***Α.Τ.Ε.Ι ΚΡΗΤΗΣ***
***ΠΑΡΑΡΤΗΜΑ ΧΑΝΙΩΝ***
***ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΗΣ***


*Διπλωματική Εργασία*

*Σταματάκη Ελένη*

***Επιβλέπων:*** *Πολυχρόνης Κουτσάκης*

*Χανιά Νοέμβριος 2004*

# *Multiple Access Control Protocols for Mobile Satellite Networks*

*Stamataki Eleni*

# Contents

*Αφιερωμένο στους*
*γονείς μου…*

# Ευχαριστίες

***Για την πραγματοποίηση της πτυχιακής αυτής θα ήθελα να ευχαριστήσω αρχικά τον επιβλέποντα καθηγητή μου, Πολυχρόνη Κουτσάκη,ο οποίος με την σωστή καθοδήγησή του και την υπομονή του με βοήθησε να φέρω σε πέρας την πτυχιακή μου εργασία.***

***Επιπλέον θα ήθελα να ευχαριστήσω όλους τους καθηγητές μου που μου στάθηκαν και με στήριξαν αυτά τα τέσσερα φοιτητικά χρόνια. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον κύριο Νίκο Φραγκιαδάκη ο οποίος με βοήθησε να πιστέψω στον εαυτό μου και να προχωρήσω δυναμικά μπροστά.***

***Επίσης, θα ήθελα να ευχαριστήσω τους συμφοιτητές μου που με βοήθησαν σε οτιδήποτε τους χρειάστηκα είτε αυτό ήταν μεγάλο ή μικρό. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Παντελή Ασκορδαλάκη που τα δυο τελευταία χρόνια είναι κοντά μου και με υποστηρίζει σε οτιδήποτε και αν κάνω.***

***Τέλος, οφείλω ένα μεγάλο ευχαριστώ στους γονείς μου και τον αδελφό μου οι οποίοι πάντα βρίσκονται κοντά μου, σε κάθε επιλογή και απόφαση που παίρνω για την ζωή μου. Τους ευχαριστώ πολύ για την ηθική και ουσιαστική βοήθεια σε όλη την διάρκεια της φοίτησης μου στο τμήμα Ηλεκτρονικής...***

*Σας ευχαριστώ όλους…..*

# Problem Statement

This literature review aims at presenting the significant body of work of the last two decades on the subject of Multiple Access Protocols for Mobile Satellite Networks.

The thesis is organized as follows. ***Chapter 1*** is an introduction into some basic elements of satellite systems. ***Chapter 2*** discusses the most important Quality of Service parameters in ATM Networks. ***Chapter 3*** presents a brief discussion on the most important Multiple Access Control protocols for satellite networks.

# Εισαγωγή

Σκοπός της βιβλιογραφίας αυτής είναι να παρουσιάσει ένα σημαντικό μέρος δουλειάς που έχει συγκεντρωθεί τις δυο τελευταίες δεκαετίες πάνω στο αντικείμενο των MAC πρωτοκόλλων (Medium Access Protocol- Πρωτόκολλο Πολλαπλής Πρόσβασης στο Μέσο).

Η βιβλιογραφία αυτή χωρίζεται σε τρία κεφάλαια στα οποία θα βρείτε πλούσιο υλικό όσον αφορά τα Δορυφορικά δίκτυα. Συγκεκριμένα, στο ***Κεφάλαιο 1*** γίνεται μια εισαγωγή στα βασικά στοιχεία που περιλαμβάνουν τα δορυφορικά δίκτυα. Συνεχίζοντας στο ***Κεφάλαιο 2*** παρουσιάζονται οι πιο βασικοί παράμετροι της Ποιότητας των Υπηρεσιών (QoS) στα ATM (Asynchronous Transfer Mode- Ασύγχρονος Τρόπος Μετάδοσης) δίκτυα. Τέλος, η πτυχιακή αυτή εργασία ολοκληρώνεται με το ***Κεφάλαιο 3,*** στο οποίο παρουσιάζονται τα νέα MAC πρωτόκολλα για τις δορυφορικά δίκτυα.

Η βιβλιογραφία αυτή περιέχει έγχρωμες εικόνες & πίνακες καθώς επίσης και μια μεγάλη λίστα των πηγών στις οποίες ανατρέξαμε για να συγκεντρώσουμε το υλικό που κρατάτε σήμερα.

# Chapter 1
# Satellite Systems

## 1.1 Introduction

A satellite is a specialized wireless receiver/transmitter that is launched by a rocket and placed in orbit around the earth. There are hundreds of satellites currently in operation. They are used for such diverse purposes as weather forecasting, television broadcast, amateur radio communications, Internet communications, and the Global Positioning System (GPS).

The concept of providing a global radio communications network using satellite was first proposed in 1945 [1], but it was only after the Sputnik satellite was launched by the Russians in 1957 that interest in satellite communications began to rise [2]. Experimental satellites were launched in the early 1960's and the feasibility of space-based communications was demonstrated, partly through successful television coverage of the Tokyo Olympics in 1964. Numerous commercial satellite systems were launched throughout the 1970's in a Geostationary Earth Orbit (GEO) to provide broadcast television and telephony services. Satellite systems had a much larger capacity for carrying telephony traffic across the Atlantic than the trans-Atlantic cable at the time, and the inherent broadcast capabilities made satellites ideal for television distribution.

Technological developments in power amplifier design and the introduction of high gain antennas enabled much higher data rate communications from small ground-based terminals. Very Small Aperture Terminals (VSATs) appeared in the 1980's combined alongside television broadcast. Recently, a significant number of proposals has been made for satellite systems in a Low Earth Orbit (LEO) [3]. These systems aim at taking advantage of the lower signal propagation delay and reduced free space path loss in LEO satellite networks, thus enabling higher data rate communications, improved Quality of Service (QoS) for real-time applications, and increased capacity due to spot-beam architectures and frequency reuse.

## 1.2 The Nature of Satellite Communications Systems

The primary differences between satellite systems and other wireless systems result from the unique location of satellite in space. Satellites are particularly suited to certain types of scenarios and services and less suited to others, although in some instances they represent the only means of providing communications access to terrestrial networks such as Internet.

### 1.2.1 Advantages of Satellite Communication Systems

➯ **Coverage:** A single geostationary satellite can cover up to a third of the Earth's surface allowing users to communicate with each other from distant locations around the globe. In sparsely populated or remote areas there may be little or no terrestrial communications provisioning, the local terrain may make it difficult to install a ground-based network or it may not be economically feasible. In these environments, satellites may represent the only viable means of communications.

➯ **Broadcast / Multicast Capability:** The unique location of satellites enables direct communication access to and from a large potential user population , ideal for broadcast or multicast applications as many users can listen to a common signal on a common channel without replication of the information for each individual user. A typical satellite transponder bandwidth is 72 MHz with recently proposed transponders having bandwidths as large as 1GHz [4]. It is therefore possible to offer very high data rates to subscribers when providing a limited number of broadcast services.

➭ **Availability:** Terrestrially-based wireless communication links can be severely affected by multipath fading, and it can be difficult to provide the direct Line of Sight (LOS) links required by the higher frequency bands. Satellite communications channels suffer to a lesser extent because the majority of the links between the ground-based users and the satellites are at high elevation angles, enabling LOS communications not blocked by hills, buildings, moving vehicles, etc.

➭ **Bypass:** Individual users can communicate over large distances via satellite without the need to involve national telecommunication companies. This may be of significant benefit in environments where the local service providers are hostile or too costly.

➭ **Configuration Flexibility:** Once the satellite network is in operation, the ground-based network configuration can be easily adapted to accommodate new users and remove others from the system. The network management software is responsible for admitting and closing connections.

➭ **Cost Independence with Distance:** The cost of providing a user with access to a satellite resource is independent of location. This is in contrast to terrestrial links where the installation costs are proportional to the distance from the service provider.

➭ **Reliability:** Satellite links only require the end user terminals to be maintained and they are less prone to disabling through accidental or malicious damage.

## 1.2.2 Disadvantages of Satellite Communications Systems

✐ **Cost:** It is very expensive to put satellites into space. The LEO based proposals are designed to incorporate as many as 80 satellites and the entire network must be deployed prior to service provision, requiring a huge investment [5]. This expense may prove to be the deciding factor in the success or failure of these ambitious LEO systems.

🖉 **Deployment Lead Time:** It takes several years to successfully plan, design, and launch a satellite system into space. The adaptability of the system must be carefully considered to ensure that changes in user demands and satellite use will not render the design inefficient or inappropriate to requirements once deployed. It can be difficult to plan far enough ahead and successfully predict the communications service requirements in the future.

🖉 **Propagation Delay:** The long propagation delay associated with satellite channels places numerous constraints on system performance, especially in the case of GEOs. The quality of service of delay sensitive applications is affected by the minimum bound of ~0.25s on the end-to-end delay of packet transmissions over a GEO satellite link. This delay represents an even greater impairment to the perceived quality of interactive applications where there is a minimum delay of ~0.5s between one user communicating and receiving a reply from the other end.

The propagation delay can also have a significant effect on protocol performance, especially in the case of window-based flow control protocols such as the Transmission Control Protocol (TCP) [6, 7]. If TCP is not configured correctly for operation over high delay/bandwidth product links, then the latency of satellite channels can result in ineffective use of the available capacity. Numerous techniques have been proposed and are continuing to be developed to improve performance of TCP over satellite [8-10]

🖉 **Limited Capacity:** If a satellite is used to support a large number of individual point-to-point connections, then the capacity available at the satellite may become a limiting factor in their economic use.

# 1.3 Satellite Systems and Networks Architectures

## 1.3.1 Satellite Frequency Bands

The main bands used for satellite communications are presented in Table 1-1. The Table presents representative uplink and downlink frequencies and the primary use of each band. The higher frequency bands suffer from greater propagation losses but have a larger capacity owing to increased spectral bandwidth. The available transmit power is generally more limited at the satellite than at the ground, hence the higher frequencies in each band are used for the uplink (ground terminals to satellite).

| Band | Representative Uplink/Downlink Frequencies (GHz) | Primary Use |
|---|---|---|
| L | 1.8/1.6 | Maritime |
| C | 6.0/4.0 | Traditional |
| X | 8.0/7.0 | Military |
| Ku | 14.0/12.0 | Current |
| Ka | 30.0/20.0 - 44.0/22.0 | Emerging |
| V | 50.0/40.0 | Military |

**Table 1.1: Satellite frequency bands and their use.**

### 1.3.2 Satellite Orbits

There are three primary classifications of satellite orbits, based on the orbit altitude:

➢ ***Geostationary Earth Orbit (GEO)***

GEO satellites orbit the Earth at an altitude of approximately 36,000 km with a corresponding orbit period of 24 hours. The satellites are situated in the same plane as the Earth's rotation (the equatorial plane). GEO satellites are able to provide continuous communications to all users within the coverage area with only three satellites required to cover the entire Earth. The majority of satellites in space today are in a geostationary orbit with adjacent satellites as close as a few degrees.

➢ ***Low Earth Orbit (LEO)***

LEO satellites orbit at a much lower altitude than GEO satellites, typically around 1,200 km above the surface of the Earth. The orbit period of a LEO satellite is of the order of 90 minutes. The orbits take the satellites over, or nearly over, the geographic poles. Each revolution takes approximately 90 minutes to a few hours. The fleet is arranged in such a way that, from any point on the surface at any time, at least one satellite is on a line of sight. In order to achieve continuous satellite access, a large network of satellites is required with regular connection handover between them. Achieving ubiquitous coverage poses a significant challenge, and the speed at which the satellites pass overhead generates rapidly changing communication channels, subject to severe Doppler spreading. If a constellation of LEO satellites is designed to provide global coverage then the satellites must be able to communicate with each other usually by incorporating Inter Satellite Links (ISLs) .

- ***Medium Earth Orbit (MEO)***

MEO satellites represent a compromise between LEOs and GEOs with orbit altitudes in the region of 18,000km. On one hand, a MEO satellite system relies on complex handover mechanisms just like LEO systems, with much greater path loss and a longer round trip propagation delay. The handover is much less frequent, however, making the system design much simpler. Compared to a GEO system, MEO satellites offer lower propagation delay and reduced free space path loss at the expense of additional complexity for global coverage.

A summary of the three orbit types with typical parameters is given in Table 1.2

| Orbit | Typical Altitude (Km) | Round trip Propagation Delay (ms) | Number Required for Global Coverage |
|---|---|---|---|
| **GEO** | 36.000 | 240 | 3.0-4.0 |
| **MEO** | 18.000 | 120 | 10.0-12.0 |
| **LEO** | 1.200 | 8 | >70 |

**Table 1.2 Satellite Orbit types.**

LEO satellites have the inherent advantage of a much lower signal propagation delay compared to GEO satellites, offering superior QoS, especially for interactive services. The price paid for smaller delays is the immense cost of deploying the large constellation of satellites required to provide global coverage, combined with the engineering challenges posed by frequent handover and incorporation of ISLs. Individual GEO satellites have inherent broad coverage capability, making it much simpler to provide global connectivity. LEO satellites are much smaller than GEOs with less capacity per satellite. However, an individual satellite covers a much smaller area of the Earth's surface, and so the available capacity is shared between a smaller number of subscribers. The total capacity of both types of systems can be increased with spot-beam architectures[1] and frequency reuse, allowing a larger number of users to be supported with higher data rate applications.

From all the above, it is clear that LEO systems provide the potential for lower end-to-end delays, despite transmission over multiple satellites, due to the significantly lower propagation delay. Transmissions are more likely to suffer large variation in delay with LEO systems, as sufficient resources need to be available across multiple satellite links to maintain low delay jitter, with fluctuations in the load on an individual link affecting the overall end-to-end performance. GEO satellite systems are arguably the best option for point-to-multipoint (broadcast) services. A broadcast signal is transmitted in a single hop[2] via a geostationary satellite, whereas many satellites may be involved in providing the service with a LEO system if the users' locations are diverse. Each satellite must provide sufficient capacity to carry the broadcast signal and so the greater the number of satellites involved in providing the service, the greater the likelihood that one satellite will not have sufficient capacity available resulting in user requests for the service being blocked. The propagation delay via a geostationary satellite is not a significant constraint for these types of service, as it is only apparent on connection set-up.

[1] A spot beam is a satellite signal that is especially concentrated in power so that it will only cover a limited geographic area. Spot beams are used so that only earth stations in a particular intended reception area can properly receive the satellite signal.

[2] A satellite hop relates to a transmission from the ground to the satellite and back down to the ground again, equivalent to one round trip.

### 1.3.3 Transparent Transponders and Processing Satellites

A transparent transponder acts as a relay station, transferring information received on an uplink channel onto a corresponding downlink channel. The architecture of a transponder is shown in Figure 1.1. The uplink signals are filtered to reject unwanted interference from adjacent channels, translated in frequency, amplified and then transmitted on the corresponding downlink channels. Transparent transponders represent the traditional satellite architecture and the majority of the current geostationary satellites are of this type.

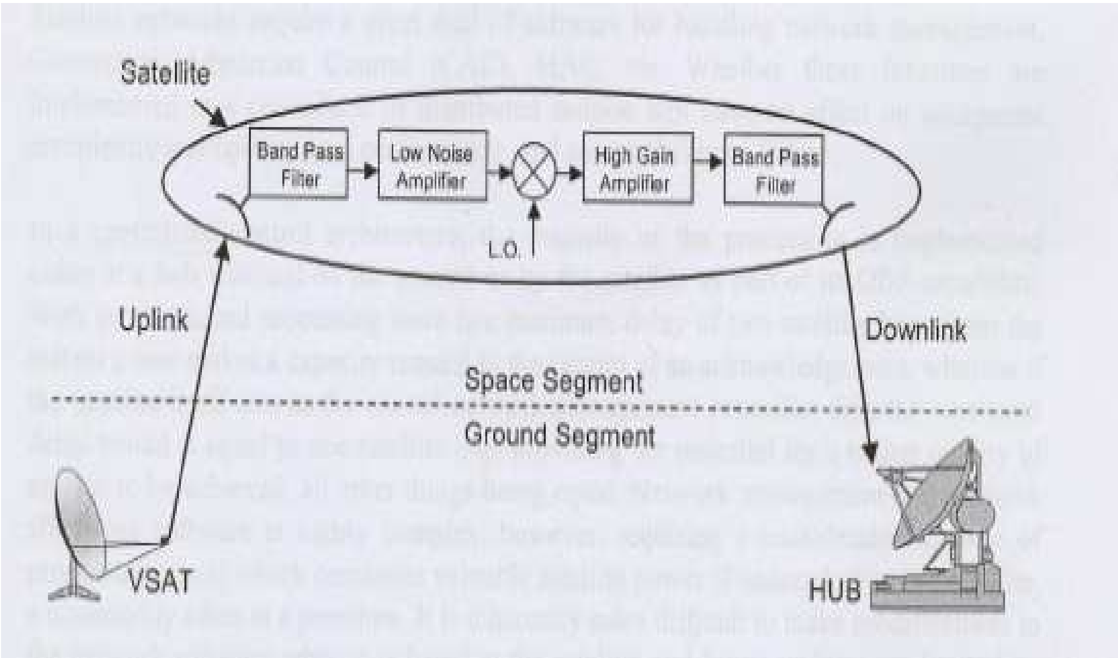


**Figure 1.1: Architecture of a transparent transponder**

Satellites usually incorporate a level of functionality within themselves, often referred to as On-Board Processing (OBP) [11]. In its simplest form, OBP provides demodulation/decoding and modulation/encoding functions to regenerate digital signals at the satellite. Signal regeneration can be of significant benefit to the communications link budgets, either reducing the transmit power requirements for the satellite and the ground terminals (resulting in lower ground terminals cost), reducing the size of the antennas, or enabling higher data rates. More complex OBP satellites may have the ability to packet switch or route traffic between a series of spot-beams (replacing the traditional single beam coverage), providing the potential for a large increase in system capacity through frequency reuse in non-overlapping beams. Many of the LEO satellite systems are designed to incorporate ISLs due to the costs, difficulties, and potential performance degradation associated with installing a ground-based hub station in every beam, focusing them to perform some form of routing at the satellite. The power available at the satellite governs the level of OBP that can be achieved, although the reduced transmit power requirements for the link budgets will help. Other advantages of OBP include: *a)* more efficient bandwidth utilization as capacity can be targeted to where it is required in spot-beam architecture, and *b)* reduced reservation delay for capacity and connection admission requests.

Despite the potential advantage of OBP there is one significant disadvantage in that the specific nature of the satellite architecture carries a financial risk. If the demand for satellite services is unsuccessfully predicted or changes significantly during the development phase, then the use of the satellite payload design may be limited. A transparent transponder is very flexible in this respect as capacity can easily be leased or sold to other service providers.

### 1.3.4. Centralized or Distributed Control

Satellite networks require a great deal of software for handling network management (i.e., Connection Admission Control (CAC), Medium Access Control (MAC) etc.) Whether these functions are implemented in a centralized or distributed fashion will have an effect on equipment complexity and cost, system performance, and susceptibility to failure.

In a centralized control architecture, the majority of the processing is implemented either at a hub terminal on the ground or by the satellite as part of its OBP capability. With ground-based processing there is a minimum delay of two satellite hops from the instant a user makes a capacity request to the receipt of an acknowledgement, whereas if the satellite itself acts as the central network management controller then the minimum delay bound is equal to one satellite hop, providing the potential for higher QoS. Network management and resource allocation software is highly complex, however, requiring a considerable amount of processing power which consumes valuable satellite power if undertaken at the satellite. It is inherently more difficult to make modifications to the network software when it is based at the satellite and future updates are limited by the technology of the satellite hardware. Another disadvantage of processing satellite architecture is the specific nature of the network, inhibiting its use as an alternative system if for example the preferred protocol standards change. Despite this, there are a number of disadvantages of satellite-based processing such as those outlined in section 1.3.3.

In distributed control architecture, processing functions such as MAC are implemented in the user terminal equipment. This configuration provides a minimum capacity request-to-acknowledgement time of one satellite hop, enabling the provision of QoS comparable to a centralized ground-based processing architecture, without increasing the complexity of the satellite and without consuming valuable satellite power. All users will receive request information from other users after a single satellite hop and implement the same protocols to determine how they should be handled. Despite the potential gains in QoS and the ability to work with a transparent transponder, there are several disadvantages to distributed control architecture. Firstly, all user terminals in the system must be compatible with each other and implement a complex set of network protocols, increasing the price of the user terminal equipment. Secondly, it is important that every user terminal is aware of every other terminal in the network; otherwise the scheduling decisions could be incorrect with, for example erroneous assignment of MAC protocol reservation requests.

Another factor in the choice between a centralized or distributed architecture is the susceptibility to failure. Centralized processing architectures are particularly exposed to software failure at the satellite or at the ground-based hub, which would result in complete network failure. The software must be extremely robust and may require multiple sets of hardware in case of failure. Malfunction of a user terminal in a distributed architecture should not affect the operation of the network as a whole. It is worth noting that any system will employ some level of centralized control for overall network management, addition of new users to the network, software updates, and user terminal synchronization.

### 1.3.5 Satellite Scenarios

There are three generic types of satellite network architecture: point-to-point, multipoint-to-multipoint, and point-to multipoint. In a point-to-point configuration a single terminal communicates directly with another. This configuration is not particularly common but arises when a satellite is used to provide a backbone connection between two large terminals, requiring a significant amount of transponder capacity. With a multipoint-to-multipoint configuration (commonly referred to as a mesh architecture), a large number of terminals communicate with each other directly. An example of this scenario is a business network with numerous offices linked via satellite. With a point-to multipoint configuration (commonly referred to as a star architecture), a group of terminals communicate with a single other terminal. This architecture is common for providing geographically dispersed users with access to a terrestrial network as well as for any kind of broadcast or multicast service provision.

# Chapter 2
# Source Modeling in ATM Networks

## 2.1 Asynchronous Transfer Mode

The ATM has been selected as the official transmission mode for BISDN (Broadband Integrated Services Digital Network). BISDN is a network architecture supporting a large applications' spectrum (voice, image, multimedia, etc.). The term ″asynchronous″ is not referring to the actual transmission, which in most cases is synchronous, but to the way the available bandwidth is reserved. The channel time is divided in fixed size time slots, which are dynamically reserved by the various network uses, depending on their needs.

The ATM technique is defined with the help of a set of principles [12]:

- The information is transmitted with fixed length data units, which are called cells. The cells consist of a header and a data field. The cell structure is described below.
- ATM uses virtual connections to transmit information.
- The main use of the cell header is the identification of the cells which belong to the same connection.
- The basic unit of an ATM network is the ATM switch. An ATM switch can be characterized as a network device with many ports. Its role is to switch the cells it receives from the input ports to the appropriate output ports.
- The cell identification headers are of local importance only. They are not absolute addresses and they are translated in each ATM switch.

The cell size is 53 bytes. Of these, 5 bytes consist the header, and the remaining 48 bytes are used for information transmission (payload). The cell structure is presented in Figure 2.1. Each row represents one byte of the cell.

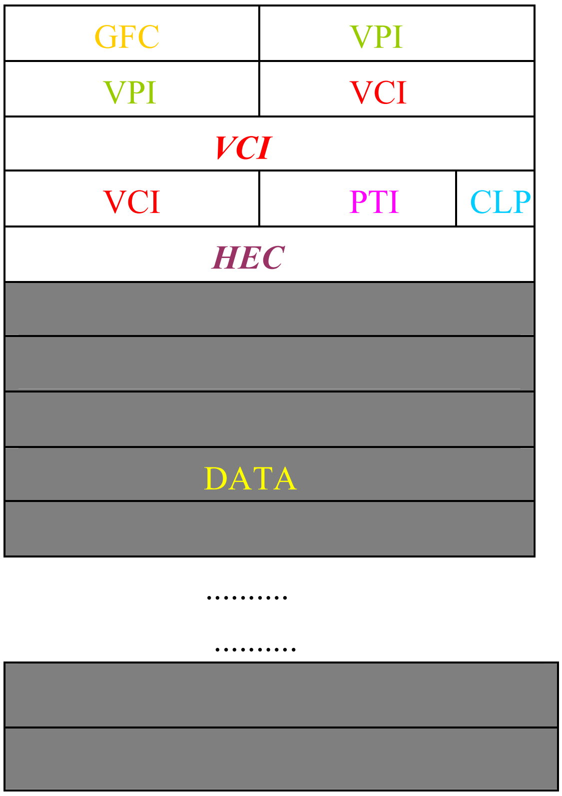


**Figure 2.1: The structure of an ATM cell.**

The use of the header fields is:

- ***GFC*** ***(G**eneric **F**low **C**ontrol):* (4 bits). It is used to regulate the information rate in an ATM network. Its full use is still an object of research.
- ***VPI*** ***(V**irtual **P**ath **I**dentifier):* (8 bits). It defines a set of virtual circuits, which are routed through the same path. VPI, together with VCI, determines the connection to which the cell belongs.
- ***VCI*** ***(V**irtual **C**hannel **I**dentifier):* (16 bits). It uniquely defines a connection between two end nodes.
- ***PTI*** (**P**ayload **T**ype **I**dentifier): (3 bits). It determines whether the data belong to a certain user or consist network administration information.
- ***CLP*** ***(C**ell **L**oss **P**riority)*: (1 bit). It determines whether the cell is of low priority and thus can be dropped in case of congestion somewhere in the network.

➢ ***HEC (Header Error Control):*** (8 bits). It is used by the end node, for error control of the cell's header.

To achieve high quality of service, the connections are classified in five classes [13]:

1. ***Constant Bit Rate (CBR).*** These are connections which demand a constant bandwidth for their entire duration. The applications which use this type of connection are usually voice or circuit emulation applications and have strict requirements on the cell transmission delay (maximum value and jitter).
2. ***Real-time Variable Bit Rate (rt-VBR).*** This type of connections is mainly used for real time applications (i.e., applications which demand very low delays). In these connections, the required transmission rates, and consequently the bandwidth demands, vary over time. Cells which fail to be transmitted within the maximum permissible delay are considered of extremely low value for the corresponding applications, and are usually dropped.
3. ***Non-real-time Variable Bit Rate (nrt-VBR).*** These are connections similar to the ones described above, but their delay requirements are loose (i.e., not strict). On the other hand, these connections have very small tolerance in packet losses, compared to the real-time VBR connections, and for this reason advanced control and error detection/ correction mechanisms are needed for these connections.
4. ***Available Bit Rate (ABR).*** The ABR connections have neither delay nor delay jitter requirements. Their only requirement from the network is that they are provided with bandwidth at least equal to a declared minimum value.
5. ***Unspecified Bit Rate (UBR).*** These connections are used for non-real-time applications, with neither delay nor delay jitter requirements, and with no minimum bandwidth demand. Classic data computer communication applications, such as file transfers and electronic mail, belong to this class. Since there is no minimum bandwidth demand, these connections have the lowest priority and use the bandwidth that is left unused by all the connections of the other classes.

As a multiplexing technique, ATM has significant advantages in comparison to the Synchronous Transfer Mode (STM), which uses static bandwidth reservation. The advantages mainly focus on the efficient use of bandwidth. According to STM, the channel time is organized in frames of equal duration, which in turn are divided into time slots. For each connection, one time slot is reserved in each frame, and the information is transmitted within this slot. An STM channel is defined by the location of its corresponding time slot within the channel frame. Clearly, STM is very efficient for CBR connections. However, its efficiency falls dramatically when it has to support VBR connections, since is has to reserve for each connection bandwidth equal to its maximum transmission rate.

In an ATM network, the support of VBR connections is facilitated via the use of ***statistical multiplexing***. With statistical multiplexing, the sum of the maximum transmission rates of all the active connections can surpass the channel bandwidth, while the available bandwidth is dynamically apportioned depending on the (variable) connection bandwidth requirements. With this method, a better usage of the network is achieved, along with the ability to support more concurrent connections. Still, the danger exists, that when the sum of the connections transmission rates surpasses the channel capacity, congestion might occur. In this case, some packets are either dropped, or temporarily kept in storage buffers, which results in increased packet delays. Some of the dropped packets must be retransmitted, which further aggravates the network congestion.

In order to deal with congestion, there is the need of traffic control algorithms, which must be capable to easily adjust to the different QoS requirements of the various connections. For example, voice applications have strict requirements regarding the maximum cell transmission delay and delay jitter. On the other hand, data applications usually demand very low error probabilities and retransmission of dropped cells, while their delay requirements are either loose or non-existent.

***The traffic control algorithm is divided in three parts***:

- **C***all* **A***dmission* **C***ontrol* ***(CAC)***
- **U***sage* **P***arameter* **C***ontrol* ***(UPC)***
- **C***ongestion* **C***ontrol*

With ***CAC***, the network decides whether it will accept a new connection or not. The decision can be based on:

- the characteristics of the traffic that the connection intends to introduce into the network,
- the connection QoS requirements
- the current network status

The traffic characteristics and the QoS requirements are defined through a set of parameters which are declared by the user when applying for the establishment of a connection. The most common parameters are the mean and peak cell rate, the burst size, the maximum cell delay tolerance and the maximum cell delay variation tolerance. Based on these parameters, the CAC algorithm decides whether it would accept the new connection or not. The algorithm should not be too strict, which would lead to low network utilization, nor too loose, thus creating congestion. If the connection is accepted, it is assumed that the network and the user agree on a traffic contract. Based on this contract, the network is bound to satisfy the user's QoS requirements, and the user is bound to operate according to its traffic description parameters [14].

However, in ATM networks bandwidth reservation is dynamic, so there is nothing that prevents a connection from violating the traffic contract and transmitting at a rate higher than the one stated. This could happen not only because of the user's bad intention, but also because of a miscalculation of the required bandwidth. Since the ATM cells corresponding to the additional rate might cause network congestion, a mechanism of usage parameter control is necessary. This mechanism controls the traffic which the user introduces into the network, and its aim is to protect the network and the other users from violations of the traffic contract, deliberate or not.

A good usage parameter control mechanism should combine [15]:

- *simplicity, so that it can be easily applied by the user*
- *responsiveness to traffic parameter violations*
- *Tolerance, because of system inaccuracies.*

Despite the use of CAC and UPC mechanisms, there still exists the possibility of congestion in some part of the network, because of temporary buffer overload. The aim of congestion control is to recognize the phenomenon and to implement mechanisms that will reduce its consequences. Congestion control, mechanisms which have been proposed for use in ATM networks are preventive. The use of reactive congestion control methods is forbidden, mainly for two reasons:

1. the response of such methods to congestion is very slow, and this could be destructive to real-time applications, and

the propagation delay is very large, compared to the very small cell transmission time. Thus the number of cells introduced in the network within a few propagation delays time after the congestion begins, can be very big.

Thus, even if the reactive congestion control mechanism reacts fast, congestion will further build up and continue for quite some time. For this reason, preventive methods have been suggested, in order to avoid the congestion problem [16-18].

## 2.2 Source Models and QoS Parameters

Source modeling is used to mimic the behavior of a source. Traffic modeling on the other hand focuses on aggregated traffic patterns. Multiplexed models will capture the effects of statistically multiplexed bursty sources and will predict to what extent the superposition of bursty streams is smoothed. Hence traffic models are used for designing connection admission control algorithms and for traffic engineering.

ATM networks are expected to support a diverse set of applications with a wide range of characteristics. Unfortunately, for the time being, there are no comprehensive measurements to satisfactorily address the characteristics of various types of B-ISDN applications realistically.

Source characterization at the macro level means defining the source traffic characteristics and its QoS requirements. The traffic characteristics of an application are the minimum set of parameters that a user can be expected to declare while providing the network with as much information as possible to effectively control network traffic and achieve high source utilization.

This Chapter first provides information about traffic parameters used to describe the traffic characteristics of a source; then QoS parameters which the user can negotiate with the network are explained.

After defining various parameters which can be negotiated between user and network, service categories supported by ATM networks are identified. Each service category can support various services depending on which traffic parameters can be declared and which QoS guarantees are required by the user. The statistical behavior of generic traffic services such as: voice, video, data, and multimedia are provided and based on this information the criteria in selecting source models are defined.

## 2.3 Source Traffic Parameters and Descriptors

Source traffic parameters are used to describe traffic characteristics of a source. They may be quantitative or qualitative (e.g. telephone, videophone). For an ATM connection, traffic parameters are grouped into ***a source traffic descriptor***, which in turn is a component of ***a connection traffic descriptor***.

***A source traffic descriptor*** is the set of traffic parameters of the ATM source. It is used during the connection set-up capture the intrinsic traffic characteristics of the connection requested by a particular source. The set of traffic parameters in a source traffic descriptor can vary from connection to connection. A **connection traffic** descriptor characterizes a connection at the User Network Interface (UNI). It consists of:

- ***Source traffic descriptor***
- ***Cell Delay Variation Tolerance (CDVT)***
- ***Conformance definition***

The ***connection traffic descriptor*** is used by the network during connection set-up to allocate network resources and derive parameters for UPC. The ***conformance definition*** is used to distinguish conforming and nonconforming cells without ambiguity.

An important issue is the set of traffic parameters to include in the source traffic descriptor. All parameters should be simple to be determinable by the user, interpretable for billing, useful to CAC for resource allocation, and enforceable by UPC. The Set should be small but sufficient for the diverse types of traffic in B-ISDN.

Some proposed source traffic parameters which will be explained in detail are:

- ***Peak Cell Rate (PCR,p) and CDVT***
- ***Sustainable Cell Rate (mean cell rate, SCR,m) and Maximum Burst Size (MBS)***
- ***Intrinsic Burst Tolerance (IBT)***
- ***Mean duration of the burst ($t_{on}$)***
- ***Maximum Frame Size (MFS)***

### 2.3.1 Peak Cell Rate and Cell Delay Variation Tolerance

The Peak Cell Rate (PCR) of the ATM connection is the inverse of the minimum inter-arrival time *T* between two cells on a transmission link. It specifies an upper bound on the traffic that can be submitted on an ATM connection [19]. The ATM Forum and ITU-T define the PCR and CDV tolerance using the Generic Cell Rate Algorithm (GCRA) and *equivalent terminal* model [19-20]. The reason for variation in the cell delay is that ATM Layer functions (e.g. cell multiplexing) may alter the traffic characteristics of ATM connections by introducing Cell Delay Variation (CDV). When cells from two or more ATM connections are multiplexed, cells of a given ATM connection may be delayed while cells of another ATM connection are being inserted at the output of the multiplexer. Similarly, some cells may be delayed while physical layer overhead cells are inserted. Consequently with reference to the *peak emission interval T* (i.e. the inverse of the contracted peak rate), some randomness may affect the inter-arrival time between consecutive cells of a connection cells of a connection. The upper bound on the 'clumping' measure is the CDV Tolerance (CDVT). The CDVT at the public UNI is defined in relation to the PCR according to the GCRA (T, $\tau_{UNI}$) where $\tau_{UNI}$ is the tolerance at the User Network Interface (UNI).

For the time being two extreme cases of characterizing the CDVT [19] have been identified.

➢ **Loose Requirements on CDV Tolerance**

A large amount of CDV can be tolerated. In this case, only the specification of the maximum value of CDV tolerance $\tau_{MAX}$ that can be allocated to a connection is envisaged. $\tau_{MAX}$ is intended as the maximum amount of CDV that can be tolerated by the user data cell stream.

➢ **Stringent Requirement on CDV Tolerance**

A connection should not be denied because of the required CDV tolerance if this CDV tolerance requirement is less than or equal to $\tau_{PCR}$ which is given by:

$$\frac{\tau_{PCR}}{\Delta} = \max\left[\frac{T_{PCR}}{\Delta}, a\left(1 - \frac{\Delta}{T_{PCR}}\right)\right]$$

Where:

**$T_{PCR}$:** is the peak emission interval of the connection (in seconds),

Δ : is the cell transmission time (in seconds) at the interface link speed,

α : is a dimensionless coefficient (suggested value is 80 [21]).

## 2.3.2 Sustainable Cell Rate and Intrinsic Burst Tolerance.

The Sustainable Cell Rate (SCR) is an upper bound on the average rate of the conforming cells of an ATM connection, over time scales which are long relative to those for which the PCR is defined. The Intrinsic Burst Tolerance (IBT) [19] specifies the maximum burst size at the PCR or in other words the maximum deviation from the *average rate*. These parameters are intended to describe VBR sources and allow for statistical multiplexing of traffic flows from such sources.

The SCR and IBT traffic parameters are optional traffic parameters a user may choose to declare jointly, if the users can upper bound the average cell rate of the ATM connection. To be useful to the network provider and the customer, the value of the SCR must be less than the PCR. The SCR and the IBT (denoted as $\tau_{IBT}$) are defined by the GCRA (Tscr, $\tau_{IBT}$). SCR and IBT belong to the ATM traffic descriptor [19]. Translation from the Maximum Burst Size (MBS) to $\tau_{IBT}$ will use the following rule:

$$\tau_{IBT} = [(MBS-1)*(T_{SCR}-T_{PCR})] \text{ seconds}$$

Where [x] stands for the first value above x out of the generic list of values defined in [19].

If the user has the knowledge of $\tau_{IBT}$ rather than of the maximum burst size, then the following rule applies:

$$\mathbf{MBS = 1+[[\tau_{IBT} /(T_{SCR}-T_{PCR})]]}$$

Where [x]stands for rounding down to the nearest integer value.

### 2.3.3 Mean Burst Period

The mean burst period ($T_{on}$) is defined at the average time the source is transmitting cells at the peak rate. This parameter is widely used for bursty sources.

### 2.3.4 Maximum Frame Size

The Maximum Frame size (MFS) is an upper bound on the number of cells in a frame or in other words an upper bound on the frame size that a user can send. The MFS has to be smaller than the MBS.

## 2.4 Quality of Service

Quality of Service (QoS) is measured by a set of parameters characterizing the performance of an ATM layer connection. These QoS parameters (referred to as network performance parameters by ITU-T) quantify end-to-end network performance at the ATM layer.

Six QoS parameters are identified by the ITU-T [21] and ATM-Forum [22] which corresponds to a network performance objective. Three of these are negotiated between the end-systems and the networks. One or more values of the QoS parameters may be offered on a per connection basis, corresponding to the number of related performance objectives supported by the network. Support of different performance objectives can be done by routing the connection to meet different objectives, or by implementation –specific mechanisms within individual network elements. The following QoS parameters are negotiated:

### ➢ 2.4.1 Cell Loss Ratio

Cell Loss Ratio (CLR) is the ratio of total lost cells to total transmitted cells in a population of interest. There are three different CLR definitions depending on the priority of the traffic. These are: ***$CLR_0$, $CLR_{0+1}$, $CLR_1$.*** The request CLR will be an upper bound on the cell loss probability.

➢ ***$CLR_0$,*** → ***this is the ratio of total lost cell with high priority to the number of CLP=0 transmitted cells.***

➢ ***$CLR_{0+1}$,*** → ***the ratio of lost cells to the total number of generated cells.***

➢ ***$CLR_1$,*** → ***the ratio of lost cells with CLP=1 to the number of transmitted cells with CLP=1.***

### ➢ 2.4.2 Cell Delay Variation (CDV)

Two performance parameters associated with CDV are the ***One-Point CDV*** and the ***Two-Point CDV*** which are defined below.

### ✎ *1- point CDV at a Measurement Point*

***One-Point CDV*** ($y_k$) for cell K at a Measurement Point (MP) is the difference between the cell's reference arrival time ($c_k$) and actual arrival time ($a_k$) at the MP:

$$y_k=c_k-a_k$$

The reference arrival time pattern ($c_k$) is defined as follows:

$$c_0=a_0=0$$

$c_{k+1}$ =

- **$c_k$ + T when ck≥ $a_k$ or when cell *k* does not arrive**
- **$a_k$ + T otherwise**

Positive values of 1-point CDV (early cell arrivals) correspond to cell clumping; negative values of 1-point CDV (late cell arrivals) correspond to gaps in the cell stream.

## *Cell Delay Variation between two MPs (2-point CDV)*

The 2-point CDV ($v_k$) for cell k between $MP_1$ and $MP_2$ is the difference between the absolute call transfer delay ($x_k$) of cell ***k*** between the two MPs and a defined reference cell transfer delay ($d_{1,2}$) between the same two MPs:

$$v_k = x_k - d_{1,2}$$

The absolute cell transfer delay ($x_k$) of cell ***k*** between $MP_1$ and $MP_2$ is the difference between the cell's actual arrival time at $MP_2$ ($a_{2k}$) and the cell's actual arrival time at MP1 ($a_{1k}$):

$$\mathbf{x_k \; a_{2k}\text{-}a_{1k}{}^{1}}$$

The reference cell transfer delay ($d_{1,2}$) between $MP_1$ and $MP_2$ is the absolute cell transfer delay experienced by cell 0 between the two MPs.

The magnitude of the delay variation produced in the satellite link is larger than that of the terrestrial network hence CDV in the satellite link is a serious problem. The most widely used method to compensate for CDV is the use of a shaper buffer at the receiver terminal.

The maximum amount of CDV to be introduced by the terrestrial network is controlled by the Usage Parameter Control (UPC). However the main component of the delay variation introduced to the cell stream is due to the satellite access technique (usually TDMA). Sources can only send a certain number of cells each frame time. Hence, the transfer interval of the user information cells at the input of the satellite network is usually equal to the access scheme frame time plus additional varying waiting delays in the buffers which can be bounded using the scheduling algorithm. More details on CDV measurements are provided in [22,21].

The following QoS parameters are ***not*** negotiated:

➢ **Cell Error Ratio (CER)**

Cell Error Ratio (CER) is the ratio of total error cells to the total of successfully transferred cells in a connection.

- **Severely Errored Cell Blocks Ratio(SECBR)**

Severely Errored Cell Blocks Ratio (SECBR) is the ratio of the total of severely errored cell blocks to the total of the cell blocks in a connection.

- **Cell Misinsertion Rate (CMR)**

Cell Misinsertion Rate (CMR) is the total number of misinserted cells observed during a specified time interval divided by the time interval duration.

Further information on ATM layer QoS may be found in ITU-T Rec.I356 [21].

# Chapter 3

# Multiple Access Control Protocols for Satellite Networks

## 3.1 Introduction

The majority of communication systems and certainly every radio communication system incorporate some level of capacity sharing, and a multiple access technique is employed to achieve this. This division of the entire radio spectrum into separate bands for different services represents multiple accesses in the broadest sense, as any radio communication system utilizes a share of the fundamental resource of the spectrum. Numerous networks require coordinated transmission from a group of users onto a common channel, with only singular point-to-point communication systems as exceptions.

A multiple access technique provides a means of dividing up capacity for simultaneous use by multiple users. The most common methods are to separate user transmissions in frequency, time, or with a set of orthogonal codes. User access to the shared capacity is then regulated and controlled by a *Medium Access Control (MAC)* protocol. MAC is a software function and corresponds to the second layer of the International Standards Organizations – Open Systems Interconnection (ISO-OSI) reference model [23]. In packet-switched networks, MAC protocols are designed to coordinate packet transmission of packet received in error, and to resolve collisions during any contention period. An example of a MAC protocol for a wired network is the MAC protocol of the Ethernet standard, which is used in numerous Local Area Networks (LANs). This standard employs the Carrier Sense Multiple Access protocol with Collision Detection (CSMA-CD) [24, 25]. The focus of this chapter is on MAC protocols for wireless systems, and more specifically, for the satellite scenario.

This chapter covers background material on multiple access techniques and medium access control, and includes a comprehensive literature review of satellite MAC schemes. Section 3.2 describes the fundamental multiple access techniques with particular emphasis on Time Division Multiple Access (TDMA) on which most of the research here is based. Section 3.3 details the alternative capacity assignment strategies and identifies their benefit and limitations in relation to the satellite channel. Important issues, constraints, and performance criteria in satellite MAC protocol design are discussed in section 3.4 and 3.5, and a literature review of pertinent satellite MAC schemes is given in section 3.6. The chapter ends with a brief summary in section 3.7.

## 3.2 Multiple Access Techniques

There are four fundamental multiple access techniques for radio systems [26]. These are:

- Frequency Division Multiple Access (FDMA)
- Time Division Multiple Access (TDMA)
- Code Division Multiple Access (CDMA)
- Random Access (RA)

The first three techniques are contention-free with user transmissions on orthogonal (or almost orthogonal) channels, separated in frequency, time or code, respectively. Random access represents contention-based multiple access with either little or no coordination of user transmissions. Some random access techniques employ a time slotted system allowing them to be integrated with contention-free TDMA-based transmissions. In these cases, random access represents an access strategy to a TDMA channel rather than a pure access technique and is treated as such in section 3.3.

### 3.2.1 Frequency Division and Time Division Multiple Access

Frequency Division Multiple Access (FDMA) is a traditional contention free-technique where users transmit simultaneously on different carrier frequencies in different bands. It is important to ensure that each user is sufficiently separated in frequency not to cause interference to other users operating on adjacent channels. This is achieved by inserting guard bands. Figure 3.1 represents the basis of FDMA. Analogue telephony and broadcast television services have traditionally been provided by FDMA, with allocation of fixed capacity channels on a connection-by-connection basis.

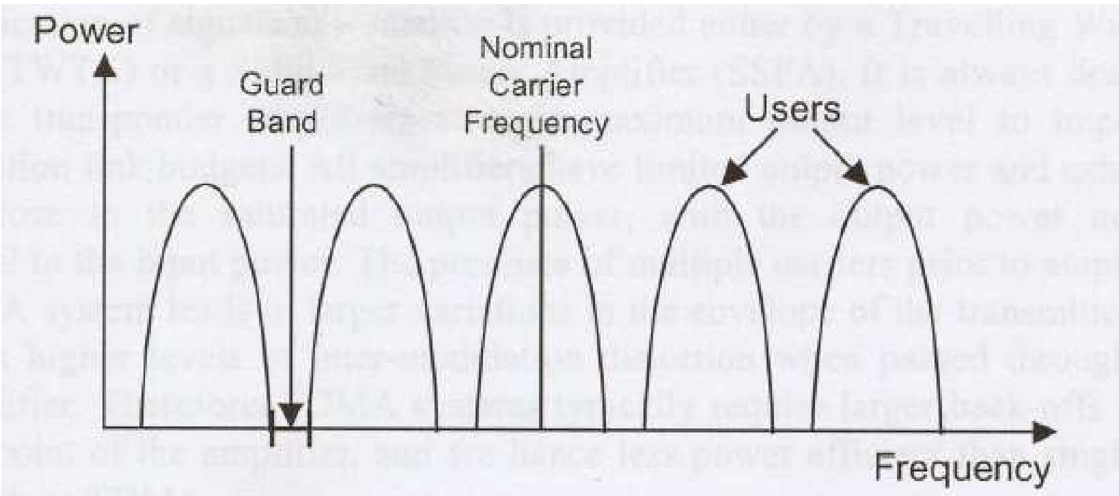


**Figure 3.1: Frequency division multiple access**

In TDMA, each user transmits on a common carrier frequency but is allocated regular time slots in which bursts of data may be transmitted in a contention-free manner. It is a popular technique as it is particularly suited to the transmission of packetized digital data, as described below. Figure 3.2 illustrates the TDMA concept.

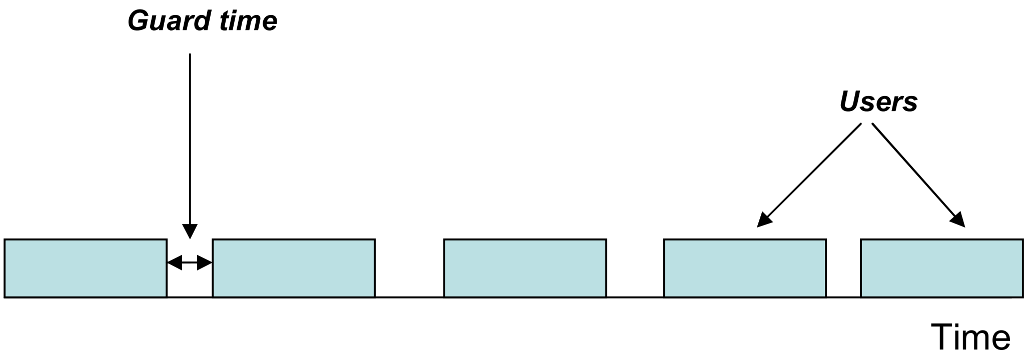


**Figure 3.2: Time Division Multiple Access**

The guard bands in FDMA and guard times in TDMA introduce a slight inefficiency to the schemes as these fractions of capacity are not used for information transmission. In ***FDMA***, the guard bands are required to compensate for any offsets in the nominal carrier frequencies and to ensure that users do not cause excessive interference to adjacent users. Accurate synchronization is important in TDMA to ensure that transmissions from different terminals do not overlap in time. This is normally achieved through regular transmission of a recognizable reference burst from a single terminal acting as reference station, used by the other terminals to synchronize their timing. Guard times are usually present in a ***TDMA*** system to ensure that transmissions do not overlap as a result of either different or changing propagation delays from user terminals, or due to their varying timing accuracies.

In an ***FDMA*** satellite system, individual users are assigned a specific frequency channel on which they can transmit at will. The satellite downlink power is shared pro-rate between all the carriers present; therefore turning off the carrier during period of no transmission can increase the power available to other frequencies. The full satellite downlink power is available to a single TDMA carrier but it is important to note that during periods when users have no data to transmit, satellite power may be wasted. The amplification of signals at a satellite is provided either by a Traveling Wave Tube Amplifier (TWTA) or a Solid State Power Amplifier (SSPA). It is always desirable to operate the transponder amplifiers at their maximum output level to improve the communication link budgets. All amplifiers have limited output power and exhibit non-linearity close to the saturated output power, with the output power no longer proportional to the input power. The presence of multiple carriers prior to amplification in an FDMA system leads to larger variation in the envelope of the transmitted signal, resulting in higher levels of inter-modulation distortion when passed through a non-linear amplifier. Therefore, FDMA systems typically require larger back-offs from the saturation point of the amplifier, and hence they are less power efficient than single carrier systems such as TDMA.

Terminals operating under a TDMA scheme must have sufficient peak power to transmit at a high data rate for short periods of time. The average power transmitted will be the same as in an FDMA system, but FDMA involves continuous low data rate transmission compared to high data rate transmission for shorts periods of time with TDMA. The peak power requirement may be too high for low power portable terminals.

The primary advantages of FDMA are that that it is simple and cheap to implement. Each user has a unique channel for transmission without interference from other users, with no coordination or synchronization required. It has been used for numerous circuit-switched satellite services including traditional analogue telephony and broadcast television transmission, and it must be noted that FDMA is always present to some extent as the division of the entire electromagnetic spectrum is the coarsest form of FDMA.

Early satellite MAC schemes were FDMA-based but the majority of modern satellite MAC schemes are TDMA-based. There are two primary reasons for using TDMA as opposed to FDMA, related to the increasing proportion of packet-based multimedia traffic that the next generation satellites will be required to support. Firstly, TDMA is particularly suited to the transmission of packetized digital data. Data packets originate from clocked computer networks at specific instants in time with transmission durations dependent on their size. TDMA systems allocate transmission time slots to users, which can be mapped directly to the generation of packets on a computer network. Secondly, FDMA lacks flexibility in both network reconfiguration and capacity assignment. In order for users to be added to or removed from an FDMA-based network, the users require either multiple transceivers or frequency agility. A single FDMA channel is capable of supporting a specific data rate, which has been designed for. It is difficult to assign different and varying amounts of channel capacity to users based on their requirements, due to the existence of fixed capacity channels with other users operating on adjacent frequencies. TDMA provides very flexible capacity assignment with the potential to vary the allocation to a user on a dynamic basis simply by assigning more or less transmission time on the channel (more or less time slots) as a function of time. Terminals can also be added and removed from the network with ease. It is for these reasons that TDMA-based schemes have been chosen over FDMA for MAC protocol designs. Despite the advantages of TDMA, allocating a large portion of the satellite transponder capacity to a large number of users by means of a single TDMA channel may result in excessive instantaneous bit rates for the ground terminals. Many MAC schemes reduce the required transmission rate by operating under a hybrid Multi-Frequency TDMA scheme (MF-TDMA).

### 3.2.2 Code Division Multiple Access

Code Division Multiple Access (CDMA) is a technique that utilizes a form of spread spectrum modulation. There are two forms of CDMA: ***direct sequence*** and ***frequency hopped.*** In ***direct sequence*** CDMA, as an example, each signal is multiplied by a unique wide bandwidth spreading code to generate a signal occupying typically a thousand times the original bandwidth. The resulting signals from many users are then modulated onto common carrier frequency. All transmissions take place simultaneously on the same carrier, with the powers from each transmitter adding. The received signal is multiplied by an identical spreading code, correctly synchronized at the receiver, to reproduce the original data. The spreading codes allocated to users must exhibit very low cross-correlation to effectively reject unwanted signals at the receiver. Other users cause some interference due to residual correlation properties between the spreading codes. This can be modeled as Additive White Gaussian Noise (AWGN) and is minimized by ensuring low correlation between the different spreading codes. As the number of users in the system increases, the total level of interference increases, degrading the channel performance. CDMA systems do not have an abrupt capacity limit as found in FDMA and TDMA; performance just gradually degrades to a level where the interference caused to users becomes a limiting factor.

CDMA offers advantages and is particularly useful in interference-limited scenarios. The signal de-spreading process at the receiver is uncorrelated. In a mobile telephony environment, fading due to multipath signals destructively interfering at the receiver can severely affect performance. With a CDMA system, any multipath components arrive out of phase with the synchronized de-spreading code and are therefore received as uncorrelated interference, which may be ignored. Channel assignment is only reliant on code synchronization, with no frequency or time coordination required.

Interference mitigation is not an issue for a single beam geostationary satellite scenario, which probably explains the abundance of TDMA-based MAC schemes in the literature.

## 3.3 Capacity Assignment Strategies

There are four primary strategies for assigning capacity to multiple users:

- Fixed Assignment
- Demand Assignment
- Free Assignment
- Random Access

### 3.3.1 Fixed Assignment

With a fixed assignment strategy, each user is provided with a quasi-permanent assignment of capacity corresponding to a periodic and regular time slot allocation in a TDMA scheme, as depicted in Figure 3.3. The static nature of the assignment can be inefficient due to the fact that whenever a user does not have any data to send, the capacity is unused and therefore wasted. Even during a period of user activity the effectiveness of this strategy is limited as the assignment is not adaptive to changing traffic requirements. With variable bit rate (bursty) traffic, sufficient capacity must be allocated to cater for the peak rate of the traffic, otherwise a significant amount of buffering would be required. By definition, the average data rate of variable bit rate sources is in many cases significantly less than the peak rate, resulting in inefficient use of the fixed capacity allocation. The regularity of the fixed assignment is more suited to constant bit rate (periodic) traffic, but even then is only fully utilized during periods of user activity. It can be an efficient strategy when the traffic demand is highly regular and constant over very long periods of time. An example scenario where this is the case is at the multiplexing point of a large number of telephone connections. The primary advantage of this strategy is that it can provide absolute guarantees on throughput and QoS as each user has exclusive rights to use a specific portion of the satellite capacity.

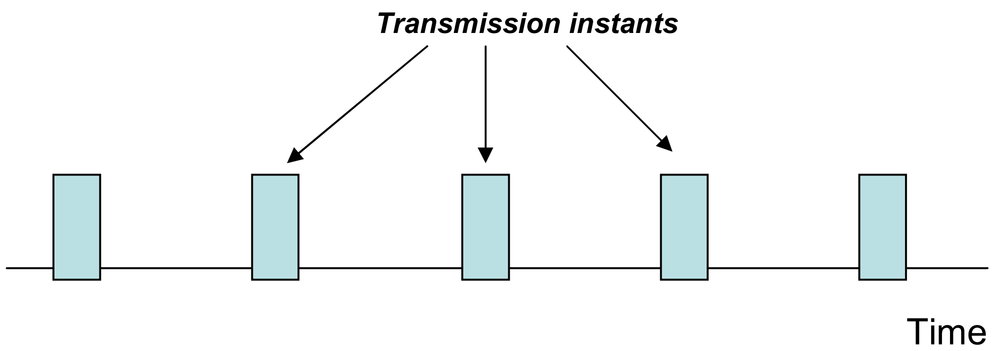


**Figure 3.3: Fixed Assignment**

### 3.3.2 Demand Assignment

A demand assignment strategy corresponds to the allocation of capacity in response to user requests. There are two types of demand assignment: ***fixed rate*** and ***variable rate***, relating to the regularity at which the capacity allocation is updated.

***Fixed rate*** demand assignment is suited to connection-oriented services with capacity allocated on a connection-by-connection basis. Figure 3.4 shows the operation of fixed rate demand assignment. At the beginning of a connection a user makes a request for capacity when it next gets the opportunity. If the connection request is accepted then the user will receive a regular and periodic allocation of time slots for the duration of the connection for exclusive use. When the connection ends, a signal is transmitted to the scheduler to release the capacity back to the network for use by other connections. It can be seen that there is an initial delay in obtaining capacity at the beginning of a connection lower bounded by one or two satellite hops for a satellite-based or ground-based capacity assignment scheduler respectively (~0.25s/0.5s over a geostationary satellite link).

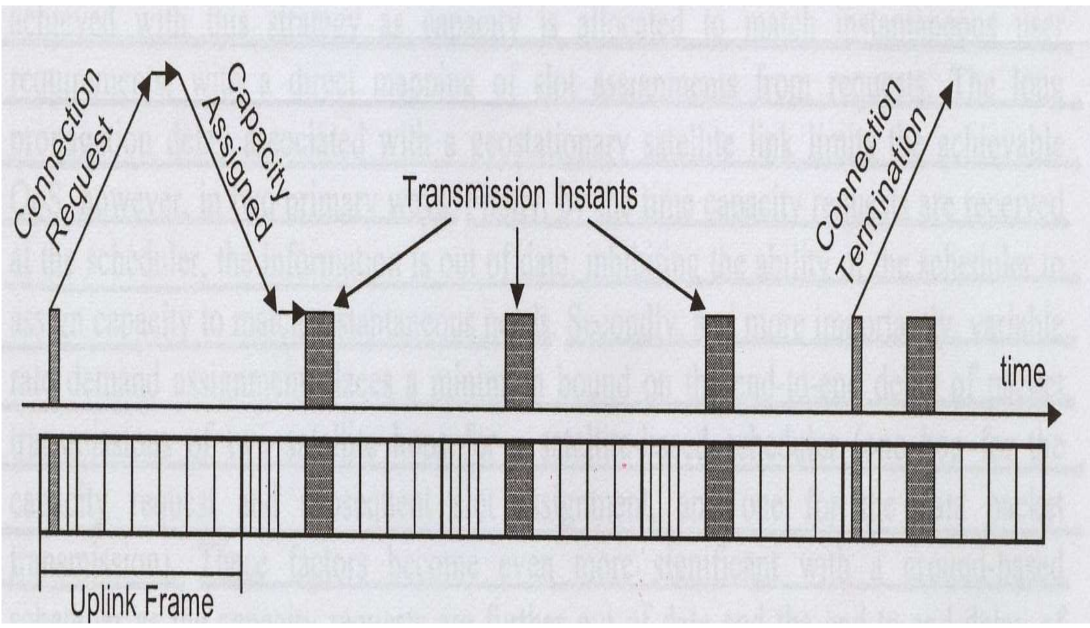


**Figure 3.4: Fixed rate demand assignment**

The strategy is inefficient for variable bit rate traffic as the fixed allocation of capacity is unable to adapt to match varying user demand. It is much more efficient for constant bit rate traffic compared with fixed assignment; owing to the principle of statistical multiplexing. *With fixed assignment,* each user is allocated individual capacity, which is wasted during periods of inactivity. *With fixed rate demand assignment*, users are also provided with separate capacity for their own use, but only when they require it (when they are active). To improve the capacity utilization efficiency in comparison with fixed assignment, the amount of capacity made available to a group of users is considerably less, based on the assumption that not all users need to communicate simultaneously, or at least it is highly unlikely that they will. If all users do wish to communicate simultaneously there will be insufficient capacity and the connection requests will be blocked. As a result, fixed rate demand assignment cannot provide absolute guarantees on capacity availability and hence QoS. It provides a statistical guarantee subject to a finite connection blocking probability, which is designed to be small. The advantage gained from the reduction in the QoS is the increase in the number of users that can be handled by a system of a given bandwidth. An analogy can be made with the Public Switched Telecommunications Network (PSTN) in which the number of trunk lines provided is not sufficient to enable every single person in the network to simultaneously make a telephone call. The number of lines required to fully interconnect a telecommunication network would be enormous and uneconomic, as each link would be in use only for a small proportion of the time. The number of trunk lines required to support a population of users is calculated based on a defined maximum rate of call arrivals and departures; it is also based on the desired call blocking probability for each traffic type supported by the network. The probability of a call being blocked is given by the Erlang-B formula [27]. An interesting feature of Erlang statistics is that for a given blocking probability, a greater amount of traffic can be supported per trunk line as the number of trunk lines is increased.

***Variable rate*** demand assignment corresponds to more dynamic allocation of channel capacity and is commonly employed to support any type of traffic where the capacity requirement varies as a function of time. An example implementation of the variable rate demand assignment strategy is shown in Figure 3.5. Individual users make regular requests for capacity based on the instantaneous requirements, requesting for a specific number of slots sufficient to clear their current queue level. A high channel utilization can be achieved with this strategy as capacity is allocated to match instantaneous user requirements, with a direct mapping of slot assignment from requests. The long propagation delay associated with a geostationary satellite link limits the achievable QoS, however, in two major ways. Firstly, by the time the capacity requests reach the scheduler, the information is out of date, inhibiting the ability of the scheduler to assign capacity to match the instantaneous user needs. Secondly, and more importantly, variable rate demand assignment places a minimum bound on the end-to-end delay of packet transmissions of two satellite hops for a satellite-based scheduler (one hop for the capacity request and subsequent slot assignment, and one for the data packet transmission) These factors become even more significant with a ground-based scheduler as the capacity requests are further out of date and the end-to-end delay of packet transmissions is lower bounded by three satellite hops.

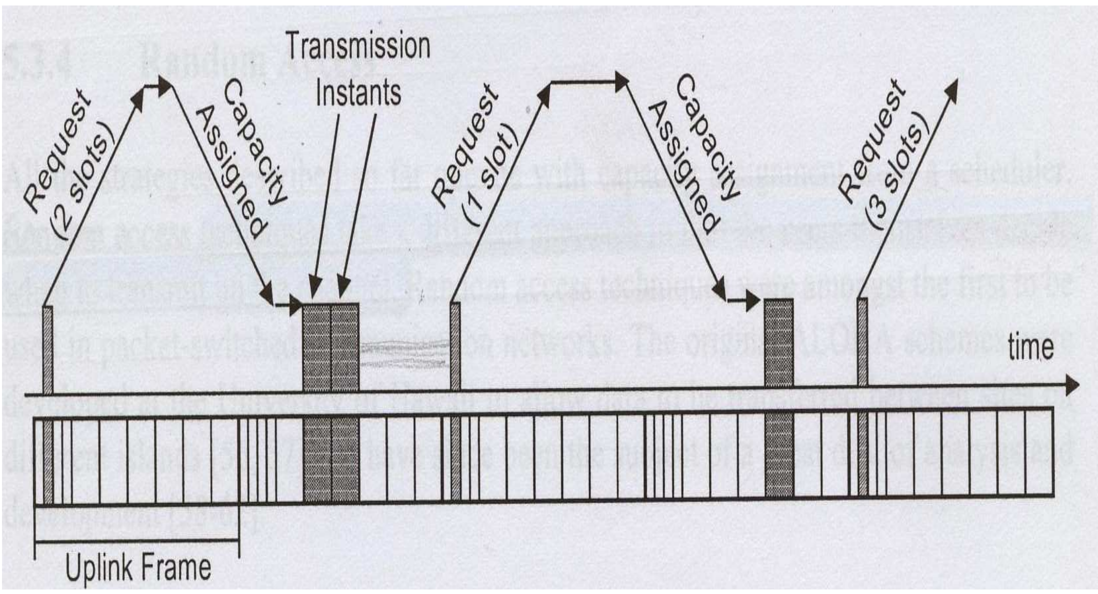


**Figure 3.5: Variable rate demand assignment**

It is clear that both demand assignment strategies require a request channel. Provision for user requests is commonly incorporated into the uplink frame as shown in Figures 3.4 and 3.5. A suitable access strategy is required for the request region of the frame which represents a secondary multiple access problem. Request slots are commonly assigned on either a fixed or round robin basis to terminals for contention-free request packet transmission, or open to random access for contention-based request in order to minimize the overhead in the uplink frame.

### 3.3.3 Free Assignment

Free assignment corresponds to the allocation of capacity ***without any form of request,*** in a similar manner to fixed assignment. The primary difference between the two strategies is that fixed assigned capacity is guaranteed to users, whereas free assigned capacity is essentially bonus capacity, useful if users happen to have data packets to send at the instants of the free assigned slots. Fixed assigned slots are provided to users at periodic and predetermined intervals, with a specific slot assigned to a user for transmission in successive frames. The availability of free assigned slots is variable and unknown to the users, with individual slot assignments identified by information transmitted in the downlink frame. Spare capacity is often allocated on a free assigned basis, commonly implemented by assigning the spare slots to a group of users, one-by-one, on a round robin basis.

The efficiency of the scheme for handling bursty traffic is difficult to determine, but it is clearly less efficient than variable rate demand assignment as there is no attempt to allocate capacity to meet varying user requirements. The primary advantage of this strategy is a minimum end-to-end delay of one satellite hop when a free assigned slot occurs immediately subsequent to the arrival of a packet in an empty terminal queue. The delay performance is heavily dependent, however, on the number of users receiving the free assigned slots. As the number of users increases, the regularity of free assigned slots to each user is reduced, increasing the average delay for packet transmissions.

### 3.3.4 Random Access

All the strategies described so far operate with capacity assignment from a scheduler. *Random Access* techniques take a different approach in that the users themselves decide when to transmit on the channel. Random access techniques were amongst the first to be used in packet-switched communication networks. The original ALOHA schemes were developed at the University of Hawaii to allow data to be transferred between sites on different islands [28, 29] and have since between the subject of a great deal of analysis and development [30- 35].

The two standard techniques are pure and slotted ALOHA. In the ***pure ALOHA*** scheme, users transmit packets onto a shared channel as soon as they arrive in their queue for transmission. There is no coordination with other users sharing the transmission medium and hence if more than one user transmits at the same time, the packets collide and data is lost. Each receiver transmits positive acknowledgment back to the appropriate sender for all correctly received packets, enabling the transmitting terminals to determine whether a collision has taken place and whether they need to retransmit a packet. Transmitting terminals wait for acknowledgments for their transmitted packets and if they do not receive them within a specified time period, they timeout and enter a retransmission mode. If the terminals involved in the collision retransmit immediately then they are likely to collide again and so a randomized back-off strategy is employed. This process is repeated until the transmission is successful. ***Pure ALOHA*** is an effective scheme for a lightly loaded communications channel. If users only need to send occasional short messages, the probability of successful transmission is high resulting in low end-to-end delay values (minimum bound of one satellite hop). As the channel load increases, either through additional users or an increase in transmission requirements, the probability of collision increases and the resulting end-to-end delay values rise as a result. The maximum channel utilization that can be achieved with pure ALOHA is approximately 18% of the channel capacity and it is the primary factor limiting its use.

***Slotted ALOHA (S-ALOHA)*** is an extension of the pure ALOHA scheme, which provides twice the maximum channel utilization in exchange for an increase in protocol complexity. In the slotted ALOHA scheme, transmission time is divided into discrete time slots with a period equal to the packet transmission duration plus an appropriate guard time. Users transmit newly arrived packets in time slots immediately subsequent to their arrival, with a collision occurring if more than one user transmits in the same time slot. The key difference between the schemes is the period in which packets are vulnerable to collision, two packet durations for slotted ALOHA. The resulting packet collision probability is halved so that the extension to slotted ALOHA generates a factor of two increase in the achieved throughput.

The Packet Reservation Multiple Access (PRMA) method is an improved form of TDMA that combines TDMA with the techniques of S-ALOHA. It can be used for mobile satellite links comprising of a BS and a number of MS [PRMA]. A TDMA satellite channel consists of multiple time slots in a framed structure. Each time slot can carry data packets belonging to one of the N users. The assignment of time slots to a channel is not fixed but is handled dynamically in real-time. Each data packet carries a VCI (Virtual Circuit Identifier) field that indicates its receiving earth station (ES). Different earth stations can recognize their packets on the down-link broadcast by checking the VCI field in the packet.

At any instant a TDMA frame consists of a fixed numbers of reserved and empty slots. This information is broadcasted to all ESs via satellite broadcasts. When an ES has reserved a slot, the packets are termed as “safe” packets. Whenever a new ES tries to establish a channel, it sends an “unsafe” data packet in one of the available slots. If the packet reaches the destination without collision, the slot is reserved and the ES can transmit “safe” packets. If a collision occurs, collision resolution schemes (for example, the binary exponential back off algorithm) are used to resolve the conflicts between the contenders. A user may reserve more than one time slots, if required. A summary of the applicability and features of the different capacity assignment strategies is given in Table 3.1. :

| Capacity Assignment Strategy | Applicability | Notes |
|---|---|---|
| Fixed Assignment | Constant bit rate traffic | Efficient for heavily multiplexed regular and constant traffic |
| Fixed rate demand assignment | Constant bit rate traffic | Very efficient. Long initial channel access delay |
| Variable rate demand assignment | Variable bit rate traffic | Very efficient. High minimum end-to-end delay |
| Free assignment | Variable bit rate traffic | Inefficient. Low minimum end-to-end delay |
| Random Access | Variable bit rate traffic | Inefficient. Low minimum end-to-end delay |

**Table 3.1: Applicability of the different capacity assignment strategies**

It appears from the above description that a single capacity assignment strategy is not able to provide both high efficiency and low delay for variable bit rate (bursty) traffic such as multimedia traffic. As a result, it is common for satellite MAC schemes to employ a hybrid strategy, combining more than one capacity assignment techniques. In many instances the schemes also incorporate a dynamic frame structure which adapts to suit instantaneous user requirements. A number of such protocol examples are described in the literature review in section 3.6.

## 3.4 Satellite MAC Protocol Design.

Satellite systems have a number of features which distinguish them from terrestrial wireless systems; the most apparent differences being the long propagation delay and the unique network topologies. These characteristics must be taken into consideration when designing MAC protocols for satellite systems. This section highlights the constraints that satellite systems place on MAC protocol design, and identifies desirable characteristics.

### 3.4.1 Satellite Constraints on MAC Protocol Design

➢ ***Propagation Delay:*** The long signal propagation delay associated with satellite channels is the primary constraint on MAC protocol design. One way packet transmissions via satellite are subject to a lower end-to-end delay bound of one satellite hop(~ 0.25s for a geostationary satellite), with a minimum delay of two hops for a transmission and subsequent response for an interactive service over satellite, such as a television news broadcast. The propagation delay inhibits the ability of demand assignment techniques to successfully allocate capacity to match instantaneous requirements, and places strict bounds on the achievable delay performance.

The performance of network protocols reliant on acknowledgements or automatic repeat request techniques is also limited by the propagation delay. The delay/bandwidth product of a communication link governs the achieved throughputs of window-based flow control protocols such as the Transmission Control Protocol (TCP). TCP sends a block of data equal to the current window size (initially very small) and then waits for an acknowledgement from the receiver before sending the next block, successively increasing the window size as each block of data is acknowledged. The throughput rate of TCP is severely limited when the acknowledgement delay is long, unless the initial window size is set large. Also, any data losses are assumed to be due to congestion on the channel, causing TCP to reduce the window size and enter a slow incremental increase once again. On a satellite channel, packet loss is much more likely to be due to packet errors than channel congestion. There is a considerable amount of ongoing research into the effects of long propagation delay links on TCP performance [6, 7] and some solutions have been suggested in [8- 10].

- ***Power and Capacity:***

The power available at a satellite is limited and must be considered during the design phase to ensure that adequate communication links can be supported. Capacity is often a limiting factor, especially with a system designed to support many point-point communication links. Satellite power and capacity are both at a premium and one or the other will limit the overall performance of the system. Adaptive modulation and coding techniques provide flexibility in this respect, allowing power and capacity to be traded off against each other as required [36]. For a processing satellite, the complexity of the software and associated hardware needs to be considered in relation to the power requirements.

- ***System Design Changes:*** In the specific case of an On-Board Processing (OBP) satellite, the difficulty in making changes to the hardware and perhaps the software at the satellite once in service must be considered. Hardware and Software reliability is a serious issue for all satellite systems due to the difficulty in maintaining them; this is an especially true for processing satellites due to their complexity.

### 3.4.2 Desirable Features of a MAC Protocol

➢ ***Channel Stability:*** The ability of the MAC protocol to cope with fluctuations in the level of traffic placed on the channel without entering an unstable state is a particularly important feature. Stability is especially important for contention-based random access schemes, where a rise in the instantaneous traffic level increases contention, which serves to place a large number of terminals into a state of retransmission with minimal achieved system throughput. A well designed system should be able to handle instantaneous channel loads greater than the maximum sustainable load when the long-term demand is less than the maximum, without exhibiting undesirable behavior.

➢ ***Protocol Scalability and Reconfigurability:*** It is important to have a MAC scheme that will provide flexibility in the number of users that the satellite can support, as well as providing the facility to add and remove users from the network with ease. A good scheme in this respect will support a wide range of user population sizes, and will provide efficient log-on and log-off procedures. The ability of the scheme to continue to provide service in the presence of terminal failures is also an important factor

➢ ***Adaptivity to Fluctuations in Channel Load:*** Satellite MAC schemes often incorporate multiple capacity assignment strategies with an adaptive uplink frame structure. In scenarios with widely varying load requirements, the scheme needs to be able to adapt the channel to suit short-term needs in order to provide good QoS and/or high efficiency.

➢ ***Low Complexity:*** In any computational system it is always desirable to keep the processing requirements to a minimum by minimizing the complexity of the underlying algorithms. If the computational complexity of a MAC protocol can be reduced then it will operate faster, consume less power, and will be easier to test.

## 3.5 Performance Measures

It is necessary to identify and select the most important performance criteria for evaluating the effectiveness of MAC protocols. Whilst some of these criteria are important in their own right, it is common for one to impact on another and so a reduction in the performance in one area is sometimes required to make improvements in another.

### 3.5.1 Channel Utilization

Channel utilization is defined as the proportion of the data carrying channel capacity that is effectively used for useful data transmissions, and is generally one of the most significant criteria in the MAC scheme's performance evaluation. It is important to use the satellite capacity as efficiently as possible, in order to support a greater number of users or to provide the same number of users with higher data rates, generating larger revenues for the service provider in either case. With TDMA-based MAC schemes the intention should be to maximize the use of the allocated transmission slots and minimize the frame overheads.

The terms ***Channel Utilization*** and ***Channel Load*** are used almost interchangeably in the literature, as they have very similar meanings and represent the same value. Load refers to the level of demand placed on a channel, expressed either as a percentage of the overall data carrying capacity of the channel, or as a fraction of the useful channel capacity. Utilization refers to the level at which the channel is being used, which is attained by placing an equivalent traffic load on the channel. Another term used for Channel Utilization is Channel Throughput, which relates to the proportion of the overall channel capacity that is effectively used to transfer new information.

### 3.5.2 End-to- End Delay

The end-to-end delay of a packet transmission is the time taken for a packet to be successfully received at the destination terminal at the other end of the satellite link from the time it was generated at the source. It consists of a number of components including the queuing delay prior to transmission, the packet duration, the signal propagation delay, and the queuing at the satellite. There is always some benefit to be gained from reducing the end-to-end delay of packet transmissions over a satellite link because even if the user-recognized improvement in the quality of service of an application is minor, any enhancement in the delay distribution of packet transmissions of a given scenario is indicative of a more efficient MAC scheme. In addition, long transmission delays may inhibit the operation of other layers of the protocol stack such as the transport layer in the case of TCP. The mean value of end-to-end delay is commonly used as a performance metric. The entire distribution of end-to-end delay values offers a greater level of detail on the effectiveness of a MAC scheme.

### 3.5.3 Variation in End-to-End Delay

Some applications are sensitive to the variation in the end-to-end delay of packet transmissions, often referred to as delay jitter, arising from fluctuations in the queuing delay at the ground terminals and the satellite. Audio and video streaming applications commonly require low delay jitter. Buffering at the receiver combined with constant buffer read out can alleviate this problem but this requires additional hardware and introduces extra complexity, serving to increase the overall delay on the received stream. Reducing the variation in the end-to-end delay of packet transmissions through good MAC protocol design is preferred.

### 3.5.4 Initial Channel Access Delay

The initial delay in obtaining access to a satellite channel may be significant in some instances; short messages in particular rely on a short access delay for rapid transmission over satellite link. This delay component can be reduced either through the choice of access strategy or through user prioritization.

### 3.5.5 Fairness of Channel Sharing

Some MAC schemes are designed to offer equitable channel access to all users in the network, whereas other schemes are designed to provide different levels of service to users based on predefined service level agreements. In all cases, the MAC protocol must be evaluated with respect to the design criteria to ensure the scheme is able to provide either equitable sharing of capacity or efficient prioritization as required.

## 3.6 Satellite MAC Schemes

### 3.6.1 Overview

MAC scheme for wireless systems have been subject to extensive research and development since the early 1970's and there are now numerous examples described in the literature. MAC schemes are commonly classified into different groups depending either on the type of network they are designed for, the underlying multiple access technique they employ, or the different capacity assignment strategies they incorporate. They are a number of articles in the literature which provide a review of many different MAC schemes for both terrestrial wireless networks [37- 39], and satellite networks [40, 41- 45]

Terrestrial communication systems have received more attention than satellites ones in recent years and there has been a lot of research carried out into protocol design for terrestrial systems. However, with the emergence of proposals for the next generation broadband satellite systems, satellites are being rediscovered as a complementary technology to terrestrial networks, an essential component in providing global access to evolving multimedia services. A number of MAC protocols designed for satellite systems have recently been proposed in the literature. The majority are unique to the satellite scenario whereas others are modifications of existing terrestrial schemes, modified to take into account the propagation delay associated with satellite links.

Early access schemes were predominantly designed to support a singular type of traffic or application such as telephony, with capacity assignment strategies to suit [46, 47]. As outlined in section 3.3, it appears that no single capacity assignment strategy is able to provide both high channel utilization and low end-to-end delay for variable bit rate (bursty) traffic such as self-similar data traffic [48]. It is therefore common to employ a hybrid MAC scheme which combines several capacity assignment strategies [49, 50].

Some of the more prevalent MAC schemes are described in this literature review, with particular emphasis on those designed to support multimedia traffic. The intention is to provide a description of the important features and concepts of each scheme, highlighting successive development. Comparative performance of the different schemes based on analytical or simulated performance is not included due to the wide range of satellite scenarios that each scheme has been evaluated for. The different schemes have been broadly classified into four categories:

- ***Combined random access and DAMA***
- ***Combined fixed assignment and DAMA***
- ***Combined free assignment and DAMA***
- ***Predictive schemes***

A plethora of schemes have been developed which combine random access with demand assignment, aiming to benefit from the high maximum channel utilization efficiency of demand assignment, along with the low end-to-end delay capability of random access. Examples of these types of schemes are described in sections 3.6.2 to 3.6.13. An alternative strategy to circumvent the propagation delay is to incorporate free assignment of capacity into a demand assignment scheme as proposed by Le-Ngoc and introduced in section 3.7.1. Combining fixed assignment with demand assignment is a useful approach to support a mixture of stream traffic and bursty traffic, with an example outlined in section 3.7.1.

MAC schemes designed for terrestrial wireless systems generally have limited applicability to satellite systems as the channel propagation delay is much smaller in terrestrial links. The limitations in performance of demand assignment techniques are not apparent and capacity can easily be allocated to closely match instantaneous requirements. The primary issues in terrestrial systems are commonly related to Dynamic Channel Assignment (DCA), with clever frequency reuse strategies employed to increase capacity whilst maintaining acceptable interference levels [51, 52]. However, a terrestrial scheme has been adapted to suit a Low Earth Orbit (LEO) satellite scenario and is presented in 3.6.13. Finally, several schemes have recently been proposed to circumvent the long capacity reservation delay associated with variable rate demand assignment by pre-empting capacity requirements. The most widely studied satellite MAC schemes from the literature review are now described briefly in turn.

### 3.6.2 Announced Re transmission Random Access

In the Announced Retransmission Random Access (ARRA) protocol, each message packet transmission is accompanied with an announcement in a control minislot, indicating the slot position of the intended retransmission in the next uplink frame should the packet collide [53]. Each user monitors the channel for message packet collisions, accumulating a list of the announced retransmissions associated with collided message slots to determine which slots will be used for retransmissions in the next frame. New packets originating in the current frame are held until the end of the frame when the permitted set of slots for new packet transmissions in the next frame can be determined. Retransmissions take place unconditionally with new packets transmitted in the remaining available slots, picked at random. If no slots are available for new packet transmissions then a retransmission announcement is made instead in a minislot belonging to a common minislot pool. This strategy completely eliminates collisions between new packets and retransmitted packets. An extended version of the protocol avoids collisions between retransmission announcements for the same slot from the collision status of the control minislots, aborting retransmissions in these cases to avoid collisions between retransmitted packets. Further retransmission announcements are made for aborted retransmissions in the common minislot pool at the start of each frame.

### 3.6.3 Generalized Retransmission Announcement Protocol

The Generalized Retransmission Announcement Protocol (GRAP) develops this strategy further [54]. The channel is monitored by the network control center which broadcasts the set of available slots for new packet transmissions on a frame-by-frame basis. If a terminal receives feedback stating that its message packet collided, then it searches the set of available slots for the slot number of its announced retransmission. Each user aborts its retransmissions if the announced slots are present in the set of available slots, as this condition indicates that the retransmission announcements also suffered a collision, and the message packet retransmissions would be sure to collide again. The scheme recognizes that packet retransmissions in reserved slots are not accompanied by anticipating announcements. The control minslots associated with the reserved slots are therefore empty and can be used for retransmission announcements of new packets or aborted retransmissions, eliminating the need for the common minislot pool and reducing the frame overhead for announcements. At high channel loads when many slots are reserved there is increased capacity to retransmit reservation requests, and at low channel loads more free slots are available for contention.

### 3.6.4 Scheduled Retransmission Multiple Access

The Scheduled Retransmission Multiple Access (SRMA) protocol behaves like *slotted ALOHA* at low channel loads and tends towards a pure reservation scheme at high channel loads [55]. The uplink frame is divided into a reserved sub-frame and a random access sub-frame. New packets arriving in the random access sub-frame are transmitted in the next slot, whilst those arriving in the reserved sub-frame are scheduled for transmission in one of the random access slots in the next frame. A retransmission announcement is made for each new packet transmission in a randomly chosen minislot associated with each random access data slot. The retransmission announcements serve as reservation requests and are honored if the message packet transmissions suffer a collision, and the reservation request packets do not. Users receive acknowledgements for successful packet transmissions along with slot reservation information from the scheduler. If a user does not receive an acknowledgement for its random access transmission, it searches for a dedicated slot allocation for the purpose of retransmission. If a reservation is not found, indicating that the reservation request packet collided, the packet is retransmitted in another random access slot after a randomized delay. There are two versions of SRMA incorporating either a fixed or dynamic frame. In the fixed frame version there is an upper limit on the number of reserved slots (equal to the specified number of slots in a frame), and so in some instances not all the required reservations can be honored. The dynamic frame version consists of a fixed length random access sub-frame, with a variable length reserved frame designed to accommodate all users requiring a retransmission reservation**.**

### 3.6.5 Controlled Multi-access Protocol

A controlled multi-access protocol is described in [56] and [57] which is an extended version of the SRMA scheme, where the relative rates of random access and reservation based transmissions are controlled to ensure optimum channel performance under all traffic conditions. A control parameter *(f)* is introduced which specifies the amount of traffic to be transferred from the ALOHA sub-frame to the reserved sub-frame and vice-versa. With $f=0$, all packets arriving in the reserved sub-frame make reservations only in minislots of the reserved sub-frame, and all packets arriving in the ALOHA sub-frame are transmitted immediately with spare reservations in case of collision. *When f is negative, a fraction of the packets arriving in the reserved sub-frame are transmitted in the upcoming slots of the ALOHA sub-frame. When f is positive, a fraction of the packets arriving in the ALOHA sub-frame will only make reservations in the headers of slots within the reserved sub-frame.*

Optimum values of *f* for minimum average delay are predetermined for a range of channel loads and stored at each user terminal for reference. As the channel load and hence throughput increases, the optimum value of *f* varies from -1 to +1. Under light traffic conditions the majority of packets are transmitted in the random access slots. Under intermediate traffic conditions, some packets make transmission reservations without attempting a random access transmission, and under heavy traffic conditions most of the packets only make reservations in the reserved sub-frame. It is indicated that the *f=-1* case corresponds to the SRMA scheme [55], and *f=0* is similar to the uncontrolled channel access strategy in the Combined Random/Reservation Multiple Access (CRRMA) scheme [58], described in section 3.6.6.

### 3.6.6 Combined Random/Reservation Multiple Access

This scheme operates on a slot-by-slot basis with transmissions in a particular slot depending on control information regarding the activity in a previous associated slot. Each slot consists of a header portion, which is divided into a number of minislots for random access reservation requests, and a data portion which is able to accommodate a single data packet. A packet is either *reserved* if the data portion is demand assigned to a specific user following a successful request, or in *contention* if it is unassigned in which case it is made available for a random access transmission by any user. If there were no request packet collisions in the header part of the previous associated slot then it is said to be *free,* otherwise it is in *rtx* (retransmission).

Two channel access strategies are described with user transmissions depending on the state of each slot. With uncontrolled channel access, a user will always transmit a request packet in one of the minislots chosen at random, and will transmit the packet if the slot is in the *contention* state. If the random access data transmission is successful then a copy of the packet held at the user terminal is discarded. If the data packet has collided but the request is successful, then the user will wait and transmit the packet in its due demand assigned slot. Finally, if the request packet and the data packet have collided, then another request is transmitted in the next associated slot. With controlled channel access, an initial reservation request for a newly arrived packet is only transmitted in a *free* slot (a slot where there were no previous request packet collisions). If more than one minislot has suffered request packet collisions, they are partitioned into groups and contention is resolved amongst one group at a time. As before, if a slot is *reserved* then only a request packet is transmitted in one of the minislots. If the data portion of the slot is in the *contention* state then a copy of the data packet is transmitted only if the header is in the *Free State*. If the slot is in the rtx-contention state then a number of request packets previously suffered collisions and so the data portion of the slot is unusable for transmission, as the data packet retransmissions would be certain to collide again. When this occurs the entire slot is divided into minislots for request packet transmissions, providing a means of quickly clearing collision backlogs.

### 3.6.7 Transmit Before Assignment Using Collision Requests

Rose and Rappaport proposed the Transmit Before Assignment using Collision Requests (TBACR) scheme in which users waiting for demand assigned capacity are able to transmit packets in unassigned slots via random access[59].When a user has a packet to send, a reservation request packet is transmitted on a slotted ALOHA request channel. If the request is successful, the scheduler attempts to demand assign a slot to the user for transmission of the data packet in a future frame. Following transmission of the reservation request, a user is also permitted to transmit a copy of the data packet in an unassigned slot of the following frame via random access, if any are available. If such a transmission is successful then the packet will experience much lower end-to-end delay than it would through waiting for its due demand assigned slot. An attempt to transmit the packet succeeds if either the random access transmission is successful and/or the user is demand assigned a slot. Otherwise the entire process is repeated after an appropriate timeout period and randomized rescheduling delay. In the instance that there are more requests in a particular frame than there are slots, a random set of requests are honored with demand assigned slots, with other requests blocked as the scheduler does not maintain a request queue.

### 3.6.8 Integrated Access Scheme

The schemes outlined so far have considered only one level of channel access with no differentiation between different types of traffic. Suda proposed a scheme incorporating random access and demand assignment, to cater for two types of terminal supporting bursty traffic and long holding (stream) traffic [60]. The uplink frame consists of a slotted ALOHA sub-frame followed by a demand assignment sub-frame, which is further divided into a reservation request region and a message packet transmission region. At the beginning, of a traffic stream, a terminal makes a request for demand assigned capacity in one of the reservation request slots chosen at random. If sufficient capacity is available, the requesting terminal is assigned one slot per frame in the same position until the scheduler receives an end-of-use flag, but if all the reservation slots are occupied the request is rejected. When the reservation request fails due to insufficient capacity or a request packet collision, the request is repeated in a randomly selected slot of the following frame. The bursty traffic terminals are permitted to transmit one packet per frame in a randomly selected slot of the following frame. The bursty traffic terminals are permitted to transmit one packet per frame in a randomly selected ALOHA slot. If a data packet collides then it is retransmitted in a subsequent frame, with new packets queued until the terminal receives confirmation that the current packet has been successfully received. Bursty terminals are not allowed to use any of the available capacity in the reserved sub-frame, primarily due to implementation complexity. It is observed that there is an optimum frame length to minimize the transmission delay of one type of traffic whilst keeping the mean transmission delay of the other type under some permissible value.

### 3.6.9 Movable Boundary Random/Demand Assignment Multiple Access

The recent trend towards convergence of communication services has resulted in a number of MAC schemes designed to support a much wider range of application/traffic types. An example scheme is Movable Boundary Random/Demand Assignment Multiple Access (MB/R-DAMA) proposed by Bohm, which employs different capacity assignment strategies to suit the needs of different types of application [61]. Connection-oriented applications such as: video, voice and file transfer are supported by fixed rate DAMA, with connectionless interactive data traffic supported by slotted ALOHA. Users make reservation requests for the fixed rate DAMA service in a randomly selected slotted ALOHA minislot. A separate region of request slots is provided for each type of connection oriented applications, limiting contention to users of the same service. Successful requests are honored in a random fashion with no priority structure and a specific slot is then assigned to the requesting terminal for periodic transmission in successive frames for the duration of the connection. If there is insufficient capacity to admit fixed rate DAMA connections, the requests are cleared and if a request packet suffers a collision then it is retransmitted in a randomly chosen slot of the next frame. A minimum number of random access slots are provided in each frame to ensure a minimum service is provided for the interactive data users, and any unreserved DAMA capacity is made available for random access to reduce delays for this type of traffic when there is spare capacity. When a terminal is informed that a data packet from a bursty source has collided, it is retransmitted in a randomly selected slot of the next frame.

### 3.6.10 Random-Reservation Adaptive Assignment

A recent example of a scheme designed to support multimedia services is Random Reservation Adaptive Assignment (RRAA), proposed by Ors [62-64]. This scheme incorporates both fixed and variable rate DAMA as well as slotted ALOHA to cater for the varying requirements of the different Asynchronous Transfer Mode (ATM) service classes [65]. Fixed rate DAMA is used for constant bit rate traffic and delay sensitive real-time variable bit rate traffic. Variable rate DAMA can be combined with fixed rate DAMA to more efficiently support non-real-time variable bit rate traffic. Unspecified bit rate traffic is sent via random access as there are no QoS requirements for this type of service. The uplink frame consists of two regions of request slots; random access slots for fixed rate DAMA reservation requests and round robin assigned slots for variable rate DAMA requests. For the fixed rate DAMA service, a terminal transmits a request packet in one of the request slots chosen at random. If the packet collides then it is retransmitted after a randomized delay. The number of random access request slots must be carefully controlled to avoid network backlog whilst minimizing the frame overhead. Successful reservation requests are queued, saving terminal from having to repeatedly contend for a request slot when a reservation cannot be immediately provided. When capacity becomes available, slots are allocated based on the priority of the traffic from each of the requesting terminals, with each terminal allocated a specific slot in successive frames for the duration of the connection. Variable rate DAMA requests consist of quantified requests for specific numbers of slots, acting as supplement to the fixed rate DAMA capacity to cater for the instantaneous requirements of variable bit rate traffic. Once all DAMA requests have been honored, the remaining part of the frame is made available for random access transmission. The advantage of the random access slots for unspecified bit rate transmissions is the absence of the lengthy reservation procedure.

### 3.6.11 Dynamic Random-Reservation Adaptive Assignment

Iera et al. have developed two enhanced versions of the RRAA scheme, designed to improve the reservation procedure and minimize frame overheads [66]. In the Dynamic Random-Reservation Adaptive Assignment (D-RRAA) schemes, real-time variable bit rate sources are supported with variable rate DAMA instead of fixed rate DAMA, and since delay is not an issue for non-real-time traffic they introduce preemptive capacity allocation to guarantee QoS for the real-time traffic with stricter delay constraints. The first version of the scheme incorporates a fixed number of random access request slots for the fixed rate DAMA Service, along with a dedicated control channel (request slot) assigned to each terminal for the entire duration of a connection to control delay and enable dynamic capacity requests without further contention. If the number of random access request slots is too low then excessive contention will generate long access delays, and if the number is too high they will represent a significant overhead in the frame. In addition, if the number of variable bit rate sources becomes particularly high, there will be a large rise in the frame overhead due to the increase in the number of dedicated control channels. The second version of the scheme incorporates a variable number of random access request slots and control channels per frame, based on the current traffic load. The control channels are then allocated on a round robin basis with an upper limit on the number that can be provided in each frame.

### 3.6.12 Interleaved Frame Flush-Out

Another example of a hybrid random/reservation family of MAC schemes is the Interleaved Frame Flush-Out (IFFO) protocols [67]. The uplink frame consists of one slot divided into minislots for terminal reservation requests (one dedicated slot per terminal), followed by a series of data slots, some of which are reserved and others which are open to random access. The frame length is variable allowing it to adapt to suit the instantaneous traffic requirements, but each successive frame must consist of at least *R* slots (where R slots constitute a longer period than the round trip propagation delay). Each user makes a request in its dedicated minislot at the start of each frame, if required, for enough slots to cater for all packet arrivals in the previous frame. If the total number of reservation requests is greater than or equal to *R-1* slots, then the number of data slots in the next frame is equal to the number of slot requests, otherwise random access slots are incorporated to make the total frame length up to *R* slots. Some packets arriving in each frame are selected for random access transmission, with the protocol designed so that the appropriate reservation requests can be cancelled when random access transmissions are successful. With the fixed contention variant, each random access slot is used to transmit a packet that arrived during the previous time slot. All packets arriving during the reserved portion of the frame are not allowed to contend due to the high risk of collision and so these packets along with arrivals in the last slot of each frame are transmitted in the subsequent reserved slots. In the controlled contention variant, a terminal transmits a queued packet in each contention slot based on a given permission probability, determined from a function based on the number of arrivals in the current frame, and the number of reserved slots in the current frame.

### 3.6.13 Packet Reservation Multiple Accesses With Hindering States.

A notable MAC protocol for wireless network design is the Packet Reservation Multiple Access (PRMA) protocol proposed by Goodman [68, 69].Closely related to the reservation ALOHA (R-ALOHA) scheme [70], it combines TDMA with slotted ALOHA and is designed to support a mixture of voice traffic and short message data traffic. It has been subject to a number of developments and is able to support a higher capacity than that achieved with a pure TDMA system [71]. The uplink frame is divided into equal sized data slots with no overhead for capacity reservation requests. The slot and frame durations are chosen to suit the requirements of the voice service, with one slot per frame providing a throughput of 32 kbits/s for a voice connection. Terminals initially contend for channel access by transmitting their first data packet at random in an unreserved slot. If two or more terminals transmit in the same slot then the colliding packets are retransmitted in subsequent unreserved slots, based on a specified permission probability. Successful transmission in a slot serves to reserve the same slot for contention-free transmissions in successive frames. At the end of a connection, a terminal releases its reservation by leaving the reserved slot empty. If voice packets remain in the terminal queue beyond a specified time limit (typically 32 ms), they are discarded.

The use of PRMA in mobile satellite systems is discussed in [72] where it is shown that the long round trip propagation delay of a geostationary satellite link prevents effective application of the PRMA protocol. The primary limitation is the upper limit on the number of possible contention attempts prior to packets being discarded because they are out of date; *this number equals one in the case of a geostationary satellite*. Most LEO systems have round trips delays low enough to allow several retransmission attempts prior to packet discard, and so PRMA can be of use in such systems. Enrico Del Re et al have carried out a feasibility study on the application of PRMA to mobile satellite systems and have developed an improved version of the scheme, optimized to cater for the propagation delay and challenges associated with LEO satellites [73, 74]. PRMA with Hindering States (PRMA-HS) overcomes the limitation in the number of contention attempts prior to packet discard. The principle is to allow user terminals to repeatedly transmit copies of the same packet during the waiting interval for the outcome of the first transmission attempt. The scheduler keeps a record of the first successful attempt and ignores any subsequent successful transmissions made by the same terminal during this period. The user terminal enters a set of hindering states after the first successful transmission as successive transmissions of the same packet provide no useful information and only serve to hinder access attempts of other terminals contending for unreserved slots. This strategy increases the likelihood that a packet will be successfully transmitted without it being discarded due to excessive delay. The hindering property serves to increase contention due to transmission of duplicate packets but simulation results have shown that the combined effect serves to reduce the probability of packet dropping, therefore improving the performance of PRMA over satellite.

## 3.7 State-of-the-Art Protocols

### 3.7.1 Combined Free/ Demand Assignment Multiple Access (CFDAMA)

The concept of combining free assignment of capacity with demand assignment was proposed by Le-Ngoc [75]. This paper took a new approach to achieving high channel utilization and good delay performance by considering the combination of free assignment with variable rate demand assignment, providing a minimum end-to-end delay of one satellite hop at low channel loads with the high channel utilization capability of demand assignment. The Combined Free/ Demand Assignment Multiple Access (CFDAMA) scheme provides significant improvements in the delay / utilization performance of geostationary satellite channels [75]. Three different schemes have been described by Le-Ngoc. These are:

- ✓ CFDAMA *with Fixed Assigned Requests (CFDAMA-FA)* [75-76]
- ✓ CFDAMA *with Piggy-Backed Requests (CFDAMA-PB)* [77-78]
- ✓ CFDAMA *with Random Access Requests ( CFDAMA-RA)* [79]

➭ CFDAMA *with Fixed Assigned Requests* ***(CFDAMA-FA)***: is the first scheme that was proposed by Le-Ngoc. The uplink frame format shown in Figure 3.6.The frame consists of a number of slots equal to the number of terminals *(N),* with each slot subdivided into a request slot and data slot. Each terminal has a dedicated request slot in the frame and so each user is able to make a contention-free request once every *N* slots.

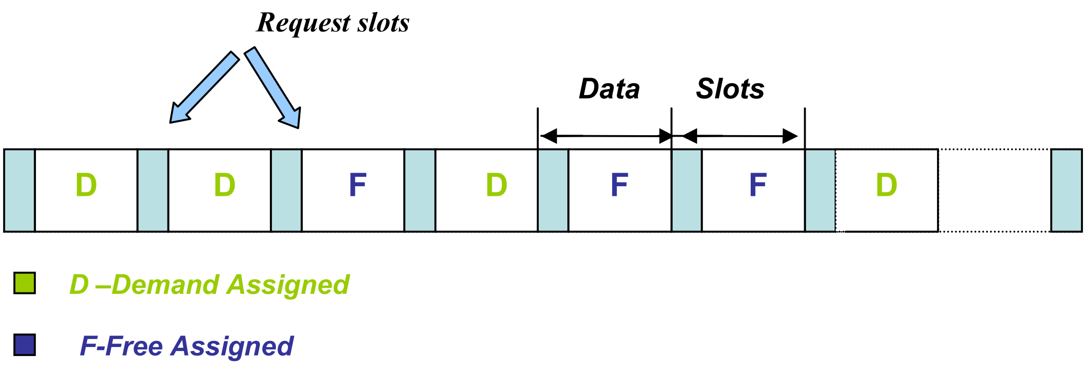


**Figure 3.6: Uplink frame format of CFDAMA**

The scheme operates with a centralized scheduling algorithm located either at the satellite or at a hub terminal on the ground. The scheduler queues capacity reservation requests and serves them on a **F**irst **C**ome, **F**irst **S**erved (FCFS) basis by demand assigning runs of successive slots in the frame based on the number requested .In the absence of any queued requests, the scheduler free assigns slots one-by-one on a round robin basis. One reservation strategy considered is to allow each terminal to make requests for all newly arrived packets for which a reservation has not been made at the instant of a request slot. This strategy may put the reserving terminal at an undue advantage with a number of demand assigned slots going unused due to transmission in free assigned slots, and demand assigned slots reserved by preceding packets of the same terminal. A controlled reservation strategy is suggested where each user keeps a count of a number of slots equal to the number of packets queued less the due slot count.

➷ CFDAMA *with Piggy-Backed Requests* ***(CFDAMA-PB):*** A second variant of CFDAMA incorporates Random Access Request slots (CFDAMA-RA). The scheme may feature request slots interleaved throughout the frame or alternatively may have a separate area of request slots at the start of the frame. The uplink frame format with interleaved request slots is shown in Figure 3.7. The request slots are available to any terminal wishing to make a request, accessed via slotted ALOHA. The scheme operates with the same scheduling algorithm and controlled reservation strategy described for CFDAMA-FA.

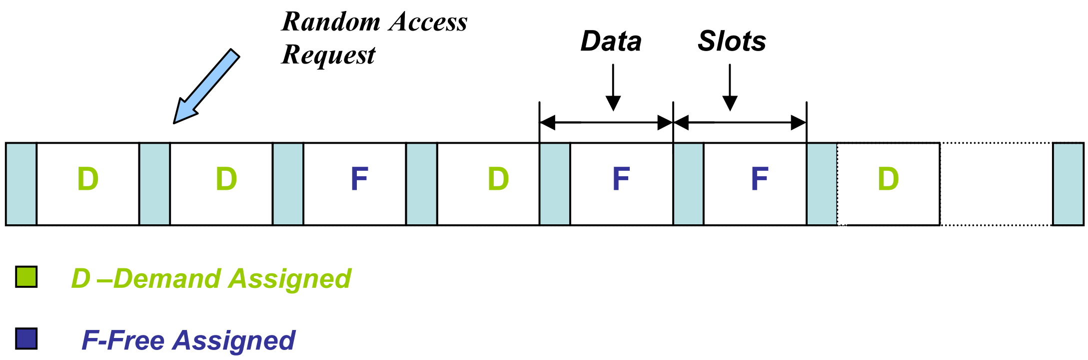


**Figure 3.7: Uplink frame format of CFDAMA-FA**

➯ CFDAMA *with Random Access Requests* ***(CFDAMA-RA):*** The third variant incorporates Piggy-Backed request slots (CFDAMA-PB). The uplink frame is shown in Figure 3.8 for illustrative purposes but Le-Ngoc does not identify a frame structure for this scheme and so in practice the frame consists of a single slot. Requests are piggy-backed onto data packet transmissions, with access rights to a particular request slot limited to the user transmitting in the associated data slot. It operates with a slightly modified version of the scheduling algorithm described above. At high channel loads a small proportion of terminals could dominate the channel with piggy-backed requests, repeatedly making requests for capacity in demand assigned slots, leaving other terminals waiting for long periods of time for a free assigned slot and associated channel access. To make the scheme fairer, terminals are moved to the end of the round robin free assignment list when they are allocated demand assigned slots, enabling those without any slot reservations to move up the free assignment list much faster. The free assignment strategy is clearly effective at low channel load levels as the regularity of free assigned slots is high, providing a significant number of packets with end-to-end delay values approaching the minimum possible bound of one round trip time. The probability of obtaining a free assigned slot is proportional to the number of terminals *(N)* and so the delay/throughput performance is heavily dependent on N, with better performance for low terminal populations. The delay performance at high channel loads is dominated by the two round trip time delay bound of variable rate DAMA.

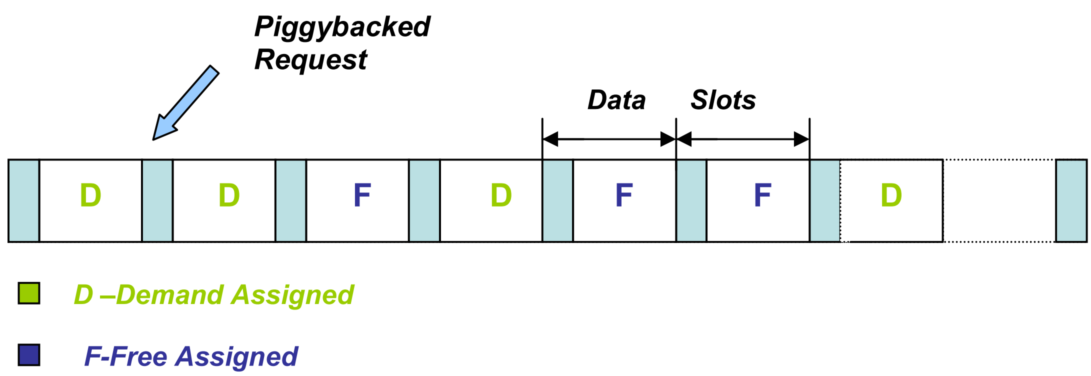


**Figure 3.8: Uplink frame format of CFDAMA-PB**

### 3.7.2 BTDAMA Schemes [82, 83]

Due to the unpredictability and statistical uncertainty of bursty data traffic, many Medium Access Control (MAC) protocols employ variable rate Demand Assignment Multiple Access (DAMA) to achieve a high channel utilization. The performance of this technique is severely constrained by the long propagation delay associated with a geostationary satellite link. *Variable rate DAMA* is often combined with other capacity assignment strategies to circumvent this delay, but these techniques are only effective in reducing delay at low channel utilization levels. The limitation of variable rate DAMA results from the need to make a request for every single packet, with a direct mapping from requests to slot assignments.

A family of schemes named Burst Targeted Demand Assignment Multiple Access (BTDAMA), which employ an original approach to implementing DAMA, have been recently designed to eliminate the limitations of traditional DAMA techniques and suited to handling traffic which can be identified by a series of bursts and inter-burst gaps (ON/OFF in nature). The ideas and concepts employed by these schemes originate through consideration of how to effectively support ON/OFF type data traffic, given the wide support and recognition that modern data traffic is ON/OFF in nature and not Poisson as traditionally assumed. BTDAMA has been subject to significant development and analysis, generating a number of different variants.

#### 3.7.2.1 Pure Demand Assigned BTDAMA

The uplink and downlink Time Division Multiple Access (TDMA) frame formats of BTDAMA with Pure Demand Assignment (BTDAMA-PD) are shown in Figure 3.9. The request slots are allocated to terminals on a round robin basis for contention-free request packet transmission with the request region in the downlink frame consisting of request slots assignment instead of acknowledgements.

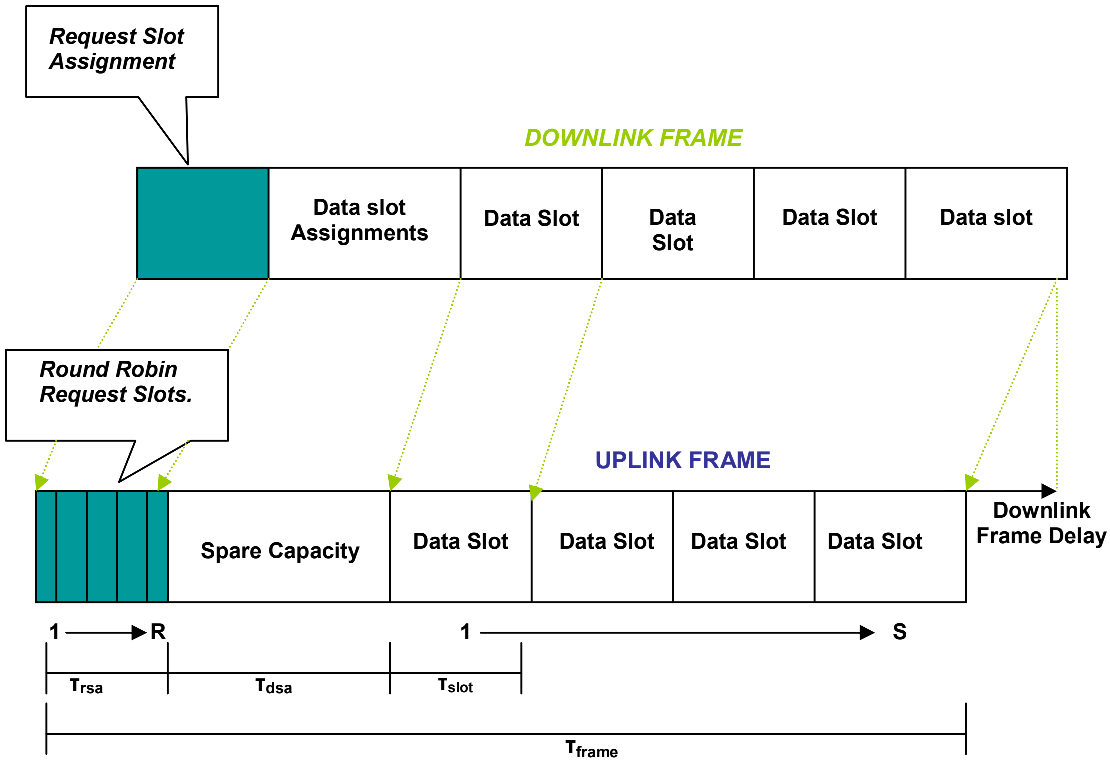


**Figure 3.9: Uplink and downlink frame formats of the BTDAMA-PD scheme**

The uplink frame is transmitted from each Very Small Aperture Terminal (VSAT) to the satellite on the multiple access uplink channel. The downlink frame is transmitted from the satellite to the VSATs and the gateway terminal on the Time Division Multiplexing (TDM) broadcast channel.

The uplink frame consists of a region of round robin request slots, followed by some spare capacity to equalize the uplink and downlink frame sizes, and the data packet transmission slots. The start of the downlink frame consists of the request slot and data slot assignment periods, which carry information regarding a frame's worth of uplink slot allocations. The remainder of the frame is used to relay the data packets received from the VSATs on the uplink to the gateway terminal on the downlink. The downlink frame is delayed with respect to the uplink frame to ensure that the data packets received at the satellite in the slots of a particular uplink frame are transferred on the corresponding downlink frame. There are no distinct guard bands in the frame formats as these are achieved indirectly in the simulations by transmitting at a slightly higher data rate to that specified. All timings and packet delays are calculated based on the specified data rates. The spare capacity is introduced to make the uplink and downlink frame sizes equal, with each channel designed to operate at the same data rate for simplicity and as an aid to effective performance evaluation with varying parameters. This capacity represents excess overhead but is an acceptable feature as capacity would be required for transmission of control information in a real implementation.

### 3.7.2.2 Key features and Characteristics

There are several key features of the BTDAMA-PD schemes, which in combination generate the significant advances in the achievable delay/utilization performance reported for the scheme. The functional operation of BTDAMA-PD for a single traffic burst is shown in Figure 3.10 to further highlight the important concepts that are described below:

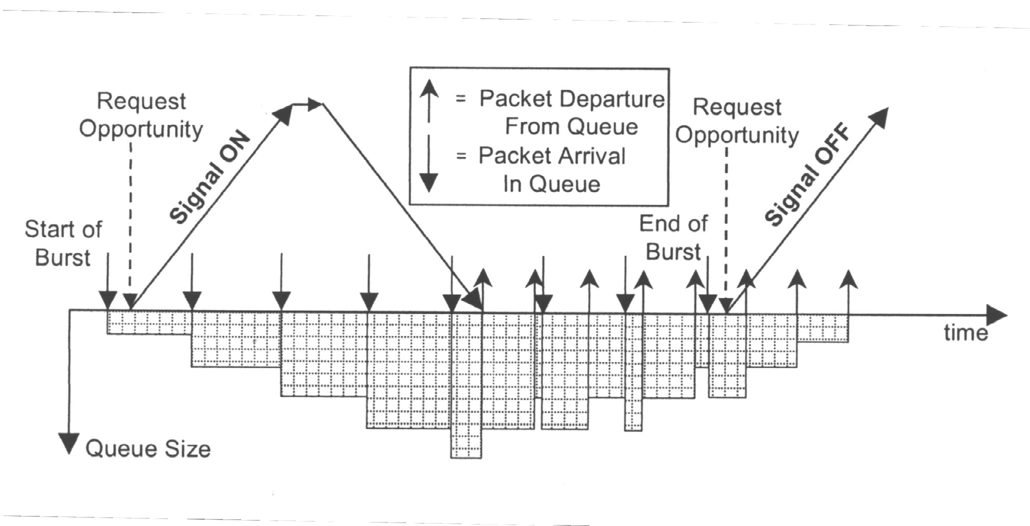


**Figure 3.10: Functional Operation of BTDAMA-PD in response to a single traffic burst**

☑ ***Separation of the request and scheduling strategies***: The delay bound limitation of conventional variable rate DAMA arises from the direct mapping from requests to slot allocations. In a pure variable rate DAMA scheme, slots requests are made for every single packet arrival, with transmission slots subsequently allocated following reception of the requests at the scheduler. In a long propagation delay environment this severely inhibits the achievable delay performance, as every packet must wait for its request to be received and subsequently acknowledged with an assigned slot. The BTDAMA-PD scheme separates the request and scheduling functions as once a terminal has signaled ON, a continual supply of demand assigned capacity is provided for as long it is required without the need for repeated requests, reducing the minimum delay bound for demand assigned slot transmissions to one satellite hop.

☑ ***Targeted and equitable capacity allocation***: The available capacity is targeted only to terminals that require it, and is shared equally between them. This enables a high maximum channel utilization to be achieved, with inherent fairness and no hard limit on capacity, just a gradual reduction in the allocation rate to each terminal as the number of terminals in the ON state increases. In some instances, it may desirable to prioritize different users based on a range of service level agreements. This can be achieved by incorporating multiple round robin assignment tables, enabling different groups of terminals to receive different slot assignment rates.

☑ ***Regular and periodic capacity assignment***: The round robin capacity assignment strategy provides terminals with a quite regular and periodic allocation of capacity, useful in reducing the variation in the end-to-end delay of packet transmissions, important for jitter sensitive applications.

MAC schemes for voice often employ a similar request mechanism with fixed rate DAMA on a call-by-call basis, but they provide a fixed assignment of bandwidth, thereby hard limiting capacity. The BTDAMA-PD requests provide terminals with access rights to the satellite capacity, which is shared equally between them. There is no hard limit on the number of terminals that can be supported, simply a gradual reduction in the capacity available to each terminal.

The simulation parameters for the BTDAMA-PD and CFDAMA-RR schemes are shown in Table 3.2.

| Parameters | Value |
|---|---|
| Number of terminals *(N)* | 100, 300 |
| Channel Load | 0.1-0.9 Erlangs |
| Channel data rate | 2.048 Mbits/s |
| Satellite altitude | 37500 km |
| Number of round robin request slots *(R)* | 50, 150 |
| Round robin request region size/ duration ($\tau_{rsa}$) | 1000bits/488μs , 3000bits/ 1.46ms |
| Number of data slots in the uplink frame *(S)* | 128 |
| Data slot assignment region size/duration ($\tau_{dsa}$) | 4096bits/2ms |
| Uplink packet header size | 0 bits |
| Uplink packet payload size | 424bits |
| Uplink slot size/duration ($\tau_{slot}$) | 424bits /207μs |
| Pareto αon/αoff | 1.2 |
| Pareto kon/koff | 1 |

**Table 3.2: Simulation Parameters for the BTDAMA-PD & CFDAMA-RR schemes**

### 3.7.3 A variable bandwidth MAC protocol for the satellite uplink channel

An efficient MAC protocol for the graceful integration of multimedia traffic on the satellite uplink channel was proposed in [80]. The protocol is based on a generic proposal being specifically designed to be capable of embedding within its structure a number of protocol standards. The exact structure of the uplink frame is developed by the satellite controller and transmitted to each station sharing the frequency band through the downlink. Obviously, the round-trip delay will influence the time response to call data-rate requests, i.e., the requests for a station's resources provided in the current downlink frame corresponds to the bandwidth which was requested in the previous uplink frame. The frame is subdivided into subframes which are further subdivided into slots as shown in Figure 3.11.

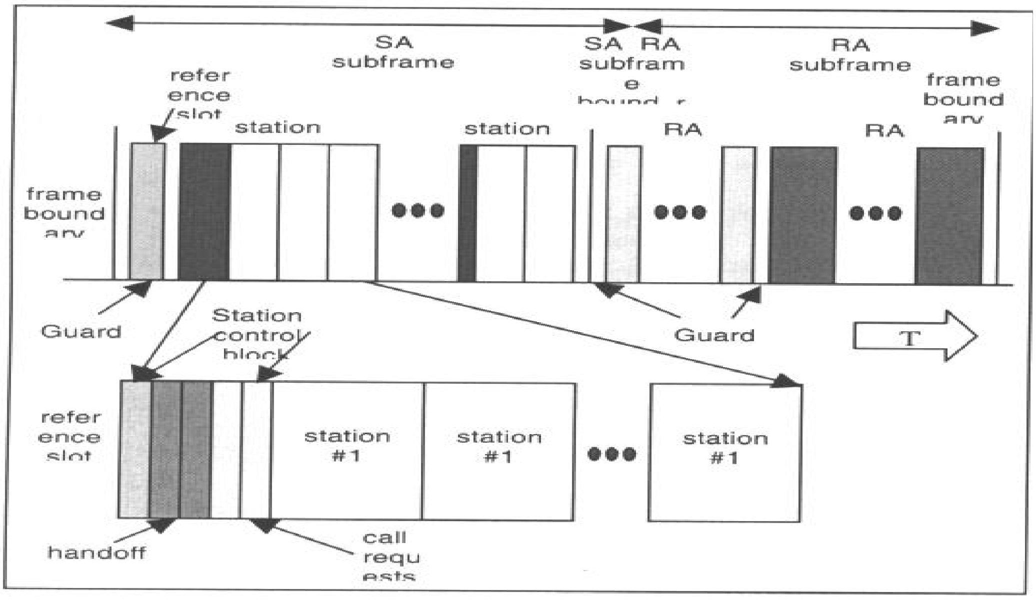


**Figure 3.11: Frame structure showing slot assignment and utilization**

To enable maximum channel efficiency, stations presently assigned to the channel are enabled to do control signaling and data transfer using the SA subframe. Each station has a beginning control slot followed by a series of data slots. Calls have been assigned slots through requests honored by the satellite as a result of previous frame requests incorporated in the first station's control field. Each station is assigned a continuous series of slots in order to reduce the need for guard slots to a minimum. After each station with active calls, a short subframe provides control slots. These slots are used by the stations that wish to communicate with the satellite but do not have any active calls. An example may be the establishment of a call reservation. Finally, a series of data slots are provided. Both control and data slots use random access so that contention can occur in either case.

The slot characteristics depend upon the network protocol supported through the satellite. For example, for ATM, the data slot size would be normally 53 bytes although additional bytes may be added to both up and down links in order to accommodate additional information needed for satellite control.

Stations indicate their need for slots assigned (or bytes) for their ongoing call(s) in the control slots or other information as, for example, call termination. Separate RA control slots are provided to handle call set-up packets, for the scheme to be able to make setting up a call easy and highly probable. Stations that want to establish a call will contend for a control sot. Random access control request slot size is designed so that data can be provided for communication with the satellite call control system. For example, in call setup, the calls characteristics will be supplied, including mean, maximum, and minimum data rates, call priority, billing information and possibly the expected duration of the call. If some data needed for control is too large for a Random Access (RA) control slot, it could use a data slot or multiple control slots. The remaining data slots are open to contention for all stations wishing to send data packets on a best effort basis. Random access data slots size depends, as with calls, upon the underlying characteristics of the network being supported. Since collisions are bound to occur, RA subframe slot size may be limited so that effects of collision do not result in a large amount of capacity being wasted.

The MAC uplink is designed to support a wide variety of low-level ground communication protocols. For example, it readily supports the QoS requirements of ATM traffic types. In ATM various traffic types such as CBR, rt-VBR, nrt-VBR, ABR and UBR are defined. In this scheme QoS is provided by resource allocation and sharing, with higher priority traffic being favored when bandwidth resources are limited. For service with varying rates, the reservation system can indicate when data rate requirements change so that the capacity can be reassigned.

### 3.7.4 Two MAC schemes for military satellite communications [81]

The United States Army is in the process of improving its information architecture by digitizing the battlefield. This encompasses examining ways in which it will connect maneuver units' Tactical Operation Centers (TOCs) within the division, so they can exchange command, control, and intelligence information (e.g., friendly and enemy positions, operation orders, collaborative planning, logistics, etc.). As the units are very mobile and can easily extend beyond line of sight (BLOS), terrestrial communication resources experience difficulties in maintaining contact. Therefore, satellite resources are being planned to bridge this shortfall.

If these limited, beyond line-of-sight resources are to be used efficiently, the communication layer (transport network, data and physical) protocols need to be examined for their performance to deliver unicast, multicast and broadcast addressed messages.

The schemes presented in this subsection focus on the data link layer and extend a previous investigation on candidate channel access protocols for broadcast satellite networks. More specifically, they concentrate on the channel access portion of the data link layer because this is the area that has the greater impact on channel bandwidth efficiency.

The focus of the protocols was on identifying the technical parameters that had the largest impact on bandwidth utilization. The three dominant parameters identified during the design were: elimination of a single point of failure (distributed operation), handling of propagation delays (delay and delay jitter), and maximizing usage of available bandwidth during both high and low loading (resource adaptation, reliability, overhead and implementation complexity).

The satellite terminals for which the schemes were deployed are intended to be used by the Army at the Brigade and Battalion TOCs. Their primary mission will be to provide continuous and extended range communication links between these command and control elements. Associated with each satellite terminal will be a subnetwork of radios and host platforms that will use the satellite services. These hosts and radios comprise the Tactical Internet connected to the satellite terminals through terrestrial radio networks such as Enchanced Position Location Radio system (EPLRS) and Single Channel Ground Airbone Radio System (SICGARS). Figure 3.12 shows an example geographical layout of the satellite network.

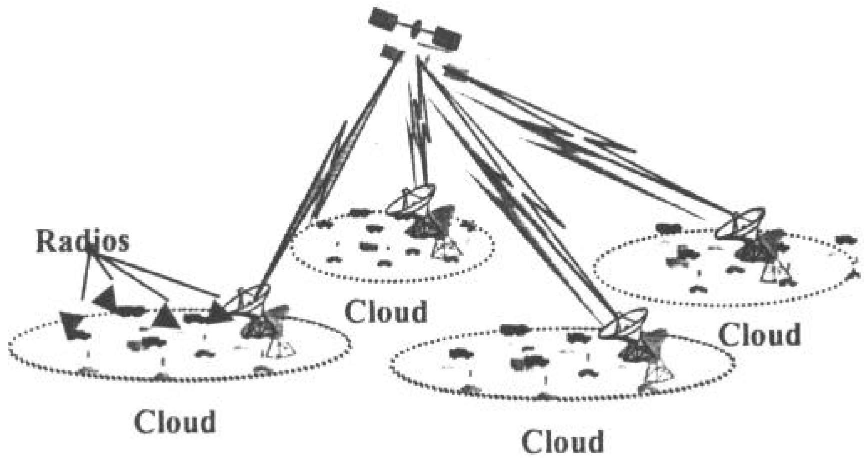


**Figure 3.12: Geographical Layout of the Satellite Network**

The satellite network is assumed to be fully connected with low mobility relative to the underlying terrestrial network. The satellite terminals serve as gateways between communicating terrestrial radios from one TOC to another. Uplink and downlink channels are separate. The uplink channel will have a maximum capacity of 128kbps information throughput at the data link level that is shared among the terminals. The satellite will be in geosynchronous orbit and have total propagation and processing delays from terminal to terminal of 500 milliseconds. The system will provide a $1* 10^{-5}$ bit error rate (BER) link quality.

The traffic transmitted through the satellite network is assumed to consist of Situational Awareness (SA), Command and Control ($C_2$) messages, and voice. The packets at the network layer will be Internet Protocol (IP) and addressed to both unicast and multicast destinations. The packet lengths assumed were SA packets of 100 bytes in length, $C_2$ packets of 1500 bytes in length and voice packets around 720 bits in length.

The two MAC layer protocols developed and analyzed are: ***a) the Windows Overlapping Reservation Protocols (WORP)*** and ***b) the Dynamic Assignment Time Division Multiple Access (DA-TDMA) protocol.*** The fundamental premise behind each protocol is that each terminal within the network will periodically transmit its bandwidth requirements in a control frame to all other terminals. Each terminal will then periodically process these requests, and run a common algorithm to assign bandwidth for the next finite duration of time.

Regarding the frame structure adopted in the schemes, time has been divided into several groupings; time slots, frames, and epochs; a time slot is the lowest increment for allocation. A frame is assumed to consist of M slots, and an epoch is the number of frames that a given terminal bandwidth allocation will remain in effect. The epoch length is required to be at least as long as the one way link delay to ensure synchronization between terminals' bandwidth/time slot allocation. Figure 3.13 depicts the Frame structure:

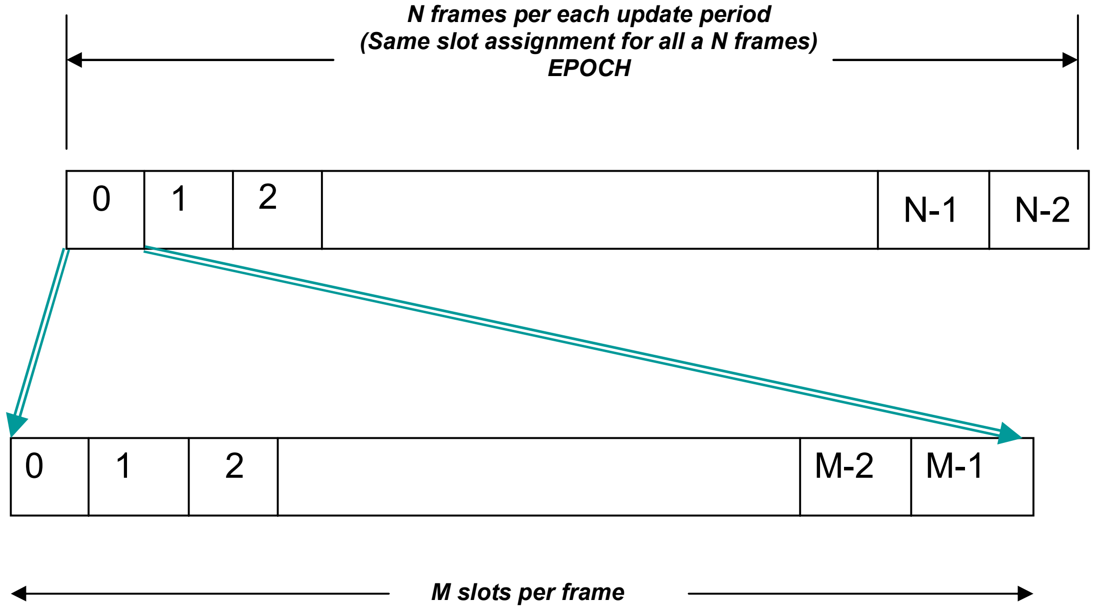


**Figure 3.13 MAC Frame Structure**

☑ ***WORP*** can be categorized as a reservation based type of channel access protocol, although it also has some characteristics of a round robin and contention resolution protocol.

The protocol has the cyclic, round robin characteristics similar to TDMA. Each terminal must periodically transmit its bandwidth requirements to every other terminal. The control packet contains the following information:

- Terminal's unique identification (i.e Terminal ID)
- First frame number that is being considered for transmitting data
- Number of time slots that it knows is used within each frame of its reservation window.

The control packet is 28-60 bytes in length.

Each terminal, at a minimum, transmits a control packet twice an epoch. Terminals also have a fixed sequential order in which these packets are transmitted. Before a terminal transmits its control packet, it updates a table that it uses to track the number of used time slots within each frame of an epoch. After it receives its preceding terminal's control packet, the terminal will determine the total number of packets within its queue and calculate the required number of time slots within its reservation window to transmit the queued data. The reservation window is the maximum number of frames it considers for transmitting data. The window size is normally the same for all terminals, such as 25 frames in a 50-frame epoch. The first frame it considers available for transmission can be designated based upon many factors, but it must be at a minimum the propagation delay to allow other terminals to receive its control packet.

The channel access protocol only attempt to reserve bandwidth based upon the information to be transmitted and does not consider packet prioritization. The data link or higher layers determine if the highest priority packet is transmitted first.

☑ ***DA-TDMA*** protocol is also a decentralized protocol (i.e., no master controller) that requires a common algorithm to be run at each terminal to allocate/deallocate bandwidth dynamically based upon user bandwidth requirements. The protocol takes advantages of the TDMA approach for an equal loading network and extends it to a dynamic protocol applicable to a non-uniform, bursty traffic load.

The DA-TDMA protocol maintains a decentralized control of channel resources that are assigned to users based upon relative demand. The resource management function, called the Network Resource Manager (NRM) is run on each terminal participating in the network, and is responsible for monitoring a terminal's queue sizes and QoS requirements, and allocating bandwidth accordingly. Each terminal maintains both a global slot table that lists the time slots assigned to each terminal for the current epoch, and a global queue table that lists the queue sizes within each terminal's buffers. Other terminals' queue sizes in the global queue table are obtained by listening to each terminal's queue size advertisement within a control packet. The entry for a terminal's own queue size must be delayed to account for the satellite delay and ensure all terminals are synchronized.

# Conclusions

This thesis presented a literature review on the subject of Multiple Access Control Protocols for Mobile Satellite Networks.
The work which has been conducted on the subject during the last two decades is significant, and the results exhibited by the most recent protocols show that they are able to efficiently integrate different types and classes of traffic over satellite networks, overcoming the inherent problems in satellite transmission.
However, to cope with the challenges related to supporting both the existing and the ever increasing new network services, future satellite technologies will need to fine tune and/or incorporate new sets of traffic control procedures, and this seems to be the future direction for Multiple Access Control Protocols for Mobile Satellite Networks.

# References

[1] **A.C. Clarke, "Extra Terrestrial Relays", Wireless World, pp.305-308, 1945**

[2] **D.J. Whalen, "Communications satellites: making the global village possible", http://www.hq.nasa.gov/office/pao/History/satcomhistory.html**

[3] **J. Farserotu and R. Prasad, "A survey of future broadband multimedia satellite systems, issues and trends",** ***IEEE Communications Magazine*****, vol.38, no. 6, pp. 128-133, 2000**

[4] **R. L, Gobbi, J. D. Grant, D.S. Rosener, and M. Liu, "Astrolink: an evolutionary telecommunications venture",** ***IEEE Seminar on Broadband Satellite:*** **The Critical Success Factors Technology, Services and Markets, vol. (ref. No. 00/067), pp. 9/1-9, 2000.**

[5] **Skybridge satellite, http:// www.skybridgesatellite.com/**

[6] **T.R. Henderson and R. H. Katz, "Transport protocols for Internet-compatible satellite networks",** ***IEEE Journal on Selected Areas in Communications (JSAC),*** **vol. 17, no. 2, pp. 326-344, 1999.**

[7] **C. Partidge and T.J. Shepard, "TCP/IP performance over satellite links",** ***IEEE Network,*** **vol.11, no.5, pp. 44-49, 1997**

[8] **M.Allman, D. Glover, and L. Sanchez, "Enhancing TCP over satellite channels using standard mechanisms" (RFC 2488), www.ietf.org/rfc/rfc2488.txt?number=24888, 1999**

[9] **N. Chani and S. Dixit, "TCP/IP enhancements for satellite networks",** ***IEEE Communications Magazine,*** **vol.37, no.7, pp.64-72, 1999**

[10] **J. Kulik, R. Coulter, D. Rockwell and C. Partridge, "Paced TCP for high delay-bandwidth networks",** ***WOSBIS at IEEE Global Telecommunications Conference (GLOBECOM)*****, pp.56-65, 1999.**

[11] **M.Wittig, "Satellite onboard processing for multimedia applications",** ***IEEE Communications Magazine*****, vol.38, no.6, pp.134-140, 2000.**

[12] **T. M. Chen and S. S. Liu, "ATM Switching Systems", 1995**

**[13] ATM Forum Technical Committee, ″Traffic Management Specification Version 4.0″, April 1996.**

**[14] W.Stallings, ″ISDN and Broadband ISDN with Frame Relay and ATM″, Prentice Hall, 1995.**

**[15] N. Passas, ″Multiple Accesses Control in Wireless ATM Networks″, *Ph.D. Dissertation, Dept. of Informatics, University of Athens*, and Dec.1997.**

**[16] K. Sriram, ″Methodologies for bandwidth allocation, transmission scheduling, and congestion avoidance in broadband ATM networks ″, *Computer Networks and ISDN Systems*, Vol. 26, No. 1, pp. 43-59, 1993.**

**[17] O.T.W. Yu and V.C.M. Leung, ″Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN″,*IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 7, pp. 1208-1225, September 1997.**

**[18] B. L. Mark and G. Ramamurthy, ″Joint source-channel control for real-time VBR over ATM via dynamic UPC renegotiation″, *IEEE Globecom '96*, Vol. 3, pp. 1726-1731, London, UK, November 1996.**

**[19] ITU-T, Recommendation I.371, Traffic and Congestion Control in B-ISDN, May 1996**

**[20] ATM-Forum, Proposed Charter, Work Plan and Schedule for Wireless ATM Working Group, 96-0712/PLEN, 1996**

**[21] ITU-T, Recommendation I.356 Draft 4R, B-ISDN ATM Layer Cell Transfer Performance, February 1996**

**[22] ATM-Forum, Traffic Management Specification Version 4.0 Technical Report 0056, April 1996**

**[23] H. Zimmermann, "OSI reference model- The ISO model of architecture for open systems interconnection", *IEEE Transactions on Communications,* vol.COM-28, no. 4,pp.425-432, 1980**

**[24] L.Kleinrock and F.A. Tobagi, "Packet switching in radio channels. I. Carrier sense multiple-access modes and their throughput-delay characteristics", *IEEE Transactions on Communications*, vol. COM-23, no.12, pp. 1400-1416, 1975**

[25] **F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels. II. The hidden terminal problem in carrier sense multiple-access and the busy-tone solution",** ***IEEE Transactions on Communications,*** **vol. COM-23, no.12, pp.1417-33, 1975**

[26] **T. C Tozer, "Satellite Access Techniques,"** ***in Satellite Communication Systems*****, B. G. Evans, 3rd ed:** ***IEEE,*** **1999**

[27] **L. Kleinrock, "Queueing Systems- Volume I: Theory",** ***John Wiley & Sons*****, 1975**

[28] **N. Abramson, "The throughput of packet broadcasting channels",** ***IEEE Transactions on Communications*****, vol. COM-25, no. 1, pp.117-128, 1977**

[29] **L. Kleinrock and S. S. Lam, "Packet switching in a multi-access broadcast channel: performance evaluation",** ***IEEE Transactions on Communications*****, vol. COM-23, no. 4, pp.410-423, 1975.**

[30] **H. Kobayashi.Y.Onozato, and D. Huynh »An approximate method for design and analysis of an ALOHA system",** ***IEEE Transactions on Communications*****, vol. COM-25, no.1, pp148-157, 1977**

[31] **D. Raychaudhuri and K. Joseph, "Performance evaluation of slotted ALOHA with generalized retransmission back off",** ***IEEE Transactions on Communications*****, vol. 38, no. 1, pp.117-122, 1990** [32]

[33] **N.B. Meisner, J. L. Segal, and M.Y. Tanigawa, "An adaptive retransmission technique for use in a slotted-ALOHA channel," IEEE Transactions on Communications, vol. COM-28, no. 9, pt.2, pp. 1776-1778, 1980. .**

[34] **Y. Birk and Y. Keren, "Judicious use of redundant transmissions in multichannel ALOHA networks with deadlines", IEEE Journal on Selected Areas in Communications (JSAC), vol. 17, no.2, pp.257-269, 1999 .**

[35] **Z. Liang, "Advances in ALOHA random multiple access techniques",** ***International Journal of Communications Systems*****, vol. 13, no.4, pp.353-364, 2000**

[36] **A. Burr, "Modulation and coding for wireless communications",** ***1st ed Prentice Hall,*** **2001**

[37] **M. Fitch and A. Fidler, "An overview of satellite access networks",** ***IEEE Communications Magazine, vol.35, no.7, pp.44-49, 1997*** **.**

[38] M. Wittig, "Large-capacity satellite multimedia systems", *IEEE Communications Magazine,* vol.35, no.7, pp.44-49, 1997

[39] A. Fidler, J,H., G. Hernandez, M.Lalovic, T. Pell, and I.G. Rose, « Satellite-a new opportunity for broadband applications", BT Technology Journal, vol.20, no. 1. pp.29-37, 2002.

[40] H. Peyravi, "Medium access control protocols performance in satellite communications", *IEEE Communications Magazine*, vol.37, no.3, pp. 62-71, 1999

[41] F. A. Tobagi, "Multiaccess protocols in packet communication systems", *IEEE Transactions on Communications*, vol. COM-28, no. 4, pp.468-488, 1980

[42] S. S. Lam, "Satellite packet communication-multiple access protocols and performance", *IEEE transactions on Communications*, vol. COM-27, no. 10p1, pp.1456-1466, 1979.

[43] C. J. Wolejsza, D. Taylor, M. Grossman, and W. P. Osborne, Multiple access protocols for data communications via VSAT networks", *IEEE Communications Magazine*, vol. 25, no. 7, pp. 30-39, 1987

[44] D. Raychaudhuri and K. Joseph, "Channel access protocols for Ku-band VSAT networks: a comparative evaluation", *IEEE Communications Magazine*, vol.26, no.5, pp. 34-44, 1988

[45] T. Nguyen and T. Suda, "Survey and evaluation of multiple access protocols in multimedia satellite networks", *IEEE SOUTHEASTCON*, vol.2, pp.408-423, 1990

[46] J. Puente and A. Werth, "Demand-assigned service for the INTELSAT global network", *IEEE EASCON*, pp.32-1A/9, 1977

[47] C. E Faflick and M. M. Mullarkey, ''Demand assigned multiple access techniques" *IEEE Spectrum*, pp.59-69, 1971

[48] K. Park and W. Willinger, "Self-similar network traffic and performance evaluation", *John Wiley& Sons, 2000*

[49] M. Mobasseri and V. C. M. Leung, "Bandwidth assignment for VBR traffic in broadband satellite networks", *IEEE Canadian Conference on Electrical and Computer Engineering,* vol.2, pp.654-658, 2000

[50] **T. Le-Ngoc, "Dynamic resource allocation schemes for multimedia satellite communications", IEEE *Personal, Indoor and Mobile Radio Communications Conference* (*PIMRC),* pp. 552-556, 1993.**

[51] **D. Grace, A. G. Burr, and T. C. Tozer, "Reducing call dropping in distributed dynamic channel assignment algorithms by incorporating power control in wireless ad hoc networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 18, no.11, pp.2417-2428, 2000**

[52] **D. Grace, A. G. Burr, and T.C. Tozer, "Comparison of a distributed dynamic channel assignment scheme with multichannel CSMA in terrestrial radio environment", *IEE Proceedings Communications*, vol. 146, no. 3, pp.191-195, 1999**

[53] **D. Raychaudhuri, "Announced retransmission random access protocols", *IEEE Transactions on Communications*, vol. COM-33, no.11, pp.1183-1190, 1985**

[54] **H. Koraitim, S. Tohme, M. Berrada, and A. Brajal, Performance of multiple access protocols in geo-stationary satellite systems", *International Conference on High Performance Networking (HPN),* pp.25-43, 1998**

[55] **T. S. Yum, "The design and analysis of a scheduled-retransmission multi-access protocol for packet satellite communications", *IEEE International Conference on Communications (ICC)*, vol. 1, pp.278-283, 1987.**

[56] **E. W. M. Wong and T. S. Yum, "The controlled-SRMA protocol for packet satellite communication", *IEEE International Conference on Communications (ICC),* vol. 2, pp.812-818, 1988**

[57] **E. W. M. Wong and T. S. Yum, "A controlled multi-access protocol for packet satellite communication", IEEE Transactions on Communications, vol. 39, no.7, pp.1133-1140, 1991**

[58] **H. W. Lee and J. W. Mark, "Combined random/reservation access for packet switched transmission over a satellite with on-board processing. I. Global beam satellite', *IEEE Transactions on Communications,* vol. COM-31, no. 10, pp.1161-1171, 1983**

[59] **S. Bose and S. S. Rappaport , " High capacity: low dealy packet broadcast multi-access", *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-16, no.6, pp.830-838, 1980**

[60] **T. Suda, H. Miyahara, and T. Hasegawa, "Performance evaluation of an integrated access scheme in a satellite communication channel", *IEEE***

***Journal on Selected Areas in Communications (JSAC),*** **vol. SAC-1, no.1, pp.153-164, 1983**

[61] **S. Bohm, A. K. Elhakeem, and V. K. M. Murthy, "Analysis of a movable boundary random/DAMA accessing technique for future integrated services satellites",** ***IEEE Global Telecommunications Conference (GLOBECOM),*** **vol.2, pp.1283-1288, 1993**

[62] **T. Ors, Z. Sun, and B. G. Evans, "Analysis of an adaptive random/reservation MAC protocol for ATM over satellite",** ***IEEE Vehicular Technology Conference (VTC),*** **vol. 2, pp.1523-1527, 1998**

[63] **T. Ors, Z. Sun, and B. G. Evans, "An adaptive random-reservation MAC protocol to guarantee QoS for ATM over satellite",** ***International Conference on Broadband Communications,*** **pp.107-118, 1998**

[64] **T. Ors, Z. Sun, and B. G. Evans, "A MAC protocol for ATM over satellite",** ***IEE Conference on Telecommunications,*** **pp. 185-189, 1998**

[65] **Teledesic, http://www.teledesic.com**

[66] **A. Iera, A. Molinaro, G. Aloi, and S. Marano, " Dynamic channel access protocol in geo-synchronous satellite networks",** ***IEEE Wireless Communications and Networking Conference,*** **vol. 2, pp.813-817, 2000**

[67] **J. E. Wieselthier and A. Ephremides, „ Anew class of protocols for multiple access in satellite networks",** ***IEEE Transactions on Automatic Control*****, vol. AC-25, no. 5,pp. 865-879, 1980**

[68] **D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet reservation multiple access for local wireless communications",** ***IEEE Transactions on Communications*****, vol. 37, no.8, pp.885-890, 1989**

[69] **D. J. Goodman and S. X. Wei, "Efficiency of packet reservation multiple access",** ***IEEE Transactions on Vehicular Technology*****, vol. 40, no. 1, pt.2, pp.170-176, 1991**

[70] **S. S. Lam, "Packet broadcast networks-a performance analysis of the R-ALOHA protocol",** ***IEEE Transactions on Computers,*** **vol C-29, no. 7, pp.385-392, 1993**

[71] **W. C. Wong, "Dynamic allocation of packet reservation multiple access carriers",** ***IEEE Transactions on Vehicular Techno*****logy, vol.42, pp.385-392, 1993.**

[72] **F. Ananasso and F. Delli Priscoli, "The role of satellites in personal communication services",** ***IEEE Journal on Selected Areas in Communications (JSAC),*** **vol.13, no.2, pp. 180-196, 1995**

[73] **E. Del Re, R. Fantacci, G. Giambene, and S. Walter, "Performance evaluation of an improved PRMA protocol for low Earth orbit mobile communication systems",** ***International Journal of Satellite Communications,*** **vol. 15, no. 6, pp.281-291, 1997**

[74] **E. Del Re, R. Fantacci, G. Giambene, and W. Sergio, "Performance analysis of an improved PRMA protocol for low Earth orbit-mobile satellite systems",** ***IEEE Transactions on Vehicular Technology,*** **vol. 48, no.3,pp.985-1001, 1999.**

[75] **T. Le-Ngoc and J. I. Mohammed, "Combined free/demand assignment multiple access (CFDAMA) protocols for packet satellite communications",** ***IEEE International Conference on Universal Personal Communications (ICUPC),*** **vol. 2, pp. 824-828, 1993.**

[76] **T. Le-Ngoc and S. V. Krishnamurthy, "Performance of combined free/demand assignment multiple-access schemes in satellite communications",** ***International Journal of Satellite Communications,*** **vol.14, no. 1, pp. 11-21, 1996.**

**[77] T. Le-Ngoc and I. M. Jahangir, "Performance analysis of CFDAMA-PB protocol for packet satellite communications",** ***IEEE Transactions on Communications*****, vol. 46, no. 9, pp.1206-1214, 1998.**

**[78] J. I. Mohammed and T. Le-Ngoc, "Performance analysis of combined free/demand assignment multiple access (CFDAMA) protocol for packet satellite communications",** ***IEEE International Conference on Communications (ICC)*****, vol.3, pp.1572-1576, 1995**

[79] **T. Le-Ngoc and S. V. Krishnamurthy, "Performance of combined free/demand assignment multiple-access schemes in satellite communications",** ***International Journal of Satellite Communications,*** **vol. 14, no. 1, pp. 11-21, 1996**

**[80] E. Foudriat, K. Maly, S. Olariu and P. Todorova, "Media Access Dynamic Bandwidth System to improve Satellite Network uplink Performances, accepted for publishing in Wireless Communications and Mobile Computing", 2003.**

[81] **Michael Snyder, Vicky Yu, Jeffrey Heissler "The MITRE Corporation Battlefield Communications & Networks Department Eatontown, New Jersey"**

[82] **P.D Mitchell, T. C Tozer and D. Grace, "Bandwidth assignment scheme for ON-OFF type data traffic via satellite",** ***Electronic Letters*****, vol. 37, no. 19, pp. 1191-1193, 2001.**

**[83] P.D Mitchell, T. C Tozer and D. Grace, "Burst targeted demand assignment multiple access (BTDAMA) for ON-OFF type data traffic via satellite",** ***IEEE Personal, Indoor and Mobile Radio Communications Conference (PIMRC),*** **2002.**