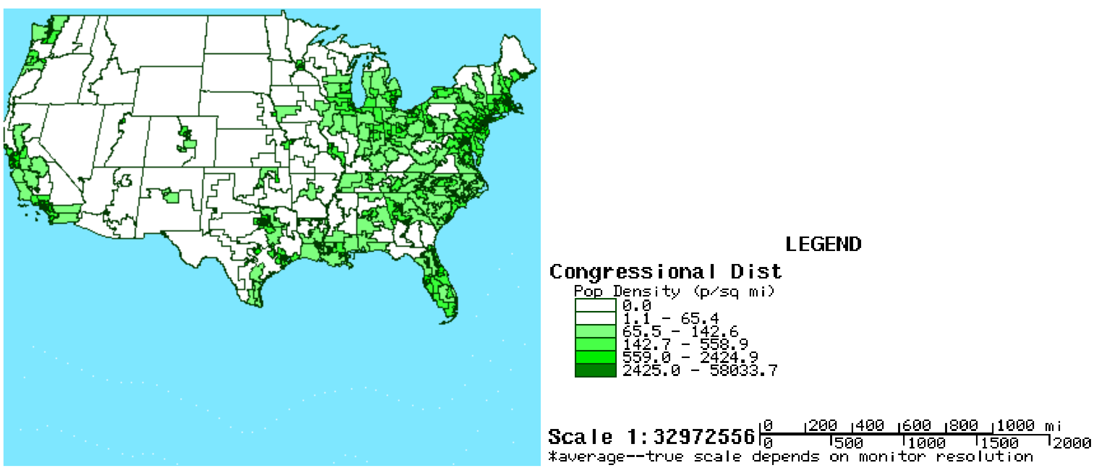


Figure 6:

Population density distribution over USA (Diagram obtained from the Dept. of Commerce, Census Bureau, http://www.census.gov)

As these areas with population density less than 100 people per square mile would typically be non-commercial (generally residential suburban) areas, it would reasonable to assume that the traffic would be web-browsing and hence generally of not very high priority.

The UL-TYPE II along with GEN II generates traffic, which would normally be expected out of areas with population density between 100 and 2000 persons per square mile. Areas of this kind cover approximately 20 percent of the USA mainland. As a result from a total of 100 uplink cells, 20 of them are modeled by this UL-TYPE. Traffic originating from a typical uplink cell of this type would be at a rate considerably higher than that from a type I cell. Also as these areas would be generally small towns (i.e. semi-commercial /semi-residential), the traffic would be of high and low priority in nearly the same proportion (slightly higher on the higher priority side during the day when businesses are in operation, while higher on the lower priority side during the evenings.).

The UL-TYPE III along with GEN III generates traffic that would be expected out of 'hot-spots', like New York City, Los Angeles, etc. These areas typically cover the remaining 10 percent of mainland USA, but are responsible for a majority of the traffic which the system operating over USA hope to support. With population density over 500 times that found in areas modeled by UL-TYPE I, the traffic rate is expected to be enormously higher than that available from any of the other kinds of cells. Here again the diurnal variation is such that higher priority traffic would dominate during the day while evenings would typically have higher proportion of low-rate, low priority traffic.

4.1.2. CELL-SWITCH

This module models the on-board switch, which switches traffic between uplinks and downlinks. It is assumed that out of *n* downlinks, 70% cover suburban USA, 20% cover small commercial towns while 10% cover 'hot-spots' like New York City. In modeling the packet-switch it was assumed that the majority of high priority traffic originating from 'hot-spots' would be destined for hot-spots. The low priority traffic is assumed to be distributed so that it follows the population distribution over the footprint. The switch switches packets between various in-streams and out-streams based on the above given assumptions of origin-destination distributions. The out-streams, out of the switch feed the *n* buffers, modeled by the next module.

4.1.3. DL-Q AND SCH

It is here that the packets accumulate, waiting to be transmitted, by the downlink antennas. The Weighted-Round-Robin algorithm is implemented here. The weights are the *aggregate-priority* of the packets (aggregated over a burst length, which is assumed to be constant). As mentioned earlier it would be impractical to assume that the on-board scheduler can look into each packet's header and get the priority information for each packet. Therefore to simulate the statistical nature of the weights available to the WRR scheduler, the aggregation is done after transmission of a *(large) random number* of bursts. For burst times between two aggregations the results of the previous aggregation is used. This is done so that the asynchronous updating from the NOCC, regarding the relative importance of the traffic for various downlink cells, is emulated. As the updates are not expected to come very frequently, and only when there is a significant change in the buffer fill-factor or the traffic mix resulting from the diurnal variation of traffic, the number of bursts between consecutive attempts at aggregation is large. The randomness ensures that the asynchronous behavior of the updates from the NOCC is incorporated.

Once the weights for the various downlink cells are available, the buffers are sorted for the next round of transmission based on these weights. After this sorting packets are sent down to the next module, which simulates the behavior of downlink phased–array antenna transmission to the ground cells. Hence the DL-Q and SCH along with the DL-BEAMS modules model the complete downlink transmission portion of a broadband satellite system, which gets its updates on the transmission plan from the NOCC.

4.1.4. DL-BEAMS

The queue size is expected to build up if the inflow into the buffers is greater than or equal to the outflow. This could happen during periods of rain, when not all antennas can burst down to the ground, *due to power limitations*. In case of excessive congestion, which would typically result during peak traffic load periods, it would become imperative to drop some packets. As mentioned before, quite often more calls than what can be handled (if most users transmit at peak permitted rate) are accepted by the system. Congestion in the downlink buffers is an expected phenomenon, and it is bound to grow during periods of rain. This module ensures that the packets dropped more often belong to the low priority downlink cells. It is implemented by *providing a* *higher time-to-live for bursts* with higher priority packets than those with lower priority packets. This ensures that if the very high congestion periods are short (short enough so that the time-to-live timer of higher priority bursts does not expire) then the bursts carrying larger proportion of higher priority cells would not be dropped. Although all bursts could have been given large time-to-live periods, this would result in enormous increase in the required buffer sizes, and this is not practically acceptable. The fact that rain could be responsible for enormously reduced service rates for the downlink antennas of satellites operating in the Ka-band, and that this would occur not more than 10% of the time of operation of the satellite, is modeled in this module. The kind of output, which would be available from the NOCC, which is expected to implement the Knapsack algorithm off-line, as discussed in the next two sections, is provided here. Again the asynchronous nature of updates and the random behavior of rain periods is ensured while building the simulation model. This module then transmits the bursts down to the ground, which is modeled by the RECEIVER.

4.1.5. RECEIVER

On receiving the bursts, the RECEIVER collects the statistics such as aggregate-priority which made it to the destination, etc. before destroying the packets.

In the next two sections, the algorithms proposed for generating burst plans in conditions of rain are described. These are the algorithms which would be used in the NOCC and when the plans are significantly different from the existing ones they would be beamed up to the satellite. As explained in the previous chapter we can consider this to be a variant of the 0-1 multiple knapsack problem. As a result, the algorithms described here are standard algorithms used for solving the 0-1 MKP.

4.2. Branch and Bound Approach for solving the 0-1 MKP

One of the algorithms that we evaluated for performance when applied to the rain-fade compensation problem was the Branch and Bound method, applied to the MKP formulation (presented in equation11 to 14, in section 3.10) attributed to Martello and Toth [15]. This is an exact algorithm for this NP-hard problem. Although the problem is NP-hard, it was claimed in [15] that for small enough input sizes (as are available here in our problem) a very close to exact solution can be obtained if the algorithm is prematurely stopped after a few steps. This was supposed to work for input, where profits and weights are random and independent of each other (as would be the case in the rain-fade compensation problem); this kind of input is called *uncorrelated items* in their work [15].

When the same algorithm was applied to our problem with inputs as would be expected of a typical satellite system with n downlink cells and k phased array antennas for providing service, it was found that the running time was enormously longer than claimed in [15]. This could be attributed to a slightly different kind of input provided by us in our experiments, than what was required by the definition of *uncorrelated items* in their work. The input we provided was such that it modeled the system as closely as possible. It should be stated that for certain input sets the performance was in accordance to results provided in their work, but these input sets were not even close to what would be typically expected from the system under study.

Table 4 below tabulates the running time for the exact algorithm. The value for B *indicates the number of back-trackings* after which the algorithm was scheduled to terminate and output the result. In their work it is claimed that for values of B greater than 5, the results were very close to the exact solution; i.e. about .01% error was reported. We run this algorithm on a Sun-Ultra-2.

S. No	M	N	Exact Algo. Time in secs(B)	Approximate Algo. (Time in secs)
1	5	100	00.05 (B = 5)	00.03
2	5	100	Stopped after 3 hours (For exact solution)	00.03
3	9	100	00.09 (B = 50)	00.03
4	5	200	41:00.00 (B = 5)	00.03
5	5	300	1:55:31.00 (B = 5)	00.03
6	20	400	No solution	00.05
7	35	700	No solution	00.07

Table 4: Running times for Branch and Bound Algorithm

Based on the above observation it was assumed that Martello and Toth's exact algorithm would be unable to give results within the desired time for our application (which is a few minutes).

For that reason an approximate scheme – A *Greedy Algorithm* was used. The results obtained by this algorithm were improved using another approach. Both algorithms are described below and the results are discussed.

4.3. THE APPROXIMATE ALGORITHMS FOR SOLUTION TO THE 0-1 MKP

As mentioned before the exact algorithm has such a long running time that it makes it impossible to implement for the problem under study. For that reason the greedy heuristic and improvements proposed by Martello and Toth [15] were applied. The algorithms are given below. In all these algorithms it is required that the items be ordered in non-increasing order of profit per unit weight.

ALGORITHM FOR FINDING THE INITIAL FEASIBLE SOLUTION [16]

Procedure GREEDY (A,c,G)

Input: A = {j : item j is available} (A Í N)

c = available capacity

Output G = {j : item j is inserted in the knapsack} (G Í A)

Set G = 0

For increasing j Î A do

If $c / w_j < 1$*, return;*

If $w_j < c$ *set* $c = c - w_j$*, G = G U {j};*

Repeat.

Return.

In the worst case the procedure requires |A| iterations, that is the number of operations is *O(n)*.

Here *n* is the total number of items, which need to be placed into the *m* knapsacks. These items are ordered in a non-increasing manner in terms of the profits per unit weight, i.e.

$$p_1/w_1 \geq p_2/w_2 \geq \ldots\ldots \geq p_n/w_n.$$

Then the following algorithm which is referred to as *MK1* in [16] can be used to obtain an initial feasible *solution* G_i $(i = 1, 2, \ldots. M)$ to the 0-1 MKP.

ALGORITHM MK

Set R=N

For increasing $i \in M$ *do*

Apply GREEDY(R, c_i, G_i)

Set $R = R - G_I$

Repeat.

Stop.

MK applies *GREEDY* *m* times, that is it requires in the worst case *O(mn)* operations.

For improving the solution obtained above from algorithm *MK* the following algorithm was used [16].

The following definitions are used:

$S = \bigcup_{i=1}^{m} G_i$;

$R = n - S$;

c_i = *current remaining capacity of the* i^{th} *knapsack;*

g_j = *knapsack where item j is currently inserted.*

Furthermore, *S* and *R* (subsets of *n*) are assumed to be in non-increasing order of p_n/w_n.

In the following algorithm *S, R,* c_i and g_j are updated whenever some G_i is updated. This is referred to as the algorithm *I1* in [16].

ALGORITHM I

For increasing $j_1 \in \left\{ j : j \in S, c_{g_j} + \max\{c_i\} \geq \min_{r \in R}\{w_r\} \right\}$ *do*

For increasing $j_2 \in \left\{ j \in S, j > j_1, g_j \neq g_{j_1}, c_{g_j} + c_{g_{j_1}} \geq \min_{r \in R}\{w_r\} \right\}$ *do*

Let $w_{j_u} = \max\{w_{j_1}, w_{j_2}\}$, $w_{j_q} = \{w_{j_1}, w_{j_2}\}$, $i_u = g_{j_u}$, $i_q = g_{j_q}$;

Set $d = w_{j_u} - w_{j_q}$;

If $d \leq c_{i_q}$ *and* $c_{i_u} + d \geq \min_{r \in R}\{w_r\}$, *set*

$p_t = \max\{p_r : r \in R, w_r \leq c_{i_u} + d\}$,

$G_{i_u} = (G_{i_u} - \{j_u\}) \cup \{j_q, t\}$, $G_{i_q} = (G_{i_q} - \{j_q\}) \cup \{j_u\}$;

repeat;

repeat.

Stop.

Algorithm *I* considers all pairs of items in S and, if possible, interchanges them whenever doing so allows the insertion of an item from R into one of the knapsacks; the algorithm's complexity is $O(n^2)$ (the search for p_t is executed at-most |R| times). It is computationally efficient to stop the execution as soon as the sum of the remaining capacity of any 2 knapsacks is less than the minimum weight of the items not yet included.

ALGORITHM A

Set c_i *(i = 1, 2, … ,m) to its original value.*

Set G_i *= 0 for all i Î m; i = 1;*

For decreasing j Î S do

Set k = 1;

While $w_j > c_i$ *and k£ m do*

Set k = k+1, i= i+1; if i > m, set i = 1;

Repeat;

If $w_j < c_i$ *, set* $c_i = c_i - w_j$ *,* $G_i = G_i$ *È {j}, i = i + 1; if i > m, set i = 1;*

Repeat.

For increasing i Î m do

Apply GREEDY (R, c_i*, L);*

Set R = R – L, $G_i = G_i$ *È L;*

Repeat.

Stop

The complexity of Algorithm A is *O(mn)* operations. A requires an extra computation, but according to [16] it improves the performance considerably. This performance improvement was not observed when the algorithm was implemented on the input set of a few hundred downlink cells and $m < 50$, as applicable for a typical system under study. However, because it did not add much to the overall complexity of the system, all results using the approximate method are obtained with *'Algorithm A'* being implemented.

Algorithm *MK* along with Algorithms *A* and *I* give solutions very close to the optimal. They were tried on the input sets used for the exact algorithms and proved to be quite efficient in terms of running times as well. As the nature of the input is statistical, because it is available at the NOCC, where it was derived from databases and other non-real-time information about the queue sizes on-board the satellite, the effect of the slight error in the solution would be minimal.

Table 5 below includes results on the performance of the algorithms on a wide variety of input sets. Also shown in this table is the effect of increased granularity of power levels. The aggregate-priority column, in the table, demonstrates this effect. When the granularity is finer (i.e. more levels (quantized values) for allocation of power) the aggregate-priority is higher. This is because with finer granularity we can fill the knapsacks better and hence make better use of the total power available. The table was put together after averaging over a large number of runs.

Power in X levels	Running time of algo.	System Power	Power required	Power used	Aggregate Priority input	Aggr. priority output	Percent Priority output
X = 3	30 msec.	28 KW	57680 W	28 KW	76131	65133	85.55
X = 6	40 msec.	28 KW	40840 W	28 KW	76131	72697	95.49
X = 10	30 msec.	28 KW	38620 W	28 KW	76131	74613	98.01

Table 5: Illustration of the performance results of the algorithms used

The relation between the two key problems (power allocation and burst scheduling) was described in the last part of the previous chapter (see section 3.12). The solution to the scheduling problem, which involves sorting and the 'timer', is implemented in the DL-Q and SCH modules of the simulation model built in OPNET. The algorithms described above for solving the rain-fade compensation problem are emulated in the DL-BEAMS module. The graphs, which show the performance of these two schemes are provided in the next section.

4.4. OPNET SIMULATION RESULTS

In the graphs given on the next few pages, we have tried to demonstrate the performance characteristics and results of the algorithms implemented in the 'DL-Q and SCH' and the 'DL-BEAMS' module of the simulation model for periods marked by different types of rain. We will use the following terminology:

A *higher intensity* of rain means more downlink cells experience rain events simultaneously.

A *heavier rain* means that transmission to the cell experiencing heavy rain would require a greater amount of power to compensate for good C/N than a cell experiencing lighter rain.

Based on these definitions, a Ka-band system could experience periods of various rain and traffic conditions as listed in Table 6 below. For each case, Table 6 catalogues the appropriate figures and tables with results and performance.

Case 1	Low Intensity, Light rain	Refer to Figure 10 & Table 9
Case 2	Low Intensity, Heavy rain,	Refer to Figure 11 & Table 10
Case 3	Higher intensity, Light rain,	Refer to Figure 12 & Table 11
Case 4	Higher intensity, Heavy rain	Refer to Figure 13 & Table 12

Table 6: Listing of figures and tables describing our results for the various conditions listed.

The systems being planned for the Ka-Band would normally accommodate for case 1. It is the occurrence of the other cases that could lead to system overload, in which case the network managers would typically want to shed load to make the system stable. One way to shed load is to allow packets to be dropped from the tail of the queues, without taking into consideration the priority of the packet that gets dropped. This is the case against which we have compared the performance of the algorithms proposed here. The performance graph for this case is shown in Figures 14, 15 and Tables 13**,** 14 on pages 71 and 72. Below each graph there is a table, which shows what percentage of packets from different priorities made it to the ground until that time when the simulation was terminated.

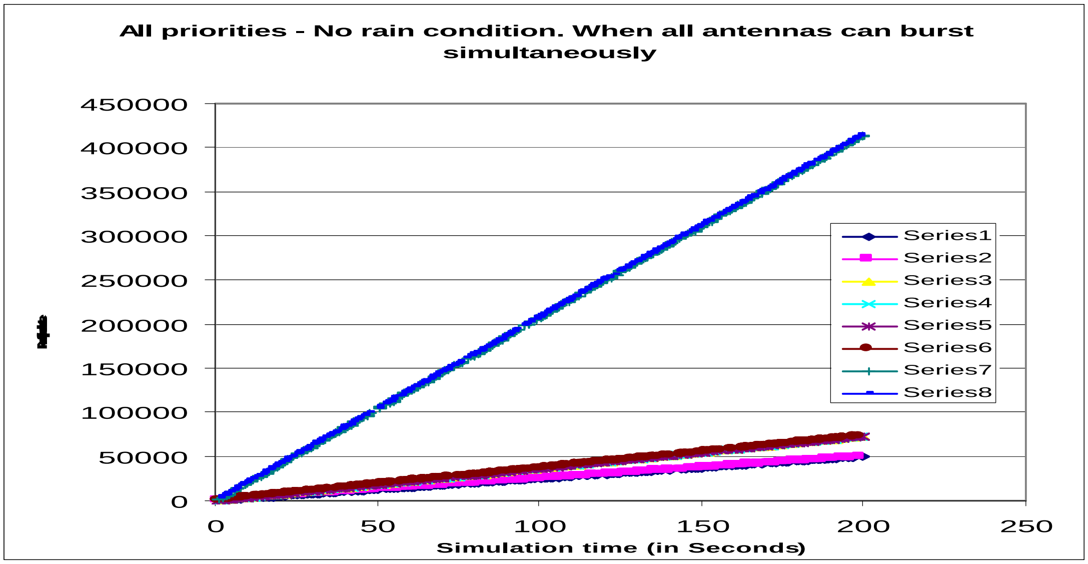


Figure 7

No Rain	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
	100.00	100.00	100.00	100.00	100.00

Table 7

As the graphs of Figure 9 and Table 7 show, all packets were received at the ground, which implies that in condition of no-rain, the system is modeled to be stable and all packets received at the beams are transmitted. The legend for the graphs in Figure 9, refers to *Series 1 through 8.* These represent the number of packets for various priority classes received at the satellite and out of those the number which are transmitted to the ground (same for all graphs on the following pages).

Series number	*Packet count for:*	*Series number*	*Packet count for:*
Series 1	Ground - Priority 4	Series 2	Satellite - Priority 4
Series 3	Ground - Priority 3	Series 4	Satellite - Priority 3
Series 5	Ground - Priority 2	Series 6	Satellite - Priority 2
Series 7	Ground - Priority 1	Series 8	Satellite - Priority 1

Table 8

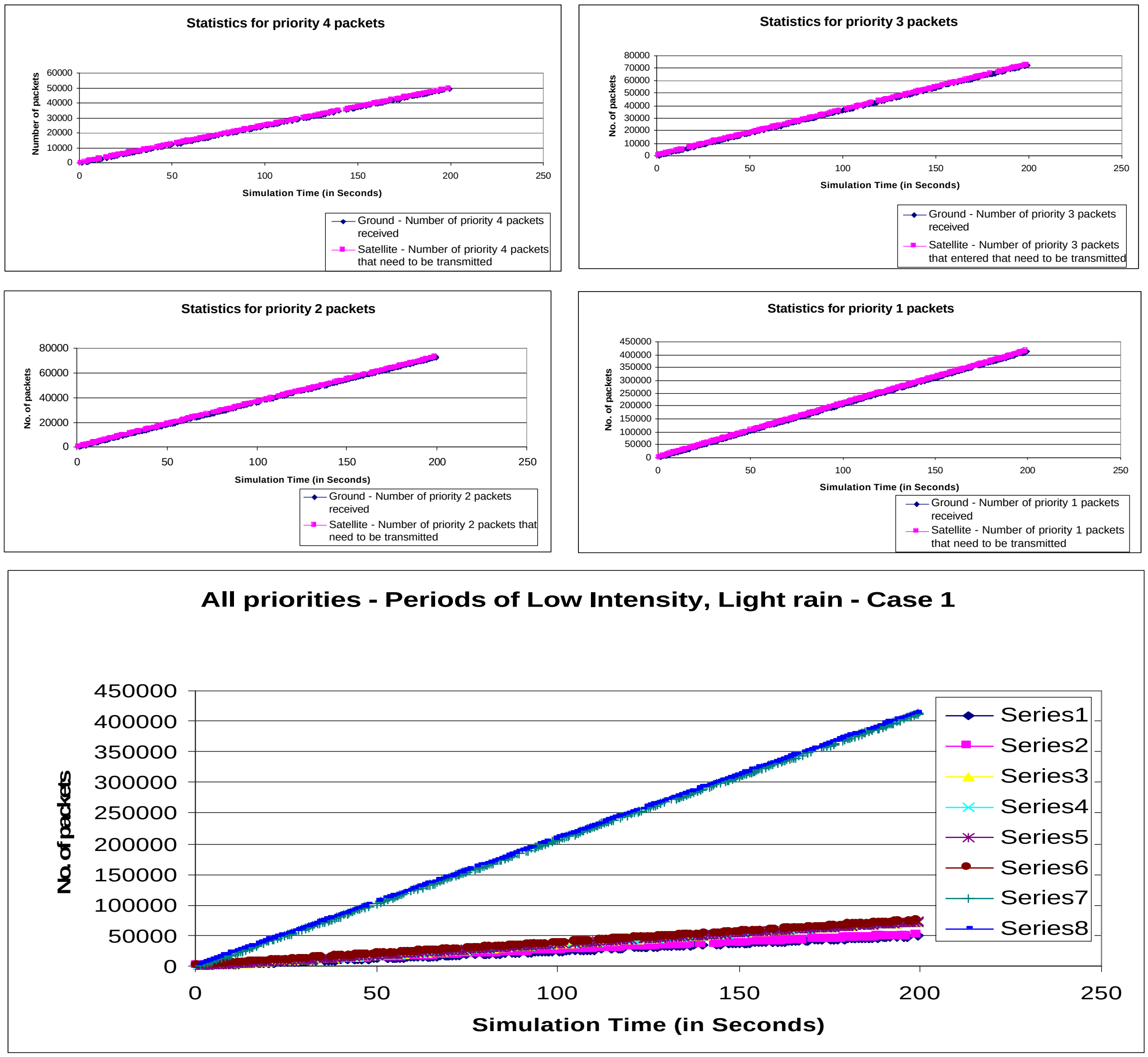


Figure 8

Case 1	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo. turned on	100.00	100.00	99.45	99.42	99.69

Table 9

Figure 10 and Table 9, describe our simulation results for the case of low intensity, and light rain periods. As can be observed almost all packets make it to their destinations.

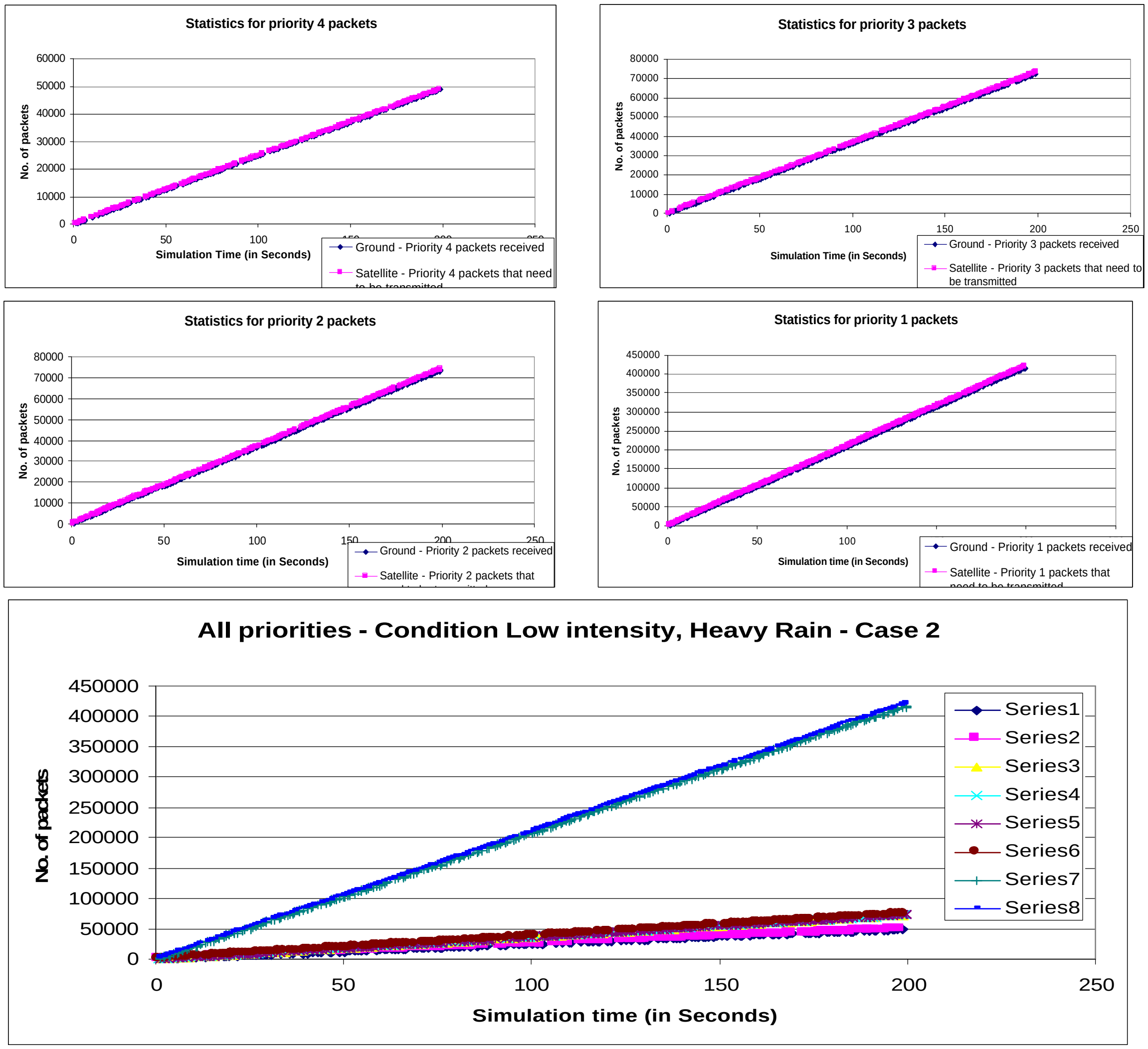


Figure 9

Case 2	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo. turned on	100.00	99.18	99.19	98.72	99.4

Table 10

In periods of low intensity, and heavy rain, the effect of the algorithm starts to become evident, as seen from our results summarized in Figure 11 and Table 10. Highest priority packets manage to reach the ground, while lower priorities begin to experience greater loss.

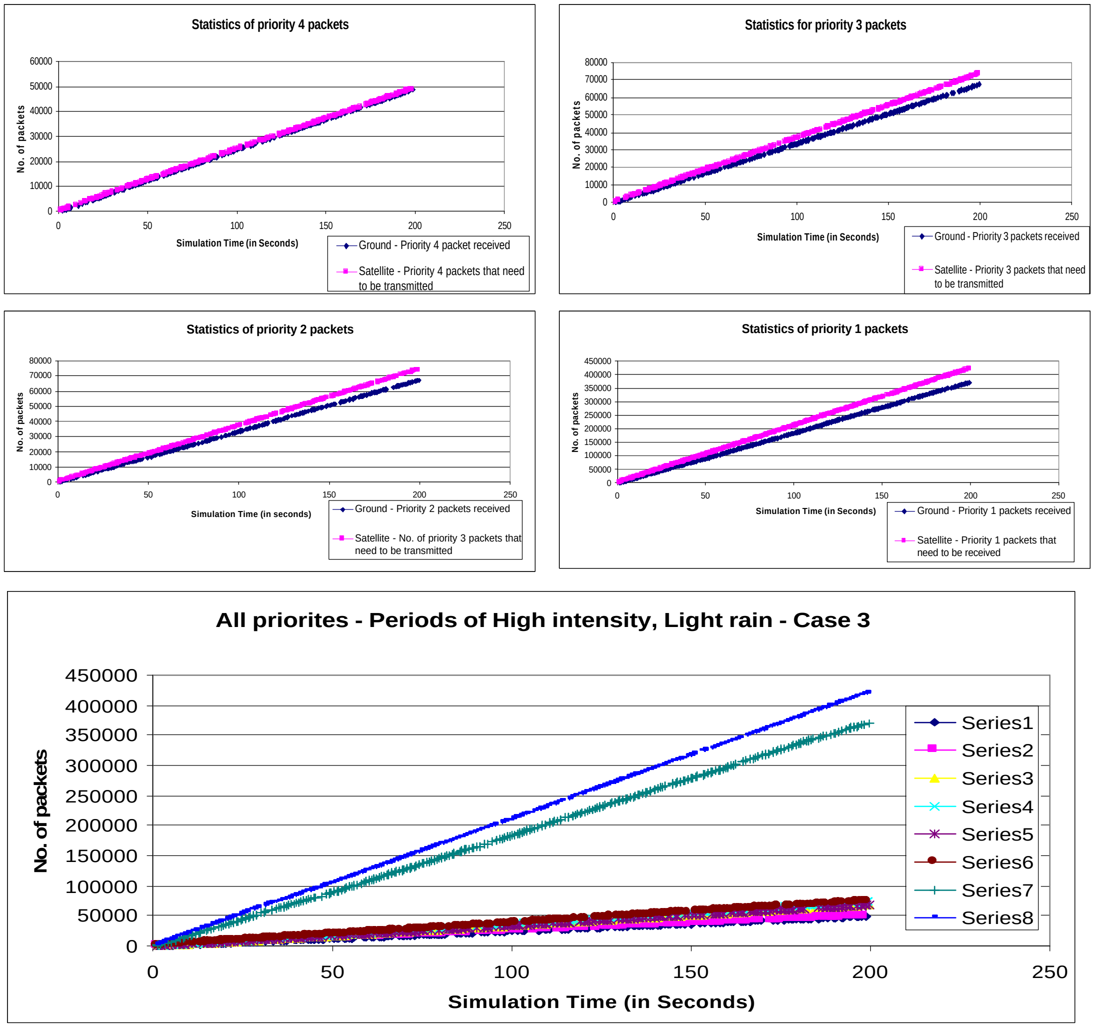


Figure 10

Case 3	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo. turned on	100.00	92.08	90.54	88.15	91.78

Table 11

During periods of high intensity and light rain, even greater percentage of lower priority packets are getting dropped on board the satellite; as evident by our results summarized in Figure 12 and Table 11. The loss for each priority type is lower than the loss for priority lower than itself. Priority 1 experiences the largest loss.

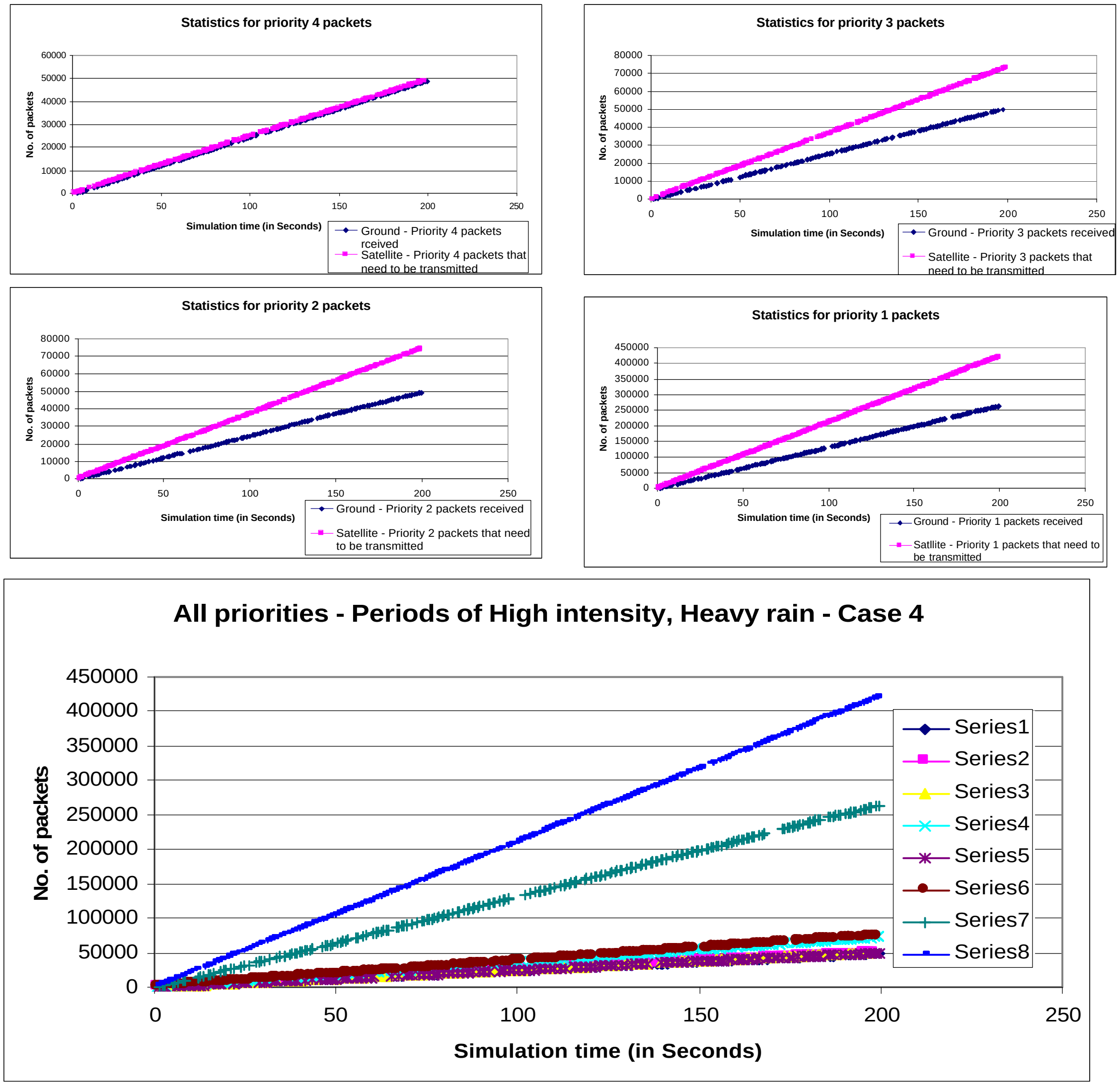


Figure 11

Case 4	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo turned on	99.59	69.06	66.49	62.46	71.88

Table 12

These periods of high intensity and heavy rain are the most severe. Even the highest priority packets experience loss, although the loss is minimal (less than 1%). Lower priority types are dropped extensively. Less than 72% percent of aggregate priority makes it to the destination. These are illustrated in Figure 13 and Table 12 summarized in our results.

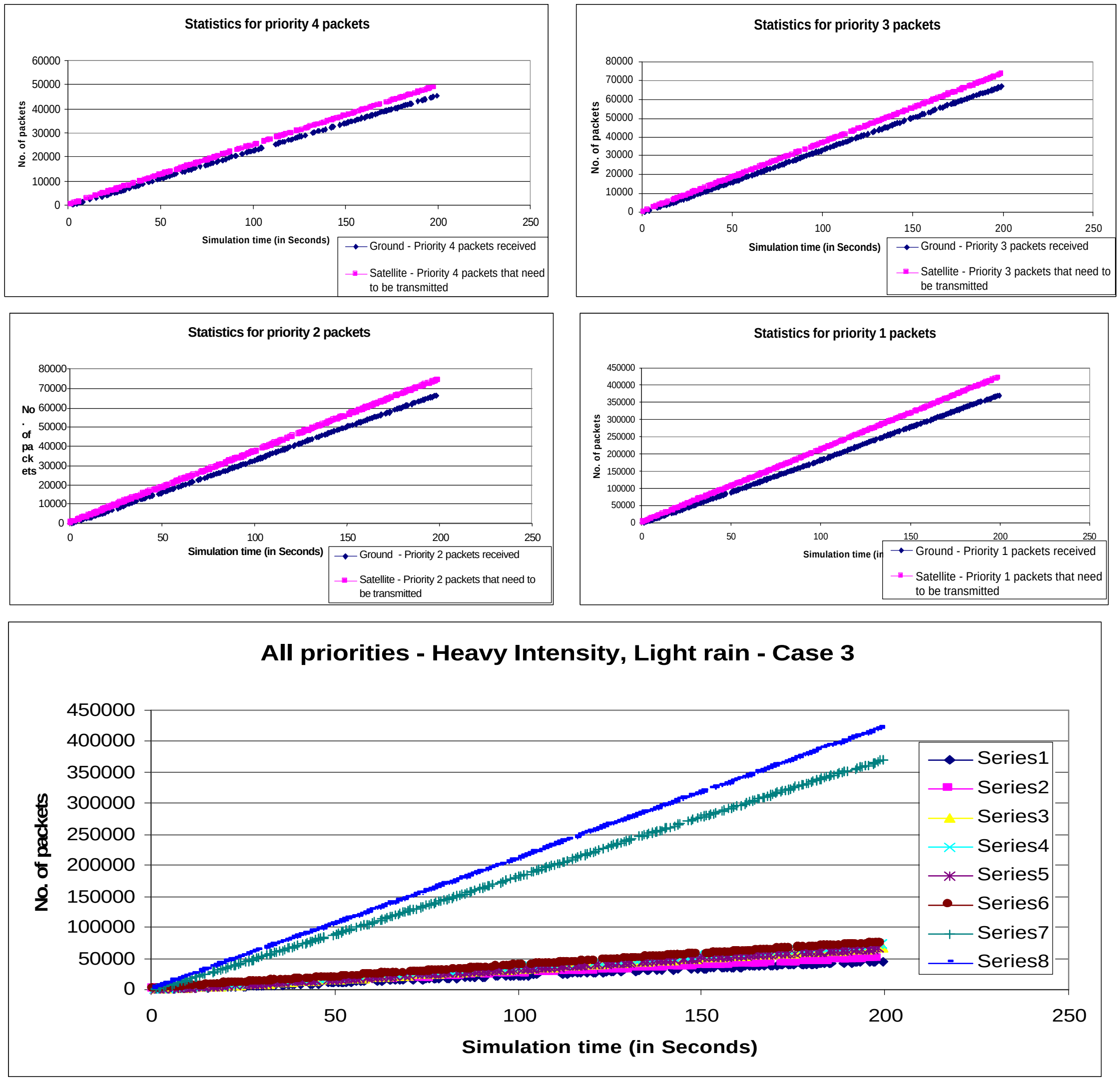


Figure 12

Case 3	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo turned off	93.03	91.53	90.16	87.86	90.05

Table 13

The graphs of Figure 14 and the results of Table 13 are for a case equivalent to case 3 shown before (i.e. high intensity, light rain) except that our algorithm is not at work. As can be seen the highest priority packets are not treated with any preference. It is also seen that the aggregate priority that makes it is slightly less than the one achieved in the case 3, when our algorithm is at work, as shown in Figure 12 and Table 11.

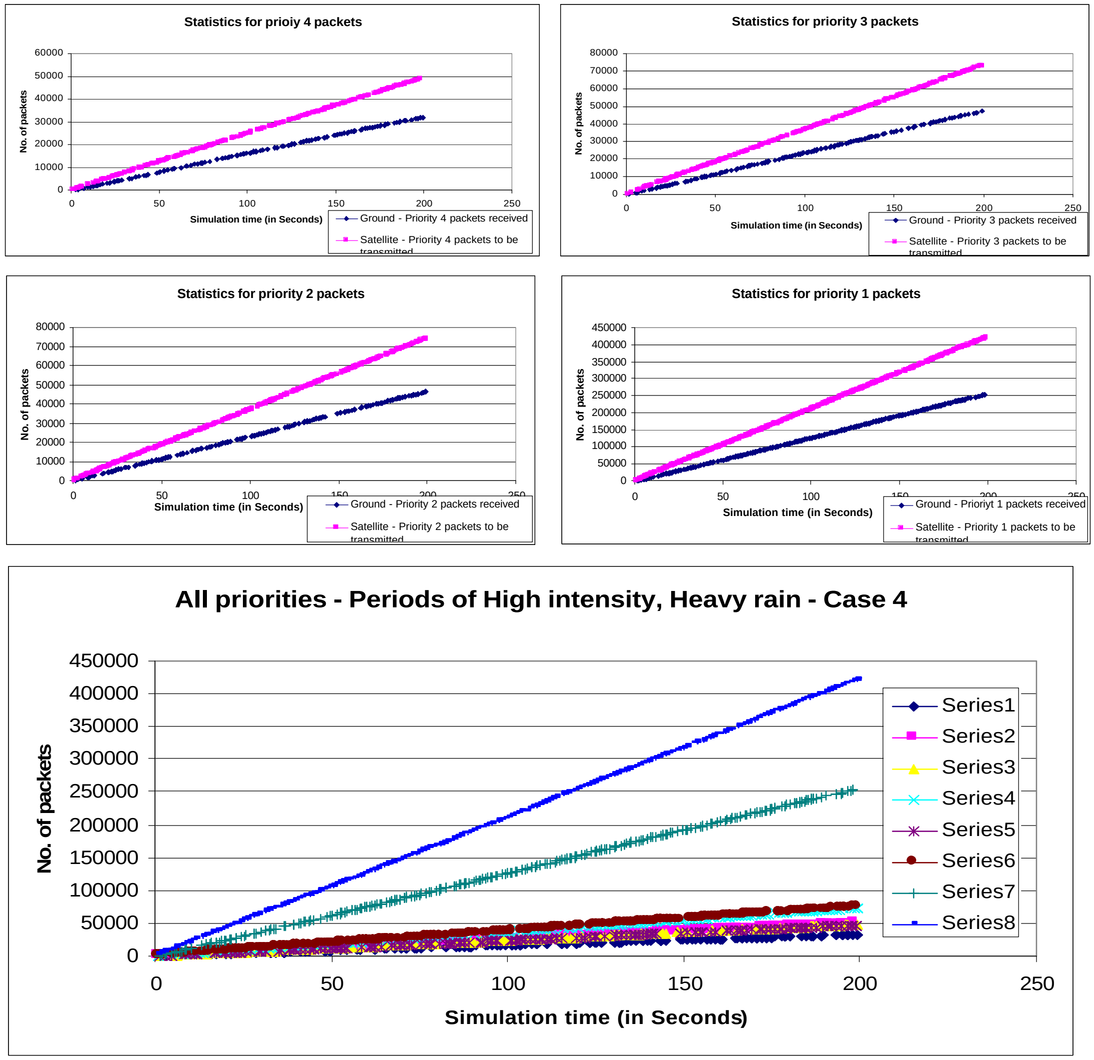


Figure 13

Case 4	Priority 4	Priority 3	Priority 2	Priority 1	All Priorities
Algo. turned off	65.57	64.48	62.77	60.47	62.76

Table 14

The graphs of Figure 15 and the results of Table 14 are for the case of high intensity and heavy rain (case 4) but when the algorithm has been turned off; the effect can be seen to be quite dramatic. The equivalent case to the one shown above with our algorithm on is summarized in Figure 13 and Table 12.

From this set of graphs, it can be seen that our effort to use and design an algorithm to increase the aggregate priority throughput, which can be defined as the sum of the number of packets for each priority multiplied by the weight attached with each priority, was successful. The aggregate priority throughput achieved is the value that is represented in the last column of each of the tables shown on the previous few pages.

At the same time our algorithm was designed to ensure that the highest priority packets achieve more than 99% throughput. Lower priority packets could be given similar, but lower, guarantees as well.

5. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

On-board resource sharing is a method wherein, for increasing effective isotropic radiated power, surplus satellite resources (radiated power, or time slots in TDMA) are used to maintain communications quality in areas affected by rain attenuation. This is done so that the communication quality is maintained.

In this work we considered satellite power as the resource which can be optimally shared to increase system capacity. We do not consider the power as a surplus that should be used only at times when the compensation is required. Rather we consider it as a resource, which that should be used to the full capacity at all times. This in turn implies that in times of rain, power reallocation would be required based on some intelligent scheme, which maximizes profit. The profit, in the simplest form, can be considered as the number of high priority packets being transmitted, or the *aggregate priority* (which is computed by multiplying the number of packets received at destination with the profit variable associated with each packet's priority) received at the ground. This is the *'figure of merit'* or performance metric we used in this work. Based on this metric we evaluated some schemes proposed here.

Rain-fade compensation schemes have been derived from the solutions to the well-known 0-1 Multiple Knapsack Problem. We also proposed a mechanism for ensuring that aggregate –priority is maximized based on a modified Weighted Round Robin scheduler with a simple implementation of a timer. This method was simulated using the OPNET simulation package for scheduling of burst-time allocation for the traffic segregated on the basis of the downlink cells/destinations. We also developed a Linear Programming formulation of this problem.

The various algorithms and schemes have been developed and described based on the assumption that they would be implemented at the NOCC. For this system to work effectively it is required that certain information be available at the NOCC as function of time. The sources of the information needed are:

1. Rain Models
2. Temporal and Spatial Databases which maintain attenuation statistics on a spot-beam basis along with the priority information of the traffic in the spot-beam on a hourly basis.
3. Polling statistics from the User Terminals.
4. The weather radar data – 6 minute data.
5. Information about the dynamics of on-board queues, provided by the satellite to the NOCC via the control channel.
6. Temporal and Spatial Databases which maintain information about variations of traffic volume and type for various downlinks.

A framework for obtaining the required information and using it to generate the power and burst allocation plans was also developed.

It was found that the algorithm designed in this thesis performed reasonably well under simulated conditions. As an attempt was made to build the simulation model as close to an actual system as possible, it would be reasonable to assume that the developed scheme would work well on the real system as well.

The forthcoming high-performance satellite constellations introduce many interesting and significant problems related to resource allocation of satellite resources. The research and problems described here addressed only a few of these problems. Much future work is needed. We describe below some such interesting problems that we have identified.

Future work in this area of resource allocation for Ka band, could be taken up in coupling the uplink resource allocation (which also includes spectrum management) with that done for the downlink channel. Such work would have to be done for specific systems, as there is a large amount of system specifications, which becomes critically important for this problem.

Another interesting problem is to jointly consider adaptive and dynamic adjustments to coding, modulation and power allocation in the various beams. Adaptive frequency allocation in the uplink is also an interesting problem.

The simulations presented in this work have been carried out using a specific traffic model. As the traffic type and traffic volume changes over the next few years the traffic models which accurately model that traffic type would also change. Future research is needed in a similar direction but with more appropriate traffic models.

In this thesis 0-1 MKP algorithms have been used without specific details about the system under study. Although it has been shown that the algorithms work well for a generic system, they might perform better if some system specifications have been accounted for in the implementation of the algorithm.

Even more interesting problems arise when we consider other types of satellite constellations like LEO or ICO networks, where satellite dynamics further complicate dynamic resource allocation.

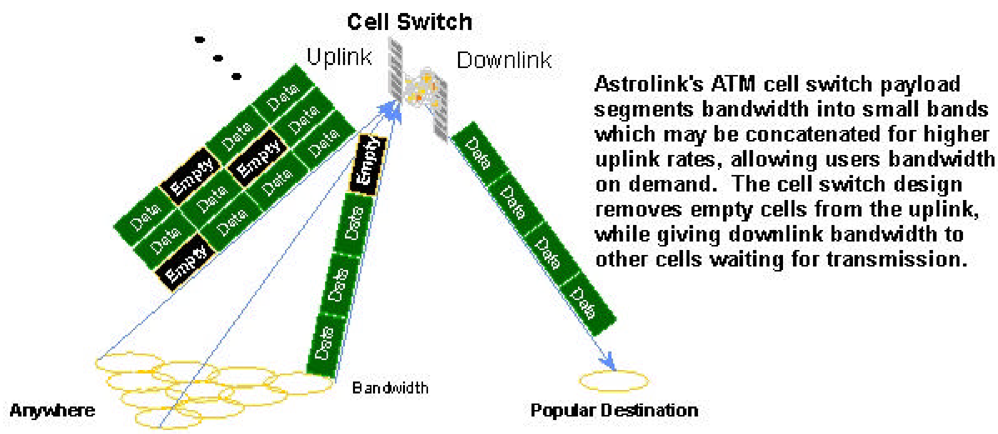
We have used weights associated with priorities in our formulation. Adjustments in predictions of these weights based on real-time data, is a very interesting problem.

The aggregate priority throughput used in this thesis is one of the possible metrics for evaluation one can consider. We could consider revenue (assuming a pricing scheme), response times etc. as alternative metrics. Investigation and evaluations of trade-off is a significant problem. Linking the weights to such metrics or trade-off is also of great value.

There is spatial and time correlation of rain events over continental USA. Such correlation can be exploited in more dynamic algorithms.

6. APPENDIX A

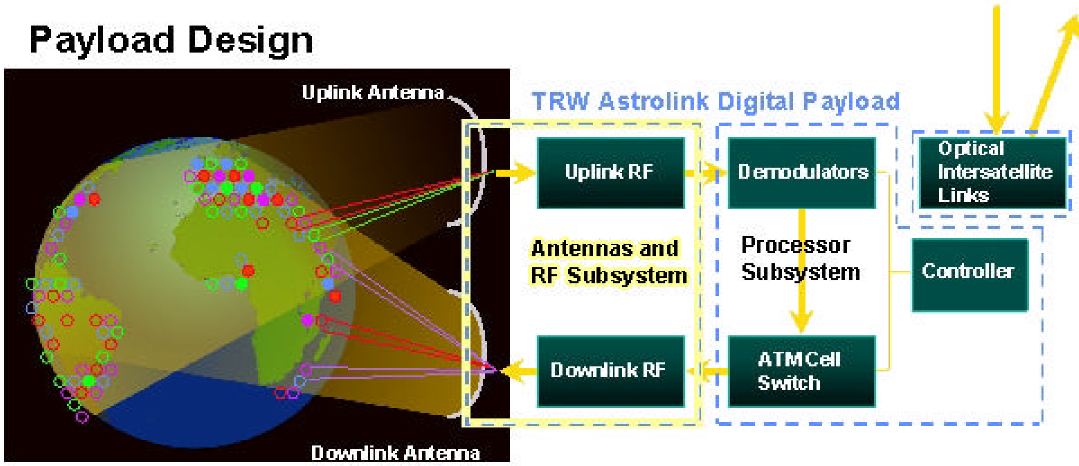
The next few figures show the high-level design details of the Ka-band Astrolink system. This system is set to launch in 2002. More details can be obtained from the web-site http://www.trw.com/astrolink/index.html.

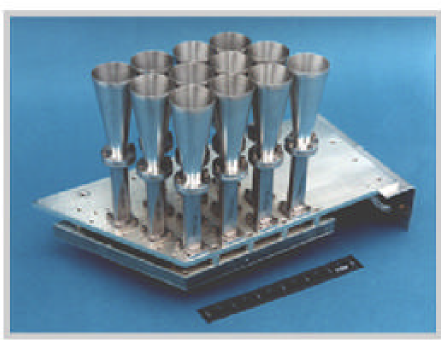


Multibeam Antenna

This TRW Ka-band antenna delivers multibeam coverage for an existing on-orbit bandwidth on demand system. The antenna features integrated recieve electronics for 13 narrow beams, providing high gain and polarization purity with low sidelobe levels. The Astrolink antenna is a next-generation version and provides coverage with many more beams.

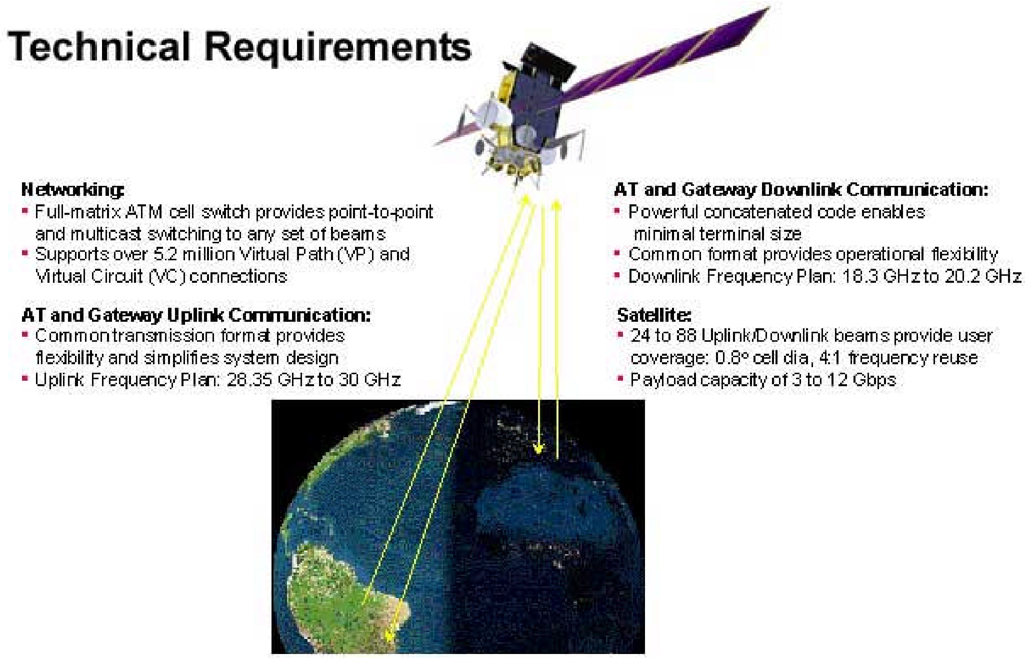
Payload Design





Multibeam Antenna Feed Assembly

TRW's feed assembly design (patents pending) combines high performance with low cost manufacturing techniques. The feed assembly produces closely packed ground cells, yet ensures high levels of isolation for users in adjacent beams.



Some of the technical specifications for this planned system are:

- Full-matrix ATM cell switch provides point to point and multicast switching to any set of beams.
- Support for over 5.2 million Virtual Path (VP) and Virtual Circuit (VC) connections
- Uplink frequency plan: 28.35 Ghz to 30 Ghz
- Powerful concatenated codes to enable minimal terminal size
- Common transmission format provides flexibility and simplifies system design
- Downlink frequency plan: 18.3 Ghz to 20.2 Ghz.
- 24 to 88 uplink/downlink beams provide user coverage: 0.8° cell diameter, 4:1 frequency reuse.
- Payload capacity of 3 to 12 Gbps.

7. REFERENCES

[1] R. J. Acosta; Rain fade compensation alternatives for Ka band communication satellites; NASA TM-107534; 1996.

[2] R. J. Acosta, et.al; Advanced Communication Technology Satellite (ACTS): Four-year system Performance; IEEE Journal on Selected Areas in Communications, Vol 17, No. 2, February 1999, pp. 193-203.

[3] I. F. Akildiz et. al. Satellite ATM networks: A survey; IEEE Communications Magazine, July 1997, pp. 30-43.

[4] T. K. Apostolopoulos, V.C. Daskalou; Temporal network Management informational model and services; Journal of Network and System management, Vol 6, No. 3, 1998, pp. 254-265.

[5] F. Barbaliscia, G. Ravaioli, A. Paraboni; Characteristics of the spatial Statistics dependence of rainfall rate over large areas; IEEE Transactions on Antennas and Propagation, Vol 40, No. 1, January 1992, pp. 8-12.

[6] A. Baiocchi, m. Listanti; An ATM like system Architecture for Satellite Communication including on-board switching; International Journal of Satellite Communications, Vol 14, 1996, pp. 389-412.

[7] M. A. Bergamo, D. Hoder; Gigabit satellite network using NASA's Advanced Communications Technology Satellite (ACTS): Features, Capabilities, and Operations; NASA TM

[8] D. P. Connors, B.Ryu, S. Dao; Modeling and simulation of broadband satellite networks: Part I, Medium Access Control for QoS provisioning; IEEE Communications Magazine, March 1999, pp. 72-79.

[9] R. K. Crane; Prediction of Attenuation by rain; IEEE Transactions on Communications, Vol Com 28, No. 9 September 1980, pp. 1717-1733.

[10] F. Gargione et.al; Services, technologies and systems at Ka band and beyond – A survey; IEEE Journal on Selected Areas in Communications, Vol 17, No. 2, February 1999, pp. 133-144.

[11] J. Gilderson, J. Cherkaoui; Onboard Switching for ATM via satellite; IEEE Communications Magazine, July 1997, pp. 66-70.

[12] J. Huan, T. Le-Ngoc, J.F. Hayes; Broadband satcom system for multimedia services; IEEE International Conference on Comm., ICC '96, pp.906-910.

[13] H. Koraitim, S. Tohme'; Resource allocation and connection admission control in satellite networks; IEEE Journal on Selected Areas in Communications, Vol 17, No. 2, February 1999, pp. 360-371.

[14] B. H. Mah; Measurement and observations of IP multicast traffic.

[15] S. Martello and P. Toth; Knapsack Problems, Algorithms and Computer Implementations;1990

[16] S. Martello, P. Toth; Heuristic Algorithm for the Multiple Knapsack Problem; Computing 1981.

[17] D. McCraw; Business model for Broadband Satellite Communication Services; International Mobile Satellite Conference, 1995, pp. 115-121.

[18] M. Moher; Power and Bandwidth tradeoffs for a third generation satellite system; International Mobile Satellite Conference, 1995.

[19] Y. Motoyama et.al.; Satellite Network Control System for SS-TDMA; IEEE Global Telecommunications Conference,1993, pp. 1257-1262.

[20] T. Matsudo et. al; Effect of compensation for rain attenuation on earth satellite links by Onboard resource sharing using AMeDAS Data; Electronics and Communications in Japan, Part 1, Vol. 77, No. 7, 1994, pp. 70-79.

[21] W.L. Stutzman, T. Pratt, A.S.Jazi; Results from the Virginia Tech propagation experiment using the Olympus satellite 12, 20 and 30 Ghz beacons; IEEE Transactions on Antennas and Propagation, Vol 43, No. 1, January 1995.

[22] K. Ueno; Phased array feed for communication satellite antennas; American Institute of Aeronautics and Astronautics, 1998, pp. 100-104.

[23] http://www.ptc.org/pubs/ptr/sept95/spaceway.html – SPACEWAY: Providing Affordable and Versatile Telecommunication Solutions; article by E. J. Fitzpatrick