UNIVERSITÀ DELLA CALABRIA

Dipartimento di Elettronica,
Informatica e Sistemistica

Dottorato di Ricerca in
Ingegneria dei Sistemi e Informatica
XXI ciclo

Tesi di Dottorato

Scalable QoS Architectures and Resource Reservation Algorithms for Broadband Satellite Networks

Mauro Tropea

UNIVERSITÀ DELLA CALABRIA

Dipartimento di Elettronica,
Informatica e Sistemistica

Dottorato di Ricerca in
Ingegneria dei Sistemi e Informatica
XXI ciclo

Tesi di Dottorato

Scalable QoS Architectures and Resource Reservation Algorithms for Broadband Satellite Networks

Mauro Tropea

Coordinatore
Prof. Domenico Talia

Supervisore
Prof. Salvatore Marano

DEIS

DEIS- DIPARTIMENTO DI ELETTRONICA, INFORMATICA E SISTEMISTICA
Novembre 2008

Settore Scientifico Disciplinare: ING-INF/03

A mia moglie,
la mia unica
ragione di vita.

Acknowledgements

I would like to acknowledge all those very helpful people who have assisted me in my work. These three years have been very important in my life, a lot of events have happened. I will remember it always with great pleasure.

First of all I have to thank my wife that in these years has always helped me, giving me the strength to go on in all 'black moments', in particular, when I was in ESA, Holand, for my period of study abroad. We will always remember Leiden and the room in which I have spent 6 months of my life.

A special thank I have to say to my parents and my sister that have been very close to me.

Then, I have to thank my prof. and university tutor Marano who in these years has been able to follow me and give me the right advice in more than one moment.

I want to thank my friend and my 'boss' Floriano who allowed me to make this PhD experience giving me a dream otherwise impossible; then, in these years helped me constantly like a big brother, despite he is younger of me.

I have to thank my colleague "sor peps", Peppino, a real brother and a real friend and my officemate in all this time. He has been able to rise me up in more than one time with his right suggestions. One of those persons that enters in your life with absolute syntony from the beginning. Thanks also for the review of this thesis.

I want also to thank my colleague Fiore, a real friend that I had the fortune to meet in my life. Franco who has left us and works in Roma. Then, I have to greet all the other colleagues of the telecommunication laboratory, the mighty AntoFraGe, a real character, Apollonia with her 'ande rating' studies, Andrea and all the guys of the laboratory.

During my stage in ESA I have been lucky to know other special people. I have to thank Domenico who allowed me of making this beautiful experience in ESTEC/ESA, Holand.

I have to thank Ana who followed me in all the stage period trying to understand my 'not so good english'.

I want to thank Alfredo who gave me some useful suggestions in order to go on with the implementation of my simulator.

And then, I have to thank my housemates in Leiden, Annalisa, Simone, Marek, Samuele with whom I have spent beautiful moments in Holand.

I have also make a thanks to the Italian Space Agency (ASI) that has funded my PhD during these three years of study and, in particular, to my ASI tutor Giuseppe.

Mauro

Summary

Life today has changed thanks to two new objects that are with us in everyday life, the computer and the mobile phone. With these objects the way of communicating has completely changed. People can now be placed in video communication, despite being separated by long distances. Everything is also possible thanks to an important network element, the satellite. During the last few years studies on satellite environments have grown exponentially and now they represent an indispensable device in a communications network.

An important architecture that has been proposed in the last few years in order to provide video services is the Digital Video Broadcasting with Return Channel Satellite (DVB-RCS) system, which exploiting the broadcast nature of the satellite medium is capable of providing many different types of services to the end-users. It is a two-way interactive system that uses a satellite link as return channel in order to provide a fast access link to customers.

Given that the number of mobile users is ever increasing in the world, in the last few years research is focusing its attention on this class of customers, who require their services everywhere. Then, also the DVB-RCS research group is standardizing an architecture capable of satisfying mobile users requirements and which is contemporarily capable of maintaining the back compatibility with the previous standard for fixed users.

This thesis' work analyzes some aspects about Quality of Services (QoS) requirements and scalable networks composed of more than one layer in order to offer the maximum availability to the users, while increasing the coverage area of the overall system. In particular, the considered networks have a common element: the satellite segment. More than one type of integration between different networks will be presented and possible solutions to QoS and scalability issues will be proposed. These issues regard the network layer and the data link layer.

The introduction of QoS architecture in terrestrial, High Altitude Platform (HAP) and satellite segment is shown and performance analyses are conducted in order to show the achieved results. Special attention has been paid to an important control in a telecommunications network, which is the call admission control. This type of algorithm allows new users to be introduced into the system without decreasing the QoS offered to the already admitted users. In the last chapter of this thesis, a new issue is studied concerning the Digital Video Broadcasting with Return Channel Satellite + Mobile (DVB-RCS+M) standard, which is the standard born for managing the mobile satellite terminal. It regards an analysis of strategies for the selection of the most adequate Medium Access Control (MAC) scheme for the return link in order to achieve the best trade-off between system efficiency and provided QoS. In particular this last chapter has been realized in collaboration with the European Space Agency (ESA), European Space Research and Technology Centre (ESTEC) site in Noordwijk, The Netherlands.

Contents

List of Figures

List of Tables

Acronyms

ACQ ACQuisition Burst
ADSL Asymmetric Digital Subscriber Line
AEED Average End-to-End Delay
AF Assured Forwarding
AR Auto-Regressive
ARSVP Aggregate Resource ReSerVation Protocol
ARTES Advanced Research in Telecommunications Systems - Multimedia Programme
ATM Asynchronous Transfer Mode
AVBDC Absolute Volume Based Dynamic Capacity
AWGN Additive White Gaussian Noise

BB Base-Band
BE Best Effort
BEF Bandwidth Expansion Factor
BER Bit Error Rate
BoD Bandwidth on Demand
BP Bent Pipe
bps bit per second
BPSK Binary Phase Shift Keying

C/N Carrier-to-Noise
C2P Connection Control Protocol
CAC Call Admission Control
CBR Constant Bit Rate
CDMA Code Division Multiple Access
CDTI Centro para el Desarrollo Tecnolgico Industrial
CJVC Core Jitter Virtual Clock
CLS Controlled Load Service
CMF Control and Monitoring Functions
CoS Class of Service

CR Core Router
CRA Continuous Rate Assignment
CSC Common Signaling Channel burst

DAMA Demand Assignment Multiple Access
dB decibel
DB Delay Bound
DBS Direct Broadcast Satellite
DBS-RCS Direct Broadcast Satellite with Return Channel Satellite
DES Discrete Event Simulation
DFD Data Flow Diagram
DiffServ Differentiated Services
DLR Deutsches Zentrum fr Luft- und Raumfahrt
DoD Department of Defense
DPS Dynamic Packet State
DSCP Differentiated Services Code Point
DTH Direct-To-Home
DULM Data Unit Labeling Method
DVB Digital Video Broadcast
DVB-RCS Digital Video Broadcasting with Return Channel Satellite
DVB-RCS+M Digital Video Broadcasting with Return Channel Satellite + Mobile
DVB-S Digital Video Broadcasting Satellite
DVB-S2 Digital Video Broadcasting Satellite - second generation
DVD Digital Versatile Disc

E2E End-to-End
EDP Excess Demand Probability
EF Expedited Forwarding
EIRP Equivalent Isotropically Radiated Power
ER Egress Router
ESA European Space Agency
ESTEC European Space Research and Technology Centre
ESW EuroSkyWay
ETSI European Telecommunications Standards Institute

F/G Feeder/Gateway
FCA Free Capacity Assignment
FCC Federal Communications Commission
FDMA Frequency Division Multiple Access
fec forwarding equivalence class
FEC Forward Error Correction
FIFTH Fast Internet for Fast Train Hosts
fps frame per second
FSMC Finite State Markov Chain

FS Fixed Service
FSS Fixed Satellite Service
FTP File Transfer Protocol

GEO Geostatioary Earth Orbit
GLR GOP Loss Ratio
GMBS Global Mobile Broadband System
GOP Group of Picture
GS Guaranteed Service
GSO Geosynchronous
GTW Gateway
G-VBDC Guaranteed VBDC

HAP High Altitude Platform
HCAC HAP Call Admission Control
HCC HAP Connection Control
HGTW HAP Gateway Station
HMCS HAP Master Control Station
HNOC HAP Network Operation Center
HPA High Power Amplifier
HP-VBDC High Priority VBDC
HRRM HAP Radio Resource Management
HSE HAP Static Efficiency
HTML Hyper Text Mark-Up Language
HTRM HAP Traffic Resource Manager
HTTP Hyper Text Transfer Protocol
HUF HAP Utilization Factor

IAT Inter-Arrival-Time
IBR In Band Request
IntServ Integrated Services
IEC International Electrotechnical Commission
IETF Internet Engineering Task Force
IHL Inter-HAP Link
i.i.d. independent and identically-distributed
IOL Inter Orbit Link
IPN InterPlaNetary
IPTV Internet Protocol TeleVision
IR Ingress Router
ISDN Integrated Services Digital Network
ISI InterSymbol Interference
ISO International Organization for Standardization
ISP Internet Services Provider
IT Information Technology
ITU International Communication Union

ITU-R International Communication Union - Radiocommunication Sector
IWF Inter Working Functions

JVC Jitter Virtual Clock

kbps kilo bit per seconds
ksps kilo symbol per seconds

LDA Loss-Delay-based Adaptation Algorithm
LEO Low Earth Orbit
LER Label Edge Router
LNA Low Noise Amplifier
LoS Line of Sight
LRD Long Range Dependent
LSP Label Switched Path
LSR Label Switching Router

MAB Maximum Available Bandwidth
MAC Medium Access Control
Mbps Mega bit per seconds
MCS Master Control Station
MF-TDMA Multi-Frequency Time Division Multiple Access
MHUT Mobile HAP User Terminal
MOWGLY Mobile Wideband Global Link sYstem
MPEG Moving Picture Expert Group
MPEG-TS MPEG - Transport Stream
MPEG2-TS MPEG2 - Transport Stream
MPL Minimum Path Latency
MPLS Multi Protocol Label Switching
MSS Mobile Satellite Service

NCC Network Control Center
NGN Next Generation Network
NGSO Non-Geosynchronous
NLoS Non Line of Sight
NOAC Number of Overall Admitted Calls
NoAd Non Adaptive
NOC Network Operator Center
nRT non Real Time
nrt-VBR Non-Real Time Variable Bit Rate

OBP On-Board Processor
OBR Out Band Request

PD Propagation Delay

PER Packet Error Rate
PHB Per Hop Behavior
PSTN Public Switched Telephone Network

QD Queuing Delay
QOAS Quality Oriented Adaptation Scheme
QoS Quality of Services
QPSK Quadrature Phase Shift Keying

RBDC Rate Based Dynamic Capacity
RCS Return Channel Satellite
RCST Return Channel Satellite Terminal
RF Radio Frequency
RL Return Link
RRM Radio Resource Management
RSVP Resource reSerVation Protocol
RT Real Time
RTT Round Trip Time
rt-VBR Real Time Variable Bit Rate

SAC Satellite Access Control
SaT Satellite User Terminal
SBBP Switched Batch Bernoulli Process
SCAC Satellite Call Admission Control
SCORE Scalable CORE
SCPC Single Channel Per Carrier
SDR Software Defined Radio
SGTW Satellite Gateway
SI Signaling Information
SLA Service Level Agreement
SMCS Satellite Master Control Station
SMDB Statistical Multiplexing based on Discrete Bandwidth Levels of GOP Rate
SMND Statistical Multiplexing based on the Normal GOP Distribution
SR Static Rate
SRD Short Range Dependence
SSE Satellite Static Efficiency
ST Satellite Terminal
SUF Satellite Utilization Factor
SYNC SYNChronization

TBSA Threshold Basic Switching Algorithm
TBTP Terminal Burst Time Plan
TCP Transport Control Protocol
TDMA Time Division Multiple Access

TFRCP TCP Friendly Rate Control Protocol
TIM Terminal Information Message
TM-RCS Technical Module - Return Channel Satellite
TOS Type-of-Service
TRF TRaFfic burst
TRM Traffic Resource Manager

UL UpLink
UMTS Universal Mobile Telecommunications System

VBDC Volume Based Dynamic Capacity
VBR Variable Bit Rate
VHS Video Home System
VoD Video on Demand
VoIP Voice over IP
VPN Virtual Private Network
VR Virtual Reality
VSAT Very Small Aperture Terminal

WiFi Wireless Fidelity
WWW World Wide Web

1

Next Generation Network (NGN) - State of the Art

As demands on the network continue to grow, it is increasingly important to upgrade the existing infrastructure in order to offer higher bandwidth and service level guarantees to the users.

The Next Generation Network (NGN) is designed to be all IP-based supporting heterogeneous core and access technologies for broadband and mobile applications. Fixed, as well as, mobile satellite systems will be fully integrated with future wireless and wired networks.

The general definition of International Communication Union (ITU) provided for NGN in [1] is: "A packet-based network able to provide telecommunication services and able to make use of multiple broadband, Quality of Services (QoS)-enabled transport technologies and in which service-related functions are independent from underlying transport related technologies. It enables free access for users to networks and to competing service providers and/or services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users."

1.1 Global Communication Network Scenario

The future generation of communication networks provides "multimedia service", "wireless (cellular and satellite) access to broadband networks" and "seamless roaming among different system".

The term NGN is commonly used to give a name to the changes to the service provision infrastructures that have already started in the telecommunication and Information Technology (IT) industry. A NGN is a packet-based network able to provide services including Telecommunication Services. It offers unrestricted access by users to different service providers. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users. The NGN is characterized by the following fundamental aspects [2]:

- Packet-based transfer;
- Separation of control functions among bearer capabilities, call/session, and application/ service;
- Decoupling of service provision from network, and provision of open interfaces;
- Support for a wide range of services, applications and mechanisms based on service building blocks (including real time/ streaming/ non-real time services and multi-media);
- Broadband capabilities with end-to-end QoS and transparency;
- Unrestricted access by users to different service providers;
- Converged services between Fixed/Mobile network;
- Independence of service-related functions from underlying transport technologies;
- Compliant with all Regulatory requirements, for example concerning emergency communications and security/privacy, etc;

In this context, the integration of heterogeneous networks that have the task of providing full access and total coverage to the users that, nowadays, exchange a lot of information whenever and wherever takes place. A possible scenario of a NGN is depicted in Fig. 1.1. A main role in networks cooperation is played by the wireless networks that are capable of reaching users in places where the normal terrestrial infrastructures do not have the possibility. Other the well-known satellite architecture there is another wireless platform that is acquiring a lot of importance in telecommunications, called High Altitude Platform (HAP). The main advantage of these types of architectures is the possibility of offering ubiquitous coverage, bandwidth flexibility, multipoint-multipoint connectivity and fast service initiation after deployment and reliability.

1.2 Satellite Networks

Satellite systems have been an important element of telecommunications networks for many years serving, in particular, long distance telephony and television broadcasting. The involvement of satellite in IP networks is a direct result of new trends in global telecommunications where Internet traffic will hold a dominant share in the total network traffic. The large geographical coverage of the satellite footprint and its unique broadcasting capabilities as well as its high-capacity channel combined with readily available Ka-band spectrum will retain satellite systems as an irreplaceable part of communications systems, despite its high cost and its long development and launching cycle. Traditionally, satellite communication systems have played a significant role in supporting services such as TV broadcasting, digital messaging, enterprise Virtual Private Network (VPN) and point-to-point telecommunications and data services.

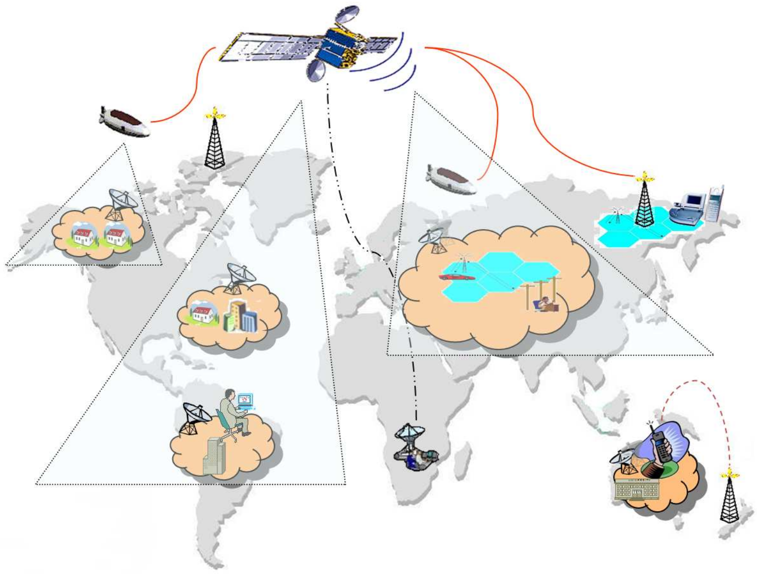

Fig. 1.1. Next Generation Network (NGN) Scenario

The recent Internet growth has resulted in a new generation of applications with higher throughput requirements and communication demands. Service, Network and Internet access providers are faced with a challenge to meet the higher capacity access to the end user and wider service offerings. Satellite network systems can be optimized to meet new service demands such as aeronautical and mobile applications. For example, trade-off studies of the frequency of operation, (Ku, Ka), processing versus nonprocessing payload, switching (IP, MPLS), QoS mechanisms, type of antennas, network protocols, transport protocols, crosslayer designs, network management, techniques have to be performed to meet the driving applications and requirements [3].

Several broadband satellite networks at Ka-band are planned and being developed to provide such global connectivity for both Fixed Satellite Service (FSS) and Mobile Satellite Service (MSS), using Geosynchronous (GSO) and Non-Geosynchronous (NGSO) satellites. Currently, GSO satellite networks with Very Small Aperture Terminal (VSAT) at Ku-bands are being used for several credit card verifications, rental cars and banking applications.

At that time, Ka-band satellite communications systems became so popular because they could provide:

- Large bandwidth: the large amount of bandwidth availability in Ka-bands is the primary motivation for developing Ka-band satellite systems since lower frequency bands have become congested;
- Small antenna size: as the frequency goes up, the size of the antenna will decrease for a given gain and beamwidth. For a fixed antenna size, this

will significantly reduce the interference from adjacent satellite systems. Obviously, the price of the smaller antenna will be lower, which makes broadband satellite service affordable to millions of commercial and residential end-users;

- Larger system capacity: Ka-band satellites provide smaller spot-beams to increase the satellite power density and allow large frequency reuses, which will lead to higher spectrum occupancy. Many user terminals can be served simultaneously;
- Ubiquitous access: services are available at any location within the satellite footprint, especially in locations where terrestrial wired network are impossible or economically unfeasible;
- Flexible bandwidth-on-demand capability: this feature maximizes the bandwidth and resource utilization and minimizes the cost to end-users.

On the other hand, Ka-band satellite links suffer degradation due to atmospheric propagation effects, which are more severe at the Ka-band than the degradation that happens at lower frequency bands. The primary propagation factors are rain attenuation, wet antenna losses, depolarization due to rain and ice, gaseous absorption, cloud attenuation, atmospheric noise and tropospheric scintillation. Among those factors, rain attenuation is the most challenging obstacle to Ka-band systems [4]

1.3 Satellite Communication Fundamentals

There are two basic types of transponders, Bent Pipe (BP) and On-Board Processor (OBP). The early satellite transponders were based on analog transmission, but most modern satellite systems deliver signals digitally to ensure reliability and accuracy in information transmission. Digital switching techniques in OBP have facilitated a large scale deployment of affordable satellite-terrestrial networks.

1.3.1 Transparent Satellite System

The BP transponder acts as a transparent repeater. It consists of receiving and transmitting antennas, a Low Noise Amplifier (LNA) receiver, a frequency converter and a High Power Amplifier (HPA). The earth station transmitter will deliver signals to the satellite receiver. The uplink signals will be received at the receiving antenna, down converted, fed to the HPA and then transmitted down to the receiving earth station via the transmitting antenna. Usually, no change is made to the signal except an amplification to overcome the large path losses and a frequency conversion to separate the up and down links. Generally, the transponder is transparent to the users since the transmitting signal from one earth station will 'bounce' and arrive to another earth station with its characteristics unchanged.

Figure 1.2 shows the basic BP satellite link. The conventional way of characterizing the satellite link behavior using BP transponders is to use Carrier-to-Noise (C/N). The C/N ratio represents the decibel (dB) difference between the desired carrier signal power and the undesired noise power at the receiver. It also indicates the received signal quality for both analog and digital transmissions. In satellite communications systems the C/N calculation is often called a power link budget.

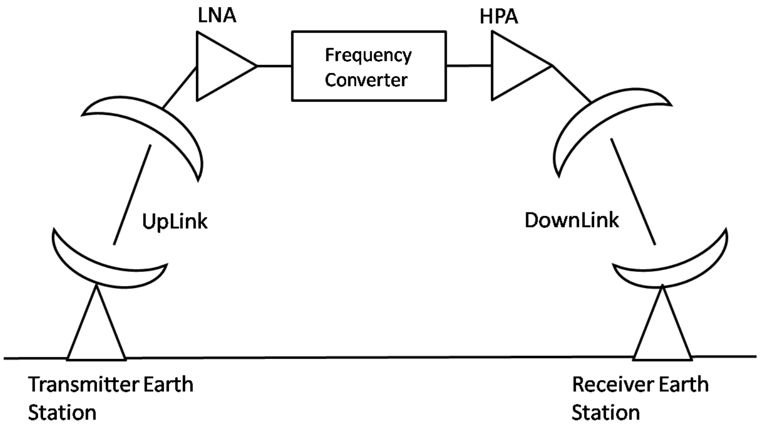

Fig. 1.2. Bent Pipe Transponder Satellite

1.3.2 Intelligent Satellite System

The conventional BP satellite delivers signals on the same route; from the receiver to the transmitter, all signals on that specific transponder will usually be together, coming from the same transmitting earth station and going to the same receiving earth station. This limits the flexibility of the satellite network application.

The OBP satellite system, consisting of regenerative transponders and on-board switching with multiple spot-beams, provides bandwidth on demand with low processing delay, flexible inter-connectivity and lowered ground station costs. In an OBP satellite system, both the uplink and the downlink of the OBP system are independent to each other. The uplink signals with distortions or noise reaching the space station receiver are down-converted, demodulated, de-multiplexed and reconstructed. The reconstructed signals are then modulated, multiplexed and up-converted to be transmitted at the downlink. Thus, the uplink degradation will have no effect on the downlink transmission. This process, called Base-Band (BB) processing, significantly improves the overall link performance at the receiving earth station.

Figure 1.3 shows the basic OBP system architecture and its links. Because demodulation is digitally applied in the regenerative transponder, it is necessary to represent the C/N ratio in terms of Bit Error Rate (BER). BER used in digital signals is to measure the probability of bit error that will occur in a given amount of time in the system.

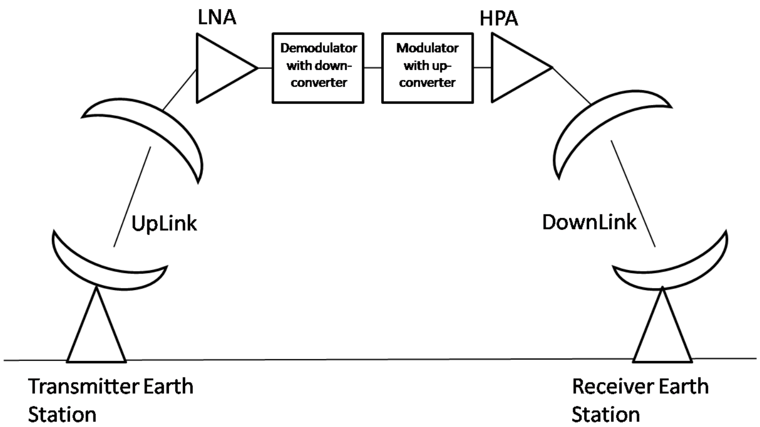

Fig. 1.3. On Board Processor Transponder Satellite

1.3.3 Satellite Link Multiple Access Techniques

Since large amounts of bandwidth are available on Geostatioary Earth Orbit (GEO) Ka-band satellites, an appropriate bandwidth management technique is necessary. One of the best ways is to use a multiple access technique. In satellite communications systems, multiple access allows many earth stations to share a transponder even thought their carriers have different signal characteristics.

Three common types of multiple access deployed in satellite communications systems are Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA). A common hybrid solution is used by combining techniques such as FDMA / TDMA well-known in literature as Multi-Frequency Time Division Multiple Access (MF-TDMA).

1.4 Digital Video Broadcasting with Return Channel Satellite (DVB-RCS) Architecture

DVB-RCS [5], [6] is a technical standard, designed by the Digital Video Broadcast (DVB) Project, that defines a complete air interface specification for two-way satellite broadband VSAT systems. Low cost VSAT equipment can provide highly dynamic demand-assigned transmission capacity to residential and commercial/institutional users (see Fig. 1.4).

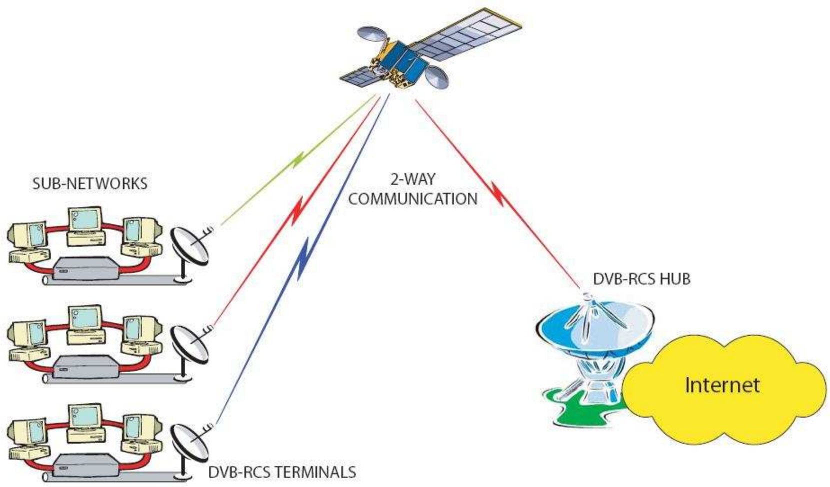

Fig. 1.4. Network Architecture for DVB-RCS

DVB-RCS provides users with the equivalent of an Asymmetric Digital Subscriber Line (ADSL) or cable Internet connection, without the need for local terrestrial infrastructure. Depending on satellite link budgets and other system design parameters, DVB-RCS implementations can dynamically provide anywhere up to 20 Mega bit per seconds (Mbps) to each terminal on the outbound link and up to 5 Mbps or more from each terminal on the inbound link. The standard is published by European Telecommunications Standards Institute (ETSI) as EN 301 790 [5].

DVB's Technical Module - Return Channel Satellite (TM-RCS) has now approved the Digital Video Broadcasting with Return Channel Satellite + Mobile (DVB-RCS+M) specification [7], providing support for mobile and nomadic terminals as well as enhanced support for direct terminal-to-terminal (mesh) connectivity. DVB-RCS+M includes features such as live handovers between satellite spot-beams, spread-spectrum features to meet regulatory constraints for mobile terminals, and continuous-carrier transmission for terminals with high traffic aggregation. It also includes link-layer Forward Error Correction (FEC) based on Raptor or Reed-Solomon codes, used as a countermeasure against shadowing and blocking of the satellite link [8].

Today there are several vendors of DVB-RCS standard compliant systems as well as proprietary VSAT systems that provide services of the same general nature as DVB-RCS, but without full compliance with the standard. DVB-RCS was developed in response to a request from several satellite and network operators who wanted to embark on large-scale deployment of such systems considering it essential to have an open standard in order to mitigate the risks associated with being tied to a single vendor. The standard was developed using state-of-the-art techniques, allowing an optimized trade-off between performance and cost. As a consensus based standard DVB-RCS also has a controlled evolutionary future, secured by global contributions to the system under an agreed and open framework.

1.4.1 How does it works

In its basic form, DVB-RCS provides hub-spoke connectivity; i.e., all user terminals are connected to a central hub that both controls the system and acts as a traffic gateway between the users and the wider Internet. User terminal consists of a small indoor unit and an outdoor unit with an antenna size not much bigger than a conventional direct-to-home TV receiver. Since the DVB-RCS terminal also transmits data, the outdoor unit includes a Radio Frequency (RF) power amplifier.

The user terminal offers an Ethernet that can be used for wired or wireless interactive IP connectivity for a local home or office network ranging from one to several users. In addition to providing interactive DVB services and Internet Protocol TeleVision (IPTV), DVB-RCS systems can thus provide full IP connectivity anywhere there is suitable satellite coverage, which in turn means most places on the earth including areas not covered by other solutions.

The core of DVB-RCS is a MF-TDMA transmission scheme for the return link, which provides high bandwidth efficiency for multiple users. The demand-assignment scheme uses several capacity mechanisms that allow optimization for different classes of applications, so that voice, video streaming, file transfers and web browsing can all be handled efficiently. DVB-RCS supports several access schemes making the system much more responsive and thus more efficient than traditional demand-assigned satellite systems.

These access schemes are combined with a flexible transmission scheme that includes state-of-the-art turbo coding, several burst size options and efficient IP encapsulation options. These tools allow systems to be fine-tuned for the best use of the power and bandwidth satellite resources.

The forward link is shared among a population of terminals using either Digital Video Broadcasting Satellite (DVB-S) (EN 300 421) [9] or the highly efficient Digital Video Broadcasting Satellite - second generation (DVB-S2) standard (EN 302 307) [10]. Adaptive transmission, to overcome variations in channel characteristics (e.g., rain fade), can be implemented in both the forward and return links. Beyond the basic hub-and-spoke architecture, the DVB-RCS air interface has also been deployed in systems that provide direct terminal-to-terminal mesh connectivity, either through satellite on-board processors that mirror the functions of a ground-based hub, or through transparent satellites, using terminals equipped with an additional demodulator.

1.4.2 Next Steps for DVB-RCS

DVB-RCS was first published in 2000 and it has been relatively stable. Until the introduction of DVB-RCS+M, changes have mainly been for maintenance, such as inclusion of support for the DVB-S2 forward link standard.

DVB-RCS+M is the first major revision of the standard. Although no further revisions are currently in progress, the DVB-RCS community is continuously assessing the commercial needs of the users, to ensure that the standard will continue to meet their objectives in the most efficient way. Successful trials and implementations have already been carried out using terminals mounted in trains, maritime vessels and aircrafts. This is being paralleled by the addition of sophisticated mobility support features to DVB-RCS. These include live handovers between satellite spot-beams, spread spectrum features to meet regulatory constraints for mobile terminals and countermeasures against shadowing and blocking of satellite link.

1.5 High Altitude Platform (HAP) Networks

HAP is a new technology of airships or planes that will operate in the stratosphere at an altitude of 17-22 km above the ground. The high potentialities of these network elements have captured the interest of academic and industrial environments in recent years. They have the potential capability to be quickly deployed and they do not need a complex infrastructure, as for the terrestrial network [11], [12].

Recently, the ITU has licensed several frequency bands for communications in HAP: 2GHz, 30 GHz, 48GHz, and mm-wave bands. These different frequency bands lead to the necessity of studying the characteristics of the physical medium and of the coding techniques in order to have an idea of the potentialities offered for new kinds of services in terms of allowable data rate, bit error rate and energy consumption. An important aspect is to understand and define which kind of antennas are suitable for this new network segment and which kind of terminals can be used in order to use new potentialities and to offer new services [13].

Many studies have been conducted in order to propose integration with the Terrestrial - Universal Mobile Telecommunications System (UMTS) network in order to reach inter-operability with the existent infrastructure; in addition, studies have been proposed considering the HAP segment as an access segment able to provide broadband and narrow-band services.

All these studies lead to a key question about the future of this technology: can the HAP be seen as a complementary technology or a disruptive technology? In the vision of this thesis, the HAP segment will represent, in areas without terrestrial infrastructure, a winning technology able to provide connection services and multimedia services. Further, it will represent in area with wired infrastructure an additional network element able to provide several benefits for new classes of users and services.

Also, if satellite networks, and in particular the GEO satellite networks, have become important in these last few years, the HAP layer will offer another complementary element able to reduce problems regarding the signal attenuation, the power of terminal and base station and the long propagation delay (GEO satellite presents high end-to-end delay).

In this context, it is important to consider the new role played by the HAP segment and to consider the potential integration of this new element with the existing networks. In particular, it is interesting to analyze the possible integration of HAP with the existing satellite network in order to obtain an efficient management and coexistence of the two networks. It is possible to use the satellite segment for services that do not need specific delay restrictions and HAP (when the segment is deployed) to give access to users that need particular delay-sensitive services.

1.6 Multi-layer Networks

The integration between different and heterogeneous networks leads to the realization of a hierarchical network scenarios, in which it is possible to see the inter-operation of more than one kind of networks [14]. A general hierarchical scenario is composed of a satellite that, with its spot beam, covers a great part of territory, i.e. a continent.

In this spot beam a certain number of HAP platforms are included, forming a constellation with a coverage area smaller than satellite one. At last, it is possible to find wireless access points or wired routers that permit to different users the network connection. This is a possible scenario that includes different network elements in order to be able to offer the maximal connectivity. An example of a heterogeneous, hierarchical, multi-layer scenario is depicted in Fig. 1.5.

It is easy to understand that satellite networks are irreplaceable for personal communications, for example in wide areas or in the case of disasters such as earthquakes, tsunamis, hurricanes or terrorist activities. In the absence of a wired infrastructure, any individual host or an ad hoc network can access the rest of the wired network through satellites. The advantages of a multi-layer satellite architecture are: a) scalability; b) low delay for real-time traffic; c) flexibility for personal communications.

1.7 Quality of Services (QoS) Issues over Next Generation Network (NGN)

Near real-time data communications (e.g. World Wide Web (WWW) browsing) has already demonstrated that the Internet and the underlying generation of packet network technologies, which operate on the basis of 'Best

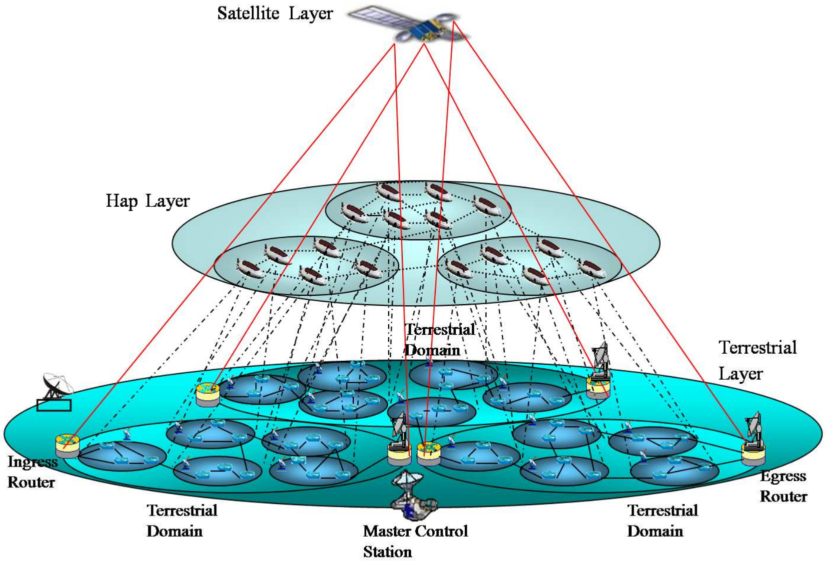

Fig. 1.5. Multi-Layer Network Scenario

Effort (BE)', provide insufficient QoS support. In the NGNs, packet network technologies will have to provide real-time services (voice, multimedia) and a improved QoS support. Packet based transport technologies will thus have to include advanced QoS management capabilities such as:

- QoS aware network equipment (Resource reSerVation Protocol (RSVP), Integrated Services (IntServ), Differentiated Services (DiffServ), Multi Protocol Label Switching (MPLS));
- Admission control mechanisms;
- Interaction between call signalling, resources management and admission control.

As far as satellite based systems are concerned, especially those including sophisticated resource management features in order to use a shared access such as DVB-RCS, no architecture or mechanisms have been defined or recommended by standardization bodies. The provision of an end-to-end QoS based service with a satellite access is thus largely undefined and will eventually lead to system specific or proprietary solutions.

Such recommendations exist in the field of cable access and mobile access networks. Their absence in the satellite field may lead to slow down the uptaking of NGN and their associated services by satellite. Moreover, due to the characteristics of those services, many of them being of real-time nature and to the specificity of geostationary DVB-RCS based satellite access systems (delay, multiple access), QoS aspects need to be analyzed specifically.

The real-time nature of interpersonal communication services poses NGNs significant challenges in terms of resource management. Typically, these services require resources to be made available on demand and expect the request to be answered almost instantly (it is unlikely that users will tolerate lengthy call set up times given their experience of todays telephony services), their QoS requirements tend to be fairly stringent and they need to be maintained for the duration of the call.

Since QoS-aware services and applications rely heavily on signalling for their operation it is important to consider the resource requirements of the signalling traffic itself. An obvious consideration is the additional bandwidth that will be consumed by the signalling traffic. Although it should still remain small compared to the user traffic, it is likely to become significant enough and should not be overlooked.

Perhaps, more important are the end-to-end delay and packet loss experienced by signalling traffic. Some services, particularly those with a real-time component, will require an almost trouble-free signalling channel for successful operation. Any problems experienced by the signalling traffic are likely to lead to degraded service performance, if not a complete service failure. Other services will be more tolerant and have less demanding targets for their associated signalling.

QoS mechanisms provide service differentiation and performance measures for Internet applications according to their requirements. Performance assurance addresses bandwidth, loss, delay and delay variation. Bandwidth is a fundamental resource for satellite communication and its proper allocation determines the system throughput. End-to-end delay is also important for several applications.

There are choices to provide the Internet QoS. These are IntServ [15], DiffServ [16] and MPLS [17]. However, the research of application of these QoS framework to broadband satellite network requires supporting IP QoS in a dynamic demand assignment capability environment. A workable QoS architecture must provide a means for specifying performance objectives for different types of packets, as well as, a means of delivering those performance objectives.

To give different packets different treatment, the network infrastructure must be capable of distinguishing between packets by means of classification, queuing packets separately as a result of the classification, scheduling packet queues to implement the differential treatments, providing means for measuring, monitoring, and conditioning packet streams to meet requirements of different QoS levels. These can be realized through the implementations of mechanisms in the packet-forwarding path. The number of ways in which a packet can be treated in the forwarding path is limited.

Aggregating individual flows according to their common packet forwarding treatment leads to a reduction of state and complexity. QoS requires the cooperation of all network layers from top-to-bottom, as well as, every network element from end-to-end. At each layer, using efficient technologies and counteracting any factors causing degradation achieve the user performance requirements. For example, at the physical layer, bandwidth efficient modulation and encoding schemes, such as concatenated coding and adaptive coding, have to be used to improve the BER and power level performance under whether conditions such e.g. rain.

Similarly, a superior QoS is achieved providing the guaranteed bandwidth at the link layer by using efficient bandwidth on demand multiple access schemes and studying the interaction of bandwidth allocation mechanisms in the presence of congestion and fading. The provision of a specific bandwidth to be offered by the physical layers to the upper layers implies the existence of a bandwidth allocation scheme that shares the bandwidth available among the different user terminals with different traffic classes.

To satisfy the different QoS service level guarantees according to Service Level Agreement (SLA)s with DiffServ at the network layer, service classification, marking, queuing, and scheduling functions can be used. Currently, most of the Internet applications use the Transport Control Protocol (TCP) and several TCP enhancements exist for satellite environment mitigating the link impairments. Eventually, these QoS parameters have to be mapped to the application QoS as required by the system design and target applications.

1.8 QoS Parameters

The QoS parameters, which can be used in order to satisfy the constraints impose by the users, are:

- Delay: the time for a packet to be transported from the sender to the receiver;
- Jitter: the variation in end-to-end transit delay;
- Bandwidth: the maximal data transfer rate that can be sustained between two end points;
- Packet Loss: defined as the ratio of the number of undelivered packets to the total number of sent packets;
- Reliability: the percentage of network availability depending upon the various environmental parameters such as rain.

To achieve an end-to-end QoS in both satellite and/or hybrid satellite/terrestrial networks is a non-trivial problem. However, end-to-end QoS objectives, including security, need considerable research. A successful end-to-end QoS model depends upon the various interfaces at each subsequent lower layer to the upper layers.

1.9 QoS Architecture

There are several mechanisms for QoS provisioning. These are IntServ [15], DiffServ [16] and MPLS [17].

1.9.1 Integrated Services (IntServ) Architecture

The IntServ architecture assumes that explicit setup mechanisms are employed to convey information to the routers involved in a source-to-destination path. These mechanisms enable each flow to request a specific QoS level. The classes of service offered by IntServ are Guaranteed Service (GS) [18], which specifies quantitative constraints in terms of maximum end-to-end data delay and Controlled Load Service (CLS) [19], expressing qualitative QoS requirements.

RSVP [20], [21], [22] is the most widely used set-up mechanism enabling the resource reservation over a specific path from source to destination. Through RSVP signaling, network elements are notified of "per-flow" resource requirements using IntServ parameters (e.g. token bucket depth b, token generation rate r, peak data rate p, etc.). Subsequently, such networks elements apply admission control and traffic resource management policies to ensure that each admitted flow receives the requested services.

The main drawback of RSVP, which limits its deployment in full services IP networks, is due to the use of per-flow state and per-flow processing which raise scalability problem for large networks. On the other hand, RSVP permits a fine control of bandwidth because it storages per-flow state in the routers [23].

Guaranteed Service (GS)

Guaranteed Service (GS) [18] provides quantitative guarantee on a certain data flow. It is capable of respecting an assured level of bandwidth and a precise end-to-end delay bound.

This Class of Service (CoS) can manage applications with stringent real-time delivery requirements, such as certain audio and video applications. Each router characterizes the GS for a specific flow by allocating a bandwidth R and buffer space B. This is done by approximating the fluid model of service [20] so that the flow effectively sees a dedicated wire of bandwidth R between source and receiver.

In a perfect fluid model, a flow conforming to a token bucket of rate r and depth b will have its delay bound by $b \setminus R$ provided $R >= r$. To allow for deviations from this perfect fluid model in the routers approximation, two error terms, C and D, are introduced [20]; consequently, the delay bound now becomes $b \setminus R + C \setminus R + D$. However, with GS a limit is imposed on the peak rate p, of the flow, which results in a reduction of the delay bound.

In addition, the packetization effect of the flow needs to be taken into account by considering the maximum packet size M. These additional factors result in a more precise bound on the end-to-end queuing delay as follows:

$$Q_{delayend2end} = \frac{(b-M)(p-R)}{R \cdot (p-R)} + \frac{(M + C_{tot})}{R} + D_{tot} \quad if\ p > R \geq r \tag{1.1}$$

$$Q_{delayend2end} = \frac{(M + C_{tot})}{R} + D_{tot} \quad if \ R \geq p \geq r \tag{1.2}$$

where C_{tot} and D_{tot} represent the summation of the C and D error terms, respectively, for each router along the end-to-end data path. In order for a router to invoke guaranteed service for a specific data flow, it needs to be informed of the traffic characteristics Tspec of the flow along with the reservation characteristics Rspec. The p, r, b, M values are the Tspec parameters of the IntServ model.

Controlled Load Service (CLS)

Under the CLS [19] model the packets of a given flow will experience delays and loss comparable to a network with no load, always assuming compliance with the traffic contract. CLS provides approximately the same quality of service under heavy loads as under light loads.

A description of the traffic characteristics for the flow desiring CLS must be submitted to the router as for the case of GS although it is not necessary to include the peak rate parameter. If the flow is accepted for CLS then the router makes a commitment to offer the flow a service equivalent to that seen by a best-effort flow on a lightly loaded network. The important difference from the traditional Internet best-effort service is that the CLS flow does not noticeably deteriorate as the network load increases.

By contrast, a best-effort flow would experience progressively worse service (higher delay and loss) as the network load is increased. The CLS is intended for those classes of applications that can tolerate a certain amount of loss and delay provided it is kept to a reasonable level. Examples of applications in this category include adaptive real-time applications.

CLS has some fairly simple implementations, in terms of the queuing systems in routers. It also works adequately for the existing Mbone applications, which can adapt to the modest (small) scale end-to-end delay, variations and jitter that it may introduce, through the use of adaptive playout buffering. It is not suited to applications that require very low latency (e.g. distributed Virtual Reality (VR) systems and so forth). These applications may assume that:

- A very high percentage of transmitted data packets will be successfully delivered by the network to the receiving hosts.
- The data packets will experience little or nothing average packet queuing delay

To ensure that these conditions are met, clients requesting CLS provide the intermediate network elements with a estimation of the data traffic they will generate: the TSpec.

1.9.2 Resource reSerVation Protocol (RSVP)

The Resource reSerVation Protocol (RSVP) is the most widely used set-up mechanism enabling the resource reservation over a specific path from source to destination. It, fundamentally, uses two message types in order to exchange signalling between terminals that want to start a communication respecting QoS requirements. These messages are Resv and Path.

An elementary RSVP reservation request consists of a “flowspec” together with a “filter spec”; this pair is called a “flow descriptor”. The flowspec specifies a desired QoS. The filter spec, together with a session specification, defines the set of data packets (the “flow”) to receive the QoS defined by the flowspec. A Path message contains the following information in addition to the previous hop address: Sender Template, Sender TSpec, Adspec [21].

In particular, Sender Template is a filter specification identifying the sender. It contains the IP address of the sender and optionally the sender port (in the case of IPv6 a flow label may be used in place of the sender port).

Sender Tspec defining the sender traffic characteristics.

Adspec is an optional object that the sender may include in its generated Path messages in order to advertise to receivers the characteristics of the end-to-end communications path. This information can be used by receivers to determine the level of reservation required in order to achieve their desired end-to-end QoS.

Each receiver host sends RSVP reservation request (Resv) messages upstream towards the senders. These messages must follow exactly the reverse of the path(s) the data packets will use, upstream to all the sender hosts included in the sender selection. They create and maintain “reservation state” in each node along the path(s).

Resv messages must finally be delivered to the sender hosts themselves, so that the hosts can set up appropriate traffic control parameters for the first hop. Each RSVP sender host transmits RSVP Path messages downstream along the uni-/multicast routes provided by the routing protocol(s), following the paths of the data. These Path messages store “path state” in each node along the way. This path state includes at least the unicast IP address of the previous hop node, which is used to route the Resv messages hop-by-hop in the reverse direction [21].

RSVP takes a “soft state” approach to managing the reservation state in routers and hosts. RSVP soft state is created and periodically refreshed by Path and Resv messages. The state is deleted if no matching refresh messages arrive before the expiration of a “cleanup timeout” interval [20], [21], [22].

1.9.3 Differentiated Services (DiffServ) Architecture

The DiffServ architecture, instead, defines mechanisms for differentiating traffic streams within a network and providing different levels of delivery service without per-flow management. These mechanisms include differentiated Per

Hop Behavior (PHB)s, as well as traffic classification, metering, policing and shaping functions that are used at the edge of a DiffServ region.

The DiffServ mechanisms manage traffic at the aggregate rather than per-flow level basing on an important design principle: pushing complexity to the networks boundary.

The internal routers in the DiffServ architecture, called core routers, do not distinguish the individual flows then it should perform fast and simple operations. They handle packets according to their PHB identifier based on the Differentiated Services Code Point (DSCP) in the IP packet header.

The border routers at the edge of the DiffServ region are called edge routers and represent the interface between the IntServ domain and the DiffServ domain. They act like IntServ-capable routers on the access network and DiffServ-capable routers in the core network.

The classes of service offered by the DiffServ architecture are the Expedited Forwarding (EF) [24], offering quantitative QoS guarantees on the aggregate path, and Assured Forwarding (AF) [25] managed with priority policies on the aggregate path. The advantages of the DiffServ architecture are:

- Larger scalability in the core network: the core routers do not maintain per-flow state;
- Bandwidth is assigned on aggregate base;
- Reduced signalling messages: there is not a signalling protocol that assigns resources as RSVP.

On the other hand, the bandwidth assignment on aggregate base can loose the guarantees assigned to each flow. For this reason, over-deployment of bandwidth is typically used for guaranteed services. Furthermore, the SLA are generally static; there are no protocol assigning dynamically aggregate bandwidth resources [23].

1.9.4 Multi Protocol Label Switching (MPLS) Architecture

Another proposed QoS architecture for constrained services provisioning is MPLS. Basically, it works in a simple way, giving a particular label to the packets in order to recognize those that belong to applications with stringent quality requirements. This new multi protocol standard permits to Internet Services Provider (ISP)s to offer in flexible and scalable manner many types of services like VPN, Traffic Engineering [1] and QoS support.

The basic idea of MPLS is quite simple: at the ingress point of an MPLS network, the ingress Label Edge Router (LER) packets are classified according to information carried in the packet (e.g., source/destination address, service class, etc.) or according to network related information (e.g., ingress port), or

[1] Traffic Engineering is the process of controlling how traffic flows through one's network in order to optimize resource utilization and network performance

combinations of both. A group of packets treated in the same way is called forwarding equivalence class (fec).

It is also possible to bundle a set of fecs and use one single label for this union. This procedure is known as aggregation. Then, a unique locally significant fixed-length label is chosen for packets belonging to a certain fec and attached to each packet. Subsequent Label Switching Router (LSR)s examine the packets labels, replace them with already specified new labels, and forward the packets according to information stored in a table to the next LSR, until the egress point (egress LER) of the network is reached. The unidirectional path along which a packet traverses the MPLS domain is called Label Switched Path (LSP) [26].

The forwarding procedure (forwarding plane) is completely decoupled from the MPLS control plane, which gives service providers a lot of possibilities to influence the networks behavior.

1.9.5 The importance of Call Admission Control (CAC)

CAC represents an important module in order to satisfy the QoS requirements: it has the role of accepting or not call requests that arrive to the network. A variety of different CAC have been presented in the literature. Some of them require an explicit traffic model while others require only traffic parameters such as the peak and the average rate.

An admission control algorithm determines whether or not a new flow can be admitted to the network such that all users will receive their required performance. Such an algorithm is a key component for multi-service and multimedia networks because it determines the extent to which network resources are utilized and whether the promised QoS parameters are actually delivered.

The CAC functionality permits a flexible handling of the bandwidth and avoids the a priori partitioning of the resources among different types of service. The CAC algorithm should be designed also to fulfill the objectives of minimizing the signaling exchange between the on-board and on-earth segments of a multi-layer system.

In order to reduce delays due to the processing of the call requests, the parameters relevant to the processed calls should be stored and elaborated in a portion of the network capable of avoid long delay. Only those parameters, which are required for the procedures of dynamic resource allocation and congestion control should be sent to the modules responsible of the handling of this algorithm. Moreover, a good CAC algorithm should have very low computational complexity and should be very flexible to any changes in the service management policy, since it is based on a set of configurable parameters.

In this thesis it will be also shown a CAC algorithm that is able to handle multimedia flows in order to manage video sources in a satellite network based on DVB-RCS standard. CAC algorithm has the important role of accepting the major number of video flows with affecting the other flows respecting the number of allowed packet loss.

1.10 Smart Terminals for Heterogeneous Architectures

The recent increase of multimedia applications lead to the design of new terminals capable of working with different technologies.

The satellite, HAP or terrestrial layer selection of the receiver can be based on many factors that determine the reason of choosing one network segment instead another one. A possible factor can be based on the end-to-end delay requested for the application and on available resources on terrestrial/satellite segments. The introduction of this smart functionality in terminals can improve the management of overall network.

It is important to make the following consideration: the HAP segment can offer an intrinsically lower delay than Satellite and according to IntServ to obtain the same delay, the satellite segment with semi-permanent connections requests higher bandwidth than the HAP segment. So, it is not suitable to use the satellite for end-to-end delay that is too low.

However, if it is possible to use the satellite segment, it is inefficient to use only the HAP, because it can be quickly saturated and further calls requesting a low delay connection would be refused. Thus, the idea, that it will be shown in this thesis, is to balance the calls on the basis of the end-to-end delay requested at the receiver. Analyzing the end-to-end delay, it is possible to understand if it makes sense to send the request on HAPs or on Satellite. It may be that, if the number of HAPs is too high or if there is some traffic congested HAP, it is better to request the connection to satellite.

Obviously, this selection depends on delay because if the delay is too low, it is not useful to request the connection to the satellite. So, through the delay evaluation and the minimum bandwidth availability (it is used to verify the bottlenecks among the path) accounting, it is possible regulate where to address the request of the receiver.

A solution for the problem of “smart receiver” capable of connecting with different technologies is Software Defined Radio (SDR), a collection of hardware and software technologies that allows reconfigurable system architectures. The new types of terminals that use the SDR technique are called multi-mode terminals. The purpose is, therefore, to use the same hardware for different functions, through a dynamic configuration according to the operational context. This is possible thanks to ”flexible” hardware and a certain level of software interface.

Currently, an univocal definition for SDR does not exist [27], [28]. Historically, the term “Software Radio” was coined by Joe Mitola in 1991, in order to underline the transition from 80’s digital radio to the new generation innovative radio: a multi-mode an multi-standard radio definable via software [2].

The Federal Communications Commission (FCC) defines the SDR as a generation of device radio that can quickly be reprogrammed in order to transmit and to receive on every frequency of a specific range; whereas the ITU has defined the SDR as a radio system in which the operative parameters (e.g. the frequency, the modulation and the power) can be set or altered by software. However, the research reference point on this area is SDR Forum [28]. It is an international non-profit organization founded in 1996 in order to accelerate the development of radio communication systems based on the concept of SDR. The SDR Forum provides another definition: the SDR is a collection of hardware and software technologies able to reconfigure the system architecture of wireless networks and user terminals.

1.11 Research Goals

After an overview on the architectures of future networks characterized principally by the concept of heterogeneity, hierarchism and multi-layer the main aim of this thesis is now introduced.

This thesis examines the potentiality of the integration of an heterogeneous network as satellite/terrestrial network and the implementation of QoS architectures for providing desired QoS to network users. As well as studying the problem of terrestrial/satellite integration, the QoS architecture introduction issue has been considered in different scenarios, trying to understand which is the better couple network/QoS architecture.

In particular, the best integration identified in the study is the implementation of IntServ scheme in the satellite network, in order to take into account the specific characteristics of the wireless platform, and the implementation of DiffServ approach in the terrestrial network, guaranteeing a better scalability in the backbone area where a major number of network elements is present. The overall identified system is capable of handling the users connection guaranteeing the QoS constraints previously negotiated in the admission phase.

The problem of integration is also faced in the context of multi-layer systems where networks with different characteristics are put together in order to guarantee an ubiquitous access to network users. In this context a new wireless element will be introduced that, in the last few years, has been acquiring most popularity in the research group, because they are capable of having in themselves the better characteristics of both satellite and terrestrial networks, the HAP. In the context of multimedia services and QoS guarantees is very important to introduce the concept of CAC, which is the pre-established part of the network responsible for resource management. In fact, thanks to CAC it is possible to admit or not in the system new connections without degrading calls that are already accepted in the network. This algorithm is particularly important when users calls are multimedia applications, which are sensitive to a set of parameters that have to be taken into account in order to avoid performance degradation.

Finally, the last chapter of this thesis concerns the new DVB-RCS architecture available to mobile customers called DVB-RCS+M. In this new standard there are a lot of critical issues that have to be resolved in order to permit the same conditions as those of the fixed one to mobile users.

1.12 Document Overview

The main body of the thesis is composed of five chapters; their content is outlined in this section.

Chapter 2 provides a detailed study of the terrestrial/satellite integration issues. It makes a careful examination of the integration of the two platforms in which the QoS architecture proposed by Internet Engineering Task Force (IETF), IntServ and DiffServ have been introduced. The greater scalability of the second architecture brought increased interest in developing DiffServ architectures for provisioning IP QoS. In order to achieve the advantage of the IntServ architecture in DiffServ one, recently a new type of network architecture called SCORE has been proposed, because it performs guaranteed services on an aggregated basis without maintaining state info in the core routers by the use of DPS technique, which allows the state info on a flow basis in the core routers to be eliminated. In this chapter, three different terrestrial admission control procedures are presented in order to show how it is possible to find a possible trade-off between QoS requirements and system scalability. A network device plays a key role in this hybrid architecture, the GTW. It is the interface between two networks that has to manage the terrestrial resource and the mapping class of service of different QoS architectures between them.

Chapter 3 deals with the proposal of intelligent terminals that are able to select the wireless segment in order to receive better connectivity. It is possible to indicate this terminal functionality as vertical switching, because the terminal can choose between the satellite or HAP segment in order to satisfy the required QoS constraints. Both wireless platforms are equipped with an IntServ QoS architecture in order to make the segment capable of providing quantitative and qualitative QoS guarantees, by exploiting the class of service that such architecture allows to manage. A key point in a vertical switching is to find a good parameter in order to perform the right choice in the platform selection. The factor to perform the correct selection can be the end-to-end delay required by the "smart terminal". In this way it is possible to exploit the difference between the two platforms in terms of different bandwidth capacity and propagation delay.

Chapter 4 introduces a detailed analysis of a main component of QoS architecture: the CAC module. In the context of multimedia applications a CAC capable of guarantee better performance results in a greater number of video flows managed by the system. The study shows how it is possible to represent a video flow with a finite state and discrete time markov chain in order to exploit the statistical characteristics of the video traffic. The proposed statistical algorithm is compared with another admission scheme used for MPEG traffic well-known in the literature.

Chapter 5, finally, gives an overview about the new DVB-RCS+M standard, showing the new functionality of the system and explaining the critical issues introduced for mobile terminals. Moreover, this chapter proposes a very new DVB-RCS+M topic, the hybrid return channel modality. It allows mobile terminals to use a continuous carrier modality in the return link beyond the classical MF-TDMA mode.

2

The Synergy of Terrestrial/Satellite Integration

In the last few years mobile radio communications have experienced great transformations and are still rapidly changing. The main "fuel" of this evolutionary process is, undoubtedly, represented by the strong user request for new conception multimedia applications (not only with a point-to-point nature, but also broadcast and even multicast) to be accessed in a "ubiquitous" fashion. An obvious solution to this challenging issue is represented by enhancing platforms based on the synergic inter-working of different technologies and algorithms, including the satellite ones. By this it is important to carry to its extremes the concept of "hybrid wireless platform" by including also satellite segments not only acting as mere bent-pipes, but acting as an additional radio access segment to be overlapped to the terrestrial ones.

The way these systems interact with the terrestrial (preferably based on the TCP/IP protocol) networks is manifold. Whichever the way is, the new generation of satellite constellations is going to become an integral part of the network. In the highlighted communication scenario it shall not be neglected that a valid prospective alternative (or a complement) to satellites is represented by stratospheric platforms, called High Altitude Platform (HAP). Several advantages derives from satellite and terrestrial system integration, but numerous as well are the issues which remain still open and need further investigations. Many projects currently aim at contributing to the design and deployment of so-called Global Mobile Broadband System (GMBS), a unique satellite-terrestrial infrastructure ensuring nomadic users access to multimedia services with a negotiated QoS.

2.1 The Integrated Terrestrial-Satellite Architecture

The recent development in telecommunication networks has carried to the major importance of the satellite studies for the provisioning of broadband multimedia services. In the optical of this development can be more interesting to analyze a system constituted by an existing integrated scalable terrestrial network and a geostationary satellite one. In the last few years the IETF has defined two architectures for QoS support: IntServ [15] and DiffServ [16]. As suggested by the IETF it is possible to use IntServ architecture for access-networks and DiffServ architecture for core networks. Therefore IntServ is used on a geostationary satellite platform and DiffServ just like the architecture on terrestrial networks.

This choice allows exploitation of the IntServ advantage of fine bandwidth control in the satellite segment where the bandwidth is still a precious resource, and DiffServ scalability in the large core networks. A recently proposed DiffServ like architecture called Scalable CORE (SCORE) [29] is used in order to offer more scalability in the core network and to offer quantitative guarantees for EF class [24]. This architecture can also offer guaranteed services [18] on an aggregated basis without maintaining state information in the core routers. This scalable architecture is integrated with the IntServ satellite architecture in order to obtain an overall architecture with end-to-end QoS provisioning. The satellite segment, instead, uses the IntServ architecture as will be shown in the following.

The integration of these segments is exploited through the Aggregate Resource ReSerVation Protocol (ARSVP). This protocol should make an aggregation of bandwidth requests but it does not specify the way in which to make the aggregate reservation.

In the proposed terrestrial-satellite architecture shown in Fig. 2.1, the Gateway (GTW) acts as a border router between the terrestrial and satellite regions, i.e., between the specific DiffServ implementation (the SCORE) and the IntServ region respectively. The GTW must be both IntServ- and DiffServ-capable; it is in charge of interworking the two network segments by:

- Selecting the effective mapping of IntServ-over-DiffServ classes.
- Supervising resource assignment in the wireless and wired networks by interacting with the satellite Master Control Station (MCS) in the satellite segment and the core routers in the SCORE terrestrial segment.

The satellite platform considered in this article is the GEO platform of EuroSkyWay (ESW) [30], [31]. This choice does not limit the scope of our proposal; the proposed framework can easily be extended to many cases of hybrid wireless-wired interworking, provided that the interworking gateway is adequately designed and IntServ-over-DiffServ philosophy is adopted. This is the great advantage of our architecture: putting complexity in the GTW and designing it as an interworking unit between the two segments, which are left unaware of each other. When, for example, GTW is the gateway between the terrestrial SCORE region and an IntServ Low Earth Orbit (LEO) satellite network, it should be burdened with the further tasks of transparent handoff management, renewal of resource reservation when the satellite goes out of sight of GTW, and routing operations, but the two network segments can be left unaffected.

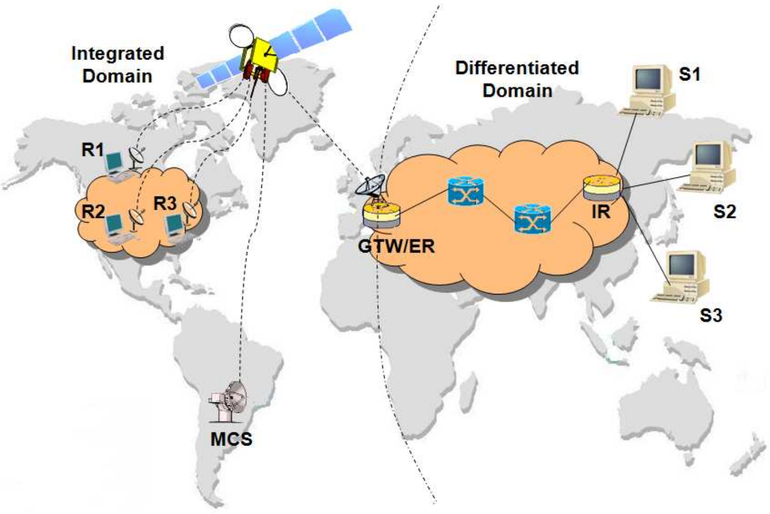

Fig. 2.1. Reference Integrated Network Architecture

2.1.1 Satellite Architecture

The reference satellite architecture is the ESW [30] platform, which is a packet-switched GEO satellite with onboard processing technology (see Fig. 2.2). The ESW network consists of the following elements:

- A GEO satellite, with a payload using the Ka band (20/30 GHz) for earth/space connections;
- Several Satellite User Terminal (SaT)s belonging to different types (SaT-A, SaT-B, SaT-C) and offering uplink data rates of 160, 512, and 2048 kb/s, respectively;
- A fixed GTW earth station, which interfaces the public terrestrial network through protocol adapting interfaces and interworking functionality;
- An MCS, which is the fixed earth station that manages and controls the whole ESW system.

Functionality of traffic and resource management is distributed in different modules of the overall system. CAC and resource management functions are distributed in the satellite user terminal and earth stations. Access control is performed in a decentralized way by the Traffic Resource Manager (TRM), in the Ka-band payload and the CAC module in the MCS.

Two bandwidth assignment modalities are foreseen: permanent and semi-permanent assignment. The permanent connections ensure that, following a CAC procedure, a number of channels are allocated to the requesting terminal on the uplink and downlink for the whole duration of the connection. In semi-permanent connections, the resources are not permanently allocated, but

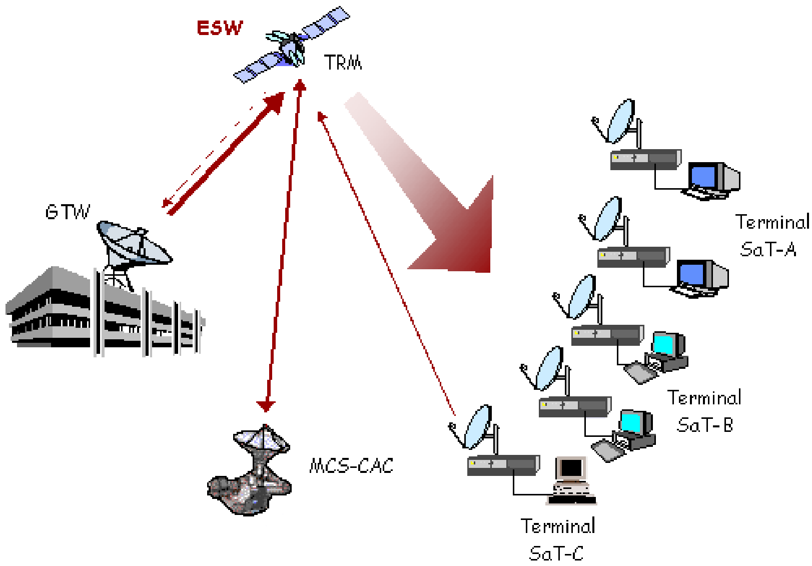

Fig. 2.2. EuroSkyWay Platform

dynamically assigned to the source on uplink and downlink on the basis of its activity. The TRM module dynamically allocates satellite resources for a fixed time interval (multiple of the frame period) each time the relevant source generates a new information burst. It is vital to decide how to map each of these connections on the IntServ classes. This policy supports the statistical multiplexing of traffic over the air interface.

Considering only the satellite segment, the two connection modes (permanent and semi-permanent) present different D values [18], [20] which impact terrestrial network in terms of a different requested rate R in order to obtain the same end-to-end maximum delay. So it is possible to divide D on satellite connection bases:

- $Dperm$ = 26.5 ms (frame delay) for permanent connections;
- $Dsemiperm$ = 270ms (round-trip delay) + 26,5ms (maximum processing time onboard the spacecraft) + 26,5ms (frame duration) + 101ms (terminal reconfiguration time) = 424ms for semi-permanent connections.

2.1.2 Scalable Terrestrial Architecture

The terrestrial network has to obey to scalability requirements due to increasing number of flows that it has to manage. So it is important on terrestrial segment to use a scalable architecture in which the control of traffic is on aggregated basis.

The considered terrestrial architecture consists of a DiffServ like architecture where the network routers are divided in two categories in order to guarantee a certain scalability.

The edge routers, which manage a lot of flows and have to classify and mark the incoming packets. The other routers are so called core routers which manage the packets performing scheduling and forwarding. Packets which have same behavior inside the network are aggregated in order to create a aggregate flow.

In particular, a recently proposed architecture called SCORE has been considered. It offers more scalability in the core routers provisioning quantitative guarantees like the IntServ architecture without storing per-flow state inside the core routers. The SCORE is based on ad-hoc technique called Dynamic Packet State (DPS). With DPS, each packet carries in its header some state that is initialized by the ingress router as shown in Fig. 2.3.

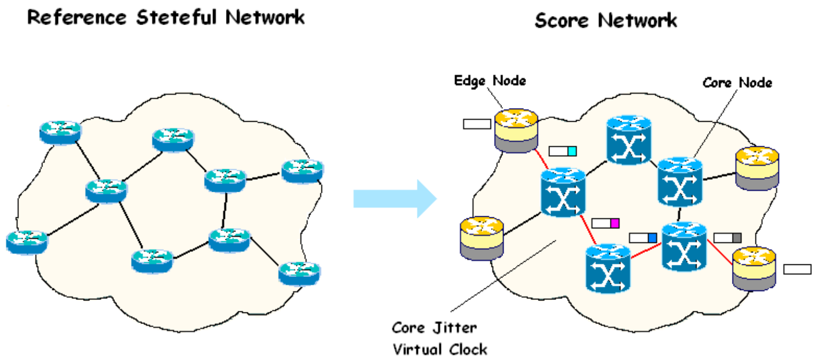

Fig. 2.3. DPS technique in SCORE architecture

Core routers process each incoming packet based on the state carried in the packet's header, updating both its internal state and the state in the packet's header before forwarding it to the next hop. At the end the egress node removes states from the packet's header. In this way no maintaining state information is requested to the core router. This technique is used in order to perform two algorithms, the scheduling called Core Jitter Virtual Clock (CJVC)and the admission control algorithm. For more details refer to [29], [32].

In SCORE terminology, edge routers are called Ingress Router (IR)s and Egress Router (ER)s; inner routers are called Core Router (CR)s. Given that the attention will focus, without loss of generality, on the case of terrestrial sources and satellite receivers, the GTW also operates as an ER for the SCORE network.

The SCORE architecture tries to reconcile the strong QoS guarantee of the IntServ model with the scalability of the DiffServ model by keeping the state information needed for QoS guarantee in the IP packet headers instead of in the routers. In this way, routers are still able to process packets on a per-flow basis, despite the fact that they do not maintain any per flow state. SCORE is similar to a DiffServ network from an architectural point of view, as only edge routers are burdened with functionality of flow management and process incoming packets on the basis of finer granularity.

On the other hand, CRs process packets on the basis of a small number of PHBs, thus increasing scalability. In this way, SCORE can approximate the service offered by a stateful network, providing end-to-end per-flow delay and bandwidth guarantees as defined in the IntServ model. The technique that allows SCORE to offer guaranteed service assurance is called DPS [29], [32].

The main idea behind DPS is very simple: instead of having routers install and maintain per flow state, the packets carry per flow state. Routers implement the scheduling algorithm called CJVC, which is the distributed derivation of Jitter Virtual Clock (JVC). JVC works as follows: each packet is assigned an eligible time and a deadline upon its arrival. The packet is held in a rate controller until it becomes eligible; the scheduler orders the transmission of eligible packets according to their deadlines. Intuitively, the algorithm eliminates the delay variation of different packets by forcing all packets to incur the maximum allowable delay. The DPS operation resembles the DiffServ scenario: routers at the edge of the SCORE region IRs differentiate between single end-to-end flows and determine the proper state (scheduling parameters) to be inserted in the packets header before entering the network; CRs process the packet and update their internal state (which does not depend on the number of flows) and the state in the packet before forwarding it; finally, an ER removes the state from the header as the packet leaves the SCORE network. This realizes a scalable mechanism that eliminates the need for maintaining per flow state at the CRs without sacrificing service granularity. For further details on DPS and CJVC, see [29], [32].

2.1.3 The Aggregate RSVP Protocol

The DPS approach is useful to eliminate per-flow state and thereby any dependence on the number of flows; thus, SCORE offers the scalability of the DiffServ model even if routers process packets on a per-flow basis. But, as pointed out in [33], the scalability of a service model is different from the inherent scalability of a signaling protocol. Infact, the main problem of the RSVP protocol in the original version is the realizability in a big network because each reservation requires a non-trivial amount of message exchange, computation, and memory resources in each router along the way.

Therefore, in order to increase scalability of the signaling protocol while maintaining fine-grained QoS control, ARSVP [34] has been introduced in the SCORE architecture. ARSVP is an extension to RSVP proposed by the IETF as a possible solution to support the IntServ-over-DiffServ paradigm. It is developed in order to support reservations for an aggregation of flows between edges of an aggregation region, rather than for individual flows as supported by RSVP. The aggregation of individual reservations yields a reduction in the amount of state to be stored and signaling messages exchanged between the routers in the aggregation region. In our approach, the considered aggregation region is the SCORE network in which DiffServ mechanisms are used for classification and scheduling of traffic supported by aggregate reservations.

The solution proposed involves the aggregation of several End-to-End (E2E) reservations that cross an "aggregation region" and share common ingress and egress routers into one larger reservation from ingress to egress. It is defined "aggregation region" a contiguous set of systems capable of performing RSVP aggregation along any possible route through this contiguous set. The routers which are "exterior" to an aggregation region are called Aggregator or Deaggregator.

In ARSVP it is defined "Aggregator" the router that for a E2E flow as the first router that processes the E2E Path message as it enters the aggregation region. It is defined "Deaggregator" router that for a E2E flow as the last router to process the E2E Path as it leaves the aggregation region (see Fig. 2.4).

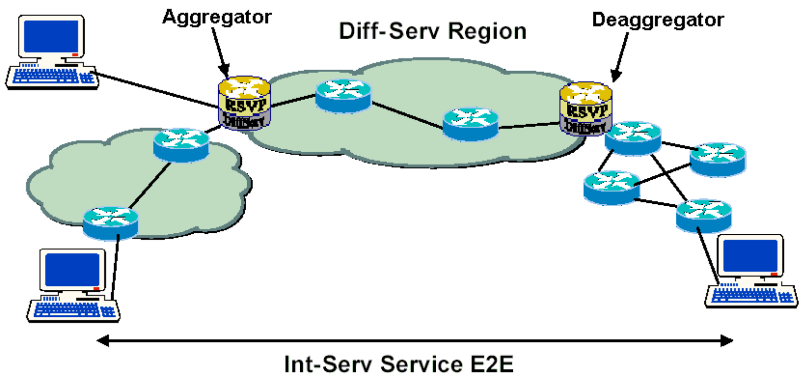

Fig. 2.4. Aggregator - Deaggregator

Aggregation depends on being able to hide E2E RSVP messages from RSVP-capable routers inside the aggregation region. To achieve this end, the IP Protocol Number in the E2E reservation's Path, PathTear, and ResvConf messages is changed from RSVP (46) to RSVP-E2E-IGNORE (134) upon entering the aggregation region, and restored to RSVP at the deaggregator point.

These messages are ignored (no state is stored and the message is forwarded as a normal IP datagram) by each router within the aggregation region. Since the deaggregating router perceives the previous RSVP hop on such messages to be the aggregating router, Resv and other messages do not require this modification; they are unicast from RSVP hop to RSVP hop anyway. Such establishment of a smaller number of aggregate reservations on behalf of a larger number of E2E reservations yields the corresponding reduction in the amount of state to be stored and the amount of signalling messages exchanged in the aggregation region.

One or more DiffServ DSCPs are used to identify traffic covered by aggregate reservations and one or more DiffServ PHBs are used to offer the required forwarding treatment to this traffic. There may be more than one aggregate reservation between the same pair of routers, each representing different classes of traffic and each using a different DSCP and a different PHB.

In conventional (unaggregated) RSVP operation, a session is identified by a destination address and optionally a protocol port. Since data belonging to an aggregated reservation is identified by a DSCP, the session is defined by the destination address and DSCP.

For simplicity and reliability, the responsibility of the mapping decision is assigned entirely to the Deaggregator. The Aggregator is notified of the selected mapping by the Deaggregator and follows this decision. The Deaggregator was chosen rather than the Aggregator because the Deaggregator is the first to have access to all the information required to make such a decision (in particular receipt of the E2E Resv which indicates the requested IntServ service type and includes information signalled by the receiver).

As described above, E2E RSVP messages are hidden from the interior routers inside the aggregation region. Consequently, the Adspecs of E2E Path messages are not updated as they travel through the aggregation region. Therefore, the Deaggregator for a flow is responsible for updating the Adspec in the corresponding E2E Path to reflect the impact of the aggregation region on the QoS that may be achieved end-to-end. The Deaggregator should update the Adspec of the E2E Path as accurately as possible. Since Aggregate Path messages are processed inside the aggregation region, their Adspec is updated by interior routers to reflect the impact of the aggregation region on the QoS that may be achieved within the interior region. Consequently, the Deaggregator should make use of the information included in the Adspec from an Aggregate Path where available. The Deaggregator may elect to wait until such information is available before forwarding the E2E Path in order to accurately update its Adspec.

To transport the information of CRs in the SCORE region toward the edge router has been used ARSVP. The E2E reservations have been considered as the reservation requests (E2E Path/Resv messages) relevant to individual flows, and the "aggregate" reservation as a request relevant to many E2E reservations (aggregate Path/Resv messages). Routers within the aggregation region forward E2E RSVP messages transparently as normal IP datagrams.

To this end, the IP Protocol Number in E2E reservation messages is changed from its normal value (RSVP) to RSVP-E2E-IGNORE upon entering the aggregation region, and restored at the egress point. Aggregate Path messages are sent from the Aggregator to the Deaggregator using RSVPs normal IP Protocol Number. Besides standard ARSVP messages, the work proposes to extend the Aggregate Path message in order to convey information on available bandwidth. This is clarified in the next section.

As for QoS control, sender, receiver, and intermediate nodes exchange several types of data. Among them, the Adspec object carries information generated or modified within the network and used at the receiver to make reservation decisions. This information might include available services, delay and bandwidth estimates, and operating parameters used by specific QoS control services. In order to generate Aggregate Path and Resv messages the Adspec object must be updated. For more details on the ARSVP protocol see [34].

Another important message used in ARSVP protocol is "aggregate Resv Confirm" that is used in signaling phase. An example of messages exchange between an Aggragator and Deaggregator node is shown in Fig. 2.5

The very first event is the arrival of the E2E Path message at an exterior interface of an Aggregator. Service on exterior interfaces is handled as defined in the classical RSVP. Service on interior interfaces is complicated by the fact that the message needs to be included in some aggregate reservation, but at this point it is not known which one, because the Deaggregator is not known. Therefore, the E2E Path message is forwarded on the interior interface(s) using the IP Protocol number RSVP-E2E-IGNORE, but in every other respect identically to the way it would be sent by an RSVP router that was not performing aggregation.

At this point, the E2E Path message traverses zero or more interior routers. Interior routers receive the E2E Path message on an interior interface and forward it on another interior interface. As such, they simply forward it as a normal IP datagram.

The E2E Path message finally arrives at a deaggregating router, which receives it on an interior interface and forwards it on an exterior interface. Again, the Router Alert IP Option alerts it to intercept the message, but this time the IP Protocol is RSVP-E2E-IGNORE and the next hop interface is an exterior interface. Before forwarding the E2E Path towards the receiver, the Deaggregator should update its Adspec. This update is to reflect the impact of the aggregation region onto the QoS to be achieved E2E by the flow.

When receiving the E2E Path, depending on the policy for mapping E2E reservation onto Aggregate Reservations, the Deaggregator may or may not be in a position to decide which DSCP the E2E flow for the processed E2E Path is going to be mapped onto, as described above. If the Deaggregator is in a position to know the mapping at this point, then the Deaggregator first checks that there is an Aggregate Path in place for the corresponding DSCP. If so, then the Deaggregator uses the Adspec of this Aggregate Path

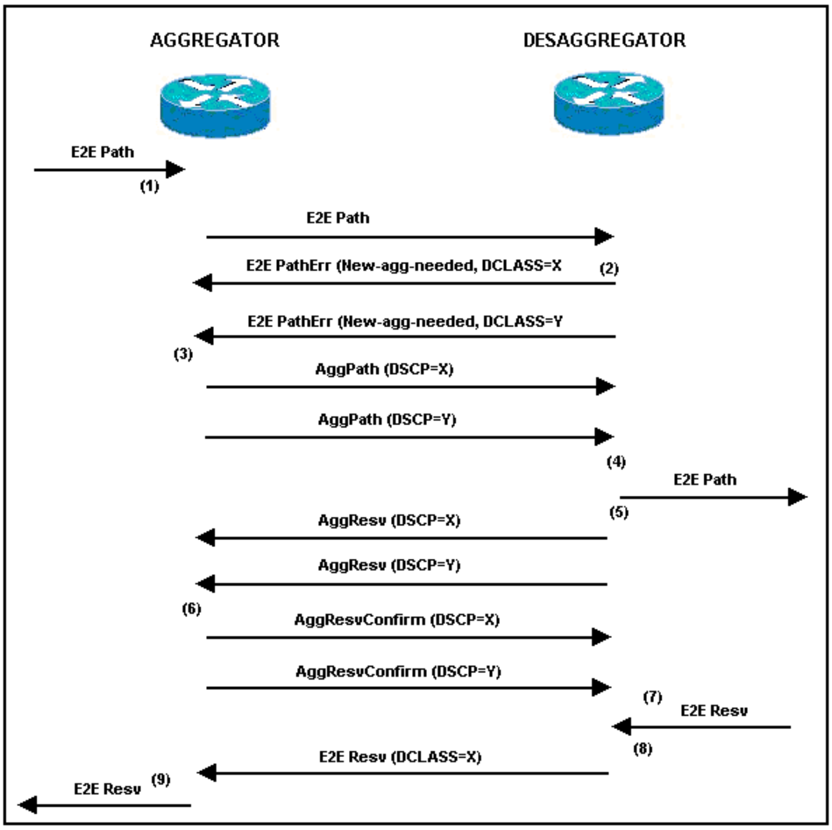

Fig. 2.5. Example Signalling Flow in an aggregate region

to update the Adspec of the E2E Path and then forwards the E2E Path towards the receiver. If not, then the Deaggregator requests establishment of the corresponding Aggregate Path by sending an E2E PathErr message with an error code of NEW-AGGREGATE-NEEDED and the desired DSCP encoded in the DCLASS Object [34], [35]. The Deaggregator may also at the same time request establishment of an aggregate reservation for other DSCPs. When receiving the Aggregate Path for the desired DSCP, the Deaggregator then uses the Adspec of this Aggregate Path to update the Adspec of the E2E Path.

The aggregating router is responsible for generating a new Aggregate Path for a DSCP when receiving a E2E PathErr message with the error code NEW-AGGREGATE-NEEDED from the Deaggregator. The DSCP value to include in the Aggregate Path Session is found in the DCLASS Object of the received E2E PathErr message. Having sent the E2E Path message on toward the destination, the Deaggregator must now expect to receive an E2E Resv for the session. On receipt, its responsibility is to ensure that there is sufficient bandwidth reserved within the aggregation region to support the new E2E reservation, and if there is, then to forward the E2E Resv to the aggregating router. The deaggregating router first makes the final decision of which Aggregate Reservation (and thus which DSCP) this E2E reservation is to be mapped onto. This decision is made according to the policy selected by the network administrator as described above.

If this final mapping decision is such that the Deaggregator can now make a more accurate update of the E2E Path Adspec than done when forwarding the initial E2E Path, the Deaggregator should do so and generate a new E2E Path immediately in order to provide the accurate Adspec information to the receiver as soon as possible. Otherwise, normal Refresh procedures should be followed for the E2E Path.

Upon receiving an E2E Resv message on an exterior interface, and having determined the appropriate DSCP for the session according to the mapping policy, the Deaggregator looks for the corresponding path state for a session with the chosen DSCP. If aggregate Path state exists, but no aggregate Resv state exists, the Deaggregator creates a new aggregate Resv.

In order to minimize the occurrence of situations where no corresponding Aggregate Reservation is established at the time of processing an E2E Resv, and in turn to avoid the delay associated with the creation of this aggregate reservation, the Deaggregator may anticipate the current demand and create the Aggregate Reservation before receiving E2E Resv messages requiring bandwidth on those aggregate reservations. The aggregate Resv message is handled in essentially the same way as defined in the classical RSVP. These routers perform admission control and resource allocation as usual and send the aggregate Resv on towards the Aggregator. The receipt of the E2E Resv message with a DCLASS Object is the final confirmation to the aggregating router of the mapping of the E2E reservation onto an Aggregate Reservation. Under normal circumstances, this is the only way it will be informed of this association. It should now forward the E2E Resv to its previous hop, following normal RSVP processing rules.

2.2 The Overall Framework For QoS Management

To provide scalable end-to-end services with QoS guarantees in the reference terrestrial-satellite system, harmonization of the call and resource management policies in the two network segments is mandatory. Specifically, a CAC procedure is designed consisting of terrestrial and satellite admission phases performed in succession. The separation of CAC phases gives the flexibility of conceiving different CAC algorithms for each network side, thus using our framework to integrate different access networks.

The GTW is in charge of coordinating the two CAC phases; this means that once the GTW is made aware of the bandwidth required by the new connection in order to satisfy its target QoS, it starts the terrestrial CAC phase, which is performed according to the mechanisms of the SCORE network adequately extended to support ARSVP messages (as described in the section 2.4.2). If terrestrial CAC is successful, GTW starts the satellite CAC phase that is performed according to proprietary algorithms managed by the satellite MCS. Satellite CAC is executed according to the statistical algorithm specified in section 2.3. The decision is made following evaluation of the Excess Demand Probability (EDP), that is, the probability that the accepted calls ask for more channels than those actually available. This means that satellite CAC accepts a given number of calls if, at a random instant, the EDP is such that a target burst loss (say ε) can be guaranteed anyway. Thus, the EDP is an upper bound for the burst loss. The choice of performing satellite CAC only after successful terrestrial CAC aims at not reserving (and eventually wasting) resources in advance in the satellite segment while waiting for the result of terrestrial CAC, given the limited bandwidth and long latency of satellite links.

2.3 Satellite Admission Control

The CAC algorithm used in the satellite segment fulfills the objectives of minimizing the signaling exchange between the on-board and on-earth segments of the system. In order to reduce delays due to the processing of the call requests on board, the parameters relevant to the processed calls are stored and elaborated within the ground segment. Only those parameters, which are required for the procedures of dynamic resource allocation and congestion control, are sent to the on-board modules, following the call admission. This type of CAC algorithm has the advantage of having very low computational complexity and of being very flexible to any changes in the service management policy, since it is based on a set of configurable parameters.

The satellite CAC algorithm takes decision on the call acceptance, following the evaluation of an EDP, that is the probability that the accepted calls ask for more atomic channels than actually available in a satellite spot beam. This means that the CAC algorithm accepts a given number of calls if, at a random instant, the EDP is such that a target service quality (expressed in terms of burst loss) can be guaranteed anyway.

Without loss of generality, a single satellite spot beam is considered, which is able to manage up to calls, each one described by a two-state Markovian model. In one state, the call is active and transmits information at its peak bit rate; in the other state, it is silent and no information is sent.

The distribution probability of the random variable X, representing the total demand for atomic channels, is found by means of a generating function $f_x(z)$, as shown in [36]. The great advantage of this method is the probability of expressing the generating function as a linear combination of terms

$$f_x(z) = C_0 + C_1 \cdot z + C_2 \cdot z^2 + ... + C_n \cdot z_n \tag{2.1}$$

in each term, C_j represents the probability that the total request of channels is equal to j. Thereby, the EDP can be easily determined by computing the sum of the coefficients $C_{L+1} + C_{L+2} + ...$, where L is the capacity of the satellite spot beam. For details see [36]. A set of calls can be safely multiplexed if the EDP is below a given bound ε, this also means that

$$C_0 + C_1 + ... + C_L \geq 1 - \varepsilon \tag{2.2}$$

The introduction of stringent bound to rule the admission of real-time calls can be explained as follows. Real-time connections have stringent delivery constraints, which can also justify the eventual preemption of non real-time calls. If too many real-time connections are accepted into the system, compared to the number of non real-time connections, it happens that: the delay constraints of the real-time sources cannot be guaranteed any longer and the non real-time connections will experience an excessive preemption probability and thus a consequent QoS degradation.

2.4 Terrestrial Admission Control

For terrestrial CAC, three possible approaches that aim at trading off between scalability and QoS support have been conceived. The first is conceived in order to "provide the best target QoS". It is a completely distributed algorithm according to which the egress router (i.e., GTW) and CRs perform admission control or the terrestrial region on a per-aggregate flow basis. This approach is mainly based on the SCORE architecture; therefore, it has been called DPS approach. It is a conservative approach, as it will see in the following.

The second considered approach aims at "reducing the signaling overhead" in the SCORE region trying to delegate the GTW/ER to decide the admission of a new connection in certain cases. To do this, the GTW/ER uses some state information carried by ARSVP, which enables it to detect a potential bottleneck in the terrestrial network, thus avoiding sending a useless reservation inside the SCORE network. This approach is called the ER-DPS approach.

The last approach gives the GTW/ER full functionality to decide call admission without involving CRs in the SCORE network. This approach is fully centralized regarding admission decisions, but it exploits periodic information contributed by the CRs and carried by ARSVP about bandwidth availability in the SCORE network. This last is called the ER approach. Its objective is "to increase both state and protocol scalability".

In the following the message exchanges for the three analyzed approaches are described. The convention of numbering the messages with a progressive identifier and using an index in order to show the temporal sequence of the messages has been used. Thus, colors and numbers define the types of messages, and indexes define the sequence order.

2.4.1 The Terrestrial Local Admission in the SCORE Region: Aggregate Reservation Estimation Algorithm

To compute terrestrial local admission, each router updates a local variable called R_{bound} which is recursively computed as $R_{bound} = R_{bound} + R$, where R is the data rate requested by the new connection in the E2E Resv message. An R_{bound} variable is stored for each managed PHB by every router. Thus, for example, $(R_{bound})_{EF}$ is computed for all flows crossing a router and carried over the same expedited forwarding (EF) PHB. If R_{bound} is lower than the output link capacity C reserved for that DSCP by the ER and each crossed CR, terrestrial CAC is successful.

In order to increase link utilization, especially in the presence of highly bursty traffic, the approach proposed in [29], [32] has been followed and periodically recalibrate R_{bound}. Note that the aim of bandwidth recalibration is to exploit the OFF periods of bursty sources in order to accommodate new calls and enjoy the statistical multiplexing gain. The drawback of multiplexing is burst loss if the admission is not regulated adequately. To perform recalibration, a new variable R_{cal} is periodically recomputed by any router to account for the actual traffic flowing in the network. R_{cal} should be an upper bound on R and over-estimation errors should be corrected and kept to the minimum.

To compute R_{cal}, an inaccurate estimate of R is performed, denoted R_{DSP}, and then adjustments to account for estimation inaccuracies are make. The estimate R_{DPS} is calculated using DPS technique.

Time is divided into intervals of dimension $T_W : (t_k, t_{k+1}], k > 0$. The IR monitors the amount of bits transmitted by any flow i in the interval T_W. $b_i(t_k, t_{k+1})$ is denominated as the sum of the bits received for flow i in the interval $(t_k, t_{k+1}]$. This information is included in the IP packet header by the IR and it is used to update the variable $B(t_k, t_{k+1})$ by all crossed CRs. $B(t_k, t_{k+1})$ is the sum of bits $b_i(t_k, t_{k+1})$ of all flows i belonging to the same DSCP crossing the router. So at the end of any interval T_W each router computes the actual rate R_{DPS} and then, based on R_{DPS}, it computes a new variable R_{Cal} defined as follows:

$$R_{DPS}(t_{k+1}) = \frac{B(t_k, t_{k+1})}{t_{k+1} - t_k} = \frac{B(t_k, t_{k+1})}{T_W} \tag{2.3}$$

From this estimate, then, an upper-bound of $R(t_{k+1})$, called $R_{Cal}(t_{k+1})$ that try to avoid the inaccuracies of the estimation process is calculate as follows:

$$R_{Cal}(t_{k+1}) = \frac{R_{DPS}(t_{k+1})}{1 - f} + R_{new}(t_{k+1}) \tag{2.4}$$

where R_{new} is a variable locally initialized by each router at the beginning of any new interval T_W and accounting for the contributions in terms of rate R of any new flow accepted $R_{new} = R_{new} + R$; f represents the frequency of rate recalibration and it is defined as follows:

$$f = \frac{T_I + T_J}{T_W} \tag{2.5}$$

with T_I is the maximum inter-departure time between two consecutive packets of a flow at the edge node, T_J is the maximum delay jitter of a flow and both T_I and T_J are much smaller than T_W.

Based on R_{cal} each router recalculates R_{bound} as the minimum between R_{bound} and R_{cal}:

$$R_{bound}(t_{k+1}) = min(R_{bound}(t_{k+1}), R_{Cal}(t_{k+1})) \tag{2.6}$$

R_{bound} is calculated by each CR and GTW/ER. The local admission control verifies that for each CR or ER, the $R_{bound} < C$ where C is the capacity associated with the link. The described DPS approach has the advantage of requesting only knowledge of the aggregate state, thus offering a good degree of scalability, but presents the drawback of low protocol scalability because of the high number of terrestrial Resv messages that are sent on the SCORE network and processed by each core router to decide local admission, as explained above. In the following it is possible to view a pseudo code of the admission algorithm.

```
---------------------------------------------
   Per-hop Admission Control
   on reservation request R
   if(Rbound+R<C)
       Rnew=Rnew+R;
       Rbound=Rbound+R;
       accept request;
   else
       deny request;
   on reservation termination R /*optional*/
       Rbound=Rbound-R;

   Aggregate Reservation Bound Computation
   on packet arrival p
       b=get_b(p);
       B=B+b;
   on time_out TW
       RDPS=B/TW;
       Rbound=min(Rbound, RDPS/(1-f)+ Rnew);
       Rnew=0;
-----------------------------------------------
```

2.4.2 The role of Gateway (GTW)

The Gateway (GTW) is the node which interfaces the two different networks. It has important functionality like internetworking between the terrestrial and satellite network, operation of Aggregator and Deaggregator of the flows, mapping of the IntServ class in the DiffServ one.

Moreover, in a hybrid network like that shown in Fig. 2.1 it has the important task of starting the CAC procedure and managing the terrestrial resource [37]. This node is designed with two different interfaces, one that is capable to operate in IntServ manner and other that manages the packets with DiffServ functionality. The gateway presents inside a module that is responsible for management of the aggregated traffic at the gateway input; it changes the aggregated DiffServ traffic into individual IntServ flows.

In Fig. 2.6 the architecture for a gateway/egress node is shown inserted in the hybrid network.

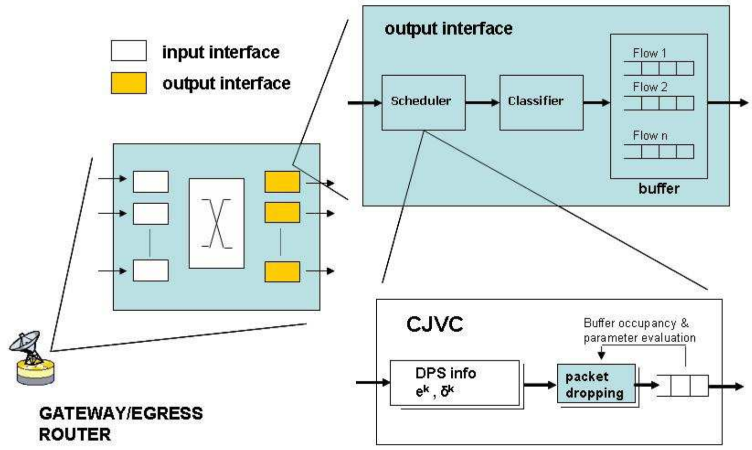

Fig. 2.6. Architecture of Gateway/Egress Router in a SCORE/Satellite Network

The figure shows in detail the composition of the output interface. The first block is the scheduler of which is presented a zoom vision. For each packet on a flow basis DPS parameters are estimated, that are eligible time (e^k) and estimated deadline (δ^k), by the CJVC algorithm [32]. The second block is the classifier that has the task of classifying the incoming packets on the basis of the DSCP codepoint [16] and insert the packets in the EF, AF or BE buffer. The last block is a pool of buffer of single flows. These packets, at the end, are sent toward the satellite segment on wireless link.

In order to evaluate the workload of the gateway node, simulation campaigns were deployed. Some parameters which are the Packets Signaling Overhead on GTW, Satellite Static Efficiency, Maximum delay on GTW have been considered. In order to show the main aspects of the GTW workload only two approaches have been considered. In the next section another approach will be presented that shown a GTW functionality that derives from the other two ones.

Fig. 2.7 shows the number of signaling packets varying an important parameter of the DPS technique called T_W. It is possible to note as the number of signaling packets is higher for a DPS approach and as for ER approach this number decreases for increasing observation window T_W. This is due to the fact that the sending of signaling packets is on time basis and in particular it is based on T_W interval

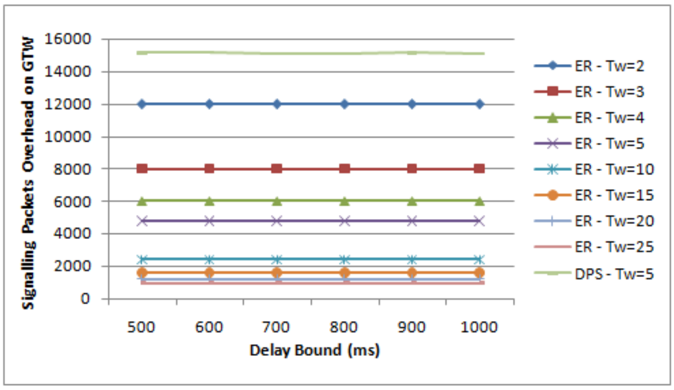

Fig. 2.7. Signaling Packets Overhead for different T_W values

The drawback of this behavior is presented in the Fig. 2.8.

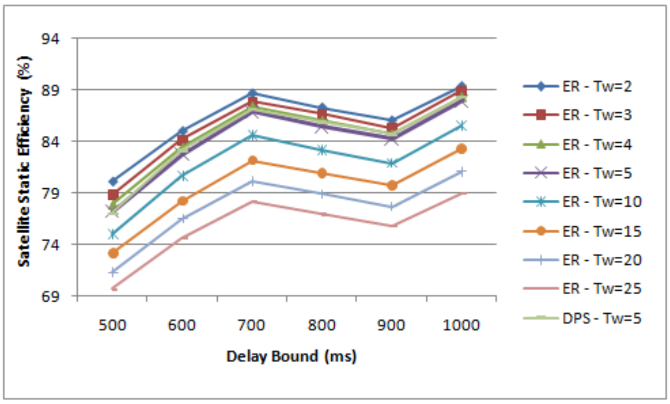

Fig. 2.8. Satellite Static Efficiency for different T_W values

It is possible to observe that the decrease of signaling packets overhead do not correspond to an increase of satellite utilization. The lower utilization corresponds to the greater value of T_W and then to better value of packets overhead. The better utilization is in correspondence of the lower value of T_W, that corresponds to a worst value of number of signaling packets overhead. Then, by observing the Fig. 2.7 and Fig. 2.8 it is possible to note what is the better trade-off between number of packets overhead and satellite utilization. This optimal value is represented by T_W equals to 5 seconds. In fact, for this value it is possible to have an utilization equals in both approaches, but a number of signaling packets in the case of ER approach are lower than DPS one. With ER approach the number of signaling packets sent inside the network is about a 32% lower than DPS approach and then the gateway need to handle a minor number of packets overhead.

At last, in Fig. 2.9 the maximum delay experimented by GTW is presented. It is possible to note that for low delay values the ER approach experiments more delay than DPS one. However, for high delay values the maximum delay experimented by GTW is very similar for two approaches.

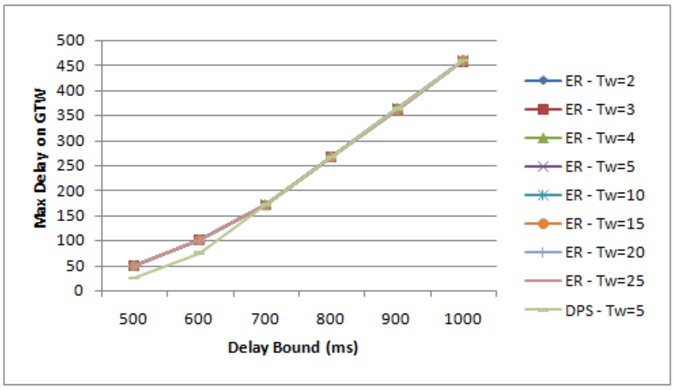

Fig. 2.9. Maximum Delay on GTW for different T_W values

2.4.3 The Distributed DPS Approach

In the DPS approach illustrated in Fig. 2.10, when sending host S1 wants to communicate with terminal receiver R1, it generates an RSVP E2E Path message (message 1_1) toward R1 to notify it of the presence of an incoming data flow. S1 describes its own traffic characteristics in terms of IntServ parameters (token bucket depth b, token generation rate r, peak data rate p).

The IR forwards the message like a normal IP datagram (IP protocol number modified to RSVP-E2E-IGNORE [34]) toward the GTW/ER (message 1_2). Core routers send it without modifying their internal state, and

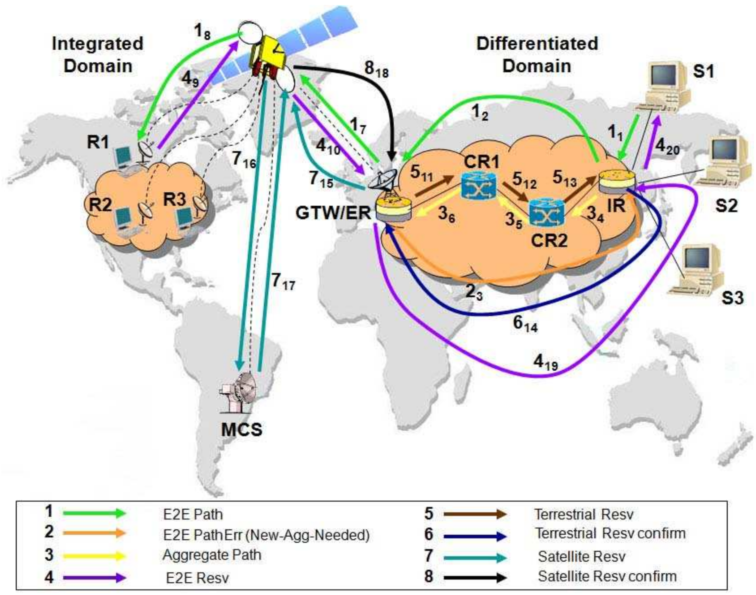

Fig. 2.10. Message exchange between terrestrial SCORE and InteServ satellite network for DPS

consequently without updating their Adpsecs. Therefore, the Deaggregator (GTW/ER) is responsible for updating the Adspec to reflect the impact of the aggregation region on the achieved end-to-end QoS.

To perform this task the GTW/ER needs the Aggregate Path state; if the aggregate state exists, the Adspec can be updated; otherwise, the GTW solicits an Aggregate Path message by sending an E2E PathErr message (message 2_3) with an error code of NEW-AGGREGATE-NEEDED and the desired DSCP to map E2E reservation onto aggregate reservations. Since Aggregate Path messages (messages 3) are processed inside the aggregation region, their Adspec is updated by interior routers. Consequently, the Deaggregator can make use of the information included in the Adspec of an Aggregate Path where available. If the aggregate state is already available at the GTW/ER, messages 2 and 3 are unnecessary, and the GTW can send the E2E Path message toward its intended destination through the satellite (messages 1_7, 1_8).

The destination host, once it has received the E2E Path message, sends a resource reservation request, called E2E Resv (messages 4_9, 4_{10}), in which it specifies the target QoS for the data flow generated by the sender. When this message arrives at the GTW/ER, the terrestrial CAC procedure starts. The GTW/ER forwards a terrestrial Resv message in the terrestrial network to the IR through the CRs (messages 5). Each CR receiving this message does local admission control based on the computation of an aggregated bandwidth (see section 2.4.1). If all CRs accept the new call, the terrestrial Resv message arrives at the IR and then a positive terrestrial Resv Confirm (message 6_{14}) is generated toward the ER/GTW; otherwise, the CR notifies the ER of admission failure through a terrestrial Resv Error. If the terrestrial CAC response is positive, the satellite CAC phase starts (messages 7). So the GTW sends the satellite Resv message toward the MCS through the TRM onboard (messages 7_{15} and 7_{16}).

The TRM, after a CAC phase executed in the MCS, forwards the admission result in the satellite Resv Confirm message (message 8_{18}) to the GTW (in the case of positive result, otherwise a satellite ResvErr is forwarded). If CAC on the satellite side is successful, the E2E Resv message (messages 4_{19} and 4_{20}) can be sent to sender S1 in order to tell it to start transmission.

2.4.4 The Hybrid ER-DPS Approach

The ER-DPS approach is proposed as a further optimization in order to reduce signaling overhead in the core network. To this aim, this work proposes to exploit the Aggregate Path messages to convey information on the Maximum Available Bandwidth (MAB) along a given aggregate route.

This information can be updated by every CR along the path with its available aggregate bandwidth calculated through the recalibration algorithm. Aggregate Path messages can be transmitted periodically after creation of the aggregate state. This periodic sending can be regulated by a timer coinciding with the refresh timer of Aggregate Path messages [34].

The proposed optimization can be exploited by the GTW/ER to make the call admission more effective. In fact, the GTW/ER decides to reject a new call without sending more messages toward the core network if the MAB on the Aggregate Path is lower than the bandwidth R requested by the new call. This dramatically reduces signaling overhead in the core network and saves processing power in CRs. The MAB information inside the Adspec of the Aggregate Path is updated by each CR_i along the path as follows:

$$MAB(t_{k+1}) = min_i(C_i - [R_{bound}(t_{k+1})]_i) \tag{2.7}$$

where $i = CR_1, CR_2, CR_n$, and C_i is the output link capacity of CRi. When the MAB information arrives at the GTW/ER, it calculates the new available bandwidth in this way:

$$[MAB(t_{k+1})]_{GTW} = min(MAB(t_{k+1}), C_{GTW} - [R_{bound}(t_{k+1})]_{GTW}) \tag{2.8}$$

If a new call requiring bandwidth R arrives at the GTW/ER and $R < [MAB(t_{k+1})]_{GTW}$, the request is forwarded to CRs (messages 5 in Fig. 2.10) to be processed; in this case the procedure continues as in the DPS approach in Fig. 2.10. Otherwise, if $R > [MAB(t_{k+1})]_{GTW}$, the call is refused by the GTW/ER and a ResvErr message is sent to the sender without requesting any processing from CRs. In other words, the ER-DPS approach takes advantage of the information carried by the Aggregate Path message to know potential bottlenecks in the core network, thus avoiding useless local admission at CRs.

2.4.5 The Centralized ER Approach

In order to further reduce the protocol signaling overhead, the ER approach gives the ER/GTW full control of the terrestrial admission decision.

The procedure, illustrated in Fig. 2.11, is the same as in the previous approaches until the gateway receives the E2E Resv message (message 4_{10}) and terrestrial CAC starts. The GTW comes to the admission decision by itself on the basis of the MAB information conveyed in the Adspec of the Aggregate Path messages.

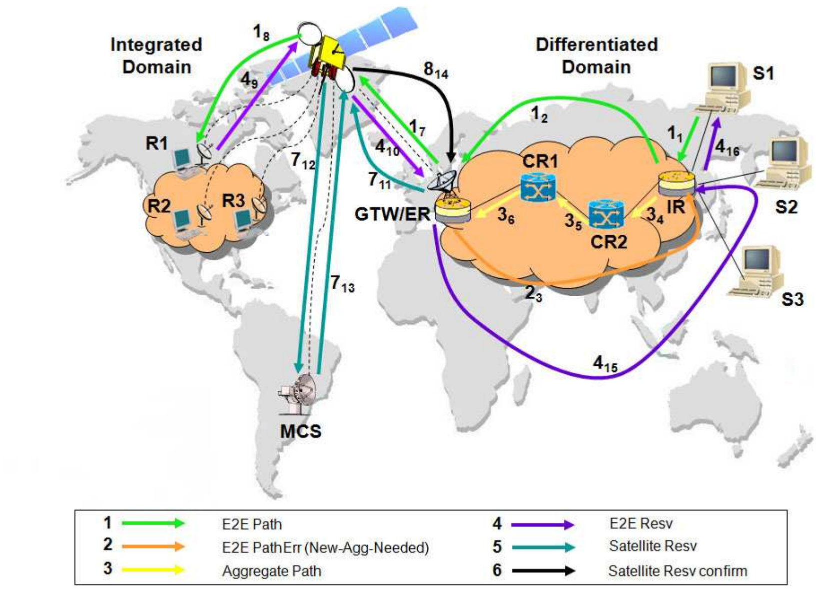

Fig. 2.11. Message exchange between terrestrial SCORE and InteServ satellite network for ER

Different from the ER-DPS approach, in this case the GTW is called to accept a new connection when $R < [MAB(t_{k+1})]_{GTW}$ without forwarding the request to CRs in order to be processed and locally admitted. The obvious advantage of this approach consists in the absence of explicit reservation messages to be sent in the SCORE network, thus dramatically reducing the signaling overhead and the processing burden in the CRs. On the other hand, this approach is highly vulnerable to wrong GTW decisions. This means that the reliability of MAB information, carried in the Aggregate Path messages, becomes the most critical issue.

To increase reliability of this approach, it is necessary that MAB information be refreshed periodically and frequently. The effects of different values of the refresh period (T_{PATH}) on system performance have been analyzed.

2.5 Performance Evaluation

A set of simulations is carried out to analyze the performance provided by the overall integrated terrestrial-satellite architecture. The main concern is to provide guaranteed quality to the most exacting class of services, that is, the GS class of the IntServ model, which specifies a quantitative bound on the end-to-end queuing delay [18]. GS traffic is mapped over semi-permanent connections in the satellite network, and over the EF PHB [24] of DiffServ in the terrestrial SCORE region. All curves report the performance of the three approaches (DPS, ER-DPS, and ER) vs. the burstiness b of the generated traffic.

Table 2.1. Simulation Parameters

Source Parameters	**Value**
Number of Terrestrial Sources	256
Burstiness β	2, 4, 8, 16
Bucket Size b	512 kbps
Peak Rate p	$\beta \cdot r$
Rate r	128 kbps
Receiver Parameters	**Value**
Number of Satellite Receivers	256
Satellite Receiver Type	SaT-C
Delay Bound (DB) requested	500, 600, 700, 800, 900, 1000 ms
Satellite Parameters	**Value**
Round Trip Time	540 ms
Medium Access Protocol	MF-TDMA
Timeout for request in OBP	26.5 ms
Satellite Link Bandwidth	16 Mbps
Atomic satellite channel	16 kbps
Target burst loss probability (ϵ)	0.01
Terrestrial Parameters	**Value**
Terrestrial Minimum Path Latency (MPL)	30 ms
Terrestrial Link Bandwidth	16 Mbps
Core Router scheduling	CJVC
T_W period	5 s
T_{PATH} period	10, 30 s

In the Table 2.1 a set of simulation parameters are reported in order to show the changing performed in the simulation scenario.

In Fig. 2.12 the burst loss is reported. The system is able to guarantee a burst loss lower than the target value ε (equal to 0.01) set by the satellite CAC algorithm. This is true independently of the burstiness value and for all algorithms but the ER approach when the T_{PATH} period lengthens. T_{PATH} is the refresh timer used for sending periodic Aggregate Path messages that allow computation of MAB information. This testifies to the critic role of T_{PATH} calibration. The ER approach obviously shows better performance when the T_{PATH} refresh timer is shortened and the MAB information more frequently updated. For all approaches, to increase the burstiness causes higher burst loss. This is due to the higher bandwidth R requested by the satellite receivers due to the higher peak data rate value p of terrestrial sources (note that simulations are performed by fixing the average data rate r at 128 kb/s, thus increasing the burstiness b means increasing the peak data rate $p = br$).

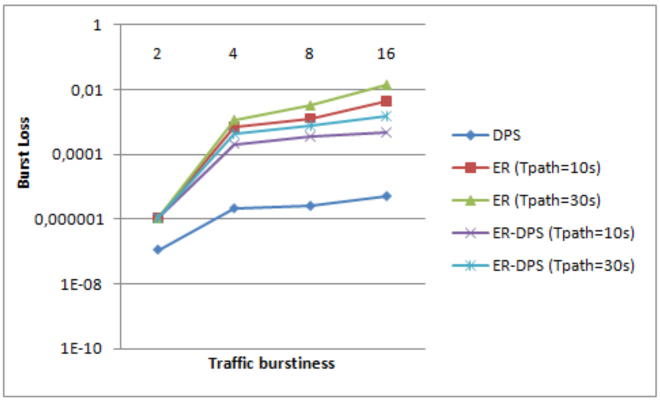

Fig. 2.12. Burst loss vs. traffic burstiness of terrestrial sources

The burst loss trend is confirmed by the trend in the number of admitted calls shown in Fig. 2.13.

The ER approach is the one admitting more calls on the terrestrial system; however, such a number may be not manageable when T_{PATH} increases ($T_{PATH} = 30$ ms), as attested by the burst loss that exceeds the bound (0.01) established by the satellite CAC.

It is very interesting to see the benefits of the ER and ER-DPS approaches for the signaling overhead shown in Fig. 2.14.

It is expected that for these approaches that use the MAB information conveyed in the Aggregate Path messages, the control overhead decreases compared to the DPS case. The measured control overhead is given by the number of Aggregate Resv messages (messages 5 in Fig. 2.10) processed by ER

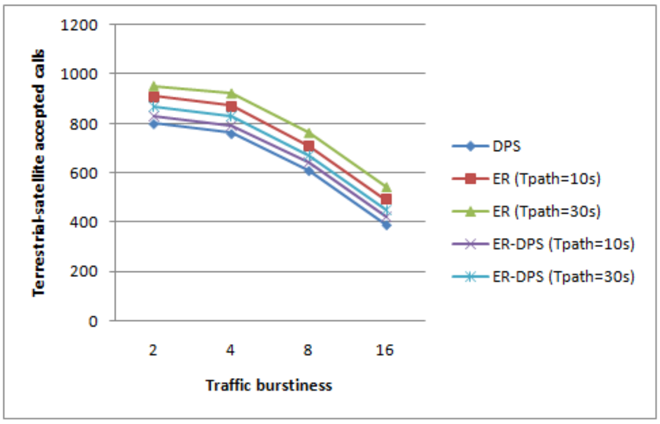

Fig. 2.13. Number of admitted calls vs. traffic burstiness of terrestrial sources

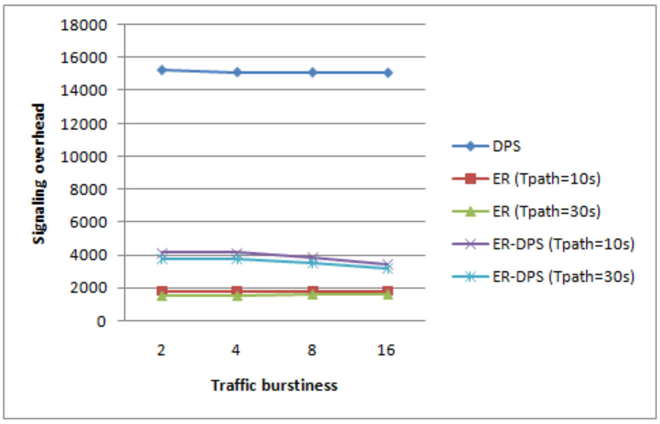

Fig. 2.14. Number of overhead packets vs. traffic burstiness of terrestrial sources

and CRs in the SCORE network in order to perform local admission control, and by the periodic Aggregate Path messages (messages 3 in Fig. 2.10 and Fig. 2.11) sent from IR to GTW to update MAB information.

In the ER-DPS approach, if there is a bottleneck in the SCORE network, the MAB information can help the ER decide not to admit the new call and thus not forward the Aggregate Resv messages into the SCORE network. The ER approach, on the other hand, avoids sending explicit reservation requests toward CRs even in the absence of bottlenecks in the terrestrial network. Obviously the additional reduction in terrestrial signaling also reduces computational overhead and CPU power consumption in the CRs. Many other simulations have been conducted for different traffic source burstiness, timer values T_{PATH} (2-50 s), and b/r token bucket values, and it has been observed that a T_{PATH} value between 5 and 10 s permits matching the QoS constraint (burst loss) while offering the best performance in reducing the protocol overhead.

2.6 Conclusions

In this chapter a careful examination of a possible integration of a terrestrial-satellite network has been made. The introduction of QoS architecture proposed by IETF has been performed in order to evaluate the behavior of a hybrid network in the managing of users requirements.

The work proposes a design specifications for the integration of an IntServ satellite access network and a DiffServ core network by exploiting the advantages of a stateless approach, per-aggregate flow management, and dynamic bandwidth recalibration within the core region.

The IntServ approach is "per-flow" based. It allows precise per-flow services provisioning and realizes this by informing routers about the necessary details of each single flow by means of the RSVP. Its main problem is that per-flow signaling and state in intermediate routers grow linearly with the number of traffic flows and require high overhead in terms of computational complexity and memory. The DiffServ approach instead is based on "aggregates" of traffic flows. Only the edge routers are burdened with processing flows and categorize traffic into separate classes in order to free core routers from per-flow management.

In order to take advantage of a new proposal about scalability and flexibility network architecture a so called SCORE platform has been introduced. The new proposal permits to exploit better the differentiation between terrestrial nodes in edge and core routers. In fact, it allows to light the work of the core routers avoiding the storage of a lot information inside the nodes and reducing the CPU utilization. The novelty of the SCORE approach is that the service offered to the aggregate flow approximates the QoS achieved with per-flow mechanisms without requiring routers to maintain per-flow state. This proposal architecture is based on a technique that suggests of insert some information useful for the router for computing of scheduling and CAC algorithm. This technique is called Dynamic Packet State (DPS).

In the considered scenario a network device plays a fundamental role, the GTW. It, in fact, is the node that interfaces the two different network. It also is responsible of the terrestrial resource management and the mapping between the class of service of IntServ and DiffServ architecture. Moreover, it has to aggregate and disaggregate the traffic flows that pass through it. In order to investigate the performance of the system in terms of QoS requirements and scalability in the large terrestrial core network, three different type of terrestrial admission controls have been proposed showing how the system improves when it moves from a distribute to centralized control. The trade-off to pay is the big responsibility of the GTW that has to decide of admitting or not a new flow in the network.

The results show that performance can be improved by delegating the GTW between terrestrial and satellite segments to make autonomous decisions on the admission of new connections in the system. Giving the gateway the autonomy to decide on rejection of a new call (the ER-DPS approach) allows increased scalability while matching target QoS requirements. However, when the offered load on the terrestrial network is high, higher scalability can be achieved by giving the gateway the autonomy to decide on either rejection or acceptance of a new call (the ER approach). In this case, it is important that the gateway relies on affordable information that come by the other network device; this means the period of updating information on available bandwidth must be adequately tuned in order to match the QoS requirements.

3

Multi-layer Network for Vertical Switching

The great interest that has grown in recent years for multimedia satellite communications has led to the integration of satellite segments with IP QoS architecture in order to obtain a full interoperation between the terrestrial backbone and the satellite segment.

In recent years, new wireless platforms have gained much interest. The recent platforms are the stratospheric platforms HAPs that offer a reduced propagation delay; they are especially suitable for interactive multimedia services. In this chapter, the interoperation between HAP and satellite segments has been considered. A new way to manage Integrated Services over a new hybrid wireless platform has been proposed. The proposed resource reservation on HAP-Satellite segments is receiver driven and it is based on maximum end-to-end delay requested by receiver and on available bandwidth of two wireless segments. The HAP platforms are introduced in order to show their characteristics and the considered element that composed the HAP segment. The satellite segment, instead, is similar to that considered in the previous chapter.

Performance evaluations of overall systems have been evaluated in terms of HAP-Satellite bandwidth utilization and on the number of admitted calls in order to give an idea of the benefits given by the HAP segment to the satellite segment. The simulations show an improvement of admitted calls, reduced end-to-end delay and increased bandwidth utilization.

In Fig. 3.1 is depicted a schematized HAP-Satellite scenario.

3.1 Wireless HAP segment

Can the HAP be seen as a complementary technology or a disruptive technology? In our vision, the HAP segment will represent, in areas without terrestrial infrastructure, a winning technology able to provide connection services and multimedia services. Further, it will represent in area with wired infrastructure an additional network element able to provide several benefits for new

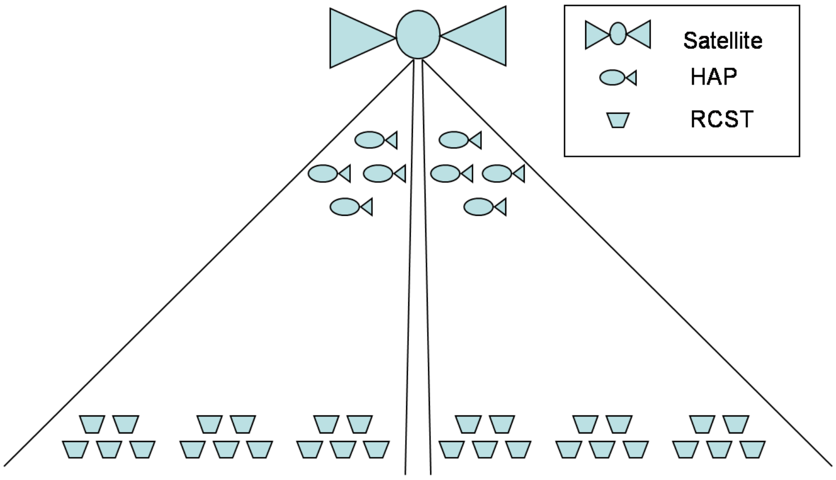

Fig. 3.1. Integrated HAP-Satellite scenario

classes of users and services. Also, if the satellite networks, and in particular the GEO satellite networks, have become important in these last few years, the HAP layer will offer another complementary element able to reduce problems regarding the signal attenuation, the power of terminal and base station, and the long propagation delay (GEO satellite presents high end-to-end delay).

In this context, it is important to consider the new role played by the HAP segment and to consider the potential integration of this new element with the existing networks. In particular, it is interesting to analyze the possible integration of HAP with the existing satellite network in order to obtain an efficient management and coexistence of the two networks. It is possible to use the satellite segment for services that do not need specific delay restrictions and HAP (when the segment is deployed) to give access to users that need particular delay-sensitive services.

Before outlining the potential interoperation of the HAP segment with the GEO satellite segment, some functionalities and architectural elements needed to manage the traffic and the admission control of the HAP segment are briefly described.

The basic elements of stratospheric architecture can be the following:

- 2GHz, 30GHz, 48GHz HAP payload: for example it is possible to use the 30GHz band, 31.0-31.3GHz in the uplink, and 27.5-28.35GHz in downlink for Fixed Service (FS) [38] or other nonadjacent bandwidth;
- Mobile HAP User Terminal (MHUT): it is suitable to have different mobile terminals that offer different data rate in uplink;
- HAP Gateway Station (HGTW): it represents the interface between existing the terrestrial network and the HAP segment. This element needs to manage different interfaces as Integrated Services Digital Network (ISDN),

Asynchronous Transfer Mode (ATM), Public Switched Telephone Network (PSTN) etc. and to perform some Inter Working Functions (IWF) to offer interoperability with other network segments;
- HAP Master Control Station (HMCS): it controls the HAP connections and the uplink and downlink channel.

Traffic and resource management functionalities can be distributed in different modules of the overall system or can be centralized for all these functions inside the HAP payload, reducing the mobile terminal and ground station. The easy deployment of these new devices permit the introduction of more complex management on payload and permit the use of HAPs not only as a bent pipe but as a switch/router with OBP capabilities.

In our architecture, the HAP Connection Control (HCC) and the HAP Radio Resource Management (HRRM) are deployed in both the HAP user terminal and the earth station. The admission control for HAP connections is performed in HMCS and the resource management in HAP Traffic Resource Manager (HTRM) on HAP payload. This framework uses the same HCC and HTRM as the ESW platform [30], [31], differentiating the performance of the HAP and the satellite platforms according to the different propagation delays.

3.1.1 Traffic Resource Manager (TRM)

The HTRM is the entity that manages transmission resources. It is equipped with different databases and with a calculator for an optimum resource management. The calculator, for each frame time (26.5 ms), memorizes the arrived requests. In the next frame time, it analyzes the requests and checks the possibility of satisfying the requests through the consulting of the occupation resources table.

If it is possible to satisfy the requests, HTRM sends a resource assignment message to the terminal requesting satellite channels. The definition and the management of the priority of the requests to be satisfied play a key rule in the resource allocation of the HAP system. It is necessary to define an adequate criterion for assigning the satellite resources to the service requests, distinguishing the low delay tolerant requests from the more delay resilient ones. For this reason, the HTRM have different priority queues that allow guaranteeing of the fairness for the real-time connections.

In particular in the HTRM are present 7 queues such as shown in Fig. 3.2 and they are subdivided as follows:

- 0: Constant Bit Rate (CBR) requests
- 1-4: Real Time Variable Bit Rate (rt-VBR) requests
- 5 Non-Real Time Variable Bit Rate (nrt-VBR) requests
- 6: Unspecified Bit Rate requests

In ESW the rt-VBR applications have to declare the maximum jitter. This information (purified of fixed propagation and processing delay and resulting

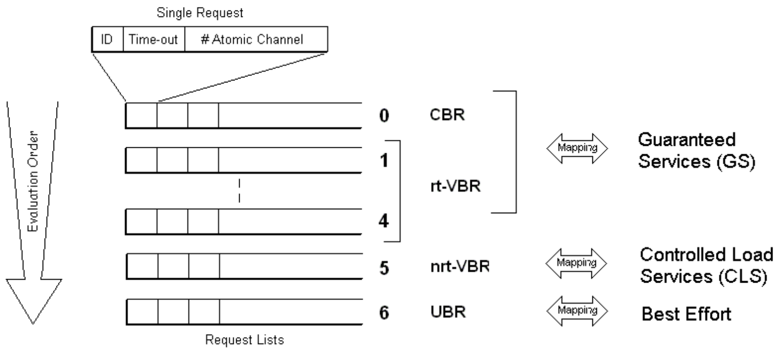

Fig. 3.2. Requests queues on TRM module for each ATM class of traffic

in a entire multiple of a frame time, 26.5 ms) is transferred from HAP Network Operation Center (HNOC) to HTRM and is called time-out: it represents the maximum allowable waiting time of a request on the HAP. For every frame interval, the HTRM performs an operation of queuing new requests and of analyzing present requests.

Once the inserting of the requests in the traffic queues is completed, HTRM begins to analyze each request starting from the high priority queue (queue 0). Only when all requests appertaining to a certain queue have been analyzed does it begin to analyze the requests in the following queue. If there are no available resources for a request, the request is queued only if the time-out is not expired; otherwise the request is deleted from the queue. The fairness for the rt-VBR connections is guaranteed by the presence of more used queues. The requests can pass from a queue to another one in dependence of the remaining queuing time.

3.2 Integrated Services on HAP

In order to provide QoS guarantees to users, the IETF proposed two main architectures: IntServ and DiffServ. The IntServ/RSVP model can be seen as a connection oriented IP service; thus, it is considered more suitable for the wireless access system. The main reasons to consider IntServ on HAP segment are fundamental two. Firstly, it permits more fine control on flow basis in order to avoid resource over-provisioning, offering deterministic guarantees (maximum queuing delay bound). Secondly, it allows capability of performing explicit bandwidth reservation for the flow, reducing the waste of bandwidth.

IntServ service classes, as it has been explained in the first chapter, have a traffic descriptor, so it is possible to account for the network state and the traffic source characteristics in order to formulate the appropriate bandwidth request. Two kinds of resource management are assumed: permanent and semi-permanent. The first management gives the resource to a call for the whole call life; the second, instead, allocates the resource on a burst basis. So, for semi-permanent connections the resource can be released during the OFF period of the sources, offering a more flexible management. This management is similar to satellite management, but the HAP approach can be more effective because of the reduced propagation delay (<1ms) compared to the GEO satellite propagation delay (250-270 ms); this assures an efficient management of bandwidth.

According to the fluid model, the bandwidth requested by a receiver (in our case an HAP receiver) is derived by fixing the DB desired by applications, the source parameters (peak rate p, average rate r and token bucket size b) and from rate dependent C and rate independent term D that accounts for the deviation from the fluid model [18], [20]. So R is expressed in the following way:

$$R = \frac{p \cdot (b - M) + (p - r) \cdot (C_{tot})}{(p - r) \cdot (DB - D_{tot}) + (b - M)} \quad if\ p > R \geq r \tag{3.1}$$

$$R = \frac{M + C_{tot}}{DB - D_{tot}} \quad if\ r \leq p \leq R \tag{3.2}$$

Where:

$$C_{tot} = \sum_{i=1}^{n} C_i \ \ and \ \ D_{tot} = \sum_{i=1}^{n} D_i \tag{3.3}$$

For a semi-permanent connection, the parameter D_i associated with fluid model for the HAP segment can be derived by the following terms:

- the time needed to the burst transmission request to arrive to satellite TRM and to the reply to come back to the sender, the two way delay. It is possible to refer to this term as Round Trip Time (RTT)/2;
- the maximum amount of time a flow data, ready to be sent, might have to wait for a slot (frame delay). This parameter can change according to MAC protocol used for the HAP segment. It is possible to call this term D_{frame_HAP};
- the maximum waiting time of the request on-board before being scheduled for transmission DB (maximum processing time). This depends on the way to manage the requests and on the HTRM in the HAP segment. It is possible to call this term D_{sched_HAP}.

The D_i term associated with semi-permanent connections that is called $D_{i_{semiperm}}$ is:

$$\lfloor D_{i_{semiperm}} \rfloor_{HAP} = RTT_{HAP}/2 + D_{frame\_HAP} + D_{sched\_HAP} \tag{3.4}$$

In semi-permanent connections, it is important to consider the RTT/2 because the requests are made on burst basis and the released bandwidth cannot be used during a RTT/2 time. So, RTT/2 can produce inefficiency in the bandwidth; yet, high burstiness and statistical multiplexing can offer some benefits as shown in [39].

For the permanent connections, instead, it is possible not to consider the delay associated with D_{sched_HAP} and RTT/2 after the admission of the call because the resources are permanently allocated for the call. So it is possible to have:

$$\lfloor D_{i_{perm}} \rfloor_{HAP} = D_{frame\_HAP} \tag{3.5}$$

The bandwidth request associated with the semi-permanent modalities on HAP is:

$$R = \frac{p \cdot (b - M) + (p - r) \cdot (M + C_{tot})}{(p - r) \cdot (DB - RTT_{HAP} - D_{frame\_HAP} - D_{sched\_HAP}) + (b - M)} \quad ifp > R \geq r \tag{3.6}$$

$$R = \frac{M + C_{tot}}{DB - RTT_{HAP} - D_{frame\_HAP} - D_{sched\_HAP}} \quad ifr \leq p \leq R \tag{3.7}$$

A similar consideration can be made for the satellite segment. Imaging to fix the parameter associated with the traffic management considering only the RTT, it is possible to have:

$$\begin{aligned} &\lfloor D_{tot_{semiperm}} \rfloor_{HAP} < \lfloor D_{tot_{semiperm}} \rfloor_{SAT} \\ &\Rightarrow \lfloor R_{semiperm} \rfloor_{HAP} < \lfloor R_{semiperm} \rfloor_{SAT} \end{aligned} \tag{3.8}$$

The equation above (eq. 3.8) shows the HAP request for the IntServ model is a lower amount of bandwidth than bandwidth on the satellite segment for the call using semi-permanent connections on the HAP segment. So, for semi-permanent management, the HAP segment is more useful than satellite, offering a high bandwidth utilization. In the following, a better utilization of HAP will be shown. However, in the view of interoperation and integration with the satellite segment, it is possible to differentiate the choice of segment according to the requirements of applications and traffic.

For the permanent connection, imaging to have the same MAC protocol (this can be true imaging to use the same MF-TDMA access scheme for the two segments), the advantage of HAP is in terms of reduced end-to-end delay. In this last case, the bandwidth request to obtain the same DB is different. The different RTT of two segments can play a key role in terms of permanent and semi-permanent management.

In Fig. 3.3, it is possible to observe the exchange of a message in the IntServ model associated with the HAP segment. The terrestrial source sends a E2E Path message to the HAP receiver, specifying the traffic parameters. The terminal, according to the fluid model and to eq. 3.5 and eq. 3.6, formulates the bandwidth request in order to obtain a maximum DB. So, an Resv message is sent to HAP, but before going back in reverse path forwarding, it is sent on HMCS to reserve the resources in uplink.

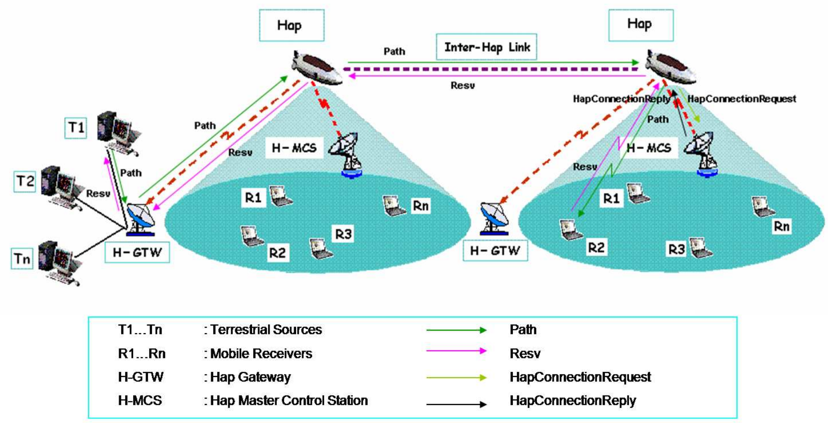

Fig. 3.3. HAP layer architecture

If this call admission control accepts the call, the Resv message can be sent to other intermediate HAPs (if these need to be crossed) through the Inter-HAP Link (IHL). If the bandwidth can be reserved along the path, the Resv arrives at a source that can start to transmit. It is important for the terminal receiver to consider the source parameter expressed in terms of the token bucket parameter, the packet size, and the availability along the path in order to request an allowable amount of bandwidth. In the Path message, as specified in [20], there is a field that can be used to transport the minimum bandwidth availability. This can be used to decide whether there is a sufficient bandwidth to request the resource or if it is better where it is possible to switch onto another network segment. In the next session, an idea of using this information with the end-to-end delay requested by receivers is presented.

3.3 Integrated Architecture for HAP-GEO Satellite infrastructure

In order to guarantee IP-QoS and to obtain a control of traffic with flow granularity in the wireless access network, the integration of an IntServ architecture in a hybrid HAP-Satellite segment has been considered. This proposal can be useful in the low propagation delay offered by the HAP segment for the provisioning of multimedia services and for the use of the satellite segment for low-interactive services. In order to obtain a full interoperation between HAP-GEO satellite segments and the terrestrial backbone, the following architecture showed in Fig. 3.4 is proposed.

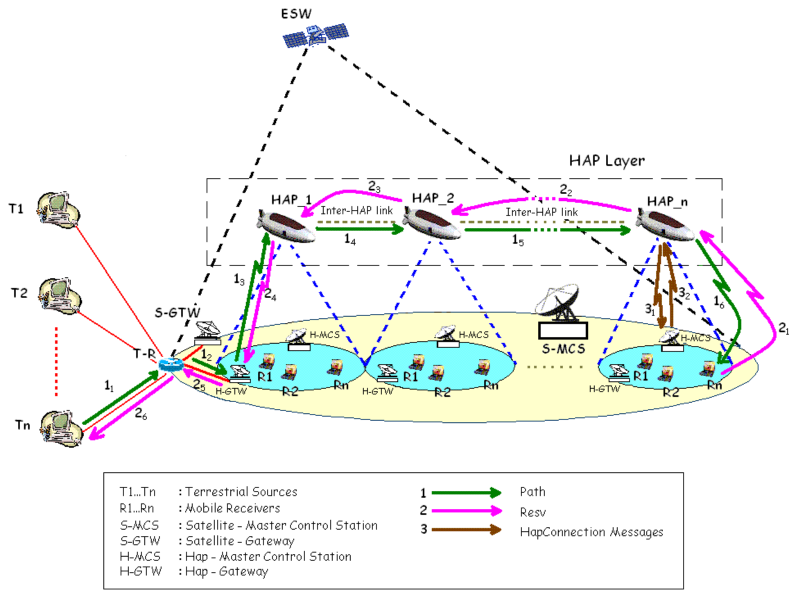

Fig. 3.4. Messages exchange for reserving resource on HAP

In Fig. 3.4, it is possible to observe a HAP segment composed of HMCS, designed to manage the call admission control on the HAP segment, and the HGTW used to aggregate traffic arriving at the terrestrial network according with the same terrestrial-satellite architecture defined in [40] [41]. The satellite platform considered is the ESW with Satellite Master Control Station (SMCS) and Satellite Gateway (SGTW) according to [30], [31]. The modalities of resource requests are receiver driven. A suitable factor for selecting the HAP or satellite segment can be the end-to-end delay requested by the user terminal. The end-to-end delay considering the fluid model can be defined in the following way:

$$D_{E2E} = QD_{tot} + PD_{tot} \tag{3.9}$$

where QD_{tot} is the overall Queuing Delay (QD) obtained in the path toward receiver and PD_{tot} is the overall Propagation Delay (PD) of the path between the source and the destination involved in the communication. For HAP layer:

$$QD_{tot} = \sum_{i=1}^{n} QD_{HAP_i} \tag{3.10}$$

$$PD_{tot} = \sum_{i=1}^{n-1} PD_{inter-HAP_{i,j+1}} + PD_{HAP\_uplink} + PD_{HAP\_downlink} \tag{3.11}$$

where n represents the number of HAPs crossed to connect receiver and transmitter and QD_{HAP} is the queuing delay on the HAP; PD_{HAP} is the propagation delay in uplink and downlink. The satellite layer presented the following characteristics:

$$QD_{tot} = QD_{SAT} \tag{3.12}$$

$$PD_{tot} = PD_{SAT\_uplink} + PD_{SAT\_downlink} \tag{3.13}$$

In particular, in the considered system the QD_{tot} coincide with DB requested by the receiver terminal. In fact, it is possible to view by eq. 3.1 and eq. 3.2 that the rate R is calculated considering a term $DB - D_{tot}$. Then:

$$DB - D_{tot} = D_R \Rightarrow DB = D_R + D_{tot} \tag{3.14}$$

3.3.1 Simple Modality Selection

Figure 3.4 shows the bandwidth reservation request, E2E Resv message exchanged in the overall considered system when the call is accepted from the HAP layer. When a source wants to communicate with a terminal receiver, it sends a Path message (Fig. 3.4 msg 1_1) towards the receiver.

The destination host, after receiving the E2E Path message (Fig. 3.4 msg 1_6), sends a message of resource reservation request, called E2E Resv (Fig. 3.4 msg 2_1), which specifies the QoS that receives the data flow generated by the sender. The RSVP Resv message arrives at the first HAP (Fig. 3.4 msg 2_1), which performs the CAC algorithm, sending the message ISConnectionRequest (Fig. 3.4 msg 3_1) towards the HMCS. After the CAC procedure, the HMCS sends the message ISConnectionReply (Fig. 3.4 msg 3_2) to the HAP.

If the CAC response is positive, this HAP forwards the Resv message (Fig. 3.4 msg 2_2) towards the other HAP. If the IHL response of all HAPs is positive, that is there are sufficient resources on the IHL for transmitting the information of this connection, the Resv message (Fig. 3.4 msg 2_6) arrives at the sender, which can start the communication.

In Fig. 3.5 the scenario that represents the failure reservation phase on the HAP segment and the reservation on the satellite segment are considered.

The HAP that cannot admit the call, because the D_{E2E} requested by the satellite receiver is lower than D_{E2E} registered on the last HAP or the HAP Call Admission Control (HCAC) response is negative due to resource deficit, sends a negative message (Fig. 3.5 msg 4_1), ResvErr to the receiver.

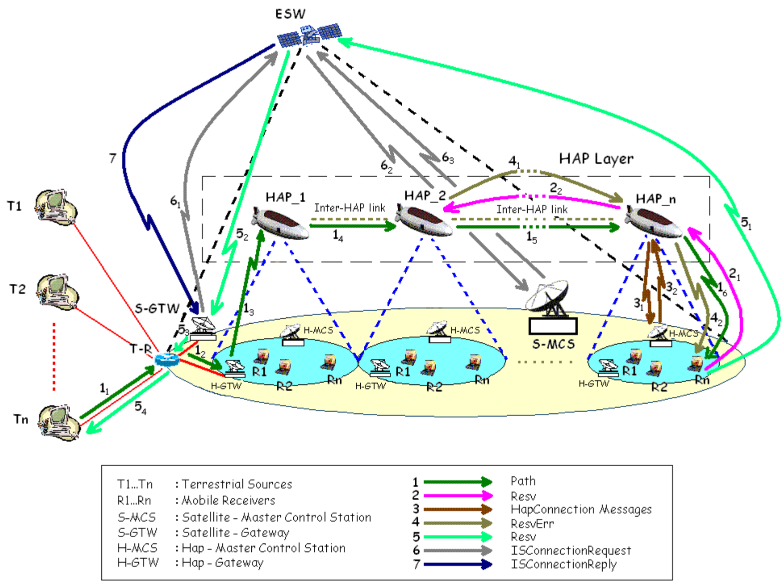

Fig. 3.5. Messages exchange for reserving resource on satellite after a HAP failure

When the receiver receives this message (Fig. 3.5 msg 4_2), it sends the channel's request (Resv) to the satellite segment (Fig. 3.5 msg 5_1) and then the satellite sends the message towards the SGTW (Fig. 3.5 msg 5_2). This last message causes the sending of the ISConnectionRequest message (Fig. 3.5 msg 6_1 - 6_2 - 6_3) to satellite MCS for performing the satellite CAC.

The bandwidth requested by the Satellite-HAP receiver is made in accordance with the fluid model [20] in order to guarantee the target D_{E2E}.

If the Satellite Call Admission Control (SCAC) implemented on SMCS according to [36] is positive, the channels have been reserved for the call and the notify message (ISConnectionReply) arrives at the TRM on the satellite (Fig. 3.5 msg 7). Now the Resv message (Fig. 3.5 msg 5_3 - 5_4) is sent by the satellite gateway towards the source host, which, now, can initiate the transmission.

3.3.2 Smart Terminal for automatic selection HAP-Satellite segment

The satellite or HAP layer selection of the wireless receiver can be based on the end-to-end delay requested for the application and on available resources available on terrestrial-satellite segments. The introduction of this smart functionality in terminals can improve the management of the two wireless segments.

It is important to make the following consideration: the HAP segment can offer an intrinsically lower delay than satellite, and according to IntServ to obtain the same delay, the satellite segment with semi-permanent connections requests higher bandwidth than the HAP segment. So, it is not suitable to use the satellite for end-to-end delay that is too low. However, if it is possible to use the satellite segment, it is impossible to think of using only the HAP, because it can be quickly saturated and further calls requesting the connection would be refused. Thus, the idea is to balance the calls in base to end-to-end delay requested at the receiver.

Analyzing the end-to-end delay, it is possible to understand if it makes sense to send the request on HAPs or on satellite. It may be that if the number of HAPs is too high or if there is some traffic congested HAP, that it is better to request the connection to satellite. Obviously, this selection depends on delay because if the delay is too low, it is not useful to request the connection to the satellite. So, through the delay evaluation and the minimum bandwidth availability (it is used to verify the bottlenecks among the path) accounting, it is possible regulate where to address the request of the receiver. The end-to-end delay for HAP segment is:

$$\begin{aligned} D_{E2EHAP} = &\sum_{i=1}^{n} QD_{HAP_i} + \sum_{i=1}^{n-1} PD_{inter-HAP_{i,j+1}} \\ &+ PD_{HAP\_uplink} + PD_{HAP\_downlink} \end{aligned} \tag{3.15}$$

considering the same HAP's altitude $PD_{HAP\_uplink} = PD_{HAP\_downlink}$, if the $PD_{inter-HAP}$ are equal and the semi-permanent connection is used on HAP segment, eq. 3.13 can be represented in this way:

$$\begin{aligned} D_{E2EHAP} = &D_R + \sum_{i=1}^{n} D_{frame\_HAP_i} + \sum_{i=1}^{n} D_{sched\_HAP_i} \\ &+ RTT_{HAP}/2 + (n-1) \cdot PD_{inter-HAP} + 2PD \end{aligned} \tag{3.16}$$

where PD is the one-way propagation delay receiver HAP then $2 \cdot PD$ is equal to $RTT_{HAP}/2$. If the HAP segment uses the same MAC protocol and the same traffic management, eq. 3.16 can be expressed:

$$\begin{aligned} D_{E2EHAP} = & D_R + RTT_{HAP} + n \cdot (D_{frame\_HAP} + D_{sched\_HAP}) \\ & + (n-1) \cdot PD_{inter-HAP} \end{aligned} \tag{3.17}$$

It is possible to observe the lower RTT value if the HAP segment is used in comparison with the satellite segment in the case of semi permanent connection. It is possible to observe as if the HAP segment is used, in the case of semi permanent connections, the RTT is lower than the satellite RTT if the satellite segment is applied. If the number of crossed HAPs is high, the second term of eq. 3.17 can give considerable contribution.

For the satellite segment, instead, the DB_{E2ESAT} is:

$$\begin{aligned} DB_{E2ESAT} = & QD_{tot} + PD_{tot} \Rightarrow DB_{E2ESAT} \\ & = QD_{SAT} + PD_{SAT\_uplink} + PD_{SAT\_downlink} \end{aligned} \tag{3.18}$$

with the assumption that only one satellite is necessary to connect source and receiver, eq. 3.18 will be:

$$\begin{aligned} D_{E2ESAT} = & D_R + RTT_{SAT}/2 + D_{frame\_SAT} + D_{sched\_SAT} \\ & + PD_{SAT\_uplink} + PD_{SAT\_downlink} \end{aligned} \tag{3.19}$$

where:

$$D_{tot_{semiperm\_SAT}} = RTT_{SAT}/2 + D_{frame\_SAT} + D_{sched\_SAT} \tag{3.20}$$

And considering also in this case $PD_{uplink} + PD_{downlink} = RTT/2$, eq. 3.19 can be expressed:

$$D_{E2ESAT} = D_R + RTT_{SAT} + D_{frame\_SAT} + D_{sched\_SAT} \tag{3.21}$$

If the delay requested by the receiver is D_{target}, it is possible to make the switching procedure represented by the Data Flow Diagram (DFD) in Fig.6.

In the DFD B_{HAP} and B_{SAT} are respectively the residual available bandwidth on the HAP and the satellite segments.

The first condition (a) depicted in Fig. 3.6 expresses the impossibility of sending a request on the satellite segment because the propagation delay is greater than tolerable delay of the terminal request. For example, in the case of semi permanent connections, the lower bound for satellite is 424ms.

The second and third condition, (b) and (c), attempt to obtain a traffic balancing for the delay range that can be supported by two segments. In this way, it is possible to avoid overloading the HAP with a higher delay that can be supported by the satellite, preserving its bandwidth for more delay sensitive services. Only if the bandwidth of the satellite is lower than the residual bandwidth of the HAP, it is possible to use the HAP segment that offers a greater factor utilization (this is due to reduced RTT).

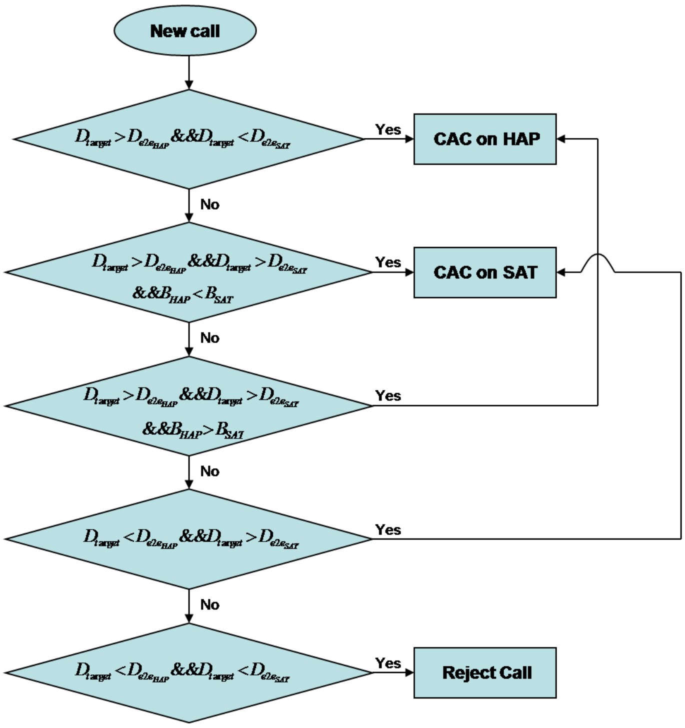

Fig. 3.6. DFD of the call admission control in the integrated HAP-Satellite network

The fourth condition (d) represents the situation in which the number of the HAP is high (the RTT gives a delay too high for the application) or there is some HAP with too high DB so it is better to go to the satellite segment, reducing the number of HAPs crossed (the satellite can reach the transmitter avoiding crossing more HAPs that can be saturated).

The fifth condition (e), instead, is relative to an impossible condition in which it does not make sense to formulate a request, because the receiver is too far from the transmitter. In this case, it is impossible to use the HAP (the number of HAPs crossed do not permit satisfying the maximum DB) and the satellite presents a high RTT for the receiver requirements.

In the HAP segment, the same TRM of the satellite segment has been maintained in order to compare the performance with the same values, except the RTT. Thus, proposed approach shows that through the RTT considerations, the end-to-end delay, the semi-permanent management, and the capability to use the HAP and satellite segments, it is possible to obtain an enhanced service.

3.4 Performance Evaluation

Different simulations have been carried out in order to show the advantage of the HAP segment introduction. DB requested by the receiver have been considered; the connection modality considered for HAP and satellite segment has been semi permanent. In the following, the simulation scenario and output indexes are presented and the performance analysis is completed (see Tables 3.1 and 3.2).

Table 3.1. Simulation Parameters for Satellite Segment

$D_{tot_{semiperm\_SAT}}$		
	Satellite Two Way Delay	270 ms
	Request Timeout	26.5 ms
	Maximum Processing Time on board the spacecraft	101 ms
	Frame Duration	26.5 ms
	Total Amount	424 ms

Table 3.2. Simulation Parameters for HAP Segment

$D_{tot_{semiperm\_HAP}}$		
	HAP Two Way Delay	0.2 ms
	Request Timeout	26.5 ms
	Maximum Processing Time on board the spacecraft	101 ms
	Frame Duration	3 ms
	Total Amount	130.7 ms

3.4.1 Simulation Scenario

Smart fixed terminals have been considered able to interconnect with HAP and satellite segments. The sources are terrestrial and are a sufficient number to represent high traffic conditions. The traffic source is the ON-OFF type. Figure 3.7 shows the simulated integrated architecture with HAP and satellite segments. In the simulation scenario an HAP layer composed by a HAP and a satellite layer composed by a GEO satellite have bwwn considered. The receiver has been considered to be a wireless device and sources are terrestrial

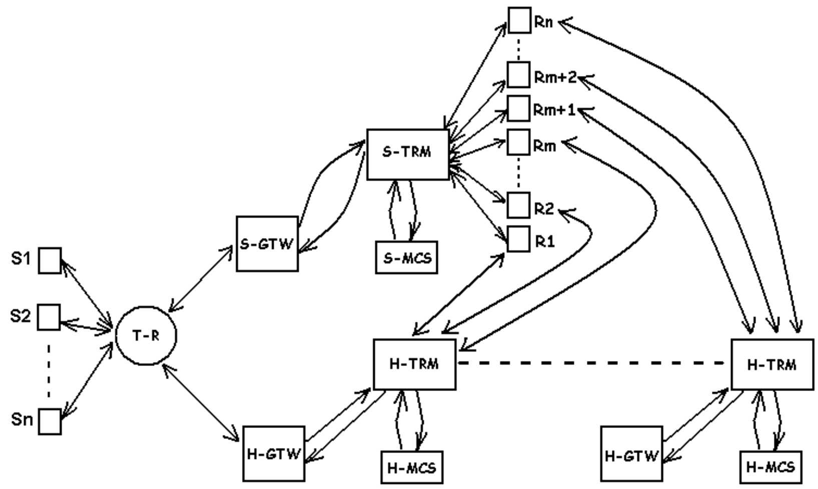

Fig. 3.7. Block diagram of simulated system

and they can communicate with HAP or satellite through the SGTW or the HGTW.

In Table 3.3, the parameters used in the simulation are shown.

The MAC protocol considered is MF-TDMA for the satellite system according to ESW architecture [31]. The HAP uses a similar MAC because in this way it is possible to associate the same D_{sched} and D_{frame} and it is possible to make a comparison associated with PD. If another MAC protocol is used, for example 802.16 [42], the validity of the approach to the network level remains (it is sufficient to change the constant). This presented analysis considers the capability of using two network segments (HAP and satellite) with two different RTT in an efficient way and analyzes the efficiency of a HAP system in comparison to the satellite system.

3.4.2 Simulation Parameters

The parameters considered for performance evaluation are:

- HAP Static Efficiency (HSE): r/R where r is the average token bucket data rate and R represents the bandwidth requested by the receiver that communicates with source via HAP;
- Satellite Static Efficiency (SSE): r/R where r is the average token bucket data rate and R represents the bandwidth requested by the receiver that communicates with source via satellite;
- HAP Utilization Factor (HUF): it is calculated as $(r \cdot T_{conn})/(R \cdot T_{poss}) = HSE \cdot (T_{conn}/T_{poss})$ where R is the bandwidth requested by the terminal that is connected to the HAP segment;

Table 3.3. Simulation Parameters

Source parameters		
	Traffic sources	rt-VBR
	Number of terrestrial sources	256
	Burstiness β	4, 8, 16
	Bucket size b	256, 512, 1024 (kbps)
	Peak rate p	$\beta * r$
	Rate r	128 kbps
Receiver Parameters		
	Number of Satellite-HAP Receivers	256
	Satellite-HAP Receiver Type	SaT-C
	Delay Bound (DB) requested	200, 300, 400, 500, 600, 700, 800 ms
Satellite Parameters		
	RTT	540 ms
	Medium Access Protocol	MF-TDMA
	Timeout for request in OBP	26.5 ms
	$D_{tot_{semiperm}}$	424 ms
	Satellite link bandwidth	32 Mbps
	Atomic satellite channel	16 kbps
	Target burst loss probability (ε)	0.01
HAP Parameters		
	RTT	0.4 ms
	Medium Access Protocol	MF-TDMA
	Timeout for request in OBP	26.5 ms
	$D_{tot_{semiperm}}$	130.7 ms
	HAP link bandwidth	16 Mbps
	Atomic satellite channel	16 kbps
	Target burst loss probability (ε)	0.01

- Satellite Utilization Factor (SUF): it is calculated as $(r \cdot T_{conn})/(R \cdot T_{poss}) = SSE \cdot (T_{conn}/T_{poss})$ where R is the bandwidth requested by the terminal that is accepted on the satellite segment;
- Number of Overall Admitted Calls (NOAC): it is calculated as the number of overall calls in the system admitted by both layers, the HAP and the satellite segments;
- Average End-to-End Delay (AEED): the average end to end delay of a data packet. This index is calculated separately for the packets that cross the satellite segment and the packets that cross the HAP segment.

The HUF represents the average amount of data generated by a source during a session $(r \cdot T_{conn})$, where T_{conn} is considered as session duration, compared to the total amount of data actually transportable when the bandwidth is not continuously allocated ($R \cdot T_{poss}$), and where T_{poss} is the time during which the calls effectively possess the resources and can transmit. The HSE and HUF index are different only for semi-permanent connections where $T_{poss} < T_{conn}$.

In this work, only semi-permanent connections are considered in order to show the contribution given by HAPs on the overall system.

3.4.3 Simulation Results

In the following, two kinds of scenarios are evaluated.

The first one considers the application of integrated architecture on the HAP segment and a performance evaluation and a comparison with the satellite segment is made.

The second scenario, instead, presents two kinds of selection of wireless segments:

1. a simple modality in which the HAP is initially selected by the receiver and in case of failure it selects the satellite segment;
2. a modality in which a selection is applied according to the DB_{E2E} and the bandwidth availability according to eq. a, b, c, d, e depicted in Fig. 3.6.

Integrated Services on the HAP segment

In Fig. 3.8, it is possible to observe the HUF and SUF for the different DB of the receivers. HUF and SUF increase for higher DB values. This is due to the lower bandwidth necessary to obtain the DB selected by receivers according to observations done in section 3.3.

The increasing burstiness values of the traffic sources reduce the utilization factor on HAP and satellite. For increasing burstiness, the bandwidth requested is higher because the bandwidth is directly proportional to peak rate; for the same average source rate, the peak rate increases for increasing burstiness values ($p = \beta \cdot r$). This increment of bandwidth for high burstiness values reduces the Factor Utilization because it has increased the number of HAP atomic channels or satellite atomic channels (they are directly proportional to peak rate) under-utilized during the release of resource in OFF period of the source. Also, if the multiplexing of the calls increases because it is possible to use the silent period of the sources, the number of channels not used during the waiting time for the release of resources reduces the effective utilization.

It is interesting to note that the HAP segment offers a higher utilization than the satellite segment. This is due to the reduced RTT of HAP that assures the time in which the resources are released and cannot used by other

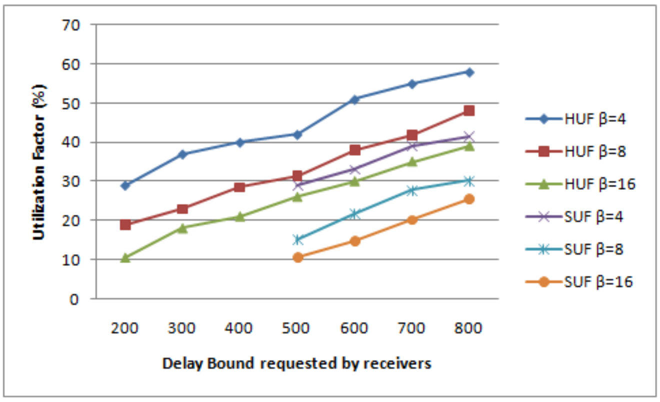

Fig. 3.8. Factor Utilization for HAP (HUF) and Satellite (SUF) vs. DB requested by wireless receivers. The terrestrial sources have $b/r = 2s$

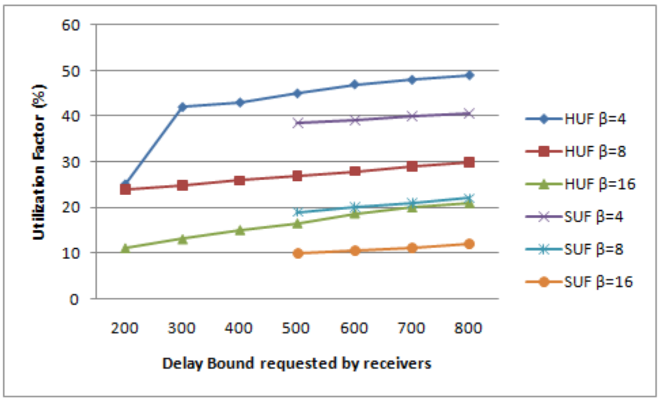

Fig. 3.9. Factor Utilization for HAP (HUF) and Satellite (SUF) vs. DB requested by wireless receivers. The terrestrial sources have $b/r = 8s$

calls is minimal. In the satellite, instead, it is necessary to wait 135 ms before using the released bandwidth for other connections.

In Fig. 3.8 and Fig. 3.9, the SUF starts from 500 ms because for satellite connections with semi-permanent management, it is not possible to obtain a lower DB. This drawback gives further importance to the traffic management of the HAP segment that for semi-permanent management can support a lower DB. The convenience of the semi-permanence of HAP connections is associated also with the number of crossed HAPs. If the number of crossed HAPs is so high that the propagation delay on the IHL is greater than propagation delay of a one-hop satellite link, the satellite node would be more useful. In this simulation only an HAP segment has been considered.

In Fig. 3.9 the HUF and SUF for a different b/r value have been shown. For $b/r = 8s$ the slope of the curve is reduced. This is due to the higher burst size to be transmitted in comparison with the average rate of the source token bucket. In this case, the traffic burstiness after the token bucket is reduced and the advantage of the multiplexing for high DB is reduced. For $b/r = 2s$, instead, for high DB values, it is possible to manage more calls and increase the Factor Utilization.

In Fig. 3.10 and Fig. 3.11, the number of HAP accepted calls is presented. Increasing DB values leads to a greater number of admitted calls. For high DB values, it has reduced the bandwidth requested by receivers so it is possible to admit more calls. Also, for the calls admitted on satellite segment and not shown here, the behavior is the same. The number of admitted calls on satellite, for the same DB value and for the same capacity, is lower. This is due to the higher bandwidth requested on satellite to guarantee the same Delay according to eq. 3.5. In Fig. 3.10, the number of admitted calls is greater than number of calls showed in Fig. 3.11 because greater ratios b/r reduce the traffic burstiness after the token bucket of the terrestrial source and reduce the potential statistical multiplexing.

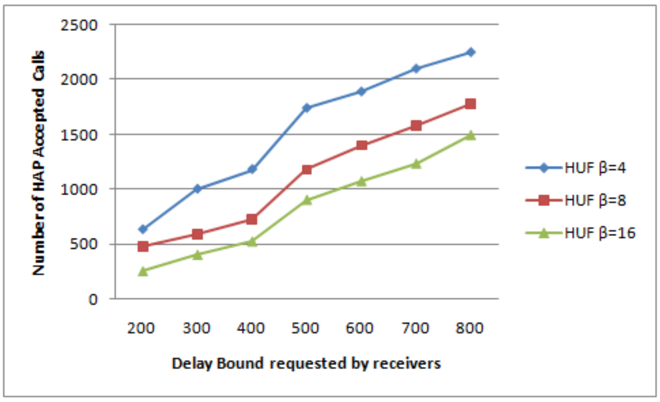

Fig. 3.10. Number of Accepted Calls on HAP segment vs. the Delay Bound requested by the wireless receivers. The terrestrial sources have $b/r = 2s$

The graphics shown in this section show the benefits introduced by HAP in comparison with satellite in terms of higher Utilization Factor. The lower propagation delay of HAP segment reduce the inefficiency due to the release of channel for semipermanent connections. The lower RTT offered by HAP also allows lower bandwidth to be requested to obtain the same DB offered by satellite admitting more calls. This last consideration can be verified by eq. 3.6 and eq. 3.7.

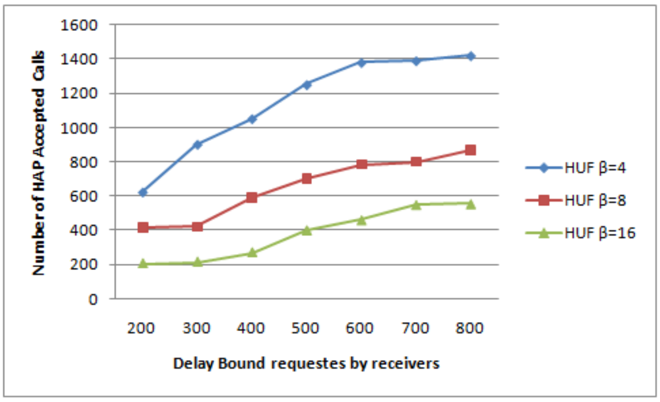

Fig. 3.11. Number of Accepted Calls on HAP segment vs. the Delay Bound requested by the wireless receivers. The terrestrial sources have $b/r = 8s$

Smart Selection of Wireless Access Segment at HAP-Satellite Receiver

In this scenario, the receiver-driven selection based on delay discrimination and bandwidth availability is proposed. This management permits a "smart selection" of the wireless segment by terminal. The performance evaluation of the proposed management against an easy management in which, in a static way, the HAP is always selected before the satellite segment is presented. This last scenario will be called "first_HAP" management.

In Fig. 3.12 it is possible to observe that the SSE is increased with the smart selection of the terminal. This is due to the better load distribution of the traffic. With the selection of the HAP before the satellite segment, the satellite platform is under utilized because the low delay tolerant services (with low DB) as well as the high delay tolerant services are served by the HAP ($DB >= 500ms$).

The smart selection, however, assures that the low delay tolerant services go on HAP and the high delay tolerant services can go also to the satellite segment according to the bandwidth availability. In this way, the HAP segment is loaded mainly with low delay tolerant calls and the satellite is loaded with high delay tolerant calls.

For increasing b/r ratios, the SSE is reduced and this is due to the reduced burstiness that decreases the possibility of multiplexing the calls. The high burstiness values reduce the SSE for similar reasons stated in the previous section. The HSE does not present an evident difference between the smart selection and the first_HAP management. The total improvement of the system is due to the better management of the satellite segment and to the effective use of the HAP.

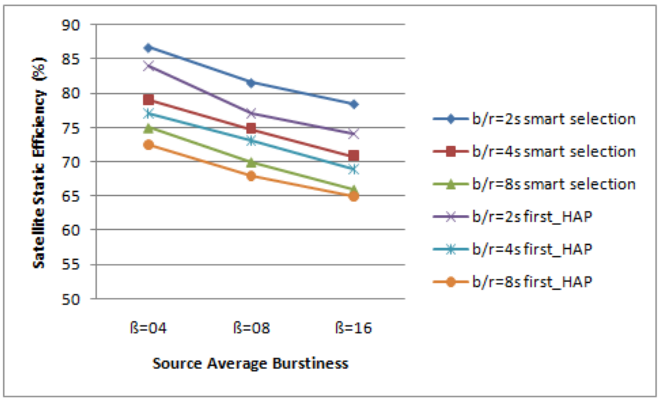

Fig. 3.12. Comparison of Satellite Static Efficiency (SSE) for receiver driven selection (smart selection) and the selection before the HAP and then the satellite segment (first_HAP) vs. the burstiness source traffic

In Fig. 3.13 the SUF for increasing burstiness values and different b/r values is shown. Also, the SUF is better for the smart selection. This is due to the reduced RTT of HAP segments that are better used. With smart selection, the high RTT of satellite is used to relay high delay tolerant traffic. With the first_HAP management, however, the satellite channel is not use too often because the HAP segment is always selected; this is also true for applications that request high DB (700ms, 800ms). So, in terms of overall management of the two segments, the satellite platform is in low use and the HAP platform is saturated. With the smart selection, the HAP is also saturated, but the satellite platform can be used for the application that requests a delay greater or equal to 500ms. High burstiness values of traffic lead to lower SUF values. The motivations are the same as discussed in the previous section.

The number of satellite admitted calls is shown in Fig. 3.14. In this case, it is possible to notice an evident increment of admitted calls. The smart selection manages in an effective way the two wireless segments. So, the applications with DBs belonging to the range [500ms, 800ms] are routed on satellite until the bandwidth availability is lower than the availability of the HAP. Obviously, the HAP is preferred to the satellite segment for the better HUF as explained in the previous section, but the satellite segment is also used in order to manage more calls. If the satellite is not used for high delay tolerant services, the HAP is saturated with applications with DB lower and higher than 500ms. So, if a new call with DB lower than 500ms arrives, it can not be served by HAP because it is saturated and cannot served by satellite because it does not support in efficient way so low a delay. The smart selection uses the appropriate segment according to appropriate delay and bandwidth, obtaining a better overall management. In Fig. 3.14, it is possible to observe also the reduction of the gain gap for increasing b/r ratios. This

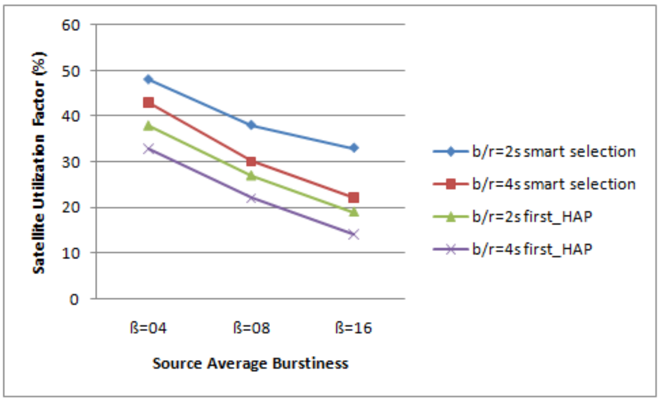

Fig. 3.13. Comparison of Satellite Utilization Factor (SUF) for receiver driven selection (smart selection) and the selection before the HAP and then the satellite segment (first_HAP) vs. the burstiness source traffic

result, according to the previous section in which the IntServ on HAP has been evaluated, shows the reduced number of admitted calls that reduce the effectiveness of satellite selection. Clearly, the smart selection continues to be more useful than the first_HAP selection.

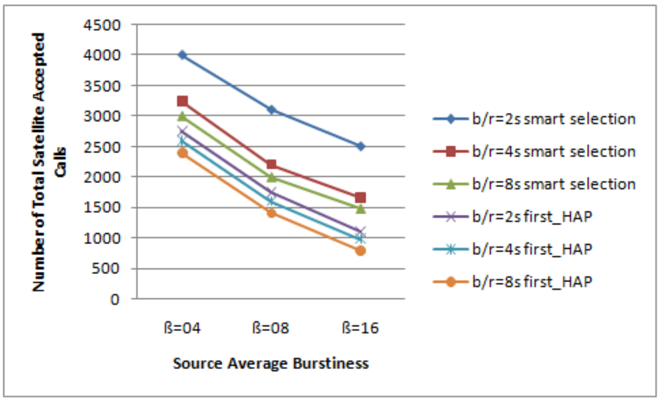

Fig. 3.14. Comparison of total satellite accepted calls for receiver driven selection (smart selection) and the selection before the HAP and then the satellite segment (first_HAP) vs. the burstiness source traffic

The effect of smart selection in terms of overall admitted calls is showed in Fig. 3.15. It is possible to observe the improvement of the overall system due to the increment of calls admitted on satellite segment. So also if the calls admitted on HAP segment is lightly reduced because the calls managed by

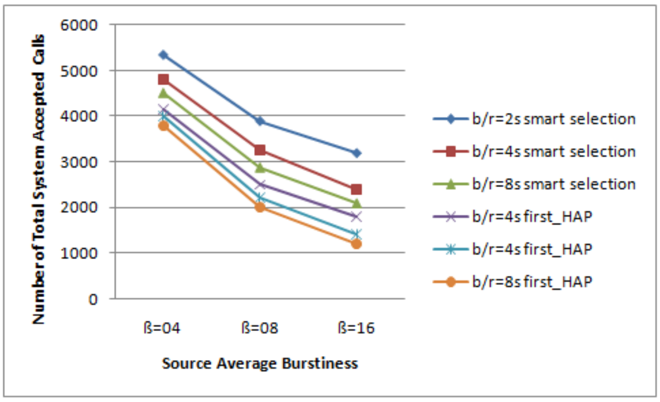

Fig. 3.15. Comparison of total system accepted calls for receiver driven selection (smart selection) and the selection before the HAP and then the satellite segment (first_HAP) vs. the burstiness source traffic

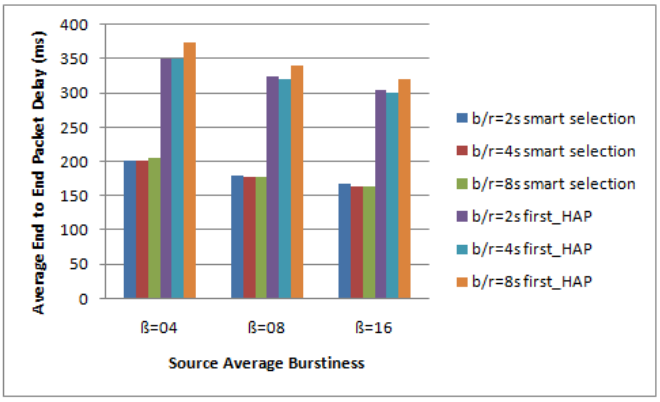

Fig. 3.16. Comparison of the AEED packet for receiver driven selection between HAP and satellite segments vs. the burstiness source traffic

HAP request more bandwidth than case in which also the high delay tolerant calls can go on HAP. The overall system performance has also improved.

The Average End-to-End Delay (AEED) is represented in Fig. 3.16. The HAP segment reduces the average end-to-end data packet delay due to the reduced propagation delay. This advantage offered by HAP can be more useful for the calls that request reduced delay because more tolerant calls can be admitted on the satellite segment that offer higher data packet delay.

3.5 Conclusions

This chapter has provided a study on an important aspect of the research in these last years. The study on heterogeneous multi-layered networks is a very actual problematic because today the telecommunication networks have to provide services with a certain QoS to users that belonging to different category. Then a services provider has to be able to provide the required application in a possible better way.

Moreover, the users can be attached to smart terminals that are able to select the best connection choosing the better platform that is able to provide the requested service. A possible way is to have a network composed of more than one layer that is able to provide services with different delay requirements. In this scenario the integration of stratospheric platform with satellite one is a winning solution. In literature it is possible to find many and many works that explain different way to perform integration also with wireless terrestrial network like UMTS, Wireless Fidelity (WiFi) solutions.

This chapter, firstly, has focused on the problem of integration of the IntServ paradigm in the HAP layer segment. It is possible to use the reduced RTT of the HAP segment to manage low delay tolerant calls. The HAP segment offers an intrinsically higher bandwidth utilization and reduced delay due to low propagation delay, but, in the view of interoperation of the HAP with the GEO satellite segment, can also be useful in managing the calls differently on the two wireless segments.

The satellite platform has been chosen because in these past few years, broadband GEO satellite communications with multimedia services provisioning have become very important. The HAP segment is considered as a complementary element, able to interoperate with the satellite segment in order to offer enhanced services. So a unified framework in which the two segments are considered is shown.

After a detailed explanation of the implementation of the IntServ architecture in the HAP segment and how to integrate this segment with the satellite one, a possible new type of receivers has been introduced in order to meet the problematic exposed above. In the chapter two modality of selection are presented and explained, one called first_HAP, in which the HAP segment is always chosen as first segment, and the second called receiver-driven, in which, on the basis of a very simple algorithm based on DB requested by receivers, the terminal can chose the better layer capable of satisfy its request.

The receiver-driven smart selection has been proposed in order also to use the satellite segment, increasing the number of admitted calls and the bandwidth utilization of the overall integrated system. The use of two wireless segments (HAP and satellite) at the same time can provide benefits to traffic management. The satellite segment can be used for calls with a high delay bound (greater or equal to 500ms for semi-permanent connections) and the HAP segment for calls with lower delay bound.

The chapter has shown criteria to select the wireless segment that can offer more effectiveness to traffic management. The receiver-driven approach permits approach to increase the overall calls admitted by the system and the total utilization. The approach attempts to balance the traffic among the HAP and satellite platforms, differentiating the services according to their maximum allowable end-to-end delay and the bandwidth availability. It is interesting to observe that by using the low propagation delay offered to HAP segments for provisioning of multimedia services, and by using the higher round trip time of the satellite segment for low-interactive services, it is possible to manage more traffic and to better utilize the bandwidth resources.

The performance of the system has been evaluated for semi-permanent connections on both wireless segments, because in these connection modalities the RTT can play a key role in determining the efficiency or the inefficiency of a system. Then, it is possible to view that the proposed selection mechanism is able to perform a good management of the network resources and so exploiting in a better way both the wireless bandwidths associated with the two layers. This mechanism is able to satisfy the users QoS requirements for applications that require stringent time constraints that, now, they can be served by the HAP layer that is able to cover the lacks of satellite segment.

4

Call Admission Control for Multimedia Traffic

Multimedia services are becoming very popular due to development of advanced applications for video and audio streaming and to the advent of broadband wireless access infrastructure able to provide ubiquitous communication. Satellite access represents a reality that has the advantages of broad and continental coverage, and quick installation [43], [44], [45].

The success of Direct-To-Home (DTH) television has proven that satellite services can compete with terrestrial alternatives in certain markets or can be integrated with terrestrial backbone for providing scalable end-to-end QoS services [46]. The development of the DVB-RCS standard, recently accepted by ETSI, is the first attempt at introducing a wide-scale satellite access standard. The Return Channel Satellite (RCS) standard is built around volume equipment delivery that can reach the price points required for market acceptance [5], [6], [44], [45].

In the quick deployment of the DVB-RCS platform that interoperates with the terrestrial Internet, a good CAC should be performed. DVB-RCS supports many classes of service and MPEG - Transport Stream (MPEG-TS). Thus, an efficient management and admission control for Variable Bit Rate (VBR) sources and for MPEG traffic is an important issue.

4.1 MPEG Standard

Recent progress in digital technology has made the widespread use of compressed digital video signals practical. Standardization has been very important in the development of common compression methods to be used in the new services and products that are now possible.

MPEG was started in 1988 as a working group within International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) with the aim of defining standards for digital compression of audio-visual signals. MPEG's first project, MPEG-1, was published in 1993 as ISO/IEC 11172 [47]. It is a three-part standard defining audio and video compression coding methods and a multiplexing system for interleaving audio and video data so that they can be played back together. MPEG-1 principally supports video coding up to about 1.5 Mbps giving quality similar to Video Home System (VHS) and stereo audio at 192 bit per second (bps). It is used in the CD and Video-CD systems for storing video and audio on CD-ROM.

During 1990, MPEG recognized the need for a second, related standard for coding video for broadcast formats at higher data rates. The MPEG-2 standard [48], [49] is capable of coding standard-definition television at bit rates from about 3-15 Mbps and high-definition television at 15-30 Mbps. MPEG-2 extends the stereo audio capabilities of MPEG-1 to multi-channel surround sound coding. MPEG-2 decoders will also decode MPEG-1 bitstreams.

MPEG-2 aims to be a generic video coding system supporting a diverse range of applications. Different algorithmic 'tools', developed for many applications, have been integrated into the full standard. To implement all the features of the standard in all decoders is unnecessarily complex and a waste of bandwidth, so a small number of subsets of the full standard, known as profiles and levels, have been defined. A profile is a subset of algorithmic tools and a level identifies a set of constraints on parameter values (such as picture size and bit rate). A decoder which supports a particular profile and level is only required to support the corresponding subset of the full standard and set of parameter constraints.

4.2 MPEG Statistical Modeling

MPEG video traffic is characterized by constant transmission rate of two Group of Picture (GOP) per second and 15 frames per GOP. Since the number of bytes in a frame is dependent upon the content of the video, the actual bit rate is variable over time. However, the MPEG video supports also the constant bit rate CBR mode.

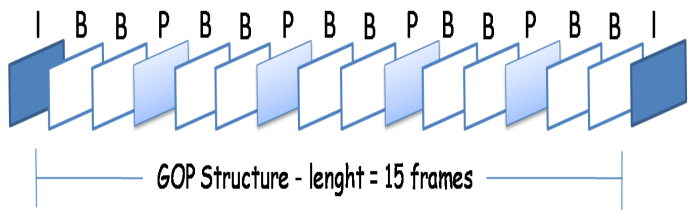

Fig. 4.1. GOP composition pattern

There are three types of frames in a GOP structure [49]:

- I-Frames (intraframes)-encoded independently of all other frames;

- P-Frames (predictive frames)-encoded based on immediately previous I or P frames;
- B-Frames (bidirectionally predictive)-encoded based on previous and subsequent frames.

Many works in literature faced the problem of analyzing and describing the structure of MPEG-2 traffic streams, trying to find a way to emulate them through statistic and stochastic streams generators, preserving the proper and intrinsic nature of the original stream.

In [50] the input process model is viewed as a compromise between the Long Range Dependent (LRD) and Short Range Dependence (SRD) models. Simulation results were found to be better than those of a self-similar process when the switch buffer is relatively small.

The MPEG video model presented in [51] is a Markov chain model based on the GOP level process rather than the frame level process. This has the advantage of eliminating the cyclical variation in the MPEG video pattern, but at the expense of decreasing the resolution of the time scale. Typically a GOP has duration of a half second, which is considered long for high speed networks. Of particular interests in video traffic modeling are the frame size distribution and the traffic correlation. The frame size distribution has been studied in many existing works.

Krunz [52] proposed a model for MPEG video, in which, a scene related component is introduced in the modeling of I frames, but ignoring scene effects in P and B frames. The scene length is independent and identically-distributed (i.i.d.) with common geometric distribution. I frames are characterized by a modulated process in which the local variations are modulated by an Auto-Regressive (AR) process that varies around a scene related random process with log-normal distribution over different scenes; i.e., two random processes were needed to characterize I frames. The sizes of P and B frames were modulated by two i.i.d. random processes with log-normal marginal. This model uses several random process and need to detect scene changes, thus complicating the modeling process.

In [53], [54] the adaptive source is modeled by means of a discrete-time queuing system representing a virtual buffer, loaded by the video source when its quantizer scale parameter is changed according to the feedback law implemented in the encoder system. The whole paper is based on the Switched Batch Bernoulli Process (SBBP) that has been demonstrated to be suitable to model an MPEG video source; in fact, being a Markov modulated model, an SBBP is able to capture not only the first-order statistics but also the second-order ones which characterize the evolution of the movie scene.

In this work a new concept of GOP-rate modeling, based on the discretization method originally proposed in [55] for the wireless channel study has been introduced. After the GOP-rate trend has been analyzed for the whole duration of the stream, it has been discretized in a certain number of states. Then the associated Markov chain parameters have been evaluated.

4.3 Multimedia Call Admission Control (CAC): Overview

Efficient Radio Resource Management (RRM) and CAC strategies are key components in heterogeneous wireless system supporting multiple types of applications with different QoS requirements. CAC tries to provide QoS to multiple types of applications with different requirements considering both call level and packet level performance measures. A CAC scheme aims at maintaining the delivered QoS to different calls (or users) at the target level by limiting the number of ongoing calls in the system. CAC schemes have been investigated extensively in each type of network. Different approaches of CAC exist in literature, centralized, distributed, Traffic-Descriptor-Based, Measurement-Based and so on [56], [57].

In satellite networks, different types of admission control have been studied. In [36] the authors have presented a novel strategy for handling ATM connections of different natures, traffic profile, and acQoS requirements in enhanced satellite systems. CAC represents a module of Network Operator Center (NOC) disposed on a terrestrial station. Its task is to regulate the access to satellite segment. It permits a flexible handling of the bandwidth and avoids the a priori partitioning of the resources among different types of service. The CAC algorithm has been designed also to fulfill the objectives of minimizing the signaling exchange between the on-board and on-earth segments of the system. In order to reduce delays due to the processing of the call requests on board, the relevant parameters of the processed calls are stored and elaborated within the ground segment. The method is based on the concept of reserving buffer resources to each virtual circuit as long as data are sent. The decision on the call acceptance is taken following the evaluation of the Excess Demand Probability (EDP), i.e., the probability that the accepted calls during their activation periods request more buffer resources than those available.

In [58], [59] the authors propose an adaptive admission control strategy, which is aimed at facing link congestion and compromised channel conditions inherent in multimedia satellite networks. They present the performance comparisons of a traditional (fixed) admission control strategy versus the new adaptive admission control strategy for a Direct Broadcast Satellite (DBS) network with RCS (Direct Broadcast Satellite with Return Channel Satellite (DBS-RCS)). Fixed admission control uses the same algorithm independent of the past traffic characteristics. The Bandwidth Expansion Factor (BEF) for VBR traffic is determined such that the probability of the aggregate instantaneous rate exceeding the fraction of the capacity assigned to the admitted VBR services will not be greater than a pre-specified probability value (ε). The dynamic approach recognizes that the admission control can only approximately estimate the statistical multiplexing and attempts to use the characteristics of past traffic streams to better estimate the gain that can be achieved. Unlike the fixed admission control, the adaptive admission control adjusts the BEF such that the actual value of is close to the desired value ε that is restricted by the acceptable QoS limits.

Concerning the Video Broadcasting delivery and scalability properties to be offered for large scale and heterogeneous networks, the authors in [60] adopted a Video on Demand (VoD) scheme where VBR videos are mapped over CBR channels and a traffic smoothing scheme with a buffering delay control are proposed. The same authors in [61], proposed novel broadcasting and proxy caching techniques in order to offer more scalability to the video delivery and to increase the overall performance of the system.

In [62], the authors proposed a scheme to reduce the waiting time of the video application client side. The video traffic considered by authors was MPEG2.

In [63] the author performs a comparison between Quality Oriented Adaptation Scheme (QOAS) against other adaptive schemes such a TCP Friendly Rate Control Protocol (TFRCP), Loss-Delay-based Adaptation Algorithm (LDA) and a Non Adaptive (NoAd) solution when streaming multiple multimedia clips with various characteristics over broadband networks.

The purpose of this study in [64] is to propose a quality metric of video encoded with variable frame rate and quantization parameters suitable for mobile video broadcasting applications.

In [65] the authors present the results of a study that examine the user's perception of multimedia quality when impacted by varying network-level parameters (delay and jitter).

In our contribution the GOP Loss Ratio (GLR) as QoS parameter to be respected and VBR traffic have been considered in a DVB-RCS architecture.

4.4 Applicative Scenario

DVB-RCS is a technology that permits to have return channel and forward channel over the same medium. In this way users can take advantages of the satellite communications. Moreover, it is obtained an interactive communications between the end-user and the service source. A DVB-RCS system with OBP on the satellite payload has been considered as reference architecture in this work such as depicted in Fig. 4.2.

DVB-RCS system was specified by an ad-hoc ETSI technical group founded in 1999 [5], [6]. It uses typical frequency bands in Ku (12-18 GHz) for the forward link and/or Ka (18-30 GHz) for the return link and it is composed by Return Channel Satellite Terminal (RCST)s, a Network Control Center (NCC), the satellite and a Feeder/Gateway (F/G). The core of the system is the NCC that has the task to manage the system, in particular it manages the connection, the system synchronization and informs all the RCST about the system. RCST consists of four different types on the basis of different bandwidth capacities: RCST A (144 kilo bit per seconds (kbps)), B (384 kbps), C (1024 kbps), D (2048 kbps).

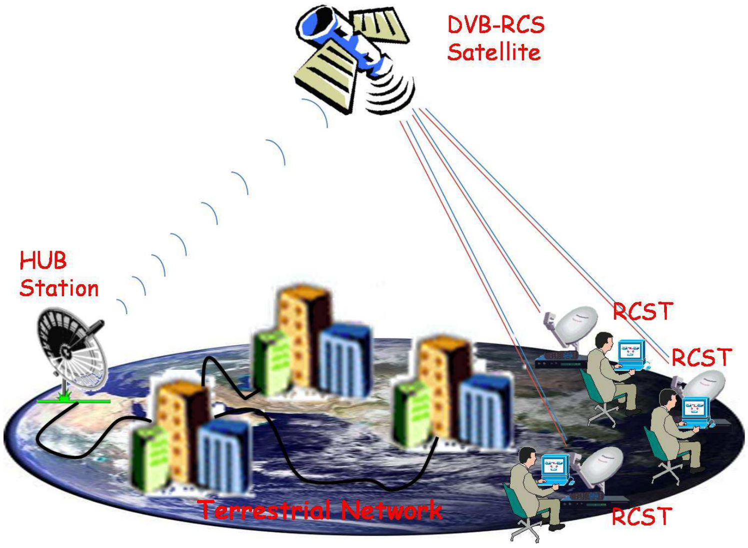

Fig. 4.2. DVB-RCS Applicative Scenario

The transmission capability uses a MF-TDMA scheme to share the capacity available for transmission by user terminals.

The mapping between source traffic type and capacity category depends on the types of provided service, on the used transmission protocols and on constraints imposed by the satellite orbit. Capacity categories supported by standard are listed below:

- Continuous Rate Assignment (CRA): it should be used for traffic that requires a fixed guaranteed rate;
- Rate Based Dynamic Capacity (RBDC): it should be used for variable rate traffic that can tolerate some delay;
- Volume Based Dynamic Capacity (VBDC): it should be used only for traffic that can tolerate delay jitter;
- Absolute Volume Based Dynamic Capacity (AVBDC): it is similar to VBDC and should be used instead of VBDC when the RCST senses that a VBDC request might be lost. This might happen when requests are sent on contention bursts or when the channel conditions (PER, Eb/N_0) are degraded. Traffic supported by AVBDC is similar to the VBDC one;
- Free Capacity Assignment (FCA): it is volume capacity which shall be assigned to RCSTs from capacity which would be otherwise unused.

The standard previews a frame structure of the duration of 47 ms and a possible use of IP packets carried via DVB/MPEG2 - Transport Stream (MPEG2-TS). A frame consists of a number of time slots on a certain number of carriers. The number and composition of time slots per frame is determined by the information bit rate to be supported by the frame. The frame structure consists of 188 bytes (4 of header and 184 of payload). For details about the frame composition see [5], [6].

The sum of allocated or requested capacity for any given RCST shall not exceed the maximum transmission capability of that RCST.

4.5 MPEG Traffic Modeling

MPEG compression has a variable bit rate and VBR traffic requires a variable bandwidth, which goes from some kbps to various Mbps. For this reason MPEG traffic on RBDC category has been mapped.

In previous work it has used a Gaussian distribution in order to characterize a MPEG traffic without considering the time dependence.

MPEG traffic sources are characterized by a variable GOP rate $r(t)$ that can change during time, so that $r(t) \in [r_{min}, r_{max}]$, where r_{min} and r_{max} are the minimum and maximum GOP rate requests. They are characterized by a Finite State Markov Chain (FSMC) that approximates the time behavior of the GOP rate. In particular, the source's GOP rate is divided into a fixed number l of discrete bandwidth levels $[r_1, r_2, ..., r_l]$. If a GOP's source requests a bandwidth $r(t)$ at a particular instant time t, and $r_{min} \leq r(t) \leq r_1$, the bandwidth request is associated with the state r_1. In the general case, if $r_{i-1} \leq r(t) \leq r_i$ with $i = 1, 2, ..., l$, ($r_0 = r_{min}$ and $r_l = r_{max}$), $r(t)$ is associated with the state i. It is possible to calculate the sojourn time ti in the l states and the transition probabilities p_{r_i, r_j} with $i, j = 1, 2, ..., l$ in order to know the steady state probabilities π_i with $i = 1, 2, ..., l$.

Through the approximation of the GOP rate's time dynamic, it is possible to get full advantage of statistical multiplexing. This can be useful to design a CAC, such as explained in section 4.6, where the probability of requesting a discrete bandwidth level i, the duration of the GOP rate and the GOP loss ratio associated with the state i can be considered in the admission phase. In Fig. 4.3 the aggregate GOP rate with and without bandwidth level discretization is shown.

In Fig. 4.4 the the Markovian approximation is shown: as it can be seen if a higher number of states is employed in the model the granularity increases, leading to a better approximation of the GOP-rate trend; as disadvantage, the computational complexity increases for the CAC scheme as will be shown in the next section.

It has been observed that for all states of considered streams, the sojourn time in a particular discrete bandwidth level is exponentially distributed for a single source as depicted in Fig. 4.5.

The stationary probability has been calculated such as shown in this section. When the aggregate GOP rate is considered, it is possible to multiplex the traffic sources considering the property of the discrete bandwidth levels. In the following a clearer idea is given.

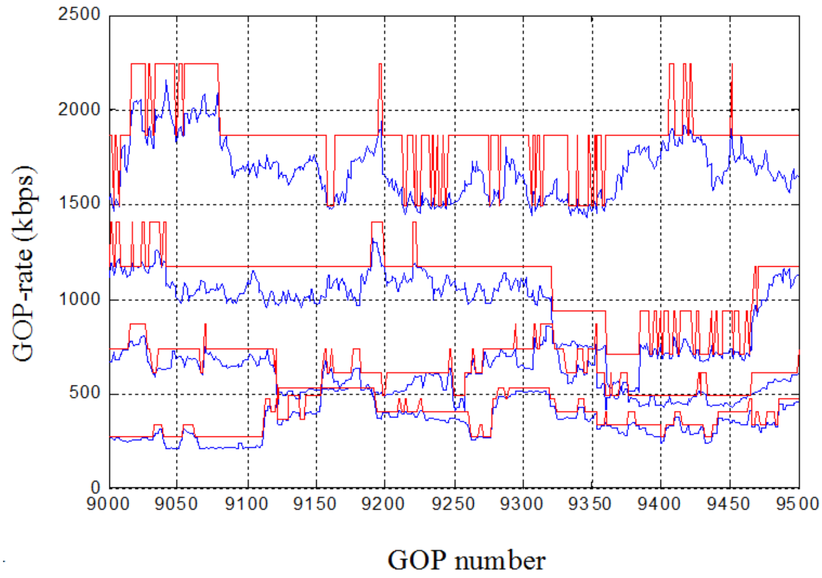

Fig. 4.3. Aggregate GOP rate with 1, 2, 3 and 4 traffic sources with 4 discrete bandwidth levels

4.6 CAC Algorithms for MPEG Traffic Sources

CAC represents a module of the NCC disposed on a terrestrial station [36]. Its task is to regulate the access to the satellite segment. It permits a flexible handling of the bandwidth and it avoids the a priori partitioning of resources among different types of service. The resource allocation algorithm is performed by DVB satellite to manage the source channel requests. The kind of service category considered to map MPEG traffic is RBDC.

In this work, the novel approach based on the discretized GOP rate has been compared with a previous work, based on the Normal distribution of aggregate GOP rate [58].

Before recalling the work in [58] and our proposal on the novel admission of multimedia MPEG traffic, some parameters applied in the math formulation of the problem are listed:

- B: aggregate instantaneous bandwidth;
- N: number of MPEG traffic sources;
- B_T: capacity assigned to VBR traffic;
- r_i: bandwidth assigned to $i-th$ MPEG traffic source;
- μ_i: average rate of the $i-th$ traffic source;
- σ_i: standard deviation around the average rate of the $i-th$ traffic source;
- $f_x(x)$: probability density function of the GOP rate associated to a specific MPEG traffic;

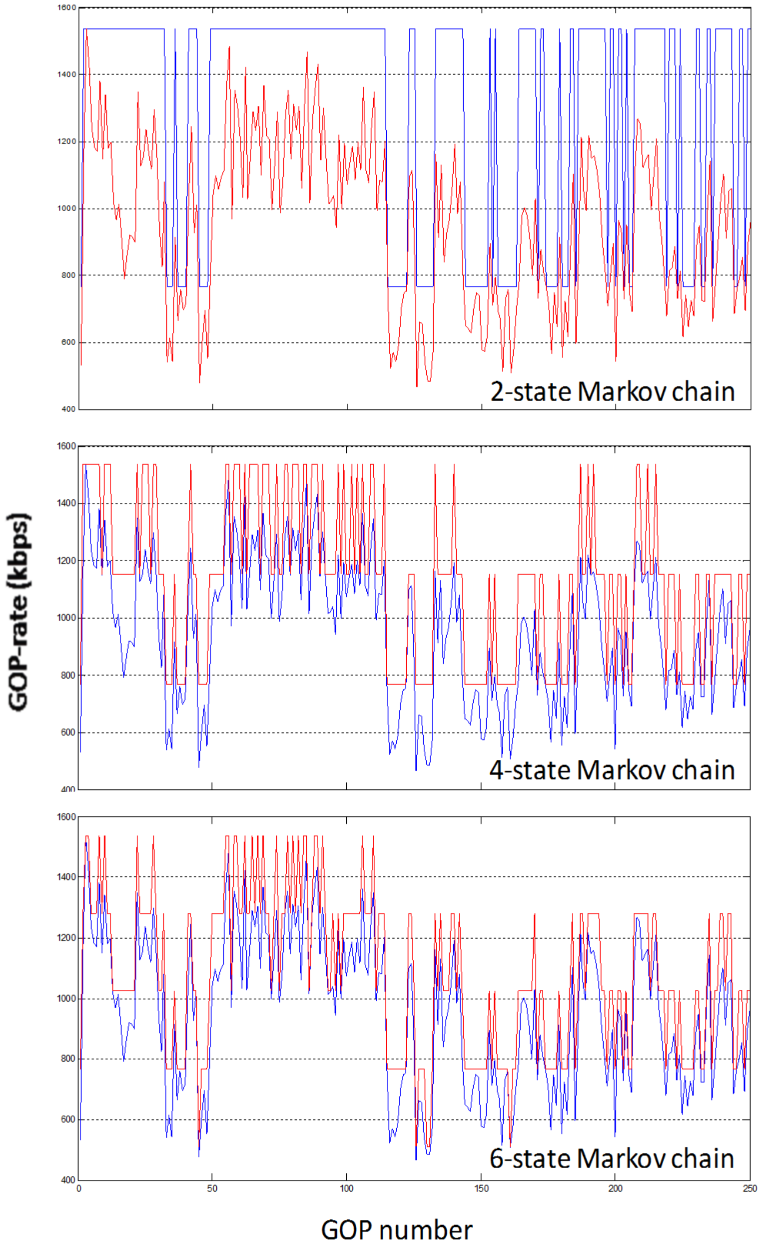

Fig. 4.4. GOP rate discretization of a specific movie with different bandwidth levels (2, 4, 6)

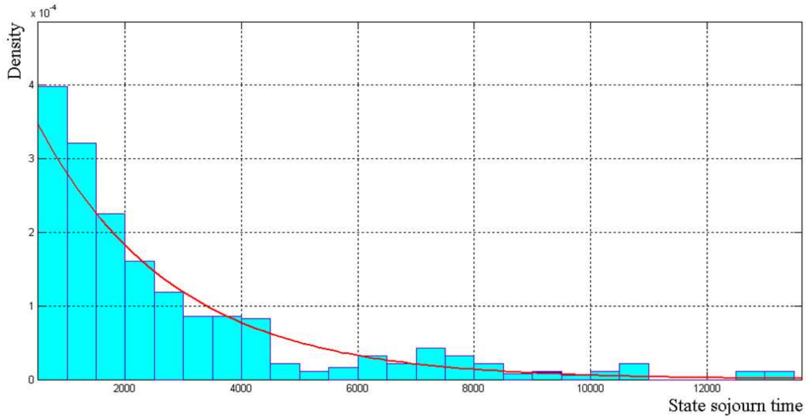

Fig. 4.5. Exponential approximation of the state sojourn time

- BEF: it is a multiplicative factor associated to GOP rate variance. More details about this term are given in the following;
- μ_{tot}: aggregate average rate associated to the aggregate traffic;
- σ_{tot}: aggregate standard deviation of the aggregate traffic. It is given by square root of the sum of single variances σ_i^2;
- $B(k)$: aggregate bandwidth when k calls are admitted;
- p_0: system outage probability. It expresses the probability to overcome the burst loss (GOP loss) ratio;
- γ: GOP loss ratio. It expresses the percentage of allowable lost GOPs;
- t_i: average sojourn time in the discretized $i-th$ GOP rate state;
- π_i: stationary probability associated to the discretized $i-th$ GOP rate state;
- p_i: probability to be in the discretized $i-th$ GOP rate;
- $\Delta_{i,j}(k)$: bandwidth gap (extra amount of bandwidth) associated with the combination of the aggregate state i of the k admitted calls and the new $(k+1)-th$ call;
- $\alpha_{i,j}$: GOP loss ratio of the aggregate traffic in the $i-th$ state with the novel call in the $j-th$ state;
- $t_i(k)$: average sojourn time of the $k-th$ call in the $i-th$ state:
- α_i: GOP loss percentage associated with the aggregate $i-th$ state:
- $P_i(k)$: probability of the aggregate traffic to be in the $i-th$ state;
- $T_{i,j}(k)$: simultaneous transmission of k calls in the aggregate state $i-th$ when the $(k+1)-th$ candidate call is in the $j-th$ state.

In the next two sections SMND and SMDB CAC algorithms will be explained.

4.6.1 Statistical Multiplexing based on the Normal GOP Distribution (SMND)

SMND takes advantage of some properties of multimedia MPEG traffic stream, but it does not account the time dependence of aggregated MPEG traffic stream. In particular, a BEF is defined for VBR traffic (e.g. MPEG) so that the aggregate instantaneous rate exceeding the fraction of the capacity of the VBR traffic will not be greater than a pre-specified threshold value γ:

$$P_r\{B > B_T\} = \int_{x=B_T}^{\infty} f_x(x)dx \leq \gamma \tag{4.1}$$

with $B = \sum_{i=1}^{N} R_i$ representing the aggregate instantaneous bandwidth rate and $B_T = BEF \cdot \sum_{i=1}^{N} \mu_i$.

Previous studies [58] have assumed that a generic multimedia MPEG traffic stream can be modeled by a statistical Normal Distribution of GOP which is characterized by a mean data rate value (μ) and a standard deviation value (σ); thus, considering independent multimedia traffic streams, the aggregate of N multimedia streams can be considered like a flow characterized by a GOP rate Normal Distribution with mean data rate as sum of the single mean data rates (μ_{tot}) and standard deviation (σ_{tot}) as the square root of the sum of single variances σ_i^2 [58], [59]:

$$\mu_{tot} = \sum_{i=1}^{N} \mu_i \tag{4.2}$$

and

$$\sigma_{tot} = \sqrt{\sum_{i=1}^{N} \sigma_i^2} \tag{4.3}$$

In general, it is enough simple to know in advance the mean GOP rate and the standard deviation of MPEG streams. For example, these parameters can be calculated during the MPEG video coding phase according to the desired video quality.

The MPEG traffic sources are statistically multiplexed so it is possible that total bandwidth request can exceed the available one. The excess demand probability (EDP) is defined as the probability that this event occurs. The objective of the CAC is maintaining the EDP below a fixed threshold (γ) value. Obviously, the fixed γ parameter coincides also with the maximum tolerated GLR. The total contribution $B(k)$ is related to parameters μ_{tot}, σ_{tot} and γ thus it is available using the inverse function of the Normal cumulative distribution F:

$$B(k) = F^{-1}(P|\mu_{tot}, \sigma_{tot}) = \{B(k) : F(B(k)|\mu_{tot}, \sigma_{tot}) = (1 - \sigma)\} \tag{4.4}$$

The integral equation of the cumulative Normal distribution

$$F(B(k)|\mu_{tot}, \sigma_{tot}) = \frac{1}{\sigma_{tot} \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{B(k)} e^{\frac{-(x-\mu_{tot})^2}{2\sigma_{tot}^2}} dx \tag{4.5}$$

does not have a closed solution then the computation of the bandwidth value $B(k)$ should be carried out using printout values.

This estimate is an approximation of the aggregate rate, but it provides reasonably good results for moderate to large number of multiplexed streams [58], [59]. The accuracy of the approximation strongly depends on the value of predefined probability parameter γ, since this value determines the number of admitted request. Executing a change of variable it is possible to evaluate a probability enclosed inside the interval between a and b values:

$$\int_a^b \frac{1}{\sqrt{2\pi} \cdot \sigma_{tot}} \cdot e^{\frac{-(x-\mu_{tot})^2}{2\sigma_{tot}^2}} dx \tag{4.6}$$

$$\begin{gathered}\int_a^b f(t)dt = \int_{\frac{a-\mu_{tot}}{\sigma_{tot}}}^{\frac{b-\mu_{tot}}{\sigma_{tot}}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{Z^2}{2}} dz = \int_{Z_a}^{Z_b} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dz = \\ P(Z_a \leq Z \leq Z_b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})\end{gathered} \tag{4.7}$$

The Z term is called *Standard Normal Variable* and the probability function $Z \approx N(0,1)$ is called *Standard Normal Distribution.*

It is possible to observe that the standard normal distribution is a particular case of the normal distribution with null mean value and unitary standard deviation.

$$Z = \frac{2}{\sqrt{\pi}} \cdot \int_0^Z e^{-t^2} dt \tag{4.8}$$

Finally, when a new MPEG call with μ_i and σ_i parameters wants to be accepted, the new total multiplexed bandwidth contribution is the following:

$$B(k) = B(k-1) + \mu_i + BEF \cdot (\sigma_i) \tag{4.9}$$

where

$$B(k-1) = \mu_{tot} + BEF \cdot (\sigma_{tot}) \tag{4.10}$$

The term BEF represents the Z value on the x axis, corresponding to the area's value of the standard normal distribution equal to $(1-\gamma)$.

The BEF is a constant term directly obtainable by fixing the desired γ value; choosing γ equal to 1% the correspondent value of BEF is 2.33.

Once obtained the new value of $B(k)$, the admission condition eq. 4.1 must be recomputed. The dismiss procedure of a generic $i-th$ MPEG call the rule is:

$$B(k) = B(k-1) - \mu_i + BEF \cdot (\sigma_i) \tag{4.11}$$

Thus the outage probability p_o of the system can be determined as follows:

$$p_o = P_r\{B \geq B_T\} = \int_{B_T}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{tot}^2}} \cdot e^{\frac{-(x-\mu_{tot})^2}{2\cdot\sigma_{tot}^2}} dx \tag{4.12}$$

And the $(k+1)-th$ call is admitted if the eq. 4.9 is verified:

$$B(k+1) = B(k) + \mu_{k+1} + BEF \cdot (\sqrt{\sigma_{tot}^2(k) + \sigma_{k+1}^2}) < B_T \tag{4.13}$$

where $B(k+1)$ is the aggregate bandwidth accounting the $(k+1)-th$ admitted call, $B(k)$ is the aggregate bandwidth at the previous step and the BEF value is obtained by the table in [66] that guarantees that the p_o does not overcome the threshold γ ($p_o \leq \gamma$). For details about the SMND approach see [58].

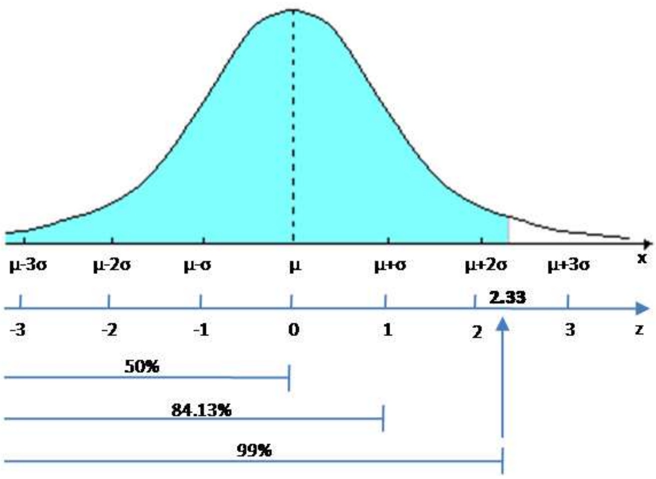

Fig. 4.6. Standard Normal Distribution

4.6.2 Statistical Multiplexing based on Discrete Bandwidth Levels of GOP Rate (SMDB)

SMND over-estimates the bandwidth for the aggregated traffic because it does not account for the time behavior of the MPEG source.

It degrades its performance when high variance for each item of video traffic is considered, as explained in section 4.7. In order to better use the time variation of MPEG traffic, a novel way to characterize the MPEG source is considered, as explained in section 4.5. A CAC, called SMDB, able to better manage the time characteristics of each video source is proposed.

SMDB is based on the characterization of MPEG sources, as presented in section 4.5. Thus, a call presents the l discrete bandwidth levels $\{r_1, r_2, ..., r_l\}$, the average bandwidth sojourn times $\{t_1, t_2, ..., t_l\}$ and the steady state probabilities $\{\pi_1, \pi_2, ..., \pi_l\}$ associated with the bandwidth levels of traffic sources.

The algorithm, through this knowledge, can calculate the GOP loss ratio of the aggregate MPEG traffic for each level and can perform statistical multiplexing as specified below.

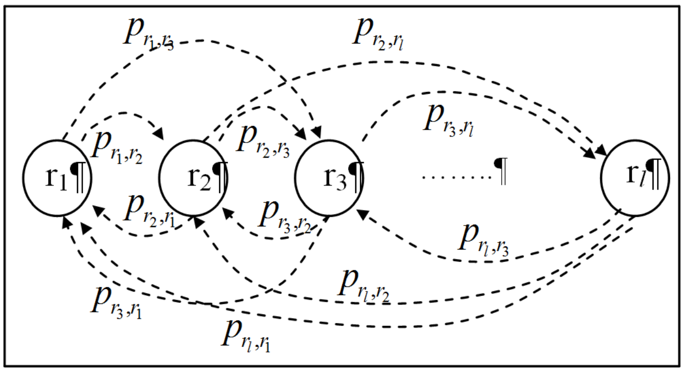

Fig. 4.7. FSMC of discrete bandwidth levels associated with the single traffic source

It is clear that, to guarantee users' QoS guaranteed service, certain admission control should be enforced to limit the number of users in the system. In this case, a greater multiplexing gain can be obtained, accounting for the time property of MPEG traffic streams associated with discrete bandwidth levels.

Given an FSMC's transition probabilities matrix A that approximates the MPEG source behavior, the steady state probability $\prod = [\pi_1, \pi_2, ..., \pi_l]$ can be calculated by solving the following vector equation:

$$A \prod = \prod \tag{4.14}$$

if state i's average sojourn time is t_i, then, at a particular instant time the probability of the GOP rate being around the discrete bandwidth level i is:

$$p_i = \frac{\pi_i \cdot t_i}{\sum_{i=1}^{l} \pi_i \cdot t_i} \tag{4.15}$$

The steady state probabilities associated with the discrete bandwidth levels of a call, together with other variables introduced below, permit to design an algorithm that is able to maintain aggregate states associated to the aggregate GOP rate and calculate new aggregate states when a new call is admitted. An idea of the proposed approach is depicted in Fig. 4.7 and Fig. 4.8.

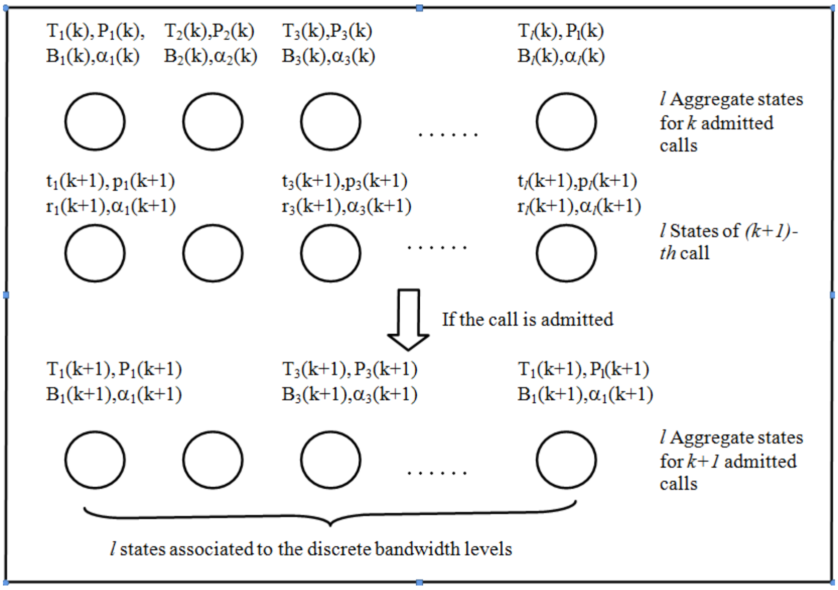

Fig. 4.8. Aggregate states of bandwidth levels of Aggregate GOP rate

The way to calculate bandwidth levels associated to the aggregate GOP rate is now explained.

Recall that when users' bandwidth levels are on a particular state i in the same time for an amount of time t, the GOP loss percentage associated with the state i (if it exists) is evaluated as follows. Define the bandwidth gap $\Delta_{i,j}$ associated with the combination of the aggregate state i of the k admitted calls and the new $(k+1)-th$ arrived call as the extra amount requested bandwidth:

$$\Delta_{i,j} = B_i(k) + r_j(k+1) - B_T \tag{4.16}$$

where $B_i(k)$ is the aggregate bandwidth reservation associated to the $i-th$ state, $r_j(k+1)$ is the GOP rate of the $(k+1)-th$ call associated to the state j and B_T is the total bandwidth assigned for the MPEG sources management. This excess demand bandwidth is accounted if $\Delta_{i,j} > 0$ otherwise a $\Delta_{i,j} = 0$ is considered. To calculate the excess bandwidth request associated with the state i, all possible combinations between new call j and the previous aggregated bandwidth reservation in the state i need to be performed. The GOP loss ratio associated with the aggregated traffic in the state i and the novel call in the state j can be represented as

$$\alpha_{i,j}(k+1) = \frac{\Delta_{i,j} \cdot min(T_i(k), t_j(k+1))}{B_i(k) \cdot T_i(k) + r_j(k+1) \cdot t_j(k+1)} \tag{4.17}$$

where $r_j(k+1)$ is the $j-th$ discrete level of the $(k+1)-th$ call, $t_i(k+1)$ is the average sojourn time of the $(k+1)-th$ call in the state j and $T_i(k)$ is the time in which all k sources will transmit in the same time. $T_i(k)$ can be calculated as follows:

$$T_i(k) = max(min\{T_i(k-1), t_j(k)\}_{j=1}^{l}) \tag{4.18}$$

with $t_j(k)$ representing the average sojourn time of the $k-th$ call associated to the state j. This calculus guarantees an overestimation of the $T_i(k)$ time.

To calculate the GOP loss ratio associated with the state i of the aggregated traffic when the $(k+1)-th$ call is admitted the following term should be calculated:

$$\begin{aligned}\alpha_i(k+1) =& \alpha_{i,1}(k+1) \cdot P_i(k) \cdot p_1(k+1) + \alpha_{i,2}(k+1) \cdot P_i(k) \cdot p_2(k+1) + ...\\ &+ \alpha_{i,j}(k+1) \cdot P_i(k) \cdot p_l(k+1) + \Theta_i(k)\end{aligned} \tag{4.19}$$

where $P_i(k)$ is the probability of the aggregated traffic to be in the state i; $\alpha_i(k+1)$ represents the GOP loss percentage associated with the aggregate state i when $k+1$ calls are managed in the satellite system, and $p_j(k+1)$ is the probability of the $(k+1)-th$ call to be in the state j. $\Theta_i(k)$ is expressed as:

$$\Theta_i(k) = \alpha_i(k) \cdot P_i(k) \tag{4.20}$$

where $\Theta_i(k)$ is the contribution given by the GOP loss ratio associated with the aggregate state i without accounting the further GOP loss contribution given by the $(k+1)-th$ call. The GOP loss percentage of the aggregate state i when $k+1$ calls are admitted is summarized as follows:

$$\alpha_i(k+1) = P_i(k) \cdot [\sum_{j=1}^{l} \alpha_{i,j}(k+1) \cdot p_j(k+1) + \alpha_i(k)] \tag{4.21}$$

However, a further observation needs to be done. Because the call can stay in a particular state i for a fraction of the overall transmission time, the outage probability p_o of the system can be calculated as follows.

The system is considered to be in outage and QoS constraints are not respected if the probability of losing GOP frames overcomes a fixed threshold γ. In particular, imaging to manage k active calls in the system, if the $(k+1)-th$ call arrives, it can be admitted if:

$$p_o = \sum_{i=1}^{l} \alpha_i(k+1) \cdot P_i(k+1) \leq \gamma \tag{4.22}$$

At this point, it is important to calculate the probability to be in the state i of the aggregated traffic when a new call is admitted. The following equation can be applied:

$$P_i(k+1) = \frac{\sum_{i=1}^{l} P_i(k) \cdot p_j(k+1) \cdot T_{ij}(k)}{\sum_{i=1}^{l} \sum_{j=1}^{l} P_i(k) \cdot p_j(k+1) \cdot T_{ij}(k)} \tag{4.23}$$

where:

$$T_{ij}(k) = min(T_i(k), t_j(k+1)) \tag{4.24}$$

If a source terminates its transmission, the index associated with the aggregated traffic needs to be recomputed. In particular, the aggregated probability $P_i(k)$, the GOP loss percentage $\alpha_i(k)$ associated to the state i and the minimum time $T_i(k)$ during which the sources can simultaneously transmit can be calculated as follows.

The minimum time period $T_i(k)$ when the $(k+1)-th$ call leaves the system is easily calculated, re-computing the minimum average interval time in which remaining calls can simultaneously transmit such as expressed in eq. 4.18. The GOP loss percentage $\alpha_i(k)$ associated with the state i can be calculated as follows:

$$\alpha_i(k) = \frac{\alpha_i(k+1) - P_i(k) \cdot \sum_{j=1}^{l} \alpha_{i,j}(k+1) \cdot p_j(k+1)}{P_i(k)} \tag{4.25}$$

The SMDB algorithm needs to store only the aggregate GOP loss percentage $\alpha_i(k)$, the aggregate Probability $P_i(k)$ of being in the aggregate state i, the aggregate bandwidth $B_i(k)$ associated with the bandwidth level i, parameters of single sources such as steady state probability π_i, discrete bandwidth sojourn times t_i and bandwidth levels b_i.

4.7 Performance Evaluation

Many simulation campaigns have been assessed to test the performance of SMDB algorithm.

Table 4.1 resumes the main characteristics of the considered streams; in particular "THE GLADIATOR", "KING KONG", "THE FANTASTIC FOUR" and "MADAGASCAR" (Fig. 4.9) have been encoded from the original Digital Versatile Disc (DVD)s at a resolution of 224x168 pixels with different bitrates, as explained later; the second column of the table illustrates the transition probabilities matrices of the 3-state Markovian models associated to each stream; in the third column the means of the exponential distributed states sojourn times are illustrated, while in the last column the steady states probabilities are evaluated.

Fig. 4.9. Four frames of the considered streams, with the mean and the standard deviation of the GOP-rate

Evaluated simulation indexes are:

- GOP Loss Ratio (GLR): it is the percentage of lost GOPs of a traffic source. A GOP is considered lost if its bandwidth request on the TRM module cannot be satisfied in a frame period (47ms);
- Admitted Calls: number of accepted calls in the overall satellite system;
- Satellite Utilization: it is the ratio between real transmitted traffic over satellite channel and potentially transmittable traffic.

In Table 4.2 the adopted simulation parameters in the DVB-RCS platform have been considered.

4.7.1 Simulation Scenario

In order to analyze the SMND limitations and the improvements of SMDB when standard deviation increases in the traffic video the scenario represented in Table 4.3 is applied. Two different MPEG sources with different average GOP rate are considered and for each source an encoding with increasing standard deviation σ in the range of [10%,30%] around its average μ is considered.

Table 4.1. Markov Chain Parameters for 4 Movies (3 State Approximation)

	S	Transition probabilities matrix			State sojourn times (ms)	Steady State Probabilities
The Gladiator						
	1	0.0938	0.9062	0.0000	615.2	0.0178835
	2	0.0182	0.8522	0.1296	3691.9	0.890439
	3	0.0000	0.2803	0.7197	1756.3	0.0916779
King Kong						
	1	0.0000	1.0000	0.0000	600	0.0029016
	2	0.0037	0.6777	0.3186	1552	0.78421
	3	0.0000	0.1773	0.8227	2973.1	0.212889
Fantastic Four						
	1	0.0000	0.9523	0.0477	600	0.0405763
	2	0.0617	0.0000	0.9383	1601.9	0.640585
	3	0.0033	0.9967	0.0000	2186.8	0.318839
Madagascar						
	1	0.0000	1.0000	0.0000	600	0.0019103
	2	0.0025	0.6350	0.3625	1419.9	0.764123
	3	0.0000	0.1864	0.8133	2809.7	0.233967

Table 4.2. DVB-RCS Simulation Parameters

Parameters	**Value**
Round Trip Time	540 ms
Return Channel's Slot	1000
Forward Channel's Slot	4000
Satellite Atomic Channel	32 kbps
Return and Forward Channel's frame period	47 ms
GOP Loss Probability (RBDC sources)	0.01
RCST Typology	D (2048 kbps)
Number of RCST	2, 4, 6, 8
Max Number of Sources for RCST	16
Number of Sources	32, 64, 96, 128
Bandwidth Expansion Factor (BEF)	2.33

The frame rate used during the coding phase is 30 frame per second (fps) and the GOP structure is IBBPBBPBBPBBPBB with a resolution of 224x168.

Table 4.3. Simulated Scenario

	Class	Traffic Profile		
Video 1	RBDC	144	14.4	10%
		144	28.8	20%
		144	43.2	30%
Video 2	RBDC	300	30	10%
		300	60	20%
		300	90	30%

For these traffic sources a light (0,6 calls/min) and heavy traffic (4 calls/min.) conditions are applied in order to verify the benefits of SMDB in many situations. Four number of sources have been evaluated on the RCS channel and a bandwidth corresponding to the number of slot necessary to admit these sources according with the SMND algorithm have been considered. For example on RCS channel 334 slot with 10% of standard deviation around an average GOP rate of 300kbps have been considered to admit 32 sources.

A number of slot equal to the slots given to the SMND algorithm has been applied for the SMDB algorithm. This permitted to evaluate the benefits offered by the SMDB algorithm in comparison with SMND.

4.7.2 Light Traffic Load

In the table of Fig. 4.10 the satellite utilization for a traffic source of 300 kbps varying the source between 32 and 128 is presented. In tables of Fig. 4.11 and Fig. 4.12 the number of admitted calls and the GOP loss percentage of SMND and SMDB algorithms are presented. SMDB outperforms SMND in light traffic condition because it can admit more calls preserving the prefixed GOP loss percentage of 1

However the improvements of SMDB are lower than benefits in heavy traffic load scenario such shown in the next subsection.

4.7.3 Heavy Traffic Load

If the heavy traffic scenario with a call arrival rate of 4calls/min is considered, the improvements of SMND are also perceptible for variance of 20% and 30% and they can be around the 15-20% for the satellite utilization such as depicted in table of Fig. 4.13. This scenario permits to verify the reduction of call block

μ Kbps	σ Kbps	Number of Sources				CAC
		32	64	96	128	
300	30	86.012	87.138	87.715	87.961	SMND
300	30	88.648	89.547	90.387	91.677	SMDB
300	60	80.928	82.859	83.686	84.207	SMND
300	60	85.962	87.882	89.803	92.684	SMDB
300	90	70.661	73.900	75.433	76.532	SMND
300	90	83.534	85.992	88.449	92.134	SMDB

Fig. 4.10. Table of Satellite Utilization (%) in Light Traffic Load (0.6 calls/min)

μ Kbps	σ Kbps	Number of Sources				CAC
		32	64	96	128	
300	30	175.500	349.750	525.750	702.750	SMND
300	30	182.750	365.875	549.000	823.688	SMDB
300	60	175.000	349.750	523.750	697.750	SMND
300	60	185.500	374.625	563.750	847.438	SMDB
300	90	162.250	323.750	488.250	653.250	SMND
300	90	193.750	386.875	580.000	869.688	SMDB

Fig. 4.11. Table of Average Number of Admitted Calls Light Traffic Load (0.6 calls/min)

μ Kbps	σ Kbps	Number of Sources				CAC
		32	64	96	128	
300	30	0.001	0.000	0.000	0.000	SMND
300	30	0.375	0.401	0.427	0.465	SMDB
300	60	0.000	0.000	0.000	0.000	SMND
300	60	0.327	0.465	0.603	0.810	SMDB
300	90	0.001	0.000	0.000	0.000	SMND
300	90	0.742	0.826	0.909	1.034	SMDB

Fig. 4.12. Table of GOP Loss Ratio Light Traffic Load (0.6 calls/min)

probability of SMND through the admission of a greater number of calls than SMDB CAC such as illustrated in table of Fig. 4.14. If the standard deviation of traffic sources is little, the improvements of SMND in comparison with SMDB are low perceptible. This is due to the reduced multiplexing capability of the traffic sources. The GLR is below the threshold value of 1% as depicted in table of Fig. 4.15.

4.8 Conclusions

This chapter introduces a important problem for the service provider that have to satisfy the stringent requirements of the users that want always more

μ Kbps	σ Kbps	Number of Sources 32	64	96	128	CAC
144	14.4	74.856	78.226	78.288	78.604	SMND
144	14.4	83.596	84.489	85.578	86.803	SMDB
144	28.8	77.513	79.675	80.818	81.277	SMND
144	28.8	83.344	83.791	84.934	86.374	SMDB
144	43.2	72.768	75.734	77.403	78.403	SMND
144	43.2	83.417	86.213	87.788	89.255	SMDB
300	30	68.922	69.978	70.174	70.761	SMND
300	30	79.145	79.553	81.438	82.232	SMDB
300	60	64.775	66.823	67.414	67.775	SMND
300	60	78.909	78.417	82.240	81.176	SMDB
300	90	63.897	65.602	66.593	67.273	SMND
300	90	79.785	82.031	84.157	84.114	SMDB

Fig. 4.13. Table of Satellite Utilization (%) in Heavy Traffic Load (4 calls/min)

μ Kbps	σ Kbps	Number of Sources 32	64	96	128	CAC
144	14.4	116.250	234.750	346.500	463.250	SMND
144	14.4	128.000	246.500	375.750	491.500	SMDB
144	28.8	118.000	234.250	353.750	474.500	SMND
144	28.8	126.750	247.000	374.250	498.000	SMDB
144	43.2	121.750	242.250	362.750	479.500	SMND
144	43.2	142.250	278.500	417.500	562.250	SMDB
300	30	141.750	284.250	424.000	567.000	SMND
300	30	164.250	325.500	495.750	664.250	SMDB
300	60	140.250	285.500	422.250	566.750	SMND
300	60	173.750	341.500	514.500	675.750	SMDB
300	90	146.250	287.500	435.500	572.000	SMND
300	90	183.000	355.500	541.750	731.250	SMDB

Fig. 4.14. Table of Average Number of Admitted Calls in Heavy Traffic Load (4 calls/min)

multimedia applications. The key aspect in order to be able to provide the services that are requested by customers is a well-working CAC.

For the satellite networks the CAC is a module that is placed in the NCC device and has the important role of deciding if it is possible or not admit a new call in the system without degrading the QoS of the already accepted call.

In literature a lot of studies on admission control algorithms exist but, at the best of our knowledge, few works have been done for satellite scenarios. Satellite access represents a reality that has the advantages of broad and continental coverage, and quick installation. The success of DTH television has proven that satellite services can compete with terrestrial alternatives in

μ Kbps	σ Kbps	Number of Sources 32	64	96	128	CAC
144	14.4	0.000	0.000	0.000	0.000	SMND
144	14.4	0.920	0.695	0.553	0.888	SMDB
144	28.8	0.000	0.000	0.000	0.000	SMND
144	28.8	0.172	0.025	0.056	0.076	SMDB
144	43.2	0.000	0.000	0.000	0.000	SMND
144	43.2	0.149	0.163	0.136	0.632	SMDB
300	30	0.000	0.000	0.000	0.000	SMND
300	30	0.404	0.916	0.016	0.649	SMDB
300	60	0.000	0.000	0.000	0.000	SMND
300	60	0.115	1.015	0.824	0.003	SMDB
300	120	0.000	0.000	0.000	0.000	SMND
300	120	0.520	0.633	0.238	0.257	SMDB

Fig. 4.15. Table of GOP Loss Ratio in Heavy Traffic Load (4 calls/min)

certain markets or can be integrated with terrestrial backbone for providing scalable end-to-end QoS services. The development of the DVB-RCS standard, recently accepted by ETSI, is the first attempt at introducing a wide-scale satellite access standard.

Differently by previous approach where a low multiplexing gain for video sources with high standard deviation around the average GOP rate and heavy traffic load is obtained, a novel CAC is proposed and it is based on the multiplexing of discrete bandwidth levels of MPEG sources. The video traffic is discretized in a finite number of bandwidth levels and it is characterized by a FSMC that tries to model the time characteristics of traffic sources

A lot of simulation campaigns are lead out in order to verify that the proposed approach is able to perform in a good way and is able of increasing the system utilization and the numebr of admitted calls without losing the QoS required.

A high multiplexing gain is obtained through the FSMC characterization of discrete bandwidth levels associated with MPEG traffic sources. The proposed solution is compared to a well-known mechanism based on the Normal distribution of the GOP rate, called SMND. SMDB outperforms SMND in all situations where the standard deviation of video is high and in heavy system traffic load conditions. A higher system utilization in the DVB-RCS platform is obtained, respecting the QoS constraints applied on the system.

5

Mobile Satellite System: MAC and scheduling design

DVB-RCS is today an open standard that offers a set of possibility for a lot of multimedia applications that make it more preferred in competition with proprietary system because of its maturity and capability. The US Department of Defense (DoD) has mandated all new deployments of star IP based satellite system be based on DVB-RCS due to the technical merits. The satellite broadband service industry is seen to have enormous potential growth, especially considering the large number of users that still lack access to wireline broadband services. The consumers are able to connect to the services thanks to VSATs that permit to use all the potentiality of a two way satellite broadband system [67]. It consists of two units - outdoor, that comprises an antenna and a radio frequency transceiver, and indoor that permits to the users to connect through a modem to an end user device.

Then, thanks to a VSAT technology the users are able to exploit all the applications served by the DVB-RCS satellite system like, Voice over IP (VoIP), access to internet access in rural areas, tele-medicine, tele-education, tele-government, as well as more conventional and generic Internet access services like email, web browsing etc.

The main task of the DVB-RCS is establishing a cost-effective VSAT solution and opening opportunities for network providers in order to deploy multimedia services and avoiding proprietary solutions controlled only by single entity. Europe is currently leading the way in Research & Development (R&D) and commercialization with ESA in order to standardize the DVB-RCS guideline. In 2007 there were more than 100 DVB-RCS deployed system capable of provide services tens of thousands of terminals [68].

Broadband Internet access is a reality for consumers both at home and at work that use a lot of applications that can require a certain QoS. Therefore, a large variety of QoS is expected to be served by satellite system and in order to compete with terrestrial broadband systems the user QoS shall be similar to the one experienced in fixed environment. The new challenge for the DVB-RCS group is to provide broadband services for another class of satellite users that is always more in growth, they are the mobile users.

Today a lot of demands for mobile interactive services using satellite exist. Many projects [69], [70], [71], [72], [73] have been conduced during the last years in order to find good guidelines for an efficient implementation of broadband services to mobile users. These reports, initiated by DVB TM-RCS ad hoc group, show the feasibility and requirements to extend DVB-RCS in order to incorporate mobility issues.

Moreover, these studies have shown the main applications for the markets of DVB-RCS mobile are those related to public transportation like cruise ships and ferries for the maritime domain, buses and trains for the terrestrial domain and planes for the aeronautical domain [68], [74], [75]. Furthermore, a small portion of users lies in the private transportation. In this new type of market it is important to consider what are the most used hours in order to receive the services for determining the users access to the networks and the data rates required in the time. As well as it is important to know, for example in the case of train transportation, the number of passengers adopting broadband that varies across Europe. The analysis has expressed the estimated number of passengers per train that the system should be capable of handling. The system should as a minimum be dimensioned to cope with the number of passengers specified in the table of Fig. 5.1 [73].

	High-Speed	Intercity	Regional*
Austria	0	47	16
Belgium	85	41	29
Cyprus	0	0	0
Czech Rep.	70	65	19
Denmark	0	44	43
Estonia	0	0	63
Finland	62	40	37
France	49	33	17
Germany	80	62	27
Greece	0	18	15
Hungary	0	11	5
Ireland	0	35	25
Italy	75	42	14
Latvia	0	0	55
Lithuania	0	0	51
Luxembourg	0	40	0
Malta	0	0	0
Netherlands	53	36	21
Poland	0	21	19
Portugal	55	35	44
Slovakia	0	0	12
Slovenia	26	0	24
Spain	92	70	60
Sweden	0	63	45
Switzerland	76	51	20
UK	118	67	58

Fig. 5.1. Maximum Numbers of Active Passengers per Country and per Train Type

Furthermore, in mobile application the antenna component in the terminal become an important aspect because the antenna size could be lower for this market. In the following section the critical issues for this new type of market are explained in order to understand what are the main aspects that the research are now studying [73].

5.1 Current Issue in the mobility support on Satellite

The reference scenario for the mobile DVB-RCS architecture is largely based on classical DVB-RCS system for fixed application. Then, more components present in mobile environment are the same that compose the classical satellite system. In the following figure (Fig. 5.2) the satellite scenario, that shows the three transportation domains served by the satellite system, is depicted [76], [77]. Normally, the recent studies report a satellite system with a forward link based on DVB-S2 standard [10], [78] in order to take advantages of the new proposed characteristics in respect to DVB-S one [68], [79], [80].

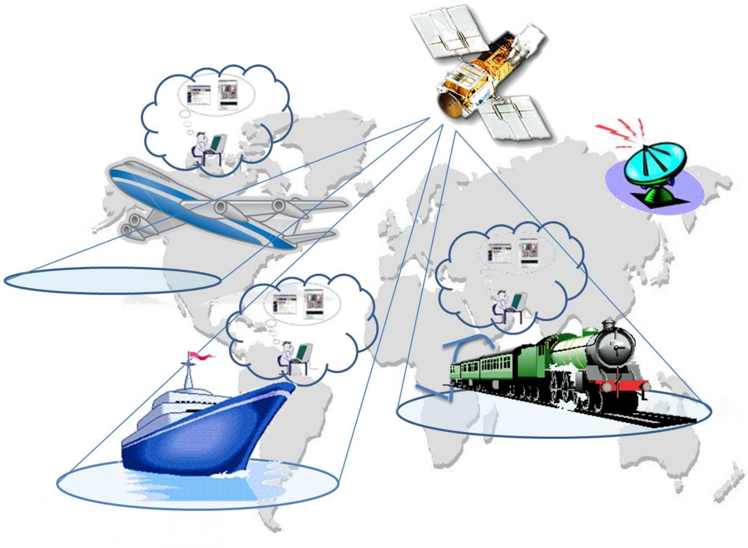

Fig. 5.2. DVB-RCS Mobile Scenario

There are several critical issues that have to be resolved in order to make the satellite system able to provide the required services to the mobile users. These issues relate all the layers of the protocol stack but the more important lie in the first three layers, physical, link and network layer. In the following of this thesis the MAC issues aspects will be detailed in order to study some aspects about MAC strategies for the selection of the most adequate MAC scheme for the Return Link (RL) of DVB-RCS+M for achieving the best trade-off between system efficiency and provided QoS [81].

Many works exist in order to minimizing the impact of the mobility on the existing standard and to achieve the maximum backward compatibility. The fundamental issues regard the use of smaller antenna in the mobile environment that conduced to the use of techniques capable of performing the spreading of the signal received by the terminals. Another important issue regards the mobile propagation channel because of the presence in the mobile environment of the Non Line of Sight (NLoS) propagation condition (e.g. railway scenario). Handover management issue is very important in order to allow a continuity of service to the mobile users.

5.1.1 Spectrum Spreading in the satellite link

The first technical issue comes from the need to introduce spectrum spreading that is a fundamental aspect when operating with small aperture antenna. The possibility to have small antenna is an important requisite for using of mobile vehicles able to host the antenna for being connected with satellite environment. In particular, in the aeronautical case where antenna size are more constrained.

Moreover, the use of small antenna complies with the stringent interference regulations that have the task to protect adjacent satellite systems. A possible interference scenario is depicted in Fig. 5.3, where both uplink and downlink interference from external systems are received. Both forward and return link transmission can be spread in bandwidth using techniques described in DVB-S2 standard [82]. The spectrum spreading can be achieved by two means either using $\pi/2$ -Binary Phase Shift Keying (BPSK) modulation, equivalent to spreading a Quadrature Phase Shift Keying (QPSK) modulated signal by a factor 2, or by using burst repetition technique [82].

A key role for resolving this issue is covered by the module that performs Control and Monitoring Functions (CMF) for mobile terminals that as primary purpose has the task of ensuring that there is no harmful interference to other mobile terminals or other services. The resulting interference scenario can be grouped in two general categories: interference to fixed services and interference to other services, including terrestrial fixed services and specialized scientific services.

In order to mitigate the occurrence of harmful interferences it is important to use interference mitigation techniques in order to reduce the Equivalent Isotropically Radiated Power (EIRP) density. A possible approach to avoid interference is to define a reference contour and adjust the size of this contour according to the actual EIRP density [82]. For the International Communication Union - Radiocommunication Sector (ITU-R) MSS a number of frequency allocations are assigned in the Ku and Ka bands for broadband mobile communications as summarized in the table of Fig. 5.4 [70].

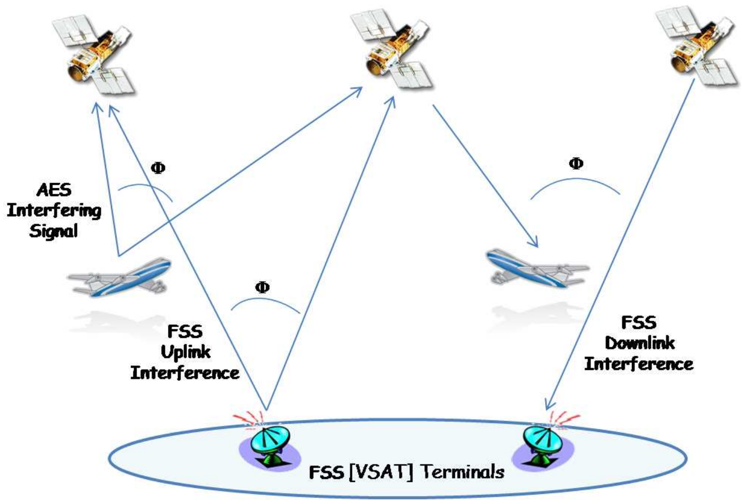

Fig. 5.3. Interference scenario

Mobile Application	K_u band	K_a band
Aeronautical	ETSI: EN 302 186 ITU-R: M.1643 ITU-R: S.728-1	ETSI: EN 301 358, EN 301 459 ITU-R: S.728-1, S.524-8
Maritime	ETSI: EN 301 427 ETSI: EN 302 340 ITU-R Resolution 902, S.728-1	
Land	ETSI: EN 301 427 ITU-R S.728-1	
High speed trains	ETSI : EN 302 448	

Fig. 5.4. ITU-R Mobile Satellite Services (MSS) Frequency Allocation

5.1.2 Mobile Propagation Channel

For most of the time the mobile terminals of DVB-RCS environment are in a Line of Sight (LoS) condition in which the channel propagation characteristics are well known and no particular problems are present for the receiver terminals (Ricean channel, Doppler effects) [68], [83].

The condition that puts the satellite terminals in a bad channel propagation environment due to some blocks or some interruption of the connection is the NLoS scenario. The main domain for the mobile users are aeronautical, maritime and terrestrial. In the case of aeronautical and maritime the satellite terminals are overall time in a LoS condition and how it is possible to view in a lot of study on the channel modeling the channel can be fairly approximated by a purely Additive White Gaussian Noise (AWGN) channel [70]. The NLoS problem is present in the land domain and in particular in scenarios with trains that have to cross tunnel in order to reach their destinations [73], [84].

Most of the studies consider only the regional coverage scenario since trains remain usually within one continent. The mobility effects, such as multipath, shadowing and blockage, encountered due to the local environment in the vicinity of the mobile RCSTs, such as adjacent buildings, vegetation, bridges, and tunnels, result in NLoS conditions of severe fading. Regarding the modeling of the NLoS channel conditions encountered, generally, in the land mobile satellite scenario, different propagation measurements at Ku and Ka bands [85], [86], [87] were performed in the last decade, based on which, reference statistical channel models exist. In recent studies it has been found that both in the case of ku and ka band the behavior of the land mobile satellite channel can be modeled using a 3 state markov channel model [74].

In conclusion, the studies have shown that, also in the case of railway scenario, it is possible to consider that the mobile terminals are in a LoS condition for most of the time and only few blockages due to the power arches present in the railway or some time long blockages due to long tunnel can put the trains in a NLoS condition [88] (see Fig. 5.5).

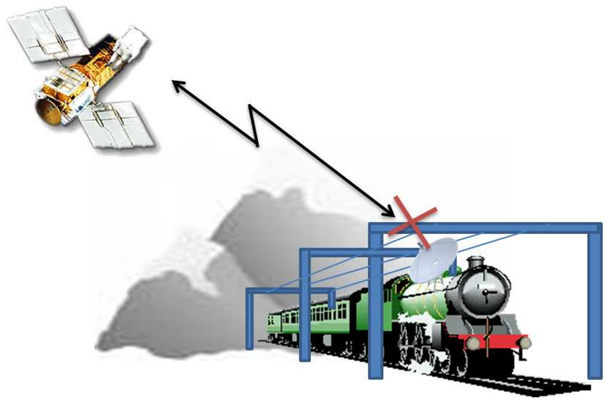

Fig. 5.5. NLOS condition in a railway environment

As the train moves along the railroad, nearly space-periodic fading events are experienced whenever the LoS between the train antenna and the satellite is shadowed by various obstacles like electric trellis bridges or overhead cable posts with bracket. This can induce peculiar fading occurrence that affects the system performance remarkably as the deep fading due to periodic fade occurrences with the rate proportional to the train speed and the power arch spacing is superimposed on the statistical mobile satellite channel factor. In fact, when the train approaches the power arches, knife-edge diffraction affects the transmitted signal producing a space-varying attenuation that depends on the structure [85], [86].

Another problem linked with the wireless channel for the mobile terminal is the logon phase. This problem can happen when the delay between terminal and satellite is not known with accuracy, and can result in a not successful logon procedure. In order to identify in an accuracy manner the position of the Common Signaling Channel burst (CSC) for performing the procedure the NCC and the terminal can exchange some correction messages [7].

5.1.3 Handover Management

The mobility aspect also introduces another important issue for the management of the mobile DVB-RCS system [80]. In fact, as the mobile terminal moves in the coverage area of the satellite network it can have the necessity of connecting under another transponder, beam, gateway or satellite.

In Fig. 5.6 is depicted a possible handover scenario. Therefore, mobile DVB-RCS system has to take into account the handover mechanisms and it has to be capable of maintaining active connections between satellite and mobile terminals in a manner completely transparent to the end-users. The handover type more important is that between beams that is required in a multi-beams platform. Another type of handover exists and it is known as Gap-Filler handover and it takes place in NLoS environment like a railway scenario when, for example, a train has to enter in a tunnel and then it loses the LoS with the satellite to which it is connected [89].

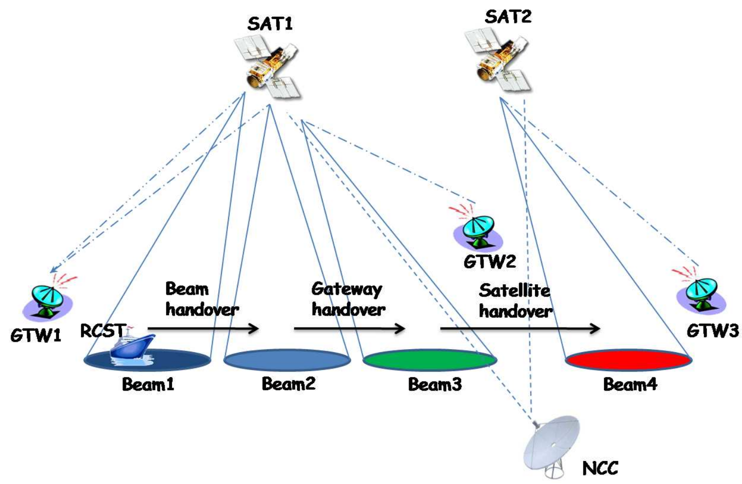

Fig. 5.6. Handover Scenario

Beam handover is the preferred technique for mobility management in the case of LoS environment specific to aeronautical applications and also applicable in some cases to maritime applications and even to land applications.

Nevertheless, it can also be used in the case of NLoS environment. However, in NLoS environment mobility management would typically also include countermeasure techniques intended to combat the mobility channel impairments, while the RCST remains within the same beam.

Beam handover involves the control and management of RCST handover from one beam to another one, while trying to preserve service continuity to the end-user. The handover process typically is composed of some steps that include handover detection, handover decision and handover execution. During the beam handover typically an RCST terminal is supposed that it remains attached to the same GTW. During the handover time the terminal has to be able to changing the carrier switching from one to another one. The handover process takes place with a large exchange of information between NCC and RCST terminals in order to check some information like what type of handover applies on forward link or on return link or on both, the availability of resource in the list of ranked beams candidates and if the beams are served by multiple transponder select the transponder. Normally in the handover procedure the SYNChronization (SYNC) burst and the Terminal Information Message (TIM) are used and exchanged between NCC and RCSTs [7].

Terrestrial Gap-Fillers are required in many situations where the propagation conditions are simply too severe for satellite signals, see Fig. 5.7. Integral Gap-Fillers can be conceptually relatively straightforward, consisting essentially of radio repeaters that re-broadcast the forward link signals from the satellite and relay return link signals to the satellite. For both regulatory and technical reasons, they will sometimes use frequencies that are different from those used for the satellite link. It may therefore be necessary for the mobile terminal to have extra transmission and reception equipment in order to make use of the Gap-Fillers [70], [90].

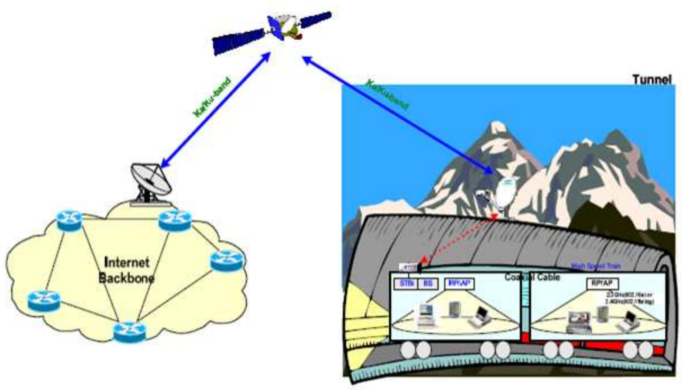

Fig. 5.7. Terrestrial Gap Filler Scenario

5.1.4 Single Channel Per Carrier (SCPC) on the reverse link

Another important feature for satellite broadband mobility concerning the SCPC in the reverse link. For a multi-passenger mobile vehicle such as trains, commercial jets and cruise ships the likelihood of traffic aggregation resulting in a fairly constant average activity of return path transmissions is great. If the resulting peak to average ratio of the return carrier is close to 1, the carrier will be fully loaded for most of the time.

In this scenario SCPC is more efficient than TDMA by avoiding burst transmission overheads and hence lower demodulation threshold. The use of a DVB-S2 format for the return path is suggested by standardized and moreover, in this case the standard suggests of reducing the DVB-S2 frame size to 4 kilo blocks. The standard proposes two modality in which a RCST terminal can be, one is called Basic RCST mode and the other one is called Enhanced RCST mode [7].

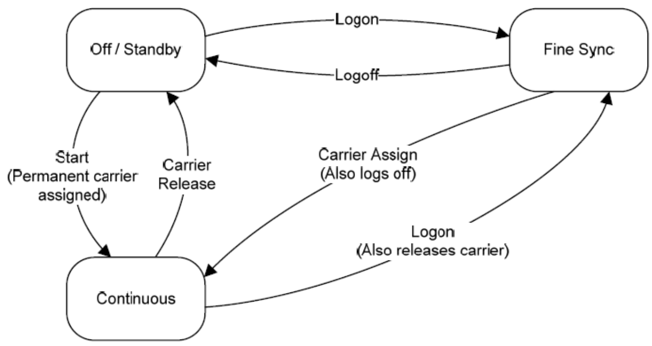

Fig. 5.8. State diagram for basic continuous carrier operation

The ability to operate in this manner shall be signaled in the CSC burst. An RCST declaring support for basic continuous carrier operation shall be capable of transmitting either a continuous carrier or an MF-TDMA signal, but will normally not be able to transmit both carriers simultaneously. An RCST declaring support for enhanced continuous carrier operation shall be able to transmit both types of signal simultaneously. Figures 5.8 and 5.9 show the RCST state diagram for both modalities.

The RCST can optionally employ a continuous carrier mode of transmission in accordance with [7].

5.2 Current and Future Projects

Broadband satellite communications can connect users located anywhere. The open standard DVB-RCS contributes greatly to increased connectivity and its

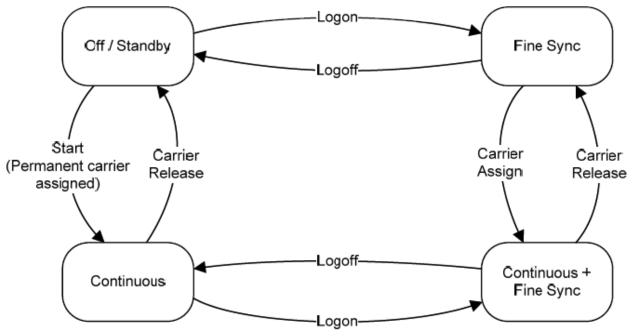

Fig. 5.9. State diagram for enhanced continuous carrier operation

rising popularity is making an impact in the market. There are a high number of projects which encourage adoption of the DVB-RCS standard, providing positive benefits for all and one of the most important space agency that works on these projects is the ESA. DVB-RCS is an open standard for two-way transmission of digital data that employs satellite transmission using Ku and Ka bands with return bandwidth up to 2 Mbps.

Interoperability is one of the main advantages of DVB-RCS. Until DVB-RCS came along, customers of two-way broadband access via satellite had no choice except to commit to propriety systems, with all its inherent inflexibility and higher cost. Interoperability gives customers the choice of purchasing from one or several vendors throughout the lifetime of their systems. The result will be improved competition among vendors, reduced costs for users and accelerated enhancement of DVB-RCS equipment. All these projects are conducted in order to perform detailed studies for proposing techniques able to allow of adapting the existing DVB-RCS mechanisms to the mobile systems case that encompasses aeronautical, maritime and terrestrial environment. These studies have investigated in vary directions like:

- Investigation of applications and markets, traffic volumes and characteristics, physical constraints;
- Investigation of the main issues for extension of the DVB-S2/DVB-RCS standard to the envisaged mobile scenarios;
- Investigation of most of the critical issues for adapting the physical layer to the mobile environment;
- Investigation of required modification at the link layer for supporting mobility;
- Investigation of impacts of mobility on network layer.

Some of these studies, which reports have been prepared from many contribution from the TM-RCS members, are:

- Mobile Wideband Global Link sYstem (MOWGLY), by Alcatel Alenia Space: It is introduced as a new research and development initiative, within the 6th EU Framework Programme, towards the commercial deployment of a global satellite system that provides high-speed data services to airplanes, trains and vessels. The intended multiple-access technique and air-interface technology are specified and the basic system architecture is overviewed. Critical parameters for the system definition and design, including market analysis, traffic modeling, telecom regulations and mobility effects are analyzed and discussed. Possible modifications to the DVB-S2/DVB-RCS standards in order to cope with the limitations of the mobile environments are also discussed [91];
- Fast Internet for Fast Train Hosts (FIFTH), by Alenia Spazio: proposes a new and challenging network solution to make available the Internet access through broadband GEO-Satellite access to high speed train passengers. A new mobile satellite terminal technology has been studied and a prototype has been designed and developed for a practical implementation in suitable demonstrator scenarios [92];
- Amerhis, a joint project of ESA, Centro para el Desarrollo Tecnolgico Industrial (CDTI) and Hispasat represents the first operational regenerative, on board processing, DVB-RCS - DVB-S satellite switching. Amerhis was designed as a response to cover the growing demand in multimedia broadband services and the adaptation of real time services to the satellite world [93];
- Applications layer QoS in DVB-RCS systems, by Indra Espacio: The main goal of this project is to provide an overall network performance characterization for different configurations of a DiffServ/DVB-RCS integration architecture in function of different architectural configurations, network loads, traffic profiles, etc. [69];
- DVB-S2/DVB-RCS broadband mobile system, by Space Engineering: The main objective is to in-depth analyze possible techniques for adapting DVB-S2 / DVB-RCS standards to mobile systems operating at Ku- or Ka-bands whilst retaining, as much as possible, backward compatibility with current version of the standards. Both LoS and NLoS application scenarios will be covered by the investigation. In particular, major emphasis will be on application in the aeronautical, maritime and railway domains [70];
- Harmonization of DVB-RCS Management and Control planes (HM&C), by Thales Alenia Space Espana: The Project "Harmonization of DVB-RCS Management and Control Planes (HM&C)" is an initiative promoted and funded by ESA in the frame of Advanced Research in Telecommunications Systems - Multimedia Programme (ARTES). Its aim is to support SatLabs activities and specifically a SatLabs working group that has been recently set up to address the next step on harmonization of management and control planes of DVB-RCS, including the incorporation of a Connection Control Protocol (C2P) [71];

- Integrated QoS and Resources Management in DVB-RCS Networks: Alcatel Space, by Thales Alenia Space France [72]: The project objectives are to define and evaluate various scenarios for QoS support in future integrated satellite/terrestrial networks. New networking and access equipment architectures shall be proposed, including support for new functions and protocols allowing enhanced QoS management. These architectures shall be defined as stepwise evolutions from the current DVB-RCS systems for operation over transparent or regenerative satellite access networks;
- Preparation for Internet to Trains Initiative: Broadband on Trains, Analysis of the Opportunity and Development Roadmap, by Deutsches Zentrum fr Luft- und Raumfahrt (DLR) [73]: This study provides train operators and ESA with a review of opportunity presented by the provision broadband services on board trains. This study provides train operators and ESA with a review of opportunity presented by the provision broadband services on board trains. It has specifically sought to address the following questions:
  - What is the size of the current market for internet to trains across the EU25 and Canada?
  - What is the optimal business structure for provision of internet to trains?
  - What combinations of business structure, market uptake and technical implementation provide a positive return on investment?
  - What are the generic technical requirements for internet to trains services to allow train operators to set up or procure services?
  - What regulatory, standardization and technical issues need to be addressed in the target markets to allow the successful implementation of Internet on Trains?

All these projects address all the possible issues specific for the mobility environment. Their main goal is to ensure interoperability between DVB-RCS terminals and systems and to achieve these low-cost solutions.

5.3 Description of DVB-RCS System Scenario

In this section a description of the considered satellite scenario for railway environment used in this work to analyze the return link of a mobile DVB-RCS is presented (as depicted in Fig. 5.10), showing all aspects and considerations done in order to implement a software simulator. The considered network scenario is composed of a transparent DVB-RCS satellite that interconnects satellite RCST terminals with a capacity in forward link and a capacity in return link. The modeled system describes a DVB-RCS based network that uses the star topology. A generic architecture comprising the RCST terminals, the GTW device separately from the NCC device has been considered. It is possible also to considered a single device called HUB that includes the functionality of both GTW and NCC. The forward channel (from GTW to the terminals) is compliant with DVB-S2 standard, but in the considered scenario no forward link simulation has been performed. While, the return channel (from the terminals to the GTW) uses the DVB-RCS standard with the extension for mobile. Then, in this scenario the NCC manages the resource allocation for the terminals instead the GTW manages the traffic queues. The NCC has the responsibility of processing the capacity requests that come from RCST and on the basis of these requests it assigns resources through the TBTP table. The propagation delay for the satellite link has been considered of 125 ms. Moreover, it has been assumed that the satellite channel is quasi error free and, in order to simplify test scenario, a zero BER has been considered during the simulation campaigns.

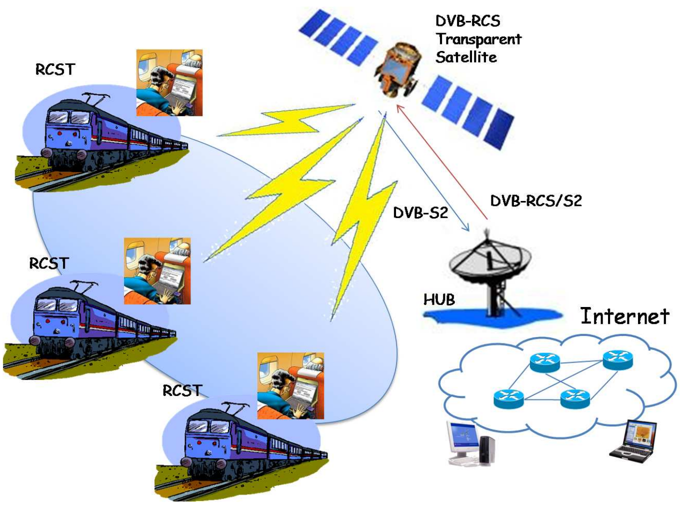

Fig. 5.10. Reference Railway Scenario

5.3.1 Return Link Structure

For simplicity, it has been assumed that the considered DVB-RCS system has a simple fixed MF-TDMA frame structure, namely, bandwidth and duration of successive slots used by RCSTs is fixed. Moreover, always for simplicity, it has been assumed that frame and superframe duration is the same. A superframe period of 45 ms has been chosen; this value is suggested by standard [6]. Additionally, it has been assumed that each timeslot carries one MPEG packet. The considered frame composition is that reported in the standard [5], [6]. In the following the MPEG traffic time slot is described and depicted in Fig. 5.11.

In the DVB-RCS system the NCC sends Signaling Information (SI) to RCSTs via forward link. Each RCST hears the SI and recognizes the system information, such as network clock and superframe composition information

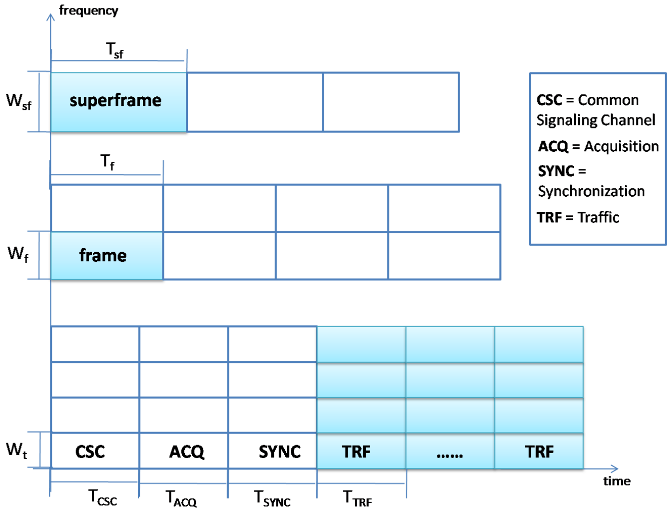

Fig. 5.11. MF-TDMA Frame Structure

for the return link. A superframe consists of a number of frames, where each frame contains a number of time slots such as CSC, ACQuisition Burst (ACQ), SYNC and TRaFfic burst (TRF) time slots. In order for each RCST to enter the system, it sends the initial access request with its MAC address to the NCC through a CSC timeslot in a contention-based manner. In case of a collision, it may retry the access in a random period of time. If the access of the RCST is authorized, the NCC assigns ACQ and/or SYNC time slots to the RCST and sends a TIM, which includes information required for fine synchronization (see [5]), to the RCST. In order to better understand the composition of frame it is important to remember some concepts. The modulation used in DVB-RCS, as suggested by standard [5], is a QPSK that uses 1 symbols for representing 2 bit of information. Another concept to remember is the roll-off factor. Before operating the modulation, the signal is filtered through a square root raised cosine filter that is able to minimize the InterSymbol Interference (ISI). This filter has a characteristic factor called roll-off factor which shapes the spectrum of the signal. The roll-off factor must be a real number between 0 and 1. The roll-off factor determines the excess bandwidth of the filter. For example, a roll-off factor of 0.5 means that the bandwidth of the filter is 1.5 times the input sampling frequency. A possible value of roll-off factor suggested by [5] is 0.35.

5.3.2 Terminal Burst Time Plan (TBTP)

This message is sent by the NCC to a group of terminals. The group is addressed by a logical Group_ID, while each individual terminal is addressed by a logical Logon_ID. Both Group_ID and Logon_ID are notified to the terminal at logon time. It contains the number of slot assigned by the NCC control mechanism for each RCST. Each traffic assignment is described by the number of the start timeslot in the block and a repetition factor giving the number of consecutive timeslot allocations. The TBTP shall be updated every superframe. In this work having considered a MPEG traffic slot a superframe time of 45 ms has been considered as suggested by standard [5]. In Fig. 5.11 the MF-TDMA frame structure is depicted. In DVB-RCS systems, the resource allocation policy is based on so-called Bandwidth on Demand (BoD). Once a logon procedure is completed, each RCST, in need of capacity, sends a request message to a scheduler which is function of the NCC. Upon receiving the capacity request messages, the scheduler in the NCC generates a TBTP table and sends it to the RCSTs. Upon receiving the TBTP table, each RCST reads the TBTP table to know which time slots are assigned. This capacity allocation procedure is performed every superframe. The ETSI Technical Report [6] shows an example of segmentation of the return link capacity, both for ATM and MPEG case. The selection of one or the other depends on many considerations pertaining to the network operator's strategy. In this work, as mentioned previously, a MPEG modality has been considered, then, in the following, an example of segmentation based on MPEG time slots is shown. The reference design is based on a symbol rate of 270 kilo symbol per seconds (ksps). The choice of 270 ksps as basic rate is motivated by its simple relationship with the clock of the system (27 MHz). Of course, it possible to use other symbols rates [6]. The frame composition for this symbol rate is shown in standard guidelines [6].

Other design objectives are given below:

- Traffic slots accommodate 1 MPEG packet (i.e. 752 modulation symbols plus preamble, guard-time and a FEC redundancy part depending on FEC rate).
- The equivalent of one traffic slot per frame is used to carry signaling. For simplicity, the beginning of the frame is used and it is divided into mini-slots.
- The frame length should be about 45 ms (to keep latency low), while the total time allocated to mini-slots should be kept below 15% of a frame.
- Mini-slots should accommodate indifferently CSC, ACQ and SYNC (for simplicity).
- The TDMA Preamble is equal to 48 Symbols, for all burst types. This does not preclude the use of shorter preambles, depending on demodulator performance and the ability of the RCST to maintain synchronization.

These design objectives can be fulfilled with the segmentation given below. An idle terminal has one mini-slot opportunity every frame in contention mode but only one pre-assigned mini-slot every 256 frames (or 12 s). An active terminal has one mini-slot opportunity every frame in contention mode and one pre-assigned mini-slot every 32 frames (or 1,4 s) for signaling purposes.

5.3.3 Digital Video Broadcasting with Return Channel Satellite (DVB-RCS) Capacity Request Signaling

In the DVB-RCS standard, terminals can use two signaling methods to forward capacity requests to the NCC:

- Mini slot method. The capacity request message is carried in the Satellite Access Control (SAC) field of a dedicated SYNC burst, used for synchronization maintenance purposes and sent about once every second (normally it is used a period of 32 frames). A SYNC burst can carry up to four different capacity requests. This mechanism is called Out Band Request (OBR);
- Prefix method. A capacity request is piggy-backed onto a traffic burst (it is inserted in a SAC field, appended as a header of the MPEG payload). One capacity request per traffic burst can be transmitted. This mechanism is called In Band Request (IBR).

DVB-RCS standard makes provision for both mechanisms, IBR and OBR. The IBR is associated with the prefix method and with the Data Unit Labeling Method (DULM), while OBR is associated with the minislot method that relies on SYNC burst.

Normally, the prefix method (IBR) is used combined with the minislot method (OBR) with the SYNC. The SYNC slots are primarily assigned to an Satellite Terminal (ST) in order to perform capacity request and for synchronization needs, typically one slot SYNC every 32 frames (1440 ms at 45 ms frame duration). The role of the OBR is therefore merely to speed up the initial access to return link capacity. In the following figure (Fig. 5.12) it is shown the time diagram of the request of capacity performed by a RCST terminal to the satellite system.

In this work it has been implemented the OBR modality, the IBR is reserved for future work.

5.3.4 Return Link Dynamic Resource Control

Dynamic resource control consists in assignment of resources (slots) to terminals based on their requests and limit values negotiated/set during connection establishment. The assignments are conditioned by the availability of resources (capacity) within defined return channels (as per system connectivity). The assignment is the responsibility of the MAC scheduler, which implements a Demand Assignment Multiple Access (DAMA) protocol. The MAC scheduler is located on the NCC terminal. The uplink scheduling consists of processes taking place in the scheduler and in terminals, namely:

- Calculation of capacity requests in terminals (once a RCST is logged it continuously monitors bandwidth related parameters, e.g. RCST queue lengths and queue input rates);

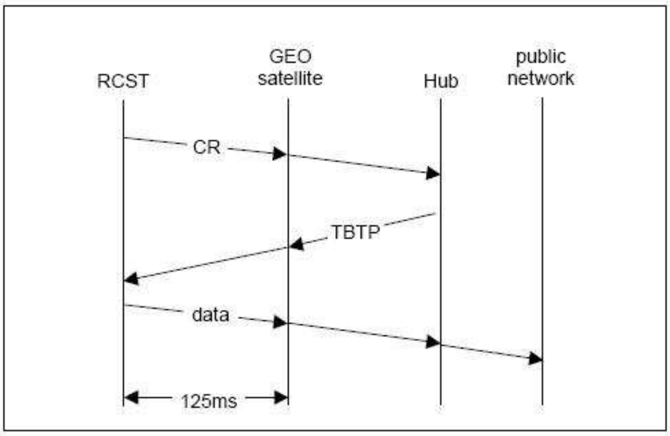

Fig. 5.12. Time diagram of capacity request

- Transmission of capacity requests to the scheduler (request signaling);
- Calculation of capacity assignment to terminals;
- Allocation (placement) of the assigned capacity;
- Generation and transmission of TBTPs carrying the allocations to terminals.

5.3.5 DAMA Scheduling Process

The overall capacity assignment process comprises a number of separate but dependent processes, some performed by the terminal and others by the scheduler, The following is an outline description of this overall process in terms of the individual processes in order of execution:

- Traffic arriving at the terminal input from the STs is queued into a number of queues. The first set of queues is at the IP layer and the second at the MAC layer. Their number depends primarily on the CoSs / types of traffic supported and the approach to QoS implementation. From the point of view of Scheduler operation only the MAC layer queues are important;
- Based on the current MAC queues status and the queues attributes, the terminal issues RBDC and/or VBDC requests. The requests are issued when needed and are frame synchronized;
- On arrival at the Scheduler, the capacity requests, tagged with the identity of the source terminal, are buffered or stored in appropriate buffer/queues. The source terminal is derived from the burst (carrying the request) assignment in the superframe;
- At the start of each superframe the Scheduler uses the buffered requests received in the last scheduling interval, plus unsatisfied requests from prior scheduling intervals, to assign capacity to terminals;

- The Scheduler builds TBTP update messages corresponding to the assignments from the previous process and transmits them on the terminal's forward link every superframe;
- On receiving the TBTP update messages, the terminal decodes them and dispatches traffic cells from the MAC priority queues for transmission in the allocated time slots of the specified UpLink (UL) superframe number. In parallel, it updates the queues status to reflect the committed traffic.

5.3.6 DVB-RCS Capacity Types

The types (categories) of capacity that can be assigned to an ST are consistent with the DVB-RCS standard. They are:

- CRA, or Static Rate (SR);
- RBDC, upper-bounded by MaxRBDC;
- VBDC;
- FCA.

The standard also defines an AVBDC, which can be used in the MAC Scheduler to recover from requests/allocation losses. Capacity types are vital to QoS support at MAC layer. Any given terminal can be assigned one or a mix of the four capacity types.

In this work the attention is focused on RBDC and VBDC capacity requests, then in the following these two types of requests will be explained.

Rate Based Dynamic Capacity (RBDC)

This capacity category is used for high priority variable rate traffic that can tolerate the MAC Scheduler dynamic response time latency. RBDC is dynamic rate capacity (in slots/frame) granted in response to dynamic requests from a ST to track the instantaneous traffic rate. A maximum rate limit (MaxRBDC) is possible to set. As with CRA, MaxRBDC can be negotiated off-line.

RBDC capacity is said to be guaranteed if the sum of CRA and MaxRBDC values configured in the MAC Scheduler for each ST are constrained to the terminal physical transmission limit, and if the total (CRA + MaxRBDC) values for all STs in the segment/area is within the segment/area capacity. If these conditions are met, the RBDC is appropriate for traffic that requires bandwidth guarantees.

In case of overbooking the latter condition above is not met and, therefore, no absolute guarantees can be provided that the requests will be granted in the expected frame. Various strategies can be used concerning the requests in excess of MaxRBDC. They can be ignored or they can be granted as VBDC (if available), but not necessarily in the expected frame. In the latter case the packets queued in the ST might eventually be transmitted, but with additional delay, so jitter will be induced. Overbooking may be used by some network operators in order to increase capacity utilization. It relies on the fact that, due to traffic variability, not all STs using RBDC transmit packets at maximum rate (i.e. MaxRBDC). In this case the capacity is guaranteed on a statistical base.

The actual amount of RBDC granted in any superframe is controlled by dynamic requests from the ST to the MAC Scheduler, each request being for the full RBDC rate currently needed. Each request overrides all previous requests from the same terminal, and is subject to a configurable time-out period, to prevent a terminal anomaly resulting in a hanging capacity.

Volume Based Dynamic Capacity (VBDC)

This capacity category is used for traffic that can tolerate delay jitter, such as the BE class of the Internet traffic. VBDC capacity is provided in response to dynamic volume requests from the ST to the MAC Scheduler (a volume request is for a given number of slots with no time constraint).

In general VBDC capacity is not guaranteed. It is assigned as best effort capacity within the available resources, after satisfying the total CRA and RBDC capacity components. The amount of VBDC capacity assigned to a terminal can be limited to a MaxVBDC value, based on rules set by the network administrator or on results from measurements on traffic.

A Guaranteed VBDC (G-VBDC) or High Priority VBDC (HP-VBDC) capacity can also be defined, by setting a minimum value for VBDC (MinVBDC) per ST and configuring it in the MAC Scheduler and ST. VBDC capacity up to MinVBDC value (whichever is less) will be granted every superframe, and so VBDC capacity up to MinVBDC is treated as a third form of guaranteed capacity along with CRA and RBDC.

Capacity Request Parameters

The Capacity Request parameters are reported briefly hereafter:

- MaxRBDC: The maximum allowed RBDC data rate (in bps). This rate is guaranteed. This parameter is considered in kbps;
- RBDC timeout: It is the time for which an RBDC request is valid. To prevent a terminal anomaly resulting in a hanging capacity assignment, the last RBDC request received by the NCC from a given terminal shall automatically expire after a time-out period, such expiry resulting in the RBDC being set to zero rate. This parameter is calculated in superframes;
- MaxVBDC: represents the limit to assign slots to VBDC traffic for a superframe. It is calculated in superframes;
- VBDC timeout: It is the time for which an VBDC request is valid, when this timer expires the slots assigned shall be de-allocated to the terminal.

5.4 Traffic Models

In order to simulate a more realistic scenario, for getting realistic and valid results, a set of real traffics (shaped realistically) have been considered in this work.

For the simulations two different kinds of non Real Time (nRT) traffic (HTTP, FTP) and one kind of Real Time (RT) traffic (streaming) have been modeled on IP layer [94], [95]. For all application classes the IP packet size is assigned in a fixed manner and the considered value is 1500 bytes.

In literature a lot of documentation exist about traffic models and the statistical distributions and the corresponding parameters are given for realizing traffic models that can be considered very realistic. In this work the considered user is a train passenger that has a regular behavior, that is he does not change his behavior because of the smaller link of the train. The next subsections show considered traffic in this work, HTTP, FTP and video streaming.

5.4.1 HTTP Traffic Model

Web traffic is nowadays the most important application used by the internet community. The term web traffic comprises all HTTP traffic generated during a session with a typical web browser like Netscape Navigator or the Internet Explorer. A session is considered to be the time between start and exit of the browser (session level) [96].

Typically, HTTP traffic is modeled as an ON/OFF source with the ON state corresponding to the request and download of the objects and the OFF state corresponding to the inactive time.

In this work the implementation is based on a work of the Communications Technology Laboratory of Intel Corporation of the 2007 [97]. The model proposed is shown in Fig. 5.13 and follows the basic ON/OFF model.

In this model the traffic was modeled as a series of user generated Web-Request. It has been considered an HTTP session that consists of a series of Web-Request. A Web-Request consists of one Hyper Text Mark-Up Language (HTML) object and zero or more embedded objects that arrive before the next HTML object. The first interval between the Web-Request and the first embedded object is called Parsing Time, the next intervals that follow a different distribution, are called Inter-Arrival-Time (IAT). At last, the time between the download of the last embedded object to the arrive of another page request is called Reading Time.

The arrival process of a Web server can be considered as superimposing individual user requests to the server. A user who is browsing a page on the Web server usually initiated a request to an HTML document. Then, it possible to have separate requests that can be automatically or by user generated during the page downloading as it is possible to view in Fig. 5.13. After all the user's requests are completed, the page is said to be completely

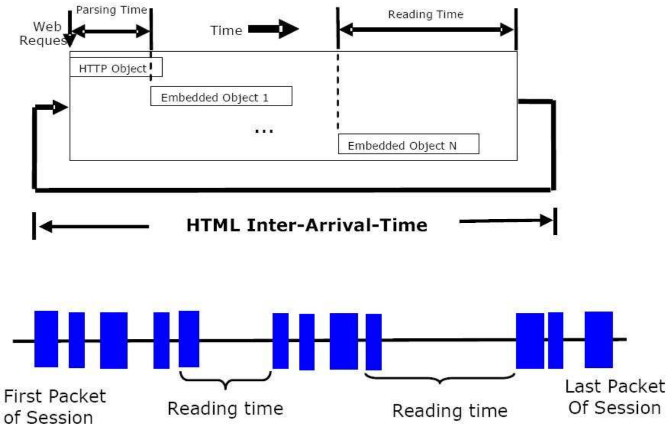

Fig. 5.13. HTTP Traffic Model Diagram

downloaded. The user, usually, takes some time to go through the page before browsing another page on the server. The delay introduced by the user is called think time o reading time. A user session may be ended if the user stops accessing the server or may be continued if the user accesses another page in the server after the reading time. Based on this user behavior, a Web arrival model can be built that can be described by the distributions of its random variables:

- n: the number of HTML documents requested per session;
- T_P: the duration of the Parsing Time;
- $NoEO$: the number of embedded objects present in the HTTP page;
- T_{IAT}: the duration of inter-arrival-time;
- $T - R$: the duration of reading time.

5.4.2 FTP Traffic Model

The importance of FTP is steadily decreasing with the upcoming of HTTP downloads. According to [73], an ESA project concerning on Broadband on Train, an FTP file transfer can be modeled like a web session with only one web request. The implementation in this work has been done creating a software model for the generation of a single HTML document, that is a HTTP page with the main object and a certain number of embedded object.

5.4.3 Video Conference Traffic Model

The traffic generation for the Video Streaming consists of two parts. First the arrival of a video session and second the behavior of the video source itself. The arrival of the video streaming sessions is modeled with a Poisson distribution, the duration of a session as an exponential distributed random variable.

The video source itself is modeled as a discrete time Markovian arrival process [98], [99]. The bit-rate that is produced by a video-conferencing terminal has been considered as the aggregated output of M independent mini-sources (see Fig. 5.14). Each of the mini-sources switches its state from ON to OFF and vice versa. A mini-source in the ON state produces traffic at a constant rate of A bps - in the OFF state no traffic is generated. The time intervals that the mini-source is in the ON or OFF state are modeled by a geometric distribution. The total amount of generated bits is then split up into single packets with a packet size as mentioned above.

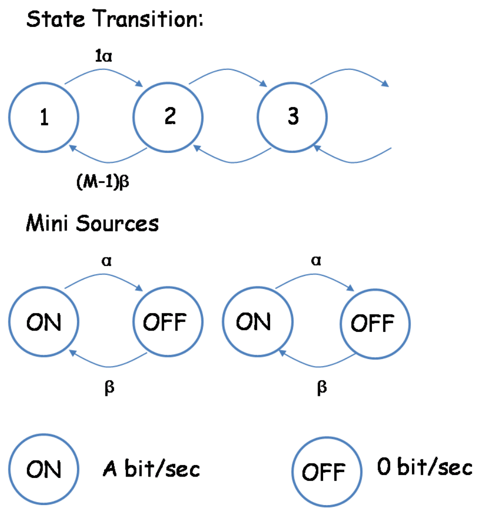

Fig. 5.14. Streaming Traffic Model

A video traffic source based on the modulating process as in Fig. 5.14 can be implemented as follows: as soon as the source enters the state i with leaving probabilities $i\alpha$ and $(M-1)\beta$ it is possible to generate two samples t_{incr} , t_{decr} from geometrically distributed random variables with mean values $1/i\alpha$ and $1/[(M-1)\beta]$ respectively. Hence, the sojourn time for state i in T_S units can be determined as $t_{sog} = mint_{incr}, t_{decr}$; correspondingly, the transition accomplished at the end of interval t_{sog} is towards state $i-1$, if otherwise, the transition is towards state $i+1$. During t_{sog} the source generating traffic is a constant bit-rate equal to iA.

5.5 Simulator Environment Description

In this section the description of the simulation environment is introduced in order to explain what is the software used for the implementation and what are the assumptions that have been done in order to simplify the realization of the simulation software. The realized software allows to model the uplink channel of a DVB-RCS satellite architecture for the mobile terminal.

It is characterized by the following features:

- Modular design: the realized simulator allows the setting of a lot of parameters useful for system testing in different conditions;
- Scalability: the tool allows the definition and implementation of as many new block types as required making feasible its extension with new functions and network models as required by the testing scenarios of this study.

The tool is realized in Java language (and it has been utilized under Windows environment). The final result is a platform able to simulate the DVB-RCS+M satellite uplink working in RCST Basic Mode and then capable of using the MF-TDMA or the SCPC modality in order to better handling system resources. The tool is able to carry out a trace file useful to debug the system. The trace is composed of text messages that give a detailed behavior of the main elements inside the system. The result, instead, are gave in a excel format in order to be processed in off-line way.

5.5.1 Discrete Event Simulation

The simulator has been implemented using the Discrete Event Simulation (DES), in which the operations of the system is represented as a chronological sequence of events. Each event occurs a time instant and marks a state change in the system.

Its main components are the clock, for keeping trace of the simulation time, the events list, in which are presents the event that will be in the time, Random Number Generators for generating number following a determinate distribution function, the Statistics module which keeps trace of the aspects of interest plus the elements related with the satellite environment.

5.5.2 Measurements

Measurements provided by real user equipment are end-to-end QoS measurements reported by test applications, which provide information on some simulation values. These measurements are:

- Goodput: defined as the application level throughput, i.e. the number of useful bits per unit of time forwarded by the network excluding protocol overhead;

- End-to-End Delay: measured from the instant an IP packet enters the RCST to the instant this packet is received by the satellite GTW. It includes the satellite propagation delay (250 ms);
- Delay Jitter: the variation in the packet arrival time belonging to the same connection caused by network congestion;
- Queue Size: the size of the queue for the MAC layer, both RBDC and VBDC queue;
- Resource Utilization Efficiency: defined as the ratio between slots really used in the system and the overall number of slots;
- Requested VBDC: indicates the average number of requested bits per superframe period. Reported units are timeslots/superframe, but they can be easily converted into kbps considering timeslot length in bytes (188 bytes) and superframe duration (45 ms);
- Requested RBDC: indicates average requested bandwidth in kbps. If no request opportunities are available, a zero sample is taken;
- Assigned Bandwidth: this monitoring result reports the number of timeslots per superframe period assigned by the DAMA Controller. As for the requested VBDC resources, units are timeslots/superframe, but they can be easily converted into kbps considering timeslot length in bytes (188 bytes) and superframe duration (45 ms). This monitoring value can be obtained per terminal and globally.

5.5.3 Physical Parameters

In order to make a fair comparison between the two modality SCPC and MF-TDMA the correct parameters have been chosen in this work. In the following it will be explained the performed choice. It is important that the obtained $E_s N_0$ is the same for the two modalities because the satellite terminal has to be able to get the signal with its antenna system. Then the choice of physical parameters has been done taking into account that the $E_s N_0$ value has to be equal (enough) both in MF-TDMA and SCPC case. The guidelines on DVB-RCS standard [6] have a section on Turbo Code Performance where the performance of the system for a Packet Error Rate (PER) of 10^{-5} and a PER of 10^{-7} are presented. In this work, a PER of 10^{-7} has been chosen. This table (Fig. 5.15) is shown in the following. Note that, the table shows the performance in terms of $E_b N_0$. It is then simple to convert $E_b N_0$ in $E_s N_0$. Note that, in this work we have chosen the MPEG encapsulation with the assumption that only one MPEG packet can be mapped in one slot.

Then, in this table there are the values of FEC associated with the values of $E_b N_0$ for the MF-TDMA mode. In the following, another table (Fig. 5.16) is presented where it is possible to view the values of coderate associated with the E_b/N_0 values in the case of SCPC mode [100].

It has been calculated that with a coderate of 2/3 for MF-TDMA mode for MPEG packets and a coderate of 3/4 for SCPC mode both with a modulation of QPSK the $E_s N_0$ value is enough similar and then it is possible to make

FEC	E_b/N_0 (188 bytes)	E_b/N_0 (53 Bytes)
1/3	2,5 dB	2,9 dB
2/5	2,7 dB	3,1 dB
1/2	3,2 dB	3,6 dB
2/3	4,0 dB	4,6 dB
3/4	4,6 dB	5,4 dB
4/5	5,3 dB	6,3 dB
6/7	6,0 dB	7,0 dB

Fig. 5.15. Turbo Code Performance for PER $= 10^{-7}$

Modulation and Code Rate	Required E_b/N_0 [dB] at PER $= 10^{-5}$	Required E_b/N_0 [dB] at PER $= 10^{-7}$
π/2-BPSK ¼	1,55	1,77
π/2-BPSK ½	1,94	2,12
π/2-BPSK ¾	3,16	3,30
QPSK ¼	1,55	1,77
QPSK ½	1,94	2,12
QPSK ¾	3,16	3,30
8PSK ¾	5,58	5,82

Fig. 5.16. Performance DVB-RCS+M - 4k

these choices for the system implementation. Both table report the E_bN_0 value then on the basis of the Annex D.2 of guidelines [6] it is possible to calculate the E_sN_0 on the basis of the following relation valid for the turbo code case:

$$\begin{aligned} &E_b/N_0 = 1/Coderate \cdot Symbol/Bit \cdot E_s/N_0 \\ &\Rightarrow E_s/N_0 = Coderate \cdot Bit/Symbol \cdot E_b/N_0 \end{aligned} \quad (5.1)$$

5.6 Threshold Basic Switching Algorithm (TBSA)

In this work a simple algorithm for switching the terminal from a working mode to another is proposed. It is a simple mechanism based on a threshold that states when a RCST terminal can move from MF-TDMA modality to SCPC one and vice versa as is depicted in Fig. 5.17. The algorithm proposed is composed of two parts: one related with the capacity request control in order to make the switching from MF-TDMA to SCPC and the other one related with the throughput control of the mobile terminal in order to make able the NCC of sending an appropriate message toward the RCST mobile terminal (TIM message) in order to make the switching from SCPC to MF-TDMA mode.

5.6.1 Moving from MF-TDMA to SCPC

If the RCST terminal is in a MF-TDMA mode the proposed mechanism controls, the capacity request performed by the terminal for the RBDC queue in

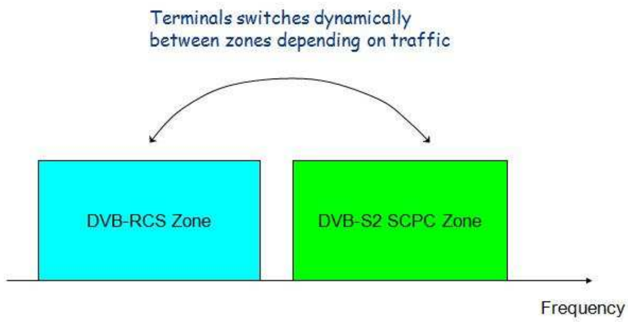

Fig. 5.17. Dynamic switching between DVB-RCS and DVB-S2 zone

order to decide if it is better to switch the RCST to a SCPC mode in which it receives a static carrier without performs on time basis a request of needed capacity.

The choice of RBDC queue has been done in order to better satisfy the guarantee to a more constrained traffic as audio or video applications. Fig. 5.18 shows a pseudo code of the proposed control mechanism performed in the NCC satellite terminal.

```
calculate CR value (bit/s);

threshold1=n1*T_SYNC
for(int=0; i<threshold1; i++)
  if (CR value > threshold2)
    perform switching to SCPC;
  else
    remain in MF-TDMA


n1 interger
threshold2= n2 * maxRBDCvalue
              n2 integer
T_SYNC=32*T_SF
```

Fig. 5.18. Pseudo code for switching from MF-TDMA to SCPC mode

The mechanism controls if the capacity request is greater than a certain threshold (threshold2 in Fig. 5.18) for a certain number of time as it is possible to view in Fig. 5.19. If the capacity request is greater than the threshold2 for threshold1 time then the NCC sends a TIM message toward the RCST in order to make the switching from MF-TDMA mode to SCPC one.

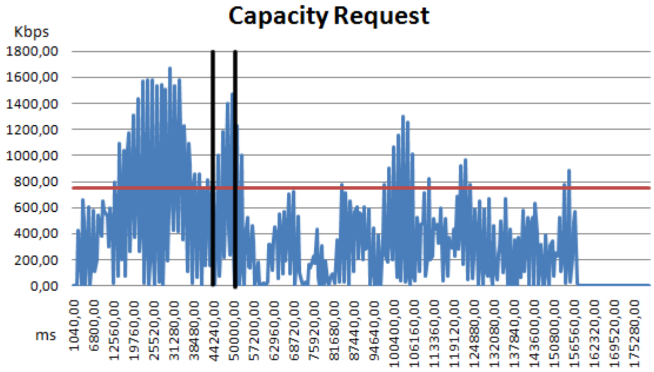

Fig. 5.19. Example of capacity request control performed by the TBSA algorithm

5.6.2 Moving from SCPC to MF-TDMA

If the terminal is in a SCPC modality the mechanism controls on periodical basis the throughput of the specific terminal in order to state if there is a wastage of bandwidth and then it is better to move the RCST to the MF-TDMA mode in order to exploit the dynamic capacity request.

In the following figure, Fig. 5.20, a pseudo code is shown that simulates the behavior of the control mechanism.

```
calculate throughput value (%);

threshold1=n1*T_SYNC
for(int i=0; i<threshold1; i++)
  if (throughput value < threshold2)
    perform switching to MF-TDMA;
  else
    remain in SCPC


n1 integer
threshold2 = n2 %
              n2 integer [0, 100]
T_SYNC=32*T_SF
```

Fig. 5.20. Pseudo code for switching from SCPC to MF-TDMA mode

As in the previous case, also for switching from SCPC to MF-TDMA mode the NCC performs a control now based on throughput of the mobile terminal. If the throughput is below a certain threshold for a certain time amount the NCC sends a TIM message toward the RCST that it will switch from SCPC to MF-TDMA modality. In Fig. 5.21 it is possible to view graphically the throughput control operated by NCC.

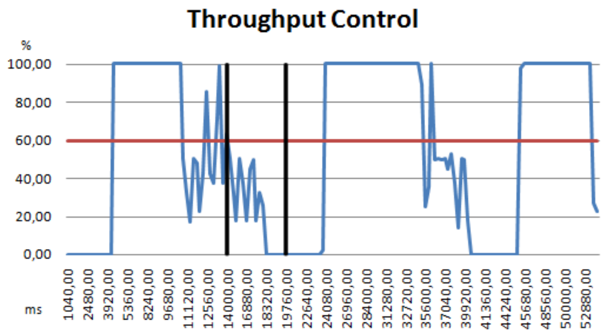

Fig. 5.21. Example of throughput control performed by NCC to switch from SCPC to MF-TDMA

5.7 Simulation Graphic Interface

In this section the graphic interface of the implemented system is shown. The software simulator is modular and scalable: it is simple to adding new module in order to perform new mechanisms and it is simple to varying the parameters setting in order to change the simulation scenarios. In Fig. 5.22 it is possible to see the initial screen of the software.

In Fig. 5.23 a configuration form is presented in order to show how it is possible to modify the parameters in the software in order to consider different satellite system scenarios with a different number of RCST terminals, different users with different distributions of applications and so on.

5.8 Simulation Results

In this section simulation results will be presented. They show the impact of the capacity request performed by the mobile terminal for the three different considered applications. Moreover, they show a comparison between a classical MF-TDMA operation scheme with a system that dedicates a part of the total return capacity to MF-TDMA and another one to SCPC and in which the proposed switching algorithm (TBSA) works. The work has focused the

Fig. 5.22. Initial screen of the simulation graphic interface

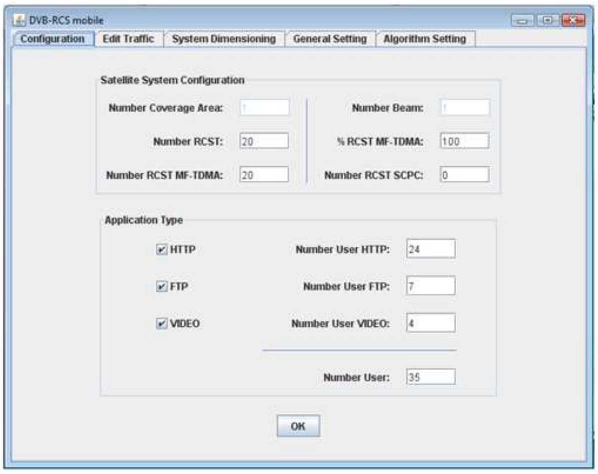

Fig. 5.23. Satellite scenario configuration

attention on the Basic RCST mode that is the modality in which the mobile terminal is able to send the information or in MF-TDMA or in SCPC modality but not simultaneously.

In order to make a fair comparison between the MF-TDMA and the SCPC modality, it has been chosen for both the same bandwidth frequency granularity. It has been considered a carrier of 500 kHz for carrying out the simulation results. In Table 5.1 the main parameters considered in the simulation are reported.

Table 5.1. Simulation Parameters

Parameters	**Value**		
Total Capacity (MHz)	10		
MF-TDMA Capacity (%)	100, 90, 85, 80, 75		
SCPC Capacity (%)	0, 10, 15, 20, 25		
Carrier MF-TDMA (kHz)	500		
Carrier SCPC (kHz)	500		
E_s/N_0 (dB)	5.2		
FEC MF-TDMA	2/3		
FEC SCPC	3/4		
PER	10^{-7}		
Number of Mobile Terminals (RCSTs)	20, 22, 24, 26, 28, 30		
Number of Users	35		
	HTTP	FTP	Video
Scenario 1	70%	20%	10%
Scenario 2	60%	25%	15%
Scenario 3	50%	30%	20%

In the system a total capacity of 10 MHz for the return link has been considered and different distributions of this capacity between MF-TDMA and SCPC mode are used, in order to find the better distribution that guarantees the better performance in a system using the proposed algorithm. Moreover, different numbers of mobile terminals have been considered, from 20 to 30 in order to load the system from 100% to 150%. Furthermore, three different types of scenario are considered for the simulations. The first has a low number of Video Conference applications that represent the heaviest traffic between those considered in the system. The latter has a greater number of Video applications in order to analyze system behavior with these different conditions.

5.8.1 Simulation on Traffic Generation

The simulations have been performed considering three types of applications that a train passenger can use during his travel. In this implementation the three types of applications are HTTP, FTP, and Video Conference traffic. For details about the modeling of these applications see section 5.4 on Traffic Model.

In this section, it will be shown how the traffic can affect system resources and then what bounds are used for the three traffic typologies in order to use the satellite capacity in a better way.

HTTP Traffic

In this section the impact of HTTP capacity request on the system is shown in order to highlight that HTTP traffic is a light traffic that requires low capacity to the system. In Fig. 5.24 the trend for a low number of users is shown; in Fig. 5.25 the trend for a higher number of users is presented.

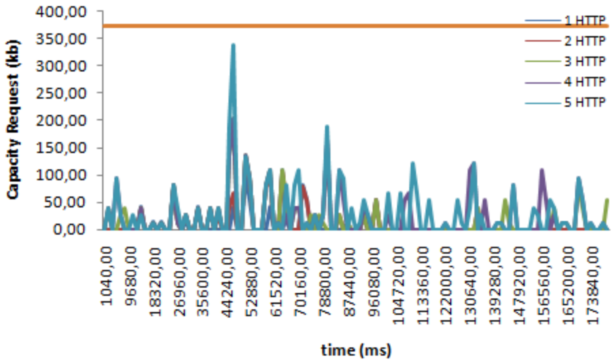

Fig. 5.24. HTTP capacity request for a small number of users

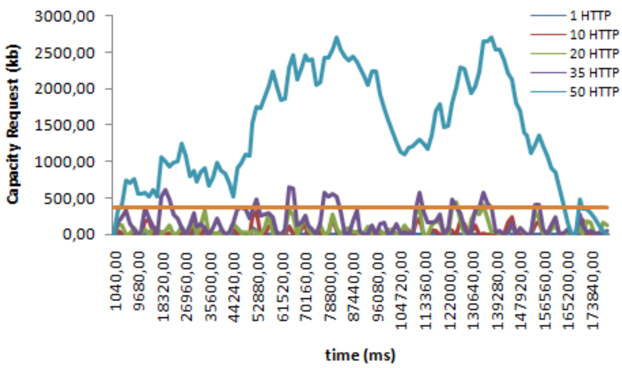

Fig. 5.25. HTTP capacity request for a great number of users

FTP Traffic

In this section the impact of the FTP capacity request on the system is shown in order to highlight that the FTP traffic is a heavier traffic than the previous one. FTP requires more capacity to the system. In Fig. 5.26 and Fig. 5.27 the trends for low and high number of users, respectively, are shown.

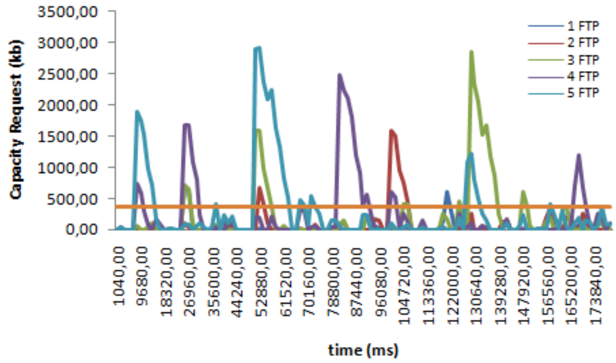

Fig. 5.26. FTP capacity request for a small number of users

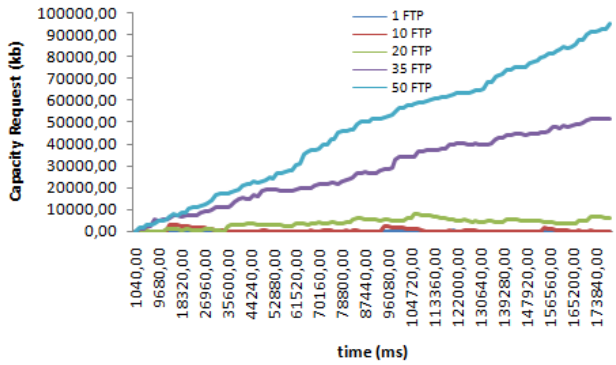

Fig. 5.27. FTP capacity request for a great number of users

Video Conference Traffic

In this section, the impact of the Video Conference capacity request on the system is shown in order to highlight that the Video traffic is the heaviest traffic considered in the system. Video Conference requires more capacity to the system than in previous cases. In Fig. 5.28 the trend for a low number of users is shown; in Fig. 5.29 the trend for a higher number of users is presented.

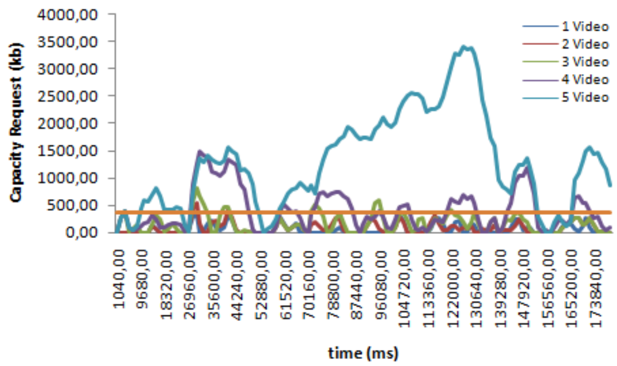

Fig. 5.28. Video Conference capacity request for a small number of users

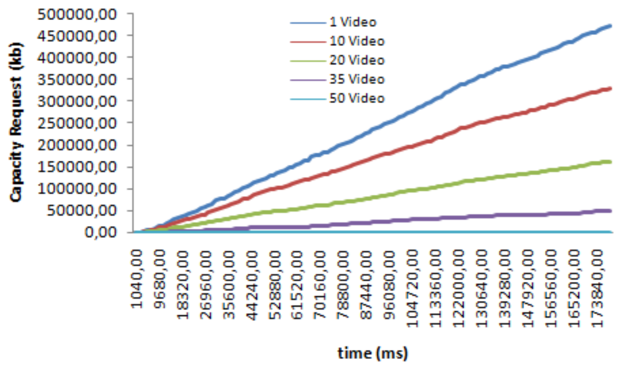

Fig. 5.29. Video Conference capacity request for a great number of users

5.8.2 Simulation results for Scenario 1

In this section, the simulation results under Scenario 1 are shown. Some simulations are performed in order to find the better capacity reparation between MF-TDMA and SCPC. The obtained results show that the repartition that allows to the system the better performance in terms of delay and queue size is a 90% of MF-TDMA and a 10% of SCPC. It is possible to note observing Fig. 5.30, Fig. 5.31, Fig. 5.32, Fig. 5.33 that this repartition guarantees lower delay and lower queue size for both considered capacity categories.

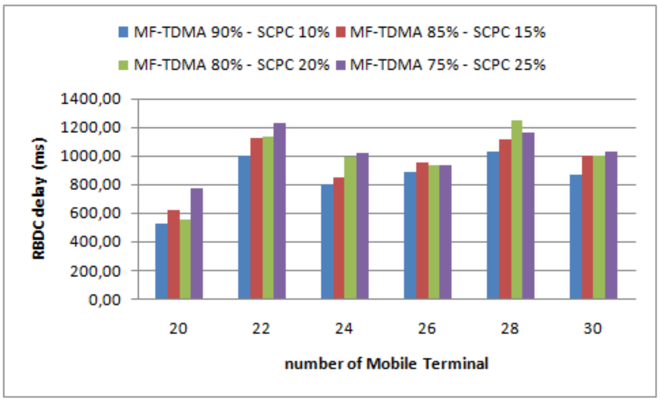

Fig. 5.30. RBDC delay varying the capacity repartition between MF-TDMA and SCPC

Then, once the better repartition has been chosen a comparison of the system so composed and with the introduction of the TBSA algorithm with a classical MF-TDMA has been performed in order to show the new system is able to manage a greater number of terminals than the classical MF-TDMA.

The following figures show the results of the comparison between the classical MF-TDMA with a system with a repartition of return link capacity of 90% MF-TDMA and 10% of SCPC.

It is necessary to highlight that this is a qualitative analysis of the system because not a real QoS management is now introduced in the simulator. The system have, now, no timing requests and no dropping performance. These results show, surely, that with a so composed system and with the introduction of the TBSA algorithm the delay and the queue size observed are lower than a system that works with a classical MF-TDMA scheme.

In Fig. 5.34 and in Fig. 5.35 it is possible to observe lower delay for both RBDC and VBDC capacity category.

Lower queue size for both RBDC and VBDC class of traffic is also obtained as shown in Fig. 5.36 and Fig. 5.37.

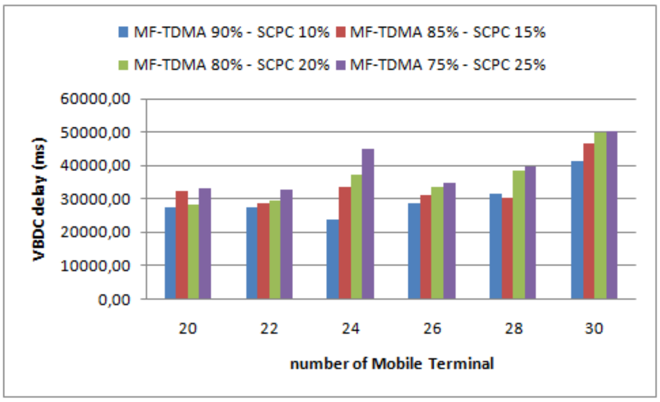

Fig. 5.31. VBDC delay varying the capacity repartition between MF-TDMA and SCPC

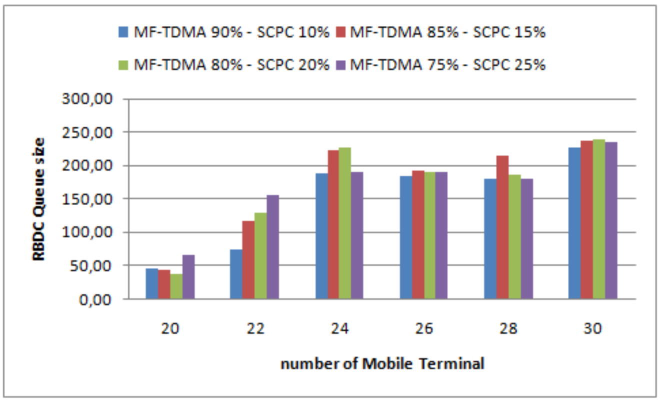

Fig. 5.32. RBDC queue size varying the capacity repartition between MF-TDMA and SCPC

At last, in Fig. 5.38 the number of switchings performed by the system in order to better use the satellite resource is shown. It is possible to make a consideration concerning the signaling overhead introduced in the system by the TBSA algorithm but it is possible to note that the average number of switching is not high and, then, not a lot of overhead is introduced in the system with the use of the switching algorithm.

5.8.3 Simulation results for Scenario 2

In this section simulation results for Scenario 2 are reported. It is possible to say that for this scenario the same considerations made on Scenario 1 are

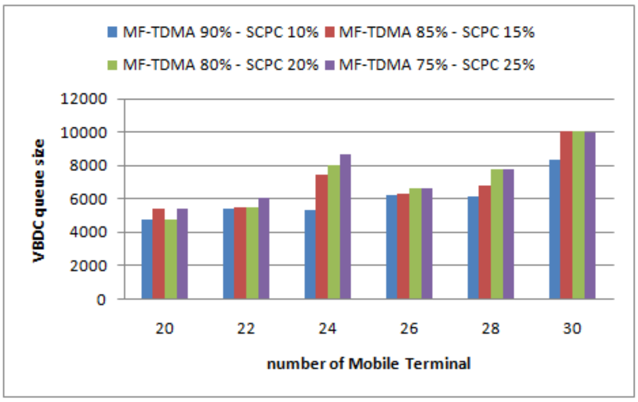

Fig. 5.33. VBDC queue size varying the capacity repartition between MF-TDMA and SCPC

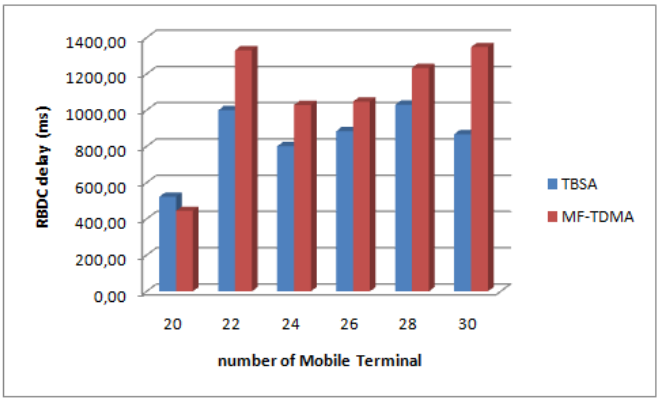

Fig. 5.34. RBDC delay comparison between classical MF-TDMA system and a system with a 90% of MF-TDMA and 10% of SCPC and the TBSA algorithm

possible. The results have shown that also in this case the better repartition of capacity between MF-TDMA and SCPC results to be 90% MF-TDMA and 10% SCPC. The comparison with the classical MF-TDMA scheme has shown that also in this case better system performance are reached with the hybrid system modality equipped with the TBSA algorithm.

Fig. 5.39 highlights the better repartition of return capacity for Scenario 2. Fig. 5.40 shows the switching number associated with this scenario.

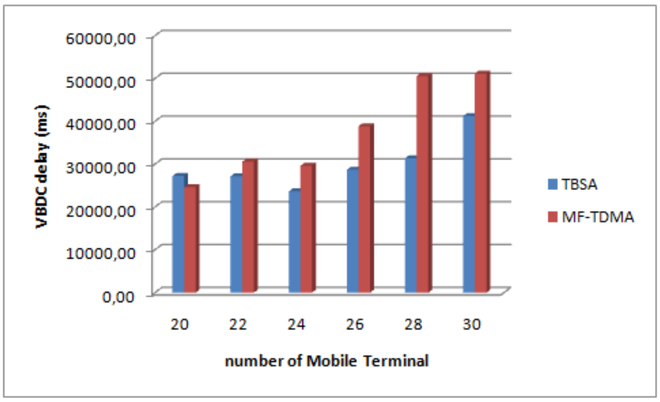

Fig. 5.35. VBDC delay comparison between classical MF-TDMA system and a system with a 90% of MF-TDMA and 10% of SCPC and the TBSA algorithm

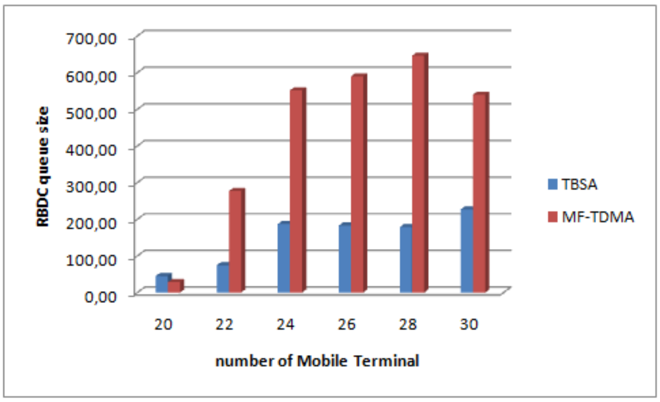

Fig. 5.36. RBDC queue size comparison between classical MF-TDMA system and a system with a 90% of MF-TDMA and 10% of SCPC and the TBSA algorithm

5.8.4 Simulation results for Scenario 3

The last considered scenario has been Scenario 3 in which a greater number of passengers using Video applications has been considered in order to understand the system behavior when a greater number of users use a heavier traffic as video conferences.

The same simulation campaign has been performed also for this scenario and the same consideration can be made. The hybrid system modality presents better performance with 90% MF-TDMA and 10% SCPC in respect to the other considered capacity repartitions. Compared with the classical

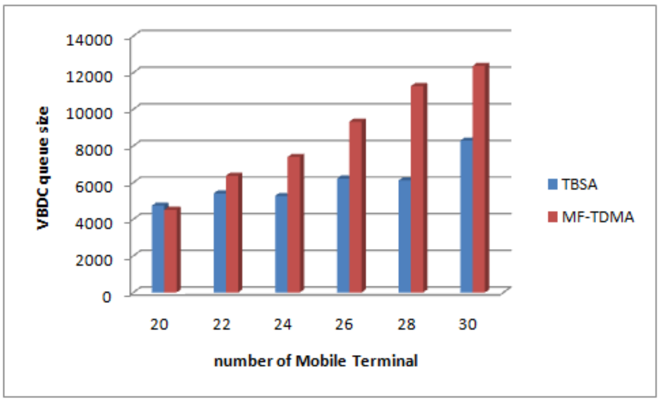

Fig. 5.37. VBDC queue size comparison between classical MF-TDMA system and a system with a 90% of MF-TDMA and 10% of SCPC and the TBSA algorithm

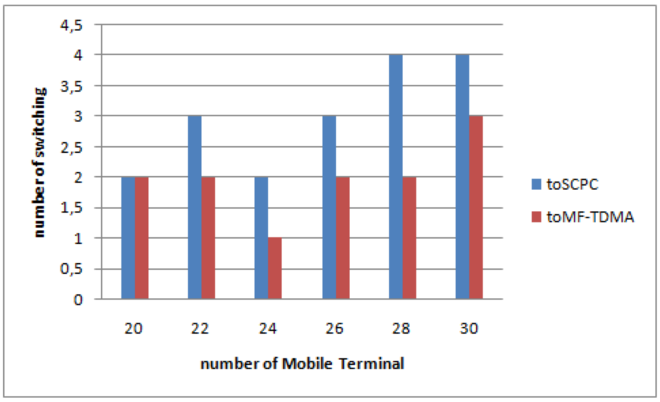

Fig. 5.38. Number of switching in the system in order to move from a modality to the other one

MF-TDMA scheme better performance in terms of delay and queue size have been found also in this case. Fig. 5.41 highlights the better repartition of return capacity for Scenario 3. The new consideration that it is possible to make is that in this case a different behavior in the switching number results (Fig. 5.42), in fact, for a great number of mobile terminals the switching is performed only in one direction, from MF-TDMA to SCPC because with a heavy traffic the throughput for the terminals is so high that no back switchings are necessary in order to have good performance in the system.

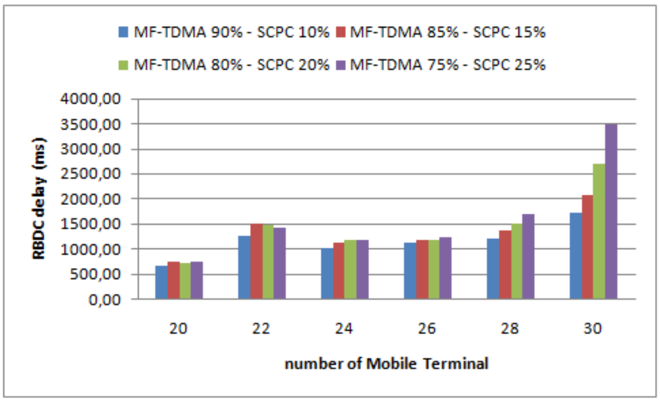

Fig. 5.39. RBDC delay varying the capacity repartition between MF-TDMA and SCPC

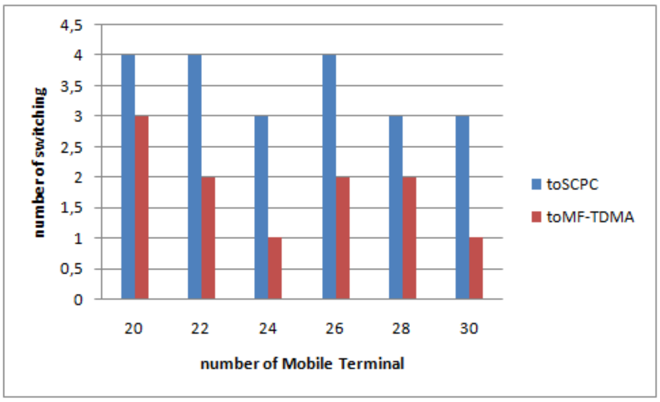

Fig. 5.40. Number of switching for scenario 2

5.9 Conclusions

In this chapter, the return link of DVB-RCS standard for mobile terminals has been analyzed. This version of the standard including the mobile feature was released in July 2008. The first part of the chapter shows the main characteristics of this new standard in order to understand the differences with the previous one for fixed terminals. Among the different mobile features included in this standard, this work has focused on the dual mode of operation of the RCST mobile terminal. This dual mode allows the terminal to switch its operation from the classical MF-TDMA assign-on-demand mode to a continuous carrier mode. The use of a continuous carrier is very useful for mobile

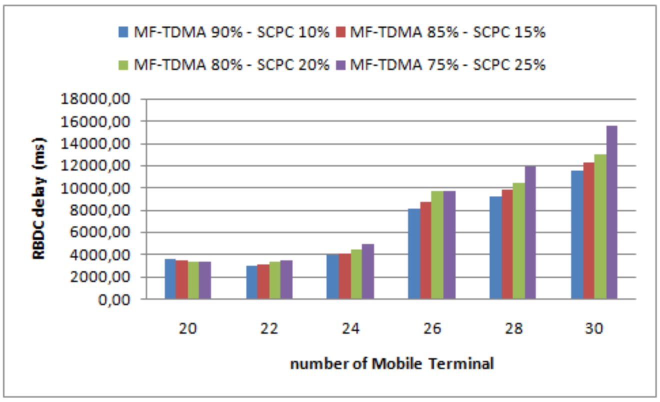

Fig. 5.41. RBDC delay varying the capacity repartition between MF-TDMA and SCPC

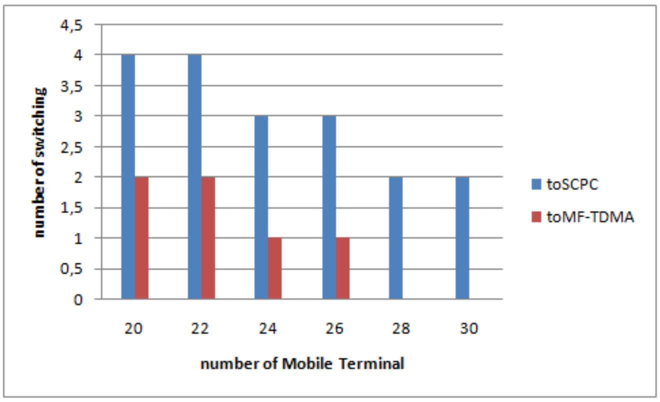

Fig. 5.42. Number of switching for scenario 3

terminals that, managing an aggregate of users, would need to perform a lot of capacity request to the satellite In such scenarios of the signaling overhead associated to the capacity assignment on demand would make the system decrease its efficiency.

In this work an algorithm for the switching between two modalities MF-TDMA and SCPC has been proposed. Threshold Basic Switching Algorithm (TBSA) meets the requirements imposed by the applications considered in the analysis. It is based on a threshold behavior and on the Basic RCST mode that is one of the two modalities suggested by the standard. The Basic RCST mode allows the mobile terminal to send the information either in MF-TDMA or in SCPC mode (although not simultaneously). This proposed algorithm (TBSA) is composed of two parts, one to make the right control for allowing the switching from MF-TDMA to SCPC mode based on the capacity request operated by the terminal and another one able to make the vice versa switching on the basis of a control on the mobile terminal throughput.

The algorithm bases its decision on two thresholds in both controls. A threshold states how many times the control has to be performed in order to get the switching decision, instead the other threshold is the term of comparison with the capacity request or throughput value. The behavior of the algorithm is simple: if the capacity-request/throughout value is greater than the considered threshold for a number of times stated by the first threshold, then the NCC, that is the satellite element that performs this algorithm, can send a TIM message toward the RCST in order to allow the changing of modality and then better exploit the system resource in order to have better efficiency in term of delay and queue size. Lower delay means better QoS provided by the network to the end users, reduced queue size means better resource management which is translated in a saving of capacity. Consequently, to the system will be able to handle a greater number of mobile terminals while satisfying the QoS required by users.

Conclusions

For a long time satellite systems have been an important communication element capable of putting in touch users who are separated by a great distance. The use of satellite platforms in cooperation with other network typologies is assuming an ever increasing importance in order to permit a wider coverage of the region and in order to manage a greater number of users. Owing to the great cost of satellite employment the transmissive resource of this platform still remains a precious resource and for this reason it has to be managed very carefully. Therefore, in this thesis the problem of the integration of heterogeneous networks and the problem of the resource management has been faced.

After showing the state of the art of the next generation network, this thesis exposes the problematic of integration of heterogeneous networks, like terrestrial, satellite and stratospheric platforms, which work all together in order to better provide services towards the end users and in order to reach most of the people in a region. It deals with the call admission control algorithm for the new multimedia applications and, finally, it explains a very new thematic, the DVB-RCS for mobile terminals whose standard is in the pipeline, only the blue book is available on the dvb.org web site.

The first chapter tries to give an exhaustive overview of the future next generation focusing in a special way on the satellite networks, which are recently becoming very important in order to provide multimedia services to users, since they are able to reach areas that the classical terrestrial networks are not able to cover. Moreover, it highlights the potential strengths of the DVB-RCS proposed by the ETSI. Furthermore, it introduces another platform, which in the last few years has captured the attention of researchers: the HAP. Finally, this chapter shows the importance, in current networks of the QoS concept, which tries to provide users the requirements in terms of delay, bandwidth and so on.

The second chapter tries to make a contribution to an important part of the current research that is attempting to study how to provide better services, using the integration of different networks, by explaining the potential role of an integrated Terrestrial/Satellite system based on QoS architecture. In this chapter it is shown how it is possible to make the integration between a terrestrial network with a satellite one and how it is possible to introduce in these networks the QoS architecture proposed in the literature, the IntServ and DiffServ scheme in order to satisfy the user requirements. Moreover, the scalability problem has been faced with the introduction of a new scalable architecture called SCORE, which diversifying the router operation, tries to facilitate the core router, thus allowing a more efficient management of the system resource.

The third chapter introduces a further integration problem giving the guidelines of an integration between a satellite networks with a HAPs mesh in order to permit a more efficient resource management. In particular, it is proposed to introduce a smart terminal capable of choosing dynamically the wireless segment in order to better use the system, by trying to avoid congesting only a wireless platform, foreclosing the possibility of managing connections that lead to particular delay requirements. In the chapter an algorithm is shown that on the basis of the delay required by the terminal and on the wireless platform capacity, allows the terminal the better choice in terms of wireless segment.

The fourth chapter focuses attention on another main problem in the management of system resource, the call admission control that avoids the introduction in the system of calls that could impact badly on the QoS constraints required by the admitted calls. This study shows how it is possible to model a video flow with a finite state and discrete time Markov chain and how this modeling permits a fine management of the aggregate video flow allowing a better resource management in the considered satellite system. The proposed model allows a great multiplexing gain for video source with high standard deviation around the average GOP rate. Thanks to this modeling a novel CAC has been proposed. This CAC takes into account the discretization of the bandwidth at a certain number of levels. This type of representation permits taking into account the time dependence of the video flow in order to better manage an aggregation of flows in the system. This algorithm was then compared with an approach note in the literature that does not take into account the time dependence and it has been shown that the algorithm proposed allows a more efficient use of the satellite transmissive resource.

The fifth and last chapter introduces a new topic for the satellite environment, the possibility of covering the mobile terminal in a satellite network. In particular, the new standard DVB-RCS+M has been studied and the hybrid modality suggested for the return channel has been analyzed in detail. This modality presented in the standard proposes the use of a continuous carrier operation in the return channel beyond the classical MF-TDMA modality. The continuous carrier operation, also called SCPC, can be very useful in the mobile environment where a satellite terminal manages an aggregate of users. The most promising market for this standard has been identified with public transportation, such as in aeronautical, maritime and terrestrial transportation. The key point of this hybrid modality is to find a mechanism in order to switch from one modality to another. In the chapter, an algorithm, called TBSA, is shown. This algorithm proposes a way in order to operate the switching between MF-TDMA and SCPC mode and vice versa. The simulation results have shown how this algorithm allows a more efficient use of the satellite return capacity in terms of delay and queue size. That is, it permits a better management of the QoS constraints required by the users and it permits a greater number of users to be managed by the system.

This thesis has shown how the integration of the satellite system with other typologies of networks, such as terrestrial networks or stratospheric platforms, is a possible way to allow a more efficient resource management and in order to coverage a wider number of users. Moreover, it shows how it is possible to introduce in hybrid platform QoS architectures in order to make the system capable of satisfying the QoS constraints required by the users. The integration with the terrestrial networks has shown how it is possible to use an aggregate reservation algorithm in order to perform an admission control of the calls in the system. Also, integration with the HAP platform has shown how it is possible to use a smart terminal capable of selecting dynamically the wireless segment. Moreover, a CAC for multimedia application has been proposed capable of taking into account the time dependence of a video flow in order to make the system able to manage a greater number of video sources. Finally, a resource management algorithm for a satellite architecture has been shown, which will be able to manage mobile users, called DVB-RCS+M showing how the satellite will play an important role in the future also for mobile applications.

References

1. ITU-T Rec. Y.2001,“General Overview of NGN”, Dec. 2004.
2. K. Knightson, N. Morita, T. Towle, “NGN Architecture: Generic Principles, Functional Architecture, and Implementation”, IEEE Communications Magazine, Oct. 2005.
3. Sastri L. Kota, “Satellite Multimedia Networks and Technical Challenges”, Microwave review, Nov. 2006.
4. SuemPingLoo, “System Design of an Integrated Terrestrial-Satellite Communications Network for Disaster Recovery”, MA Thesis, VirginiaPolytechnicInstituteandStateUniversity, Virginia, 2004.
5. ETSI, “Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems” ETSI EN 301 790 v.1.4.1, (2005-09).
6. ETSI, “Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790”, ETSI TR 101 790 V1.3.1, (2006-09).
7. DVB BlueBook, “Interaction channel for Satellite Distribution Systems (draft EN 301 790 V1.5.1 - DVB-RCS+M)”, 07 2008.
8. DVB-RCS Fact Sheet, “Return Channel Satellite - The open standard for two-way satellite broadband VSAT systems”, April 2008.
9. ETSI , “Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services”, ETSI EN 300 421 V1.1.2, (1997-08).
10. ETSI , “Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications”, Draft ETSI EN 302 307 V1.1.1, (2004-06).
11. D. Grace, J. Thornton, G. Chen, G.P. White, T.C. Tozer, “Improving the System Capacity of Broadband Services Using Multiple High-Altitude Platforms”, IEEE Transactions on Wireless Communications, vol. 4, no. 2, March 2005.
12. T.C. Tozer, D. Grace, “High-altitude platforms for wireless communications”, Electronics & Communication Engineering Journal, June 2001.

13. S. Karapantazis, F.-N. Pavlidou, "Broadban Communications via High-Altitude Platforms: a Survey", IEEE Communications Surveys & Tutorials, First Quarter 2005.
14. A. Durresi, S. Kota, "Real-Time Communications over Multilayer Satellite Networks", Vehicular Technology Conference (VTC), Sept. 2005.
15. R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.
16. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.
17. E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
18. S. Shenker, C. Partridge, R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.
19. J. Wroclawski, "Specification of the Controlled Load Quality of Service", RFC2211, September 1997.
20. P. White, "RSVP and Integrated Services in the Internet: A Tutorial", IEEE Communications Magazine, May 1997.
21. R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource Reservation Protocol (RSVP) Version 1 Functional Specification", RFC2205, September 1997.
22. J. Wroclawski, "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.
23. F. De Rango, A. Molinaro, S. Marano, "Guaranteed Services in Integrated Terrestrial-Broadband Satellite Networks", WPMC Symposium, Japan, Oct. 2003.
24. V. Jacobson, K. Nichols, K. Poduri, "An Expedited Forwarding PHB", RFC2598, June 1999.
25. J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.
26. A. Donner, M. Berioli, M. Werner, "MPLS-Based Satellite Constellation Networks" IEEE Journal on Selected Areas in Communications, vol. 22, no. 3, April 2004.
27. Z. Salcic, C.F. Meeklenbrauker, "Software radio - architectural requirements, research and development challenges", The 8th International Conference on Communication Systems (ICCS), vol. 2, pp. 25-28, Nov. 2002.
28. http://www.sdrforum.org
29. I. Stoica, "Stateless Core: A Scalable Approach for Quality of Service in the Internet", Ph.D. thesis, Carnegie Mellon Univiversity, Pittsburgh, PA, Dec. 2000.
30. G. Losquadro, "EUROSKYWAY: Satellite System for Interactive Multimedia Services", Proceedings of Ka-Band Utililization Conference, pp. 13-20, Sept. 1996.

31. G. Losquadro, M. Marinelli, “The EuroSkyWay system for interactive multimedia and the relevant traffic management”, Proceedings of the third Ka-band Utilisation Conference, pp. 17-24, September 1997.

32. I.Stoica, H.Zhang, “Providing Guaranteed Services without Per Flow Management”, Proc.ACM SIGCOM, Sept. 1999.

33. Y. Bernet, “The Complementary Roles of RSVP and Differentiated Services in the Full-Services QoS Network”, IEEE Communication Magazine, Feb. 2000.

34. F. Baker, C. Iturralde, F. Le Faucheur, B. Davie, “Aggregation of RSVP for IPv4 and IPv6 Reservations”, RFC3175, Sept. 2001.

35. Y. Bernet, “Format of the RSVP DCLASS Object”, RFC 2996, Nov. 2000.

36. A. Iera, A. Molinaro, S. Marano, “Call Admission Control and Resource Management Issues for Real-Time VBR Traffic in ATM-Satellite Networks”, IEEE Journal on Selected Areas in Communications, vol. 18, no. 11, Nov. 2000.

37. A.Iera, A.Molinaro, “Designing the Interworking of Terrestrial and Satellite IP-based Networks”, IEEE Communication Magazine, pp. 136-44, Feb. 2002.

38. M. Oodo, R. Miura, T. Hori, T. Morisaki, K. Kashiki, M. Suzuki, “Sharing and compatibility study between fixed service using high altitude platform stations (HAPs) and other services in the 31/28 GHz bands”, Wireless Personal Communications Journal, Vol. 23, pp.3-14, 2002.

39. E. Faletti, M. Mondin, F. Dovis, D. Grace, “Integration of a HAP within a terrestrial UMTS network: Interference analysis and cell dimensioning”, Wireless Personal Communications Journal, Vol. 24, pp. 291-325, 2003.

40. A. Iera, A. Molinaro, A. Marano, “IP with QoS guarantees via Geosatellite channels: Performance issues”, IEEE Personal Communication Magazine, pp. 14-19, June 2001.

41. F. De Rango , M. Tropea, S. Marano, “Aggregated Resource Reservation Protocol in Integrated Scalable-Terrestrial and Int-Serv Satellite Network”, Atti del convegno IEEE Wireless Communications and Networking Conference (WCNC2004), Atlanta, Georgia, USA, 21-25, March 2004.

42. C. Eklund, R.B. Marks, K.L. Stanwood, S. Wang, “IEEE Standard 802.16: A technical overview of the wirelessMAN air interface for broadband wireless access”, IEEE Communication Magazine, Vol. 40, No. 6, pp. 98-107, 2002.

43. GA.Jamalipour, “Satellites in IP networks”, Wiley Enciclopedia Telecommunications, Jan. 2003.

44. J.Neal, R,green, J.Landovskis, “Interactive channel for multimedia satellite networks”, IEEE Communications Magazine, vol.39, pp. 192-198, Mar. 2001.

45. D.J.Bem, T.W.Wieckowski, R.J.Zielinski, “Broadband satellite systems”, IEEE Communications Surveys & Tutorials, vol.3, no.1, 2000.

46. F. De Rango, M. Tropea, P. Fazio, S. Marano, “A Scalable Approach for QoS Management in Next Generation Multimedia GEO-Satellite Networks”, ASSI Satellite Communication Letter, Vol.6, pp.10-21, 2006.

47. ISO/IEC 11172: “Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s”.

48. ISO/IEC 13818: “Generic coding of moving pictures and associated audio (MPEG-2)”.
49. B.Haskell, A.Puri, A.Netravali, “Digital MPEG: An Introduction to MPEG-2”, London, U.K.: Chapman and Hall, 1997.
50. M. M. Krunz, A. M. Makowski, “Modeling video traffic using M/G/1 input processes: a compromise between Markovian and LRD models”, IEEE Journal on Selected Areas in Communications, vol. 16, no. 5, pp. 733-749, Jun. 1998.
51. O. Rose, “Simple and efficient models for variable bit rate MPEG video traffic”, Performance Evaluation, vol. 30, pp. 69-85, 1997.
52. M. Krunz and S. K. Tripathi, “On the characterization of VBR MPEG streams”, Measurement and Modeling of Computer Systems, pp. 192-202, Jun. 1997.
53. A.Lombardo, G. Morabito, G. Schembra, “Modeling intramedia and intermedia relationships in multimedia network analysis through multiple timescale statistics”, IEEE Transactions on Multimedia, Vol.6, no.1, pp. 142-157, February 2004.
54. A.Lombardo, G.Schembra, “Performance evaluation of an adaptive-rate MPEG encoder matching IntServ traffic constraints”, IEEE/ACM Transactions on Networking, Vol.11, no.1, pp. 47-65, Feb. 2003.
55. X. S. Wang, Moayeri, “Finite-state Markov channel - a useful model for radio communication channels”, IEEE Transactions on Vehicular Technology, Volume 44, Issue 1, pp. 163-171, Feb. 1995.
56. L.Huang, C.-C.Jay Kuo, “Joint Connection-Level and Packet Level Quality-of-Service Support for VBR Traffic in Wireless Multimedia Networks”, IEEE Journal on Selected Areas in Communications, vol.23, no.6, pp. 1167-1177, June 2005.
57. D.Niyato, E. Hossain, “Call Admission Control for QoS Provisioning in 4G Wireless Networks: Issues and Approaches”, IEEE Network, Sept./Oct. 2005.
58. F.Alagoz, B.R. Vojcic, D. Walters, A. AlRustamani, R.L. Pickholtz, “Fixed versus Adaptive Admission Control in Direct Broadcast Satellite Networks with Return Channel Systems”, IEEE Journal on Selected Areas in Communications, Vol.22, No.2, pp. 238-249, Feb. 2004.
59. F.Alagoz, “Approximation on the aggregate MPEG traffic and their impact on admission control”, Turkish J.Elec.Eng.Comput.Sci., vol.10, pp. 73-84, 2002.
60. W.-F. Poon, K.-T. Lo, J. Feng, “Interactive Broadcasting System for VBR Encoded Videos”, IEEE Transaction on Broadcasting, Vol.53, No.2, pp. 459-467, June 2007.
61. K.-M. Ho, W.-F. Poon, K.-T. Lo, “Performance Study of Large-Scale Video Streaming Services in Highly Heterogeneous Environment”, IEEE Transaction on Broadcasting, Vol.53, No.4, pp. 763-773, Dec. 2007.
62. T. Yoshihisa, M. Tsukamoto, S. Nishio, “A Broadcasting Scheme Considering Units to Play Continuous Media Data”, IEEE Transaction on Broadcasting, Vol.53, No.3, pp. 628-636, Sept. 2007.

63. G.-M. Muntean, "Efficient Delivery of Multimedia Streams Over Broadband Networks Using QOAS", IEEE Transaction on Broadcasting, Vol.52, No.2, pp. 230-235, June 2006.

64. R. Feghali, F. Speranza, D. Wang, A. Vincent, "Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate", IEEE Transaction on Broadcasting, Vol.53, No.1, pp. 441-446, March 2007.

65. S.R. Gulliver, G. Ghinea, "The Perceptual and Attentive Impact of Delay and Jitter in Multimedia Delivery", IEEE Transaction on Broadcasting, Vol.53, No.2, pp. 449-458, June 2007.

66. J.Banks, J.S.Carson, B.L.Nelson, D.M. Nicol, "Discrete Event System Simulatiom", Prentice Hall, Aug. 2000.

67. G. Boccolini, M. Luise, B. Garnier, J.-M. Merour, A. Brunelle, S. Titomanlio, V. Mignone, M.A. Sasse "A two-way interactive broadband satellite architecture to break the digital divide barrier", 16th Ka and broadband communications conference, Turin, Italy, 24-26 September, 2007.

68. H. Skinnemoen, "DVB-RSC: Revitalizing the Open Standard for Mobile Applications", 16th Ka and broadband communications conference, Turin, Italy, 24-26 September, 2007.

69. ESA Project, "Applications Layer QoS in DVB-RCS Systems", Nov. 2007.

70. ESA Project, "DVB-S(2)/DVB-RCS broadband mobile system", Feb. 2008.

71. ESA Project, "Harmonization of DVB-RCS Management and Control planes (HM&C)", Nov. 2007.

72. ESA Project, "Integrated QoS and Resources Management in DVB-RCS Networks", May 2004.

73. ESA Project, "Preparation for Internet to Trains Initiative: Broadband on Trains, Analysis of the Opportunity and Development Roadmap", Sept. 2006.

74. G. Acar, C. Kasparis, P. T. Thompson, "The Enhanced of DVB-S2 & DVB-RCS by adding additional mobile user capability", The Institution of Engineering and Technology Seminar on Digital Video Broadcasting Over Satellite: Present and Future, IEEE, Nov. 2006.

75. G. Matarazzo, P. Karouby, V. Schena, P. Vincent, "IP ON THE MOVE FOR AIRCRAFT, TRAINS AND BOATS - Innovative satellite telecommunications systems offer true broadband services to travelers", Technology White Paper, Alcatel Telecommunications Review - 2nd Quarter 2006.

76. M. Cot, L. Erup, M. Lambert, N. McSparron, "Implementation Challenges and Synergistic Benefits of DVB-S2 and DVB-RCS", The Institution of Engineering and Technology Seminar on Digital Video Broadcasting over Satellite: Present and Future, IEEE, Nov. 2006.

77. D. Gil Oh, P. Kim, Y.J. Song, S. Ik Jeon, H.-J. Lee, "Design Considerations of Satellite-Based Vehicular Broadband Networks", IEEE Wireless Communications, October 2005.

78. A. Morello, V. Mignone, "DVB-S2: The Second Generation Standard for Satellite Broad-band Services", Proceedings of the IEEE, vol. 94, no. 1, January 2006.

79. Ho-Jin Lee, P. Kim, T. Kim, D. Oh, “Broadband systems based on DVB-S2 and mobile DVB-RCS, and their future applications to broadband mobiles”, International Workshop on Satellite and Space Communications, IEEE IWSSC ’06, Sept. 2006.

80. A. Juoras, C. Morlet, “Network and Mobility Management for Mobile DVB-S2/DVB-RCS systems”, International Workshop on Satellite and Space Communications, IEEE IWSSC ’07, Sept. 2007.

81. H. Peyravi, “Medium Access Control Protocols Performance in Satellite Communications”, IEEE Communications Magazine, Mar. 1999.

82. C. Morlet, A. Bolea Alamanfac, A. Ginesi, G. Gallinaro, L. Erup, P. Takats, “Implementation of Spreading Techniques in Mobile DVB-S2/DVB-RCS systems”, International Workshop on Satellite and Space Communications, IEEE IWSSC ’07, Sept. 2007.

83. C. Morlet, A. Ginesi, “Introduction of Mobility Aspects for DVB-S2/RCS Broadband Systems”, International Workshop on Satellite and Space Communications, IEEE IWSSC ’06, Sept. 2006.

84. B. Matuz, G. Liva, C. Parraga Niebla, N. Riera Diaz, S. Scalise, “Link Layer Coding for DVB-S2 Interactive Satellite Services to Trains”, Vehicular Technology Conference, IEEE VTC Spring May 2008.

85. S. Scalise, J. Huguet Guasch, V. Schena, and F. Ceprani, “Link Performance for a Satellite-Based Communications System for Fast Trains: Analysis of Trials and Measurements”, proceedings of the 6th European Mobile and Personal Satellite Workshop & 2nd Advanced Satellite Mobile Ssystems Conference, Noordwijk, Holland, 2004.

86. M. lvarez Daz, S. Scalise, G. Sciascia, R. Mura, P. Conforto, H. Ernst, “DVB-S Air Interface over Railroad Satellite Channel: Performance and Extensions”, Sixth Baiona Workshop on Signal Processing in Communications, Sept. 2003.

87. S. Scalise, H. Ernst, G. Harles, “Measurement and Modeling of the Land Mobile Satellite Channel at Ku-Band”, IEEE Transactions On Vehicular Technology, vol. 57, no. 2, March 2008.

88. S. Scalise, R. Mura, V. Mignone, “Air Interfaces for Satellite Based Digital TV Broadcasting in the Railway Environment”, IEEE Transactions on Broadcasting, Vol. 52, No. 2, June 2006.

89. G. Pasolini, A. Giorgetti, “DVB-S Gap Fillers for Railway Tunnels”, IEEE 64th Vehicular Technology Conference, VTC-Fall 2006, 2006.

90. Liva, N. Riera Diaz, S. Scalise, B. Matuz, C.Parraga Niebla “Gap Filler Architectures for Seamless DVB-S2/RCS Provision in the Railway Environment”, IEEE Vehicular Technology Conference, VTC Spring 2008, 2008.

91. P. Vincent, A. Arcidiacono, N. Chevet, L. Audounet, G. Naym, J. Alvarez, L. Babarit, A. Vaccaro, M. Le Saux, M. Holzbock, R. Lo Forti, C. Charatsaris, R. Alvarez, P. Bates, M. Smith, B.G. Evans, “ ’Mobile Wideband Global Link sYstem’ (MOWGLY) - Aeronautical, Train and Maritime Global High-Speed Satellite Services”, AIAA, 2005.

92. V. Schena, F. Ceprani, “FIFTH Satellite Project Solutions Demonstrating New Broadband Communication System for High Speed Train”, IEEE 59th Vehicular Technology Conference, VTC Spring 2004, 2004.

93. A. Yun, J. Prat, "Amerhis: DVB-RCS meets mesh connectivity", White Paper, Satlabs, 2007.

94. D. Sivchenko, B. Xu, G. Zimmermann, S. Hischke, "Internet Traffic Performance in High Speed Trains", Performance Modelling and Evaluation of HETerogeneous NETworkS (HET-NETs), 2004.

95. D. Staehle, K. Leibnitz, and P. Tran-Gia, "Source Traffic Modeling of Wireless Applications", Report No. 261, Universitt Wrzburg, June 2000.

96. Bruce A. Mah, "An Empirical Model of HTTP Network Traffic", IEEE Infocom, 1997.

97. J.J. Lee, M. Gupta, "A New Traffic Model for Current User Web Browsing Behavior", Intel Coorporation, 2007.

98. A. Andreadis, G. Giambene, "Protocols for High-Efficiency Wireless Networks", Kluwer Academic Publishers, 2002.

99. C. Parraga, C. Kissling, E. Lutz, "Design and Performance Evaluation of Efficient Scheduling Techniques for Second Generation DVB-S Systems", Proceedings ICSSC-2005, 23rd AIAA International Communications Satellite System Conference (ICSSC-2005), Rome, Italy, 25 - 28 September 2005.

100. TM-RCS0922, "Very Short Frame Size Transmission Scheme: Performance Results for all Applicable Modulation Schemes and Use of BCH Code", May 2008.

List of Publications

Mauro Tropea graduated in computer engineering at the University of Calabria, Italy, in 2003. Since 2003 he has been with the telecommunications research group of D.E.I.S. in the University of Calabria. In 2004 he won a regional scholarship on Satellite and Terrestrial broadband digital telecommunication systems. Since November 2005 he has a Ph.D student in Electronics and Communications Engineering at University of Calabria. His research interests include satellite communication networks, QoS architectures and interworking wireless and wired networks, mobility model.

International Journal

J1. De Rango F., Tropea M., Fazio P., Marano S., "Call Admission Control for Aggregate MPEG-2 Traffic over Multimedia Geo-Satellite Networks," to be published on IEEE Transaction on Broadcasting.

J2. De Rango F., Veltri F., Tropea M., Santamaria A.F., Fazio P., Malfitano A., Marano S., "Interdisciplinary issues for the management of next generation autonomic wireless systems: nature-inspired techniques and organic computing", to be published on International Journal Mobile Network Design and Innovation, 2008.

J3. De Rango F., Tropea M., Veltri F., Marano S., "GS Burst Loss Percentage Analysis over an IntServ Satellite System with a Mixed GS-CLS Traffic". IETE Journal of Research, 2008.

J4. De Rango F., Tropea M., Santamaria A., Marano S., "An Enhanced QoS CBT Multicast Routing Protocol based on Genetic Algorithm in a Hybrid HAP-Satellite System". Computer Communication Journal (Elsevier), 2007, pp. 1-21.

J5. Loscri' V., Tropea M., Marano S., "Voice and Video Telephony Services in Smartphone". EURASIP Journal, 2006, pp. 1-8.

J6. De Rango F., Tropea M., Fazio P., Marano S., "Overview on VoIP: Subjective and Objective Measurement Methods". International journal of computer science and network security, 2006, Vol. 6, n. 1B, pp. 140-153.

J7. De Rango F., Tropea M., Fazio P., Marano S., "A Scalable Approach for QoS Management in Next Generation Multimedia GEO-Satellite Networks". ASSI Satellite Communication Letter, 2006, Vol.6, pp.10-21.

J8. De Rango F., Tropea M., Marano S., "Integrated Services on High Altitude Platform: Receiver Driven Smart Selection of HAP-Geo Satellite Wireless Access Segment and Performance Evaluation". International Journal of Wireless Information Networks, 2006, pp. 1-18.

J9. Molinaro A., De Rango F., Marano S., Tropea M., "A scalable framework for end-to-end QoS assurance in IP-oriented terrestrial-satellite networks". IEEE Communications Magazine, 2005, Vol. 43, pp. 130-137.

International Conference

C1. De Rango F., Tropea M., Provato A., Santamaria A.F., Marano S., "Multiple Metrics Aware Ant Routing over Hap Mesh Networks," in IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2008), Niagara Falls, Ontario, Canada, May 4-7, 2008.

C2. De Rango F., Tropea M., Laratta G.B., Marano S., "Hop-by-Hop Local Flow Control over InterPlaNetary Networks based on DTN Architecture", IEEE International Conference on Communication (ICC), 2008-01-24.

C3. De Rango F., Santamaria A.F., Tropea M., Marano S., "Meta-heuristics methods for a NP-Complete networking problem," 66th IEEE Vehicular Technology Conference (VTC Fall 2008), Calgary, Alberta, Canada, 21-24 Sept. 2008.

C4. De Rango F., Tropea M., Provato A., Santamaria A., Marano S., "Multi-Constraints Routing Algorithm based on Swarm Intelligence over High Altitude Platforms", International Workshop on Nature Inspired Cooperative Strategies for Optimization (NICSO), Acireale, Sicilia, November 8-10, 2007.

C5. De Rango F., Tropea M., Santamaria A., Veltri F., Fazio P., Marano S., "Multi-Mode DVB-RCS Satellite Terminal with Software Defined Radio". Atti del convegno "IEEE Wireless Communication and Networking Conference (WCNC 2007)", Hong Kong, China, March 11-15, 2007.

C6. De Rango F., Santamaria A., Veltri F., Tropea M., Fazio P., Marano S., "Multi-Satellite DVB-RCS System with RCST based on Software Defined Radio". Atti del convegno "65th IEEE Vehicular Technology Conference (VTC Fall 2007)", Baltimore, USA, Oct. 22-25, 2007.

C7. De Rango F., Tropea M., Provato A., Santamaria A., Marano S., "Routing Algorithm based on Swarm Intelligence over a Hap Constellation". Atti del convegno "10th International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2007)", San Diego, CA, USA, July 16-18, 2007.

C8. De Rango F., Santamaria A., Tropea M., Milanese G., Marano S., "QoS Multicast Routing Protocol for Broadband hierarchal network". Atti del convegno "13th Ka and Broadband Communication Conference (KaBand 2007)", Turin, Italy, Sept. 24-26, 2007.

C9. De Rango F., Veltri F., Fazio P., Santamaria A., Tropea M., Marano S., "FER Regression Analysis of DS-UWB-based WPAN". Atti del convegno "Wireless Telecommunication Symposium (WTS 2007)", Pomona, CA, USA, Apr. 27-29, 2007.

C10. De Rango F., Veltri F., Santamaria A., Tropea M., Marano S., "Software Defined Radio-based Multi-Mode DVB-RCS Terminals". Atti del convegno "2007 Military Communications Conference (MILCOM 2007)", Orlando, Florida, USA, Oct. 29-31, 2007.

C11. De Rango F., Tropea M., Gentile A.F., Provenzano A., Marano S., "A Novel Proposal of TCP Protocol based on Bandwidth Estimation over Satellite Networks". Atti del convegno "13th Ka and Broadband Communication Conference", Turin, Italy, Sept.24-26, 2007.

C12. De Rango F., Laratta G.B. Tropea M., Codispoti G., Marano S., "Dynamic Routing over Interplanetary Networks based on DTN approach and on the Link Failures prediction". Atti del convegno "13th Ka and Broadband Communication Conference", Turin, Italy, Sept. 24-26, 2007.

C13. De Rango F., Veltri F., Santamaria A., Tropea M., Fazio P., Marano S., "Software Defined Radio based Multi-Mode Satellite Terminal over DVB-RCS Platform". Atti del convegno". 12th Ka and Broadband Communications Conference", Naples, Italy, Sept. 27-29, 2006.

C14. De Rango F., Tropea M., Marano S., "The Important Role of Gateway in a Hybrid HAP/DVB-RCS Satellite Platform". Atti del convegno". 13rd International Conference on Telecommunications (ICT 2006)", Madeira Island, Portugal, May 9-12, 2006.

C15. De Rango F., Tropea M., Santamaria A., Marano S., "QoS Multicast Routing Algorithm over Hybrid HAP-Satellite Networks". Atti del convegno". 12th Ka and Broadband Communications Conference", Naples, Italy, Sept. 27-29, 2006.

C16. De Rango F., Tropea M., Santamaria A., Marano S., "A QoS Multicast Genetic Algorithm in a Hybrid HAP/DVB-RCS Satellite Platform". Atti del convegno "13rd International Conference on Telecommunications (ICT 2006)", Madeira Island, Portugal, May 9-12, 2006.

C17. De Rango F., Tropea M., Santamaria A., Marano S., "QoS Core Based Tree Multicast Routing based on Genetic Algorithm in a Hybrid HAP-Satellite Architecture". Atti del convegno "International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2006)", Calgary, Canada, July 31 - August 2, 2006.

C18. De Rango F., Tropea M., Fazio P., Marano S., "Call Admission Control with Statistical Multiplexing for Aggregate MPEG Traffic in a DVB-RCS Satellite Network". Atti del convegno "IEEE Global Telecommunication Conference (Globecom'05)", St.Louis, MO, USA, 28 Nov. - 2 Dec., 2005.

C19. De Rango F., Tropea M., Fazio P., Loscr V., Marano S., "Scalable QoS Management in Next Generation GEO-Satellite Networks," Int. Symposium on Performance Evaluation of Comp. and Telecomm.Systems (SPECTS'05), 24-28 July, 2005.

C20. De Rango F., Tropea M., Marano S., "Controlled Load Services in IP QoS Geostazionary Satellite Networks". Atti del convegno "11th International Conference on Telecommunications (ICT 2004)", Fortaleza, Brasil, 1-6 August, 2004.

C21. De Rango F., Tropea M., Marano S., "Controlled Load Services Management based on Smoothing Factor and Request Timeout on Satellite Systems". Atti del convegno "15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC04)", Barcelona, Spain, 2004.

C22. De Rango F., Tropea M., Marano S., "HAP layer with Satellite Systems for Enhanced Integrated Services over IP Wireless Networks". Atti del convegno "Vehicular Technology Conference 2004 Fall (VTC Fall 2004)", Los Angeles, CA, USA, 2004.

C23. De Rango F., Tropea M., Marano S., "Controlled Load Sercive Management in Int-Serv Satellite Access Network". Atti del convegno "Canadian Conference on Electrical and Computer Engineering (CCECE 2004)", Ontario, Canada,, May, 2004.

C24. De Rango F., Tropea M., Marano S., "Controlled Load Traffic Management through bandwidth smoothing factor". Atti del convegno "7th International Symposium on Wireless Personal Multimedia Communications (WPMC'04)", Albano Terme, Italy, 2004.

C25. De Rango F., Tropea M., Marano S., "Predictive Aggregate Resource Reservation in an Integrated Scalable Terrestrial-Geostazionary Satellite Network". Atti del convegno "7th International Symposium on Wireless Personal Multimedia Communications (WPMC'04)", Albano Terme, Italy, 12-15 Sept., 2004.

C26. De Rango F., Tropea M., Marano S., "Aggregated Resource Reservation Protocol in Integrated Scalable-Terrestrial and Int-Serv Satellite Network". Atti del convegno "IEEE Wireless Communications and Networking Conference (WCNC2004)", Atlanta, Georgia, USA, 21-25 March, 2004.

C27. De Rango F., Tropea M., Marano S., "Receiver driven Adaptive Selection of HAP-Satellite Segment". Atti del convegno "7th International Symposium on Wireless Personal Multimedia Communications (WPMC'04)", Albano Terme, Italy, 12-15 Sept., 2004.

C28. De Rango F., Tropea M., Marano S., "Call Admission Control for Integrated Diff-Serv Terrestrial and Int-Serv Satellite Network". Atti del convegno "IEEE Vehicular Technology Conference (VTC2004 Spring)", Genova, Italy, 17-19 May, 2004.