International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R S.1897
(01/2012)

Cross-layer QoS provisioning in IP-based hybrid satellite-terrestrial networks

S Series

Fixed-satellite service

ITU

International
Telecommunication
Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at **http://www.itu.int/publ/R-REC/en**)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: *This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2012

© ITU 2012

RECOMMENDATION ITU-R S.1897

Cross-layer QoS provisioning in IP-based hybrid satellite-terrestrial networks

(2011)

Scope

This Recommendation provides guidelines about implementing cross-layer design approaches for improving the performance of multimedia applications over satellite networks (either stand-alone or hybrid).

The ITU Radiocommunication Assembly,

considering

a)   that satellite systems are increasingly being used for providing broadband applications directly to users in addition to their role as backbone links;

b)   that fading events ultimately impact the application layer of services provided by such satellite systems;

c)   that cross-layer design consists in allowing interactions and sharing state variables among different protocol layers (including non-adjacent ones) in order to achieve network capacity and performance gain;

d)   that a cross-layer approach conceived to achieve a better adaptation to the satellite transmission dynamics has the potential to mitigate the effect of fading events;

e)   that cross-layer approaches may also be used to adapt the transmission requirements in case of an event affecting the available bandwidth;

f)   that the overall performance of a satellite link may be affected by a variety of factors (delays, delay variations, throughput, etc.) that can be monitored through a cross-layer approach,

noting

a)   that Recommendation ITU-R S.1711 provides guidelines to implement a number of transmission control protocol (TCP) performance enhancements over IP-based satellite networks;

b)   that various cross-layer applications have been studied and experimented (see Report ITU-R S.2222),

recommends

1   that a cross-layer design approach should be used for satellite networks (either stand-alone or hybrid);

2   that the reference architectures, set out in Annex 1 of this Recommendation, should be considered as a basis when implementing cross-layer design approaches;

3   that, when assessing the impact of fading events into different layers, Annex 2 should be considered;

4   that cross-layer design approaches in satellite links employing TCP should be considered in assessing the throughput and delay performance (see Annex 3).

NOTE − Report ITU-R S.2222 provides background material on cross-layer design concepts and methodologies.

TABLE OF CONTENTS

List of acronyms

3G	Third generation
4G	Fourth generation
AAL	ATM adaptation layer
ACELP	Algebraic code excitation linear prediction
ACK	Acknowledgment
ACM	Adaptive coding and modulation
AMR-WB	Adaptive multirate wideband
APP	Application
ARQ	Automotive repeat request
ASN	Access service network
ATM	Asynchronous transfer mode
AWGN	Additive white Gaussian noise
BDP	Bandwidth delay product
BIC	Binary increase congestion control
BOD	Bandwidth on demand
BPSK	Binary phase-shift keying
CCM	Constant coding and modulation
CMR	Code model request
CQI	Channel quality information
CRA	Continuous-rate assignment
C-TCP	Compound TCP
DAMA	Demand assignment multiple access
DLSR	Delay since last sender report
DVB	Digital video broadcast
DVB-RCS	Digital video broadcast – return channel by satellite
DVB-S	Digital video broadcast by satellite
DVB-S2	Digital video broadcast – satellite transmission 2nd generation
EF	Expedited forward
ETH	Ethernet
ETSI	European Telecommunications Standards Institute
FEC	Forward error correction
FSS	Fixed satellite service
FTP	File transfer protocol
GEO	Geostationary earth orbit
GSM	Global system for mobile communications

GSM-AMR	GSM adaptive multi-rate
GSM-EFR	GSM enhanced full rate
GT	Gateway terminal
GW	Gateway
HTTP	Hypertext transfer protocol
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet protocol
ISO	International Organization for Standardization
ITU-R SG	ITU-R Study Group
ITU-R WP	ITU-R Working Party
LAN	Local area network
LOS	Line of sight
MAC	Medium access control
MAC-CPS	MAC common part sublayer
MAC-CS	MAC convergence sublayer
MF-TDMA	Multiple-frequency time-division multiple access
MIMD	Multiplicative increase multiplicative decrease
MODCOD	Modulation and coding
MPEG	Moving picture experts group
NCC	Network control center
OFDM	Orthogonal frequency division multiplexing
OSI	Open system interconnect
PDU	Packet data unit
PER	Packet error rate
PHY	Physical layer
PSK	Phase shift keying
QoS	Quality of service
QPSK	Quadrature phase-shift keying
RCS	Return channel satellite
RCST	Return channel satellite terminal
RR	Receiver report
RRA	Radio resource allocation
RRM	Radio resource management
RT	Real time
RTCP	Real time transport control protocol
RTO	Retransmission on time out

RTT	Roundtrip time
SACK	Selective acknowledgement
SNIR	Signal to noise interference ratio
SR	Sender report
S-TCP	Scalable TCP
TCP	Transmission control protocol
TDM	Time division multiplexing
TDMA	Time division multiple access
TLSR	Time of last sender report
ToS	Type of service
TV	Television
UDP	User datagram protocol
UE	User equipment
VCM	Variable coding and modulation
VoIP	Voice over Internet protocol
VR-JT	Variable rate real time traffic jitter tolerant
VR-RT	Variable rate real time traffic
WiFi	Wireless fidelity (products based on IEEE 802.11 standards)
WiMAX	Worldwide interoperability for microwave access

Annex 1

Reference architectures for implementing cross-layer design approaches
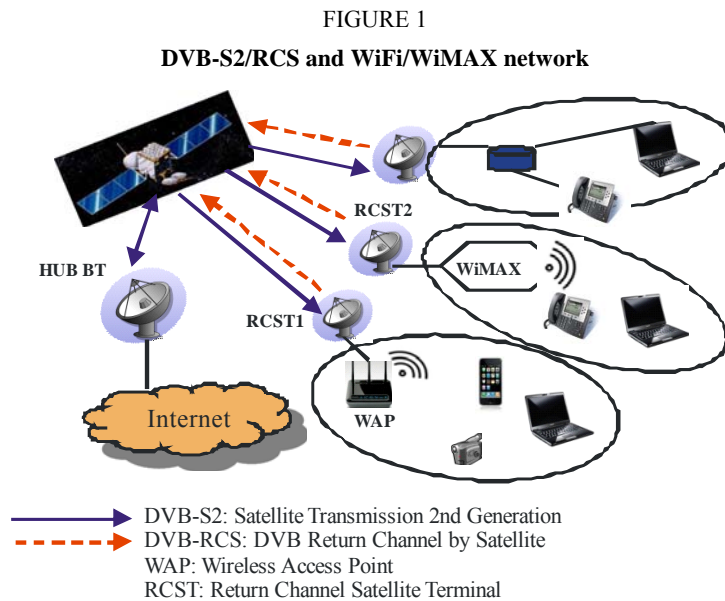
1 Scope

This Annex presents reference architectures for hybrid satellite – wireless networks including satellite links and either a WiFi or WiMAX terrestrial segment. This will be followed by a description of quality of service (QoS) improvement in multimedia networks using cross-layer design approaches.

2 Reference architectures

Figure 1 shows a hybrid satellite-wireless network operating at Ka-band to support multimedia applications. Various scenarios could include a Geostationary Earth Orbit (GEO) satellite system with digital video broadcast (DVB-S2/return channel satellite (RCS)) air interface connected with either a WiFi and/or a WiMAX terrestrial segment. As shown in this figure, DVB-RCS terminals (RCSTs) could also directly support applications such as VoIP, streaming multimedia, video conferencing, and bulk data transfer. The system is composed of gateway terminals (GTs), RCSTs

and network control and management center. The forward link, i.e., from the gateway to the user terminal (solid blue arrows) follows DVB-S2 with adaptive coding and modulation (ACM). The return link from the terminal to gateways (dashed red arrows) is based on DVB-RCS.

FIGURE 1

DVB-S2/RCS and WiFi/WiMAX network

DVB-S2: Satellite Transmission 2nd Generation
DVB-RCS: DVB Return Channel by Satellite
WAP: Wireless Access Point
RCST: Return Channel Satellite Terminal

S.1897-01

The DVB-S2 features two main enhancements compared with its predecessor, DVB-S. First, it introduces an improved physical layer, offering several higher order modulation waveforms with more powerful forward error correction (FEC). Secondly, it supports real-time adaptation to link and propagation conditions. It supports 28 combinations of modulation format and coding schemes to guarantee a low packet error rate across a wide range of signalled noise plus interference ratio (SNIR). The three operational modes supported include (a) constant coding and modulation (CCM) (b) variable coding and modulation (VCM) and (c) adaptive coding and modulation (ACM).

DVB-RCS used on the return link implements multi-frequency − time-division multiple access (MF-TDMA) and adaptive coding. The return link MF-TDMA enables it to have bi-dimensional framing in which every time-frequency window is portioned into carriers, super frames, frames and slots. The MF-TDMA return link is coded with concatenated Convolution and Reed Solomon codes. The data may be encapsulated in Asynchronous Transfer Mode (ATM) cells using ATM Adaptation layer 5 or it may use native IP encapsulation over MPEG-2. These protocols allow IP traffic transmission over the physical layer which is used in simulation experiments. Rain becomes the most affecting atmospheric event for transmission in the Ka-band. Therefore, the effect of fading on various MAC and application layer parameters using cross-layer design approach must be evaluated. The differentiated services (DiffServ) model is assumed to prioritize IP-based networks interfacing with WiFi and WIMAX segments.

2.1 Hybrid satellite-WiFi network architecture

To provide broadband connectivity to areas of both low population density network (i.e. rural areas) and high population density (i.e., suburban and urban areas), hybrid networks are composed of both satellite and terrestrial radio access technologies.

FIGURE 2

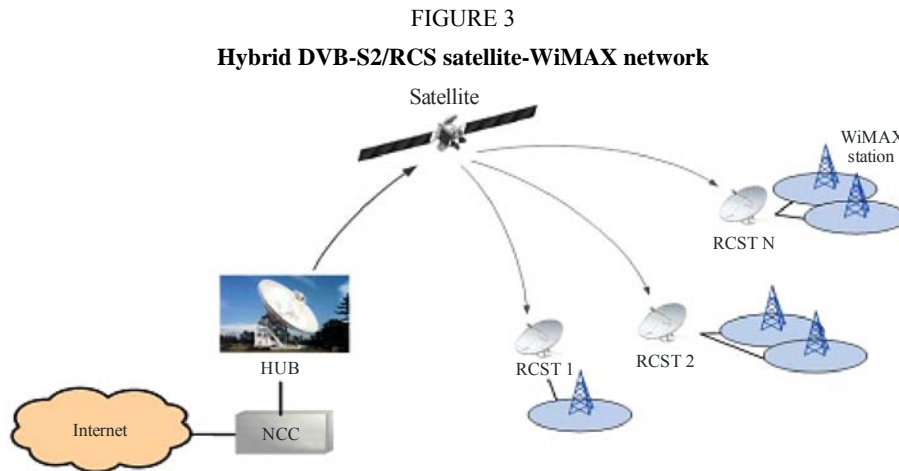**Satellite – Wireless network protocol architecture**
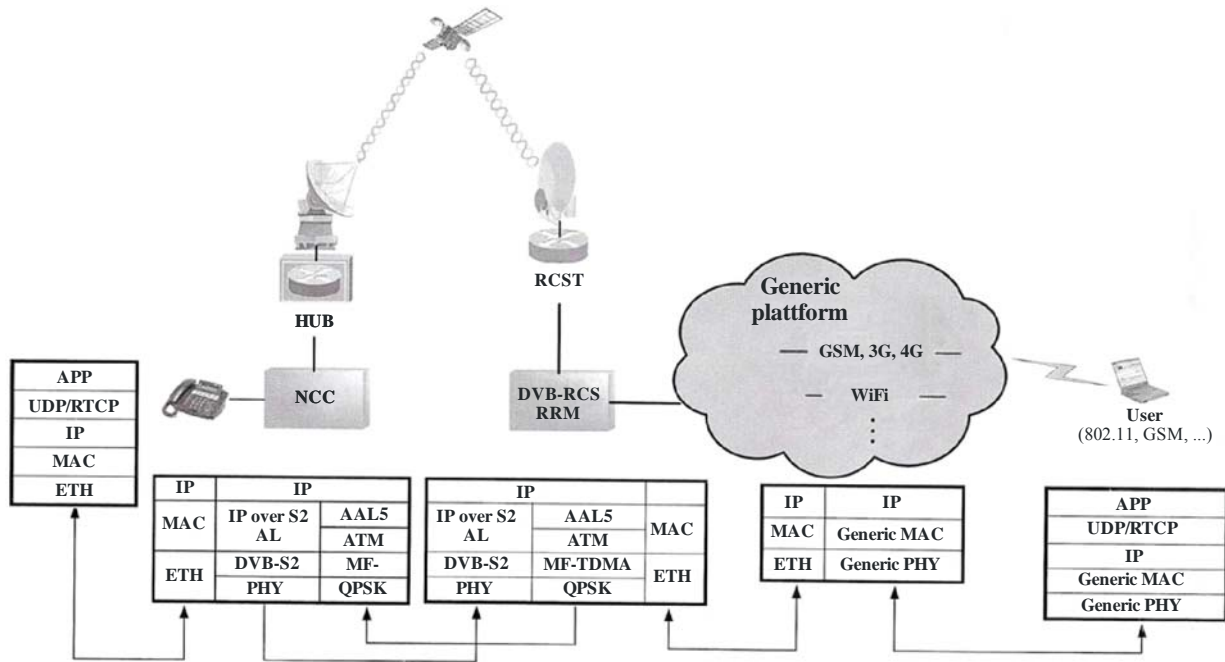
S.1897-02

Figure 2 shows the satellite-terrestrial wireless network protocol architecture. The wireless segment can use protocols such as GSM, 3G, WiFi, WiMAX, and 4G. Both the satellite segment and the terrestrial network will provide resource allocation algorithms and a control management system.

Other components in the network architecture include the access service network gateway, and the DVB-RCS Radio Resource Management (RRM). The RRM of the satellite terminal checks if enough resources are available to enable admission of new user equipment requesting services from the gateway to the satellite link. The RCST communicates with the Hub, which is associated to a network control center (NCC). The NCC controls the interactive network, user service request via satellite access and manages the satellite spectrum depending on the satellite terminals requests.

2.2 Hybrid satellite-WiMAX network architecture

Figure 3 shows hybrid satellite network using DVB-S2/RCS terminals connected with a WiMAX network.

FIGURE 3

Hybrid DVB-S2/RCS satellite-WiMAX network

S.1897-03

2.2.1 Hybrid satellite-WiMAX network protocol architecture

Figure 4 shows the hybrid satellite-WiMAX network protocol architecture. The terrestrial link is based on IEEE 802.16, where a user in the core network directs traffic (e.g. VoIP conversation) to a mobile user, called User Equipment (UE). The UE is placed in an area serviced by the WiMAX network. The Base Station (BS) is responsible of the IEEE 802.16 connectivity through the radio link to the UE located inside its coverage area. An adaptive physical layer maximizes the data rate by adjusting transmission modes to channel variations while maintaining a prescribed Packet Error Rate (PER). The WiMAX Radio Resource Management (RRM) is in charge of utilizing the limited radio spectrum resources and radio network infrastructure of its associated BS efficiently using strategies and algorithms for controlling parameters. Figure 4 includes the Access Service Network (ASN) gateway, and the DVB-RCS RRM. The RRM of the satellite terminal checks if enough resources are available to enable admission of new UEs requesting services from the gateway to the satellite link. The RCST communicates with the Hub, which is associated with a Network Control Center (NCC). The NCC controls the interactive network, user service requests via satellite access and manages the satellite spectrum depending on the satellite terminals requests. The protocol stack of the WiMAX RRM consists of three sub-layers forming the whole MAC layer. The Convergence Sublayer (MAC-CS) provides the transformation or mapping of external network data (e.g. Ethernet, IP). The Common Part Sublayer (MAC-CPS) performs packing into MAC packet data unit (MPDU) of information coming from MAC-CS, and the Privacy Sub-layer, provides authentication, key exchange and data encryption. Main features of the IEEE 802.16 Physical layer are the utilization of Orthogonal Frequency Division Multiplexing (OFDM), Time Division Multiplexing (TDM) and power control in the S-Band (principally around 3.5 GHz). MAC-CPS is the core of MAC layer. It provides QoS, manages bandwidth, multiplexes VoIP flows directed to the BS, establishes and maintains the connection, performs Forward Error Correction (FEC), and enables Automatic Repeat Request (ARQ) mechanisms.

FIGURE 4

Reference architecture and protocol stack for IEEE 802.16e-2005 / DVB-RCS network
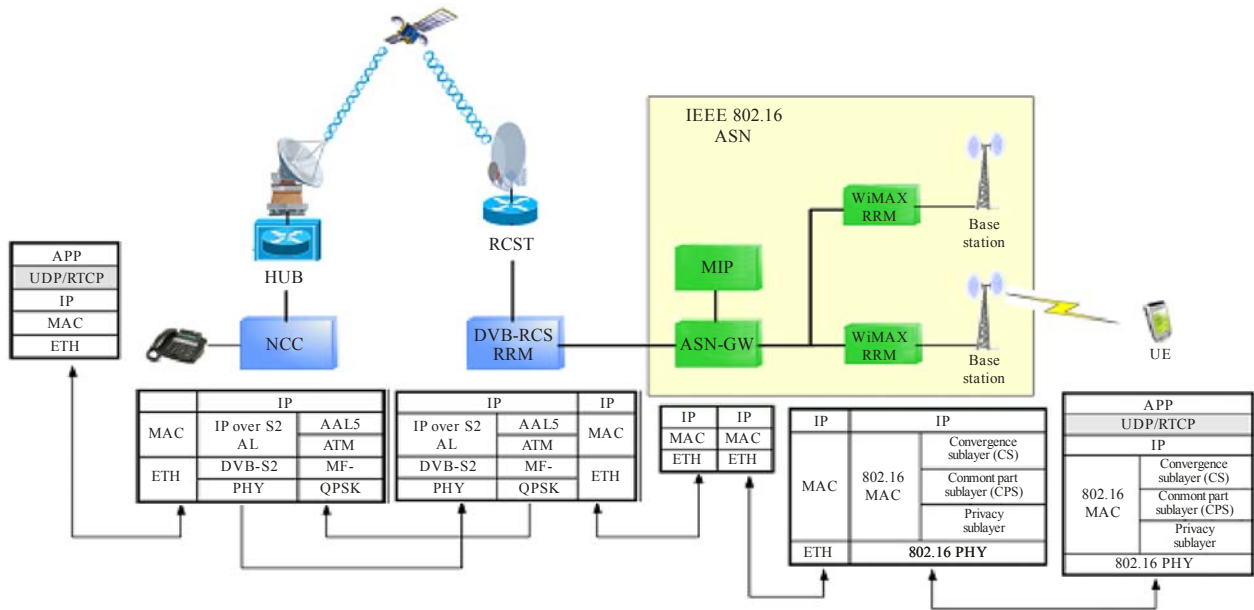
S.1897-04

Several ASN profiles have been specified in WiMAX as a tool to manage diversity node usage and implementation:

–   Profile A: Centralized ASN model with ASN-GW and BS in separate platforms with split RRM. Radio Resource Allocation (RRA) in BS and Radio Resource Control (RRC) in the gateway.

–   Profile B: BS and ASN-GW functionalities implemented in a single platform.

–   Profile C: Separate platforms, with the RRM controlled by the BS.

Profile A is suitable for soft handover, used in high speed mobiles, where the typical users are mobiles inside a rural area. Although profile B is the most simple, operators prefer separate platforms, as it is easier to customize IP and wireless functions. Profile C includes the ASN-GW between the two RRM, i.e. in the satellite terminal and the WiMAX network. It allows an interaction between both and manages the resources in a friendly way. This option allows each BS to manage the IEEE 802.16 service within its area, while the RCST carries out the resource assignment of all ASNs. The ASN-GW incorporates the mobile IP, to provide an efficient and scalable mechanism for roaming within the Internet.

Annex 2

Cross-layer based QoS performance of hybrid satellite-terrestrial networks

1   Scope

This Annex presents the results of simulations experiments conducted on hybrid satellite-terrestrial networks (e.g. WiFi and WiMAX) demonstrating the advantages on QoS performance based on cross-layer design methodologies. The applications include video teleconferencing, VoIP, and best effort data transfer (HTTPs, FTPs). The impact of fading on QoS parameters measured includes delay, delay variation, jitter, and packet loss rate which provides valuable information for satellite system designers. For further details on cross layer methodologies, see Report ITU-R S.2222.

2   Rain fade mitigation and cross-layer QoS design

The rain fade mitigation technique described in this section follows DVB-S2. One of the prominent features of DVB-S2 is its coding and modulation for normal (64 800 bits) and short (16 200 bits) FEC block lengths. An interesting aspect of the coding and modulation is that it can be adaptively changed on a frame by frame basis for certain types of transport streams. The adaptation depends on receiving signal to noise + interference information at the sending station from the destination station(s). This feature is intended to help mitigate the effects of rain induced fading, especially for Ka and higher frequency bands. The rain fade mitigation involves cross-layer interaction between the physical layer at the ground terminals whose received signal is affected by the rain fade and the MAC layer in the gateway which controls the coding and modulation of the forward stream of traffic directed to the affected ground terminal(s). During a rain fade event as the received Es/(N+I) changes at a ground terminal which is noted by the physical layer and reported to the gateway MAC by sending special messages in the reverse direction. The gateway MAC responds by changing the coding and modulation of the traffic stream directed to the affected ground terminals in such a manner as to maintain the ground terminal's bit error rate at an acceptable level. This is possible because at a given Es/(N+I), within a suitable range, it is known that a certain FEC code rate and modulation will result in a certain bit error rate at the ground terminal(s). This relationship is developed to give an MPEG packet error rate of less than 10^{-7}.

The current protocol suites for ISO/OSI and TCP/IP are based on a layering paradigm at each layer. In which protocols address a specific function using the service provided by modules below and give a new service to upper layers. QoS provisions based on a strict modularity and layer independence may lead to non-optimal performance in IP-based satellite networks. Therefore, an optimized cross-layer approach is required to account for interactions between non-adjacent protocol layers.

3   Fading effects on QoS for satellite-WiFi multimedia networks

3.1   Simulation network model

Figure 5 shows a multimedia satellite network with an onboard processing satellite operating to provide IP-based satellite services for multimedia content from VoIP to video conferencing and best effort traffic composed of gateway terminals (GTs) and Return Channel Satellite Terminal (RCST). The forward link follows DVB-S2 with adaptive coding and modulation (ACM) and the return link is based on DVB-RCS which follows multiple frequency time-division multiple access (MF-TDMA).

FIGURE 5

Multimedia satellite network
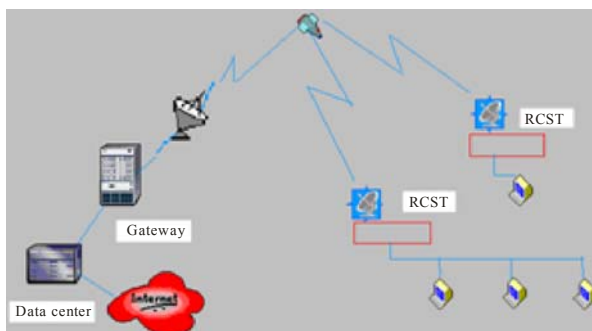
S.1897-05

DVB-S2 is the second generation specification used on the forward link and provides for audio broadcasting, Internet access, and other services such as TV contribution feeds and digital satellite news gathering. It utilizes adaptive coding and modulation (ACM) that approach the Shannon bound for channel capacity. The coding and modulation can be applied in an adaptive manner for one-to-one links to provide mitigation against signal fades due to rain.

DVB-RCS, used on the return link, specifies a satellite terminal which supports a two-way DVB satellite system in which the forward path conforms to DVB-S2. The MF-TDMA return link is coded with concatenated Convolution and Reed Solomon codes. The data may be encapsulated in Asynchronous Transfer Mode (ATM) cells using ATM Adaptation Layer 5 or it may use a native IP encapsulation over MPEG-2.

The traffic has been classified into conversational services, interactive services, and streaming services with recommended QoS parameters, e.g. packet delay, delay variation and packet error rates. The service classes are grouped in terms of error tolerant and intolerant applications with various delay bounds.

3.2   Bandwidth On Demand (BOD) scheme used in OPNET simulations

The BOD scheme involves operations at the ground terminals and at the resource controller in the satellite. The basic sequence of operations for BOD allocation is as follows.

–   Ground Terminal(i) (GT[i]) makes a request at frame n for dynamically allocated uplink data slots based on conditions at the terminal (size of class queues and number of packet arrivals during frame n-1). See Fig. 6 for frame definition.

–   At frame n+5 (at GT[i]) the satellite receives the request as well as requests from all other GTs. On the basis of the request information and stored information including a constant rate allocation (CRA) for each GT, the Sat makes a bulk allocation of uplink data slots for each GT and "broadcasts" the allocations to all the GTs.

–   During downlink frame n+9 (at GT[i]) GT[i] receives a bulk allocation, including the constant rate allocation, to be used on its next uplink frame. GT[i] apportions the bulk allocation among its classes of traffic based on its constant rate allocation and on the queue sizes at frame (n+9) using a class weighted algorithm for the dynamically allocated portion. The apportioned allocation is then used on its next uplink frame (n+10).

3.2.1   Return channel bandwidth allocation algorithm

The bandwidth allocation to the ground terminals consists of fixed allocation and dynamic allocation as below:

–   Constant bit rate traffic – The allocation for real-time traffic is constant and persistent.

–   Dynamic allocation – The allocation for IP traffic is based on packet arrival rate and sizes of three class queues using a class weighted priority algorithm.

FIGURE 6

Uplink and downlink frames at the satellite
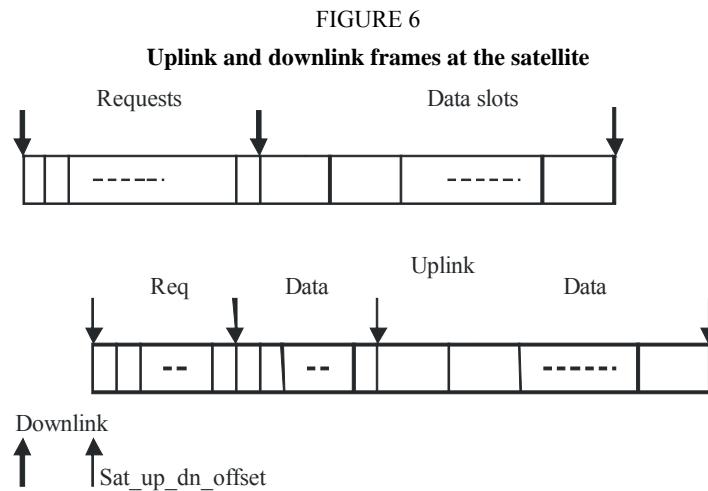
S.1897-06

Figure 6 shows the TDM frame structure consisting of the request slots and the data slots corresponding to uplink and downlink.

The allocations are aggregated and made in bulk to each ground terminal on a frame by frame basis. Each GT maintains class queues 0-3. Traffic with Type of Service (TOS) marking of 4 and above is directed to class queue 0. Traffic with TOS marking of 3 is directed to class queue 1. Traffic with TOS marking of 2 is directed to class queue 2 and traffic with TOS 1 and 0 is directed to class queue 3. Each GT apportions its bulk allocation to its four class queues. The constant rate allocation (CRA) following the DVB-RCS bandwidth allocation specifications is made first to class queue 0 based solely on the value computed at initialization time. The remainder of the bulk allocation is made sequentially to class queues 1, 2 and 3 as follows:

Initialization:

total_avail = bulk_alloc – CRA;

score[i] = queue_length[i]*class_weight[i], i=1..3;

total_score = sum(score[i]), i=1..3;

total_length = sum(queue_length[i]), i=1..3;

for(i = 1; i < 4; i++)

{

if(total_avail >= total_length) then

alloc = queue_length[i];

else alloc = total_avail*score[i]/total_score;

total_avail -= alloc;

total _length -= queue_length[i];

}

The class queues may be operated with either unbounded size or with some maximum size. In the latter case, the traffic is somewhat variable and since the DAMA scheme cannot respond in less than a round trip time the maximum sizes are normally set to the mean flow rate through the queue times equivalent to a couple of round trip times.

3.3 Simulation experiments and performance results – Scenario 1

3.3.1 Simulation model

Figure 7 shows an initial simulation framework for IP-based satellite DVB network to support multimedia applications. Router A serves as a gateway to the subnet composed of switch A and the four servers. It also serves as a ground terminal and the subnet represents the Internet.

The satellite represents the bandwidth bottleneck of the network. The clients request services from the servers and the communications are conveyed by means of the reverse direction traffic.

FIGURE 7

OPNET model for simulation network
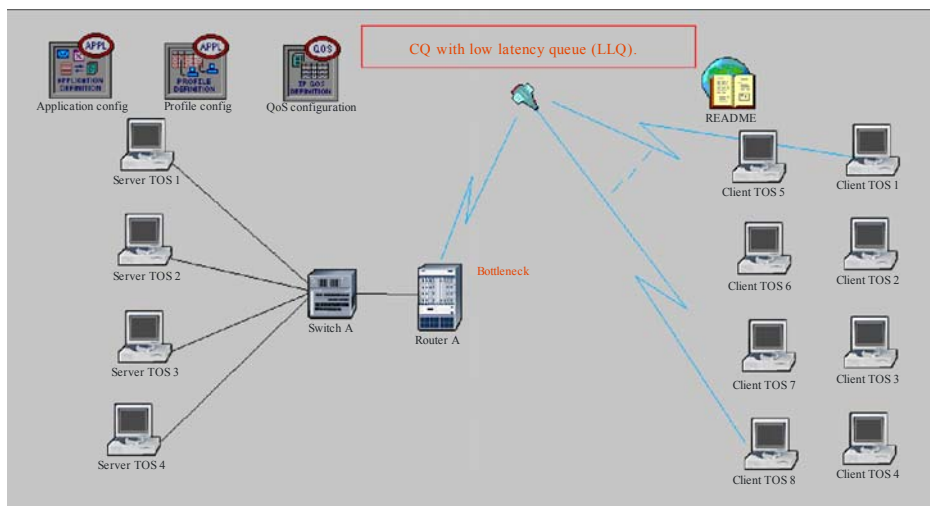
S.1897-07

3.3.1.1 Traffic model

The traffic is assumed to start at 110 sec and persists to the end of the simulation. A summary of the baseline traffic flows and boundary conditions follows:

Video Conferencing: Between clients TOS 1 and client TOS 8 via TOS servers 1 and 4 is represented in the following figures as cases 1-4 respectively. The forward direction was exponentially distributed with mean of 1 000 bytes per frame and reverse direction was exponentially distributed with mean of 250 bytes per frame at 30 frames/sec in both directions.

VoIP: G.729 Silence suppressed: Between server 1 and clients 1 and 5 with TOS = 1. The talk spurts were exponentially distributed with mean duration of 0.352 sec and silence intervals exponentially distributed with mean duration of 0.65 sec. During the talk spurt, an 80-byte packet including headers was generated 100 times per second. No data was sent during the silence intervals.

HTTP: Between server 2 and clients 2 and 6 with TOS = 1. Forward traffic consisted of one 1 000 byte packet and five packets uniformly distributed between 500 and 2 000 bytes occurring with exponentially distributed mean interval of 30 sec. Reverse direction traffic was six 350-byte packets occurring with exponentially distributed mean of 30 sec.

Data Base Query: Between server 3 and clients 3 and 7 with TOS = 1. Forward and reverse consisted of a 512-byte packet with exponentially distributed mean interval of 12 sec.

FTP: Between server 4 and clients 4 and 8 with TOS = 3. Forward and reverse consisted of one 512 byte control packet and one 8 000 byte data packet (50% get and 50% put) exponentially distributed with mean interval of 60 sec.

3.3.1.2   Simulation parameters

The major parameters for the time division satellite link are as follows:

Basic bit rate	2.048 Mbit/s
Number of request slots	150
Number of data slots	128
Downlink request slot (bits)	20
Uplink request slot (bits)	47
Data slot assign (bits)	32
Basic data slot (bits)	424

Fade parameters

A 10 dB rain fade event occurs during the simulation on the downlink for all clients with the following characteristics:

Start time	140 sec
Attack rate	0.25 dB/s
Attack duration	40 sec
Recovery start time	180 sec
Recovery rate	0.125 dB/s
End of rain fade	260 sec

3.3.2   Simulation experiment results

This section provides the simulation results for bandwidth allocation and PHY-Application layer and PHY-Transport layer interactions specifically considering the effect of rain fade.

3.3.2.1   PHY-MAC layer interaction: Bandwidth on demand with rain fade

Figure 8 shows the bulk allocation of data slots to the gateway and the terminals, as a function of simulation time. The bandwidth on demand algorithm is used to provide the data slot allocations for the router and the clients.

Each curve shows the fraction of the total available data slots that is assigned to the node identified by the curve label. (For instance in router A DAMA 80% of the slots are available at 200 msec) Traffic starts after 150 sec, so the allocation before that time is just random allocation of free slots.

After the traffic starts, it can be seen that an equitable time averaged allocation based on relative demands of the individual nodes is achieved. (At 200 ms in Clients 1, 2 and 3, 5% slots are available.)

FIGURE 8

Moving average of data slot allocation for router A and clients 1-3, with rain fade
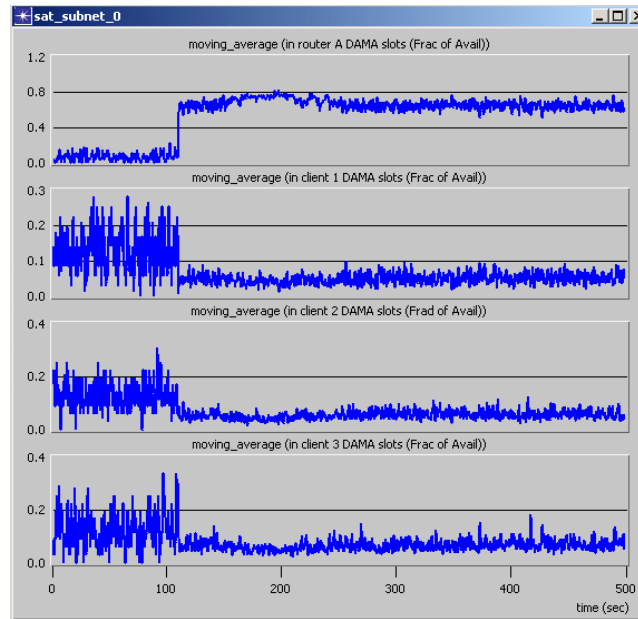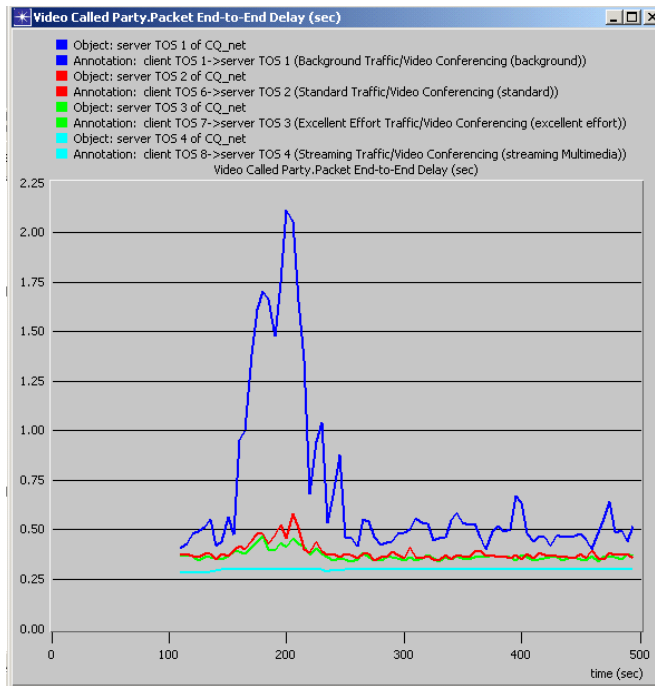

S.1897-08

3.3.2.2 PHY-application layer interaction

This subsection describes the delay performance of the applications video conferencing, VoIP, and HTTP due to fading.

Figure 9 shows the end-to-end delay for the reverse direction video conferencing traffic from clients 1-4 measured at servers 1-4. This shows the higher delay and variability of the traffic passing through the lower class queues. Also, when there is congestion, the lower priority traffic suffers more than the higher priority traffic.

These curves show the spread of delays for the four classes of traffic. During the rain fade the lowest priority traffic experiences a much larger delay than the highest priority traffic. The highest priority traffic was unaffected by the rain fade because its assignment was from constant rate allocation, i.e. pre-negotiated, and not dynamic traffic.

FIGURE 9

**End-to-end delay for reverse direction video conferencing
traffic from clients 1-4 measured at servers 1-4**

S.1897-09

Figure 10 shows aggregate video conferencing traffic, sent, received, end-to-end delay and dropped packets. A total of 287 dropped packets occurred during the rain fade. Of these, 165 occurred at router A class 3 queue and 122 at client 1 class 3 queue. During the rain fade the aggregate peak delay reached approximately 0.859 sec. At other times, it remained at about 0.4 sec.

The two top curves indicate that the send and receive rates for the aggregate video conferencing traffic were nearly equal except during the rain fade where the receive rate dropped slightly below that of the send rate. Immediately after the rain fade the receive rate exceeded the send rate for a short time. This was sufficient to work off the queued backlog. The third curve from the top shows the build-up of aggregate packet delay. The bottom curve indicates the time at which packets were dropped.
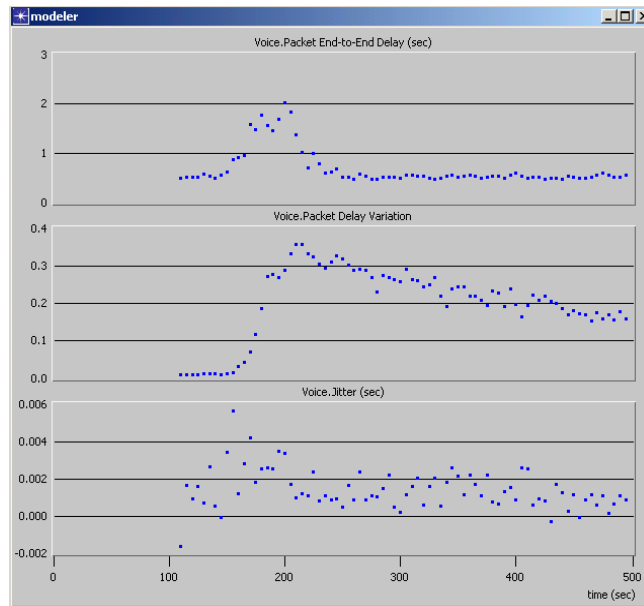
FIGURE 10

**Aggregate video conferencing traffic, sent, received, end-to-end delay
and dropped packets, equal traffic levels at TOS 1-4, rain fade**
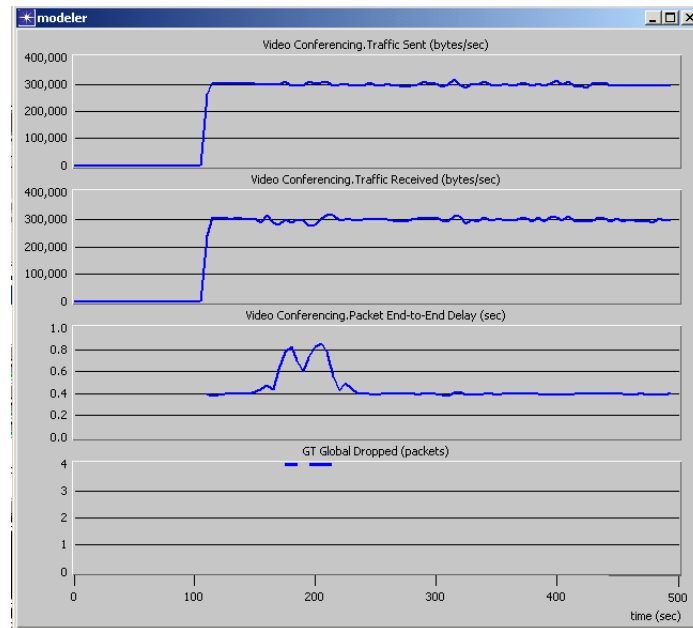
S.1897-10

Figure 11 shows aggregate voice packet end-to-end delay, packet delay variation and jitter. The top curve shows the aggregate end-to-end delay for the four audio streams (two forward and two reverse). The delay reached just over two seconds during the rain fade. Such delay would be unacceptable for an interactive conversation. After the rain fade ended the delay was approximately 0.5 sec, still undesirably high. The middle curve shows the variance of the delay. It reached a peak of approximately 0.36 s just after the end of the rain fade and then declined to about 0.17 s at the end of the run. The bottom curve shows the jitter in the packet delay, i.e. the amount of difference in delay of consecutive packets. This had a range of nearly 6 ms.
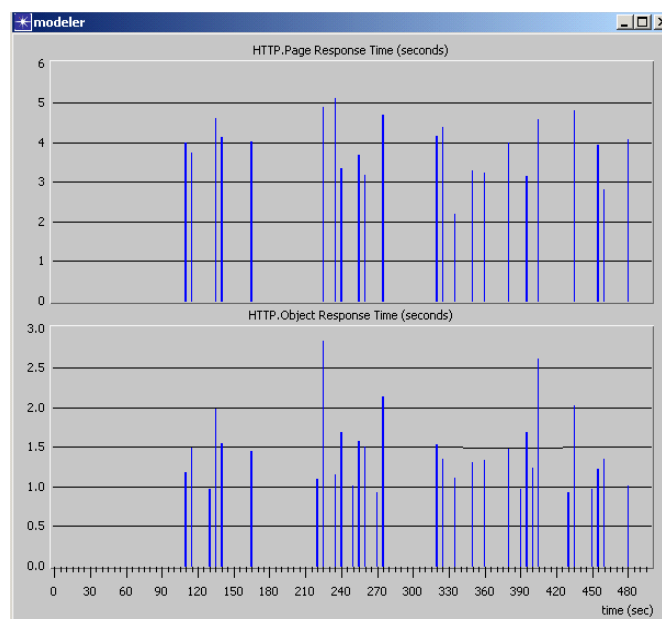
FIGURE 11

**Aggregate voice packet end-to-end delay, packet delay variation
and jitter at TOS = 1, rain fade**

S.1897-11

Figure 12 shows aggregate HTTP response times. The top plot shows the aggregate of the response time for pages. These times ranged from just over 2 seconds to about 5 seconds. Each transaction consisted of two pages: a header page and a result page. The header page contained a single object of size 1 000 bytes. The result page contained 5 objects ranging from size uniformly distributed from 500 to 2 000 bytes. The response time for the objects ranged from just under 1 s to about 2.75 s. The averages of these times were 3.907 s for page and 1.436 s for object.

FIGURE 12

Aggregate HTTP page and object response times at TOS =1, rain fade
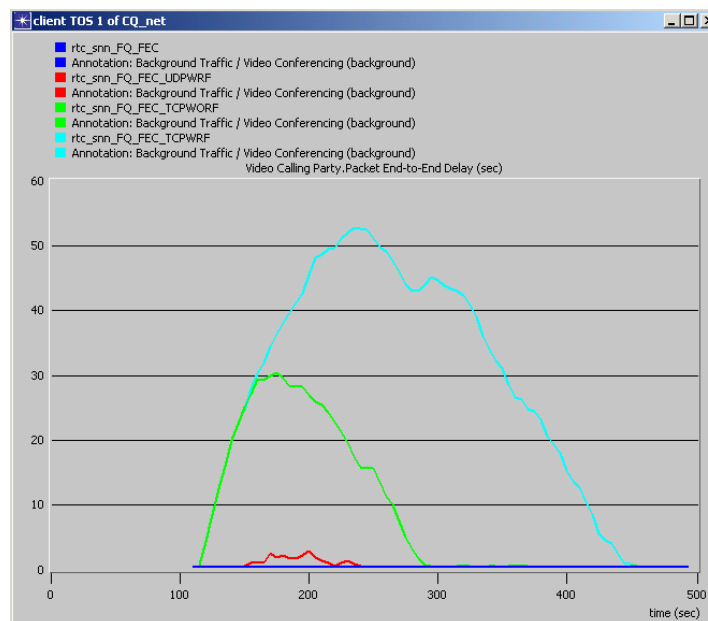

S.1897-12

3.3.2.3   PHY-Transport layer interaction: TCP/UDP performance

TCP poses special problems when congestion or transmission errors occur. Figure 13 shows four cases of end-to-end delay for video conferencing traffic in the forward direction between server 1 and client 1. The four cases are for UDP transport without rain fade, UDP transport with rain fade, TCP transport without rain fade and TCP transport with rain fade. In both TCP cases, TCP RENO SACK was used. TCP experienced a peak delay of approximately 30 seconds at about 175 seconds of simulation time due to start-up. In the TCP rain fade run it had not recovered from the start-up delay when the rain fade started. This resulted in a peak delay of approximately 53 seconds at 240 seconds of simulation time.

FIGURE 13

**End-to-end delays for video conferencing traffic in forward direction between server 1
and client 1 for UDP without rain fade, UDP with rain fade, TCP without
rain fade and TCP with rain fade**

S.1897-13

Figure 14 shows the same four cases except for traffic in the reverse direction, i.e. from client 1 to server 1 where the level of traffic is only one-quarter that of the forward direction. In both TCP cases there is a peak start-up delay of approximately 4.2 seconds occurring at about 130 seconds followed by recovery. Then in the rain fade case a second peak of approximately 4.1 seconds occurs at about 210 seconds simulation time but again it quickly recovers.

In both Figs 13 and 14 for the TCP runs no traffic was dropped at the terminal MAC class queues. TCP simply stopped the application from filling its input buffer which caused the application to queue it at that level. In all these cases, all the traffic was eventually delivered.

Comparison of Figs 13 and 14 shows the effect of the volume of traffic on delay using TCP as a transport protocol with the same parameters.

FIGURE 14

**End-to-end delays for video conferencing traffic in reverse direction between client 1
and server 1 for UDP without rain fade, UDP with rain fade, TCP without
rain fade and TCP with rain fade**
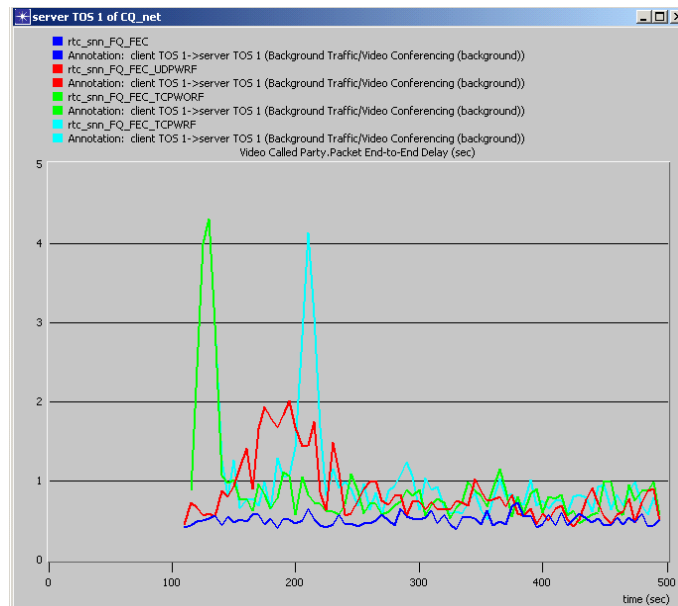
S.1897-14

3.3.3 Summary

Next generation IP-based satellite networks supporting multimedia applications demand QoS guarantees. The non-adjacent protocol layer interactions impact the QoS performance and the capacity. The simulation results considering the fade mitigations techniques due to DVB-S2 /DVB-RCS shows the influence of fading over QoS parameters such as bandwidth allocation, delay and delay variation for video conferencing, VoIP, data and TCP/UDP applications.

3.4 Simulation experiments and performance results – Scenario 2

This section provides simulation results for hybrid network model using cross-layer design with special emphasis on fading effect on dynamic resource allocation, i.e. interaction between physical-media access control (PHY-MAC) layers and interaction of physical and transport layer protocols such as transmission control protocol/user datagram protocol (TCP/UDP). A system model using DVB-S2 for forward link and DVB-RCS for return link connected to a WiFi terrestrial network. OPNET was used for source to sink network simulations.

3.4.1 Simulation network model

The simulation network model shown in Fig. 15, which has been adapted from OPNET, consists of a base station and four wireless clients. Router A is the gateway to the requests internet which is represented by the server nodes on the left. Router A also acts as a ground terminal for satellite access. Router B acts as a ground terminal to the satellite for the client nodes in the wireless LAN on the right. It is connected to the base station of the wireless LAN by an Ethernet link. The clients in the wireless LAN communicate with the servers in the return direction by sending packets over the air interface to the base station.

FIGURE 15

Simulation network model
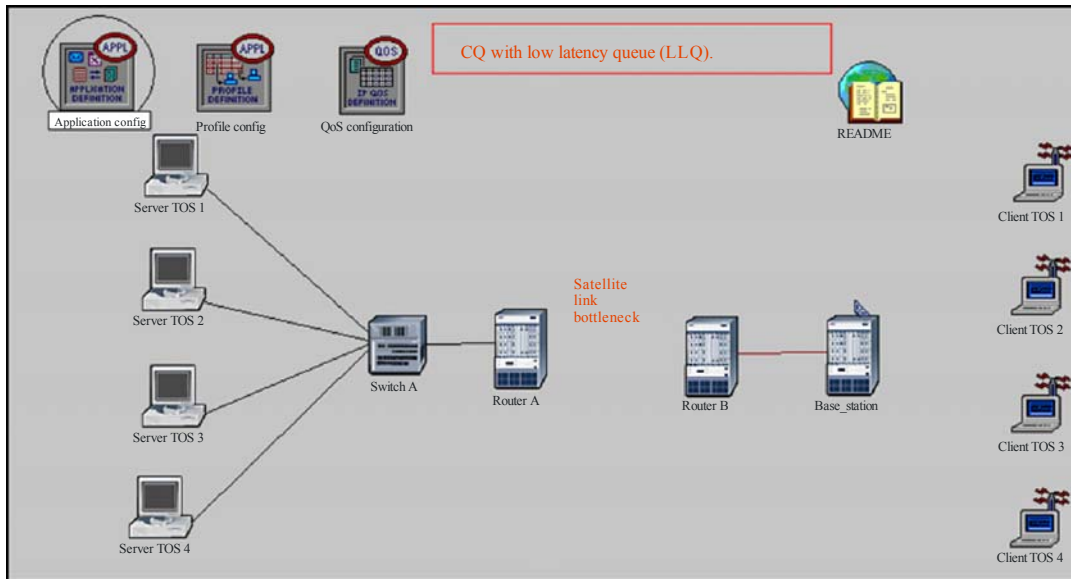
S.1897-15

The base station forwards the packets to router B which in turn forwards them via the satellite to router A. Router A forwards packets received from the satellite to the appropriate server. Traffic in the forward direction (server to client) follows the reverse of the return path just described.

3.4.2 QoS provisioning

In order to provide the quality of service for various applications such as Video Conferencing, Voice over IP, HTTP, Data Base Query and FTP the traffic has been categorized, and each ground terminal MAC will consist of four input traffic buffer queues. The DiffServ classification of the traffic is performed and according to the type of service field, the traffic class is entered into the appropriate class of buffer, i.e. RT, VR-RT, VR-JT and JT. The mapping of the applications level and the four queues is shown in Table 1.

3.4.3 Traffic model

The traffic model used for the simulation experiments described in § 3.3.1.1 includes five classes namely: Video Conferencing, Voice over IP, HTTP, Data Base Query and FTP as shown in Table 2. This traffic is provisioned into four classes of service including four Queues in the ground terminal MACs.

3.4.4 Satellite link loading

The traffic model results in a satellite link loading of 88.52% of capacity. This load takes account of the data slot size, the various applications packet sizes, the overhead of the forward link ACM and return link overhead.

3.4.5 Simulation parameters

Rain fade event

A 10 dB rain fade event occurs during the simulation on the downlink for client Type of Service (TOS) 1 and Client TOS 2 with the following characteristics:

TABLE 1

QoS Provisioning (DiffServ)

Application level		Course queues	Mapping
Best effort	0	RT – for real-time constant bit rate	0,1 -> JT
Background	1	VR-RT – for variable rate real-time traffic	2 -> VR-JT
Standard	2	VR-JT – for variable rate real-time traffic that is jitter tolerant	3 -> VR-RT
Excellent effort	3	JT – for jitter tolerant and all other traffic	4,5,6 -> RT
Streaming multimedia	4		
Interactive multimedia	5		
Interactive voice	6		

Fade parameters

Start Time	140 sec
Attack rate	0.25 dB/s
Attack duration	40 sec
Recovery start time	180 sec
Recovery rate	0.125 dB/s
End of Rain Fade	260 sec

Frame parameters

The major parameters for the time division satellite link are as follows:

Basic Bit Rate 2.048 Mbit/s (Scaled by power of 2)

Number of Request slots	150
Number of Data Slots	128
Downlink Request slots	20 Bits
Uplink Request Slot	47 Bits
Data Slot Assign	32 Bits

Basic Data Slot 424 Bits (Scaled by power of 2)

The application level QoS mapping, fade event and simulation link formats discussed above are used to conduct the simulation experiments described in § 5.

TABLE 2

Traffic model

| Application	From	To	Size	Priority (TOS
Video Conference	Client 1	Server 1	15 kbit/s	Background (1)
VoIP	Client 1	Server 1		Best Effort (0)
Video Conference	Server 1	Client 1	60 kbit/s	Best Effort (0)
VoIP	Server 1	Client 1		Best Effort (0)
Video Conference	Client 2	Server 2	15 kbit/s	Standard (2)
Video Conference	Server 2	Client 2	60 kbit/s	Standard (2)
HTTP	Client 2	Server 2	Request Packet	Best Effort (0)
HTTP	Server 2	Client 2	Response Page	Best Effort (0)
Video Conference	Client 3	Server 3	15 kbit/s	Excellent Effort (3)
Video Conference	Server 3	Client 3	60 kbit/s	Excellent Effort (3)
Database Query	Client 3	Server 3	Query Packet	Best Effort (0)
Database Response	Server 3	Client 3	0.512 kbit/s	Best Effort (0)
Video Conference	Client 4	Server 4	15 kbit/s	Streaming Multimedia (4)
Video Conference	Server 4	Client 4	60 kbit/s	Streaming Multimedia (4)
FTP	Client 4	Server 4	8 kbit/s file in 1 500 byte packets	Excellent Effort (3)
FTP	Server 4	Client 4	8 kbit/s file in 1 500 byte packets	Best Effort (3)

3.4.6 Simulation experiments results

This section provides simulation results for bandwidth allocation in the presence of fading. The impact of fading on the multimedia application QoS parameters, e.g. end-to end, packet delay variation, jitter, response time and traffic received are described. The applications considered for simulation includes: Video Conferencing, VoIP (G.729 Silence), HTTP, Data Base and FTP. The network model consists of the SATCOM network connected to the WiFi network through a base station and four nodes and the SATCOM network consists of DVB-S2 for the forward link protocol and DVB-RCS for the return link.

3.4.6.1 Bandwidth allocation with fading

Figure 16 shows the bulk allocation of data slots to gateway terminal A (GTA) and gateway terminal B (GTB) with rain fade as a function of simulation time. The bandwidth on demand algorithm discussed in § 3.2 is used to determine the data slot allocation for the routers. Each curve shows the fraction of the total available data slots assigned to the nodes identified by the curve label.

FIGURE 16

Moving average of DAMA data slot allocation for GT_A and GT_B with rain fade
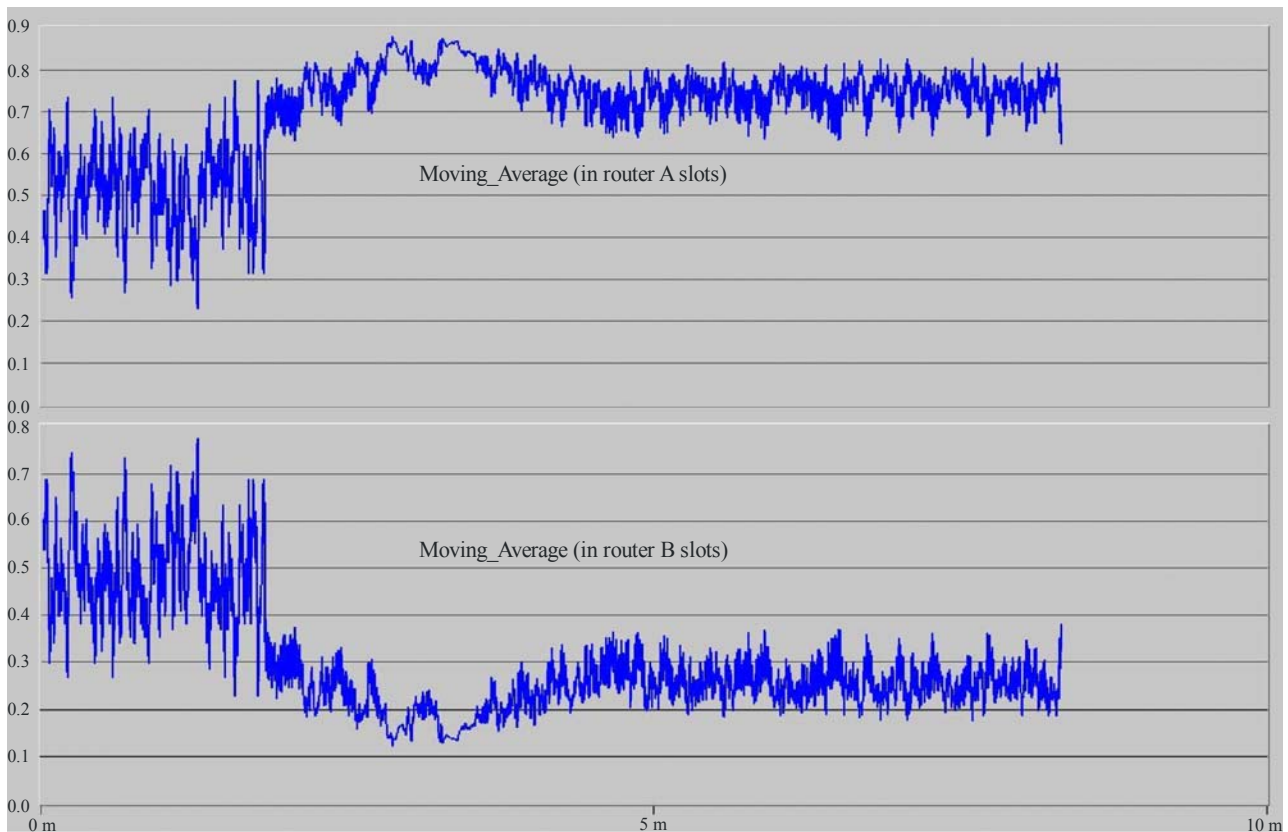
S.1897-16

3.4.6.2   Video conferencing

This subsection describes the end-to-end delay performance and traffic sent and received for video conferencing application. Four different TOS classes considered include background, standard, excellent effort, and streaming multimedia traffic. Figure 17a shows the video conferencing traffic TOS of background class between client 1 and server 1 without fading and Fig. 17b with fading. The end-to-end delay for both calling and called party increases with fading. Figure 17c shows the video conferencing for streaming multimedia, classes respectively with rain fade. These curves show the spread of the delay for different classes of traffic. During rain fade, the lowest priority traffic experiences much larger delay than the highest priority traffic.

FIGURE 17a

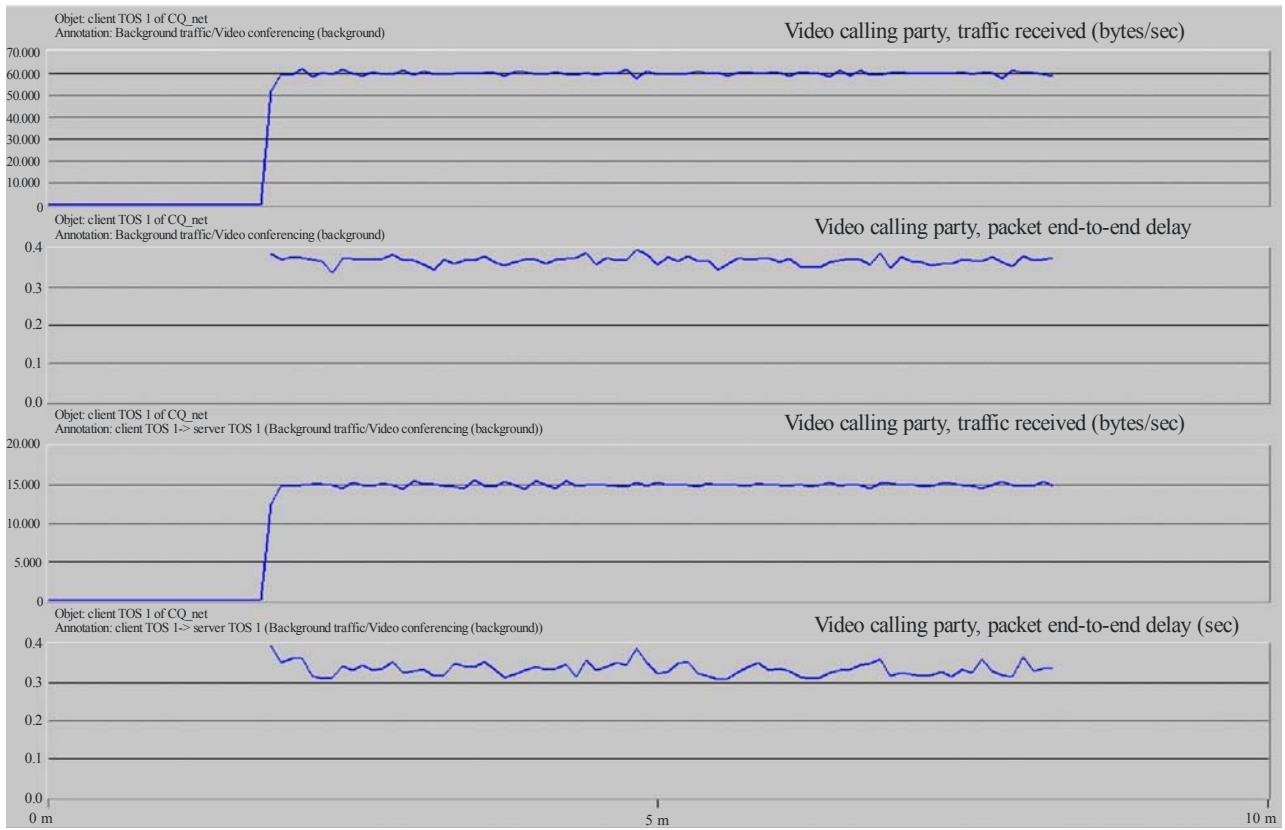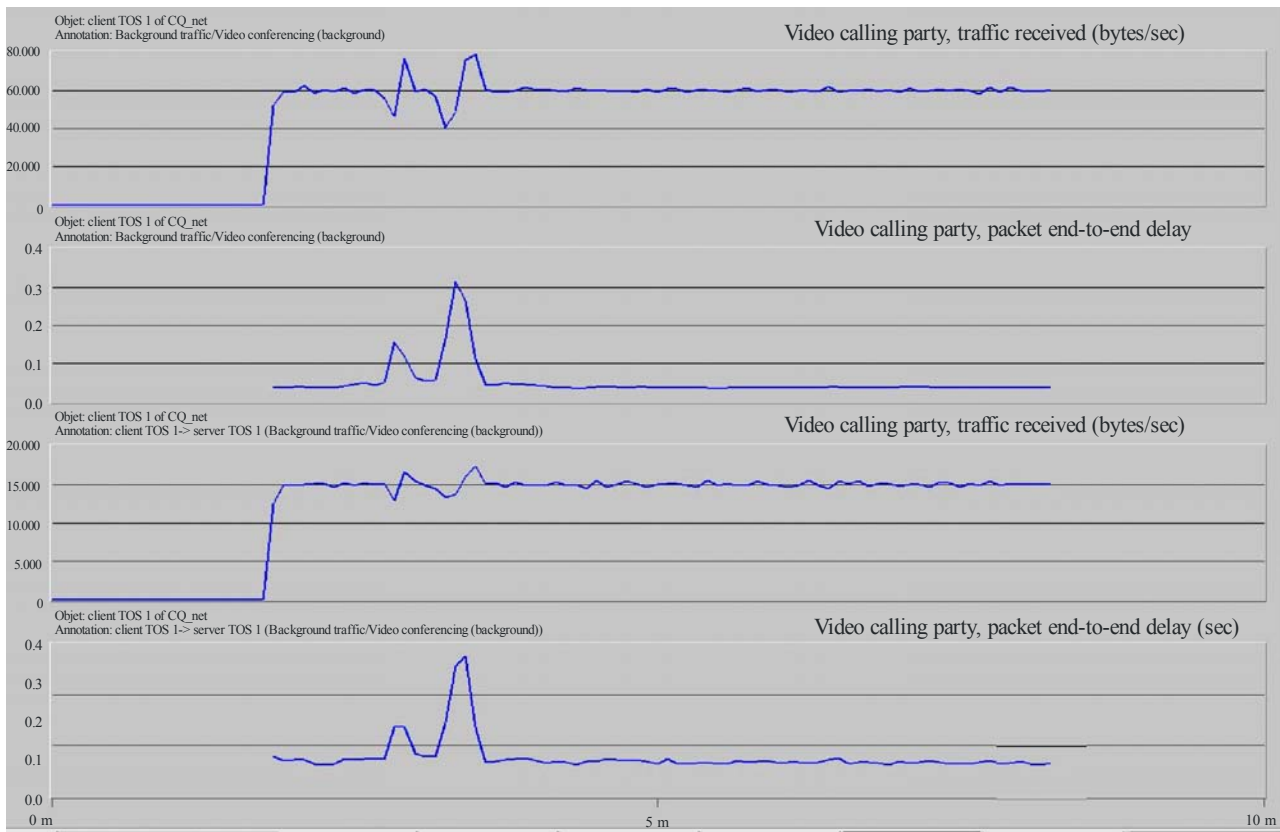**Video conferencing between client 1 and server 1, TOS = Background, no rain fade**

S.1897-17a

FIGURE 17b

**Video conferencing between client 1 and server 1, TOS = Background with rain fade**

S.1897-17b

The highest priority traffic like streaming multimedia or interactive voice was unaffected by the rain fade because its assignment was from a pre-negotiated allocation. For example, the end-to-end delay for called party drops from 3 seconds (Fig. 17b) in the background class to 0.58 seconds in the excellent effort traffic to 0.32 seconds for streaming multimedia with rain fade.

FIGURE 17c

**Video conferencing between client 4 and server 4,
TOS = Streaming multimedia with rain fade**
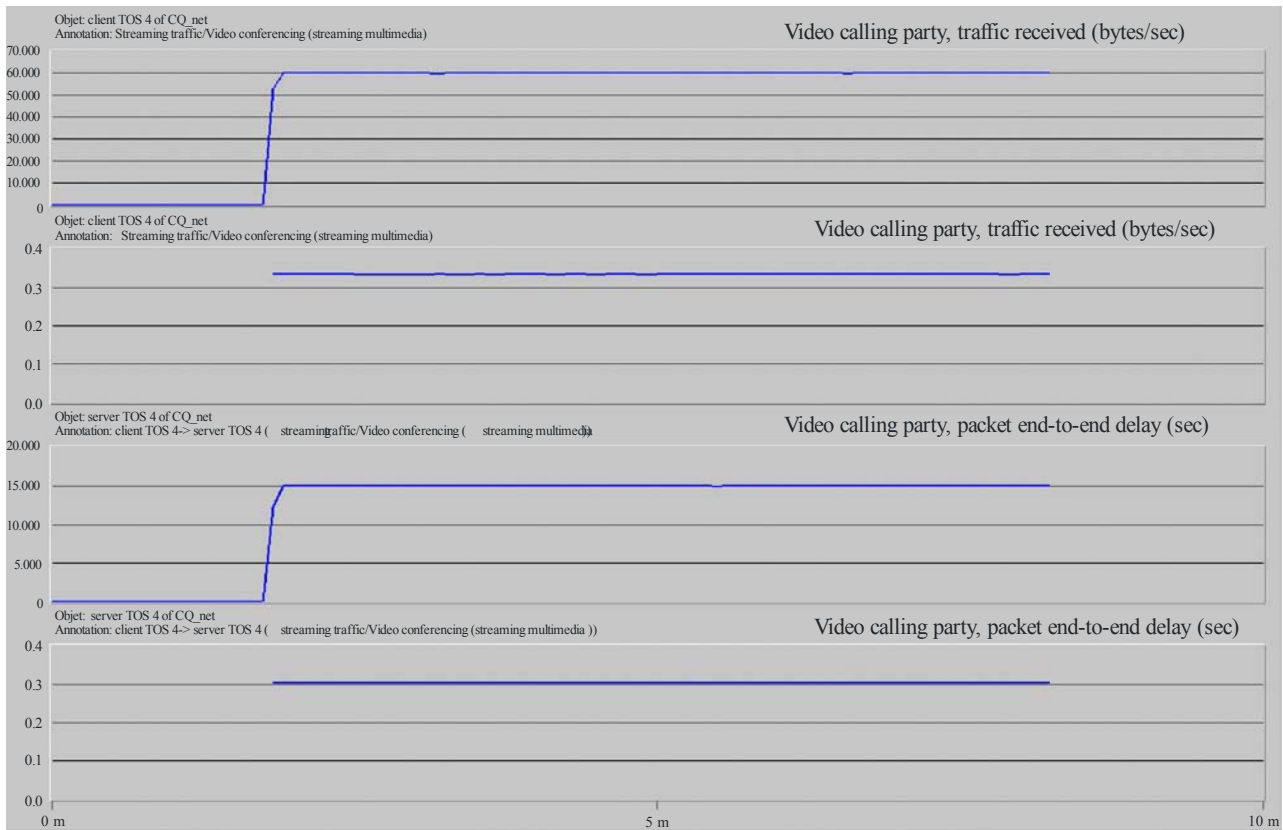
S.1897-17c

3.4.6.3   Voice over IP (VoIP) delay and jitter: Best effort class with fading

Figure 18 shows aggregate voice packet end-to end delay variation and jitter with rain fade. The end-to-end delay reached 1.5 seconds during the rain fade and remained close to 0.5 seconds at other times. The packet delay variation had a peak of close to 0.08 seconds squared and then declined to 0.025 seconds by the end of the simulation. The jitter had a peak of about 0.0045 seconds during the rain fade. At other times it ranged from –0.0025 to 0.002.

FIGURE 18

**VoIP (G.729 Silence) received traffic at server 1, TOS = Best effort with rain fade**
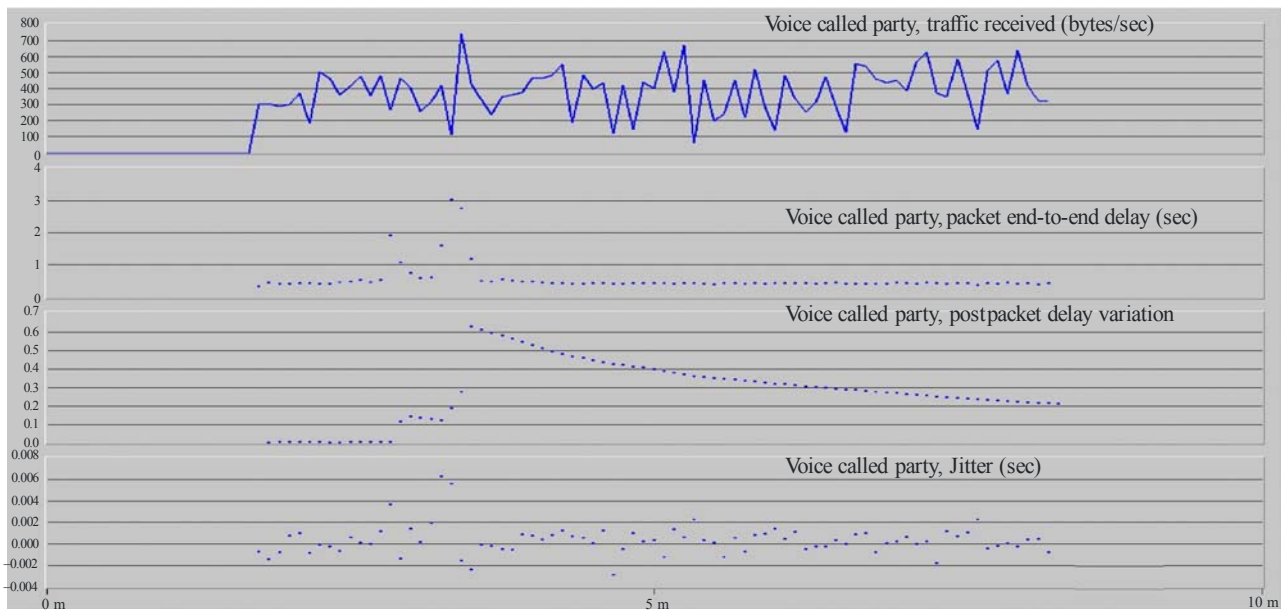
S.1897-18

3.4.6.4 HTTP: Page and response time

Figure 19 shows traffic sent and received, downloaded pages, and objects. The page response time ranges from 2 seconds to about 6 seconds. The object response time varies from about 1 to 4 seconds under rain fade conditions assuming the best effort traffic class.

FIGURE 19

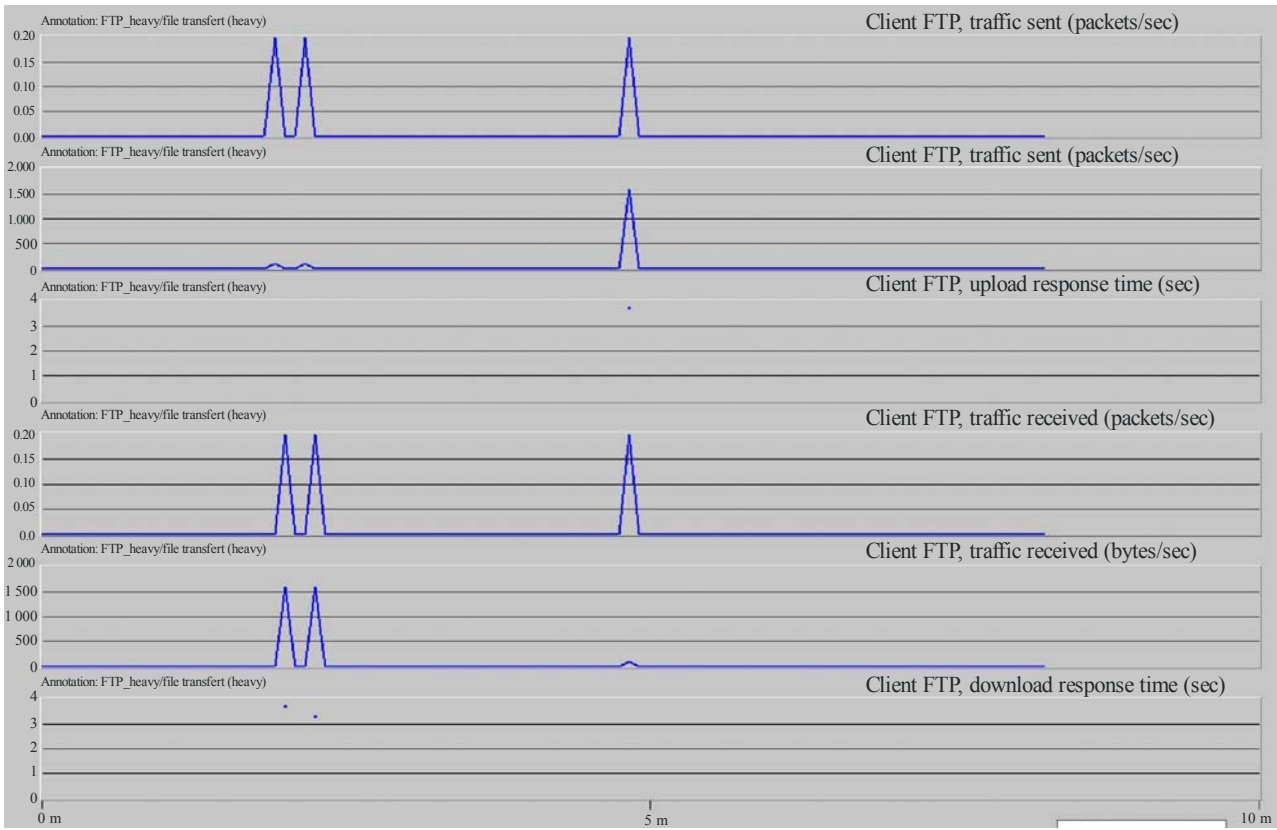**HTTP traffic between client 2 and server 2, TOS = Best Effort, rain fade**
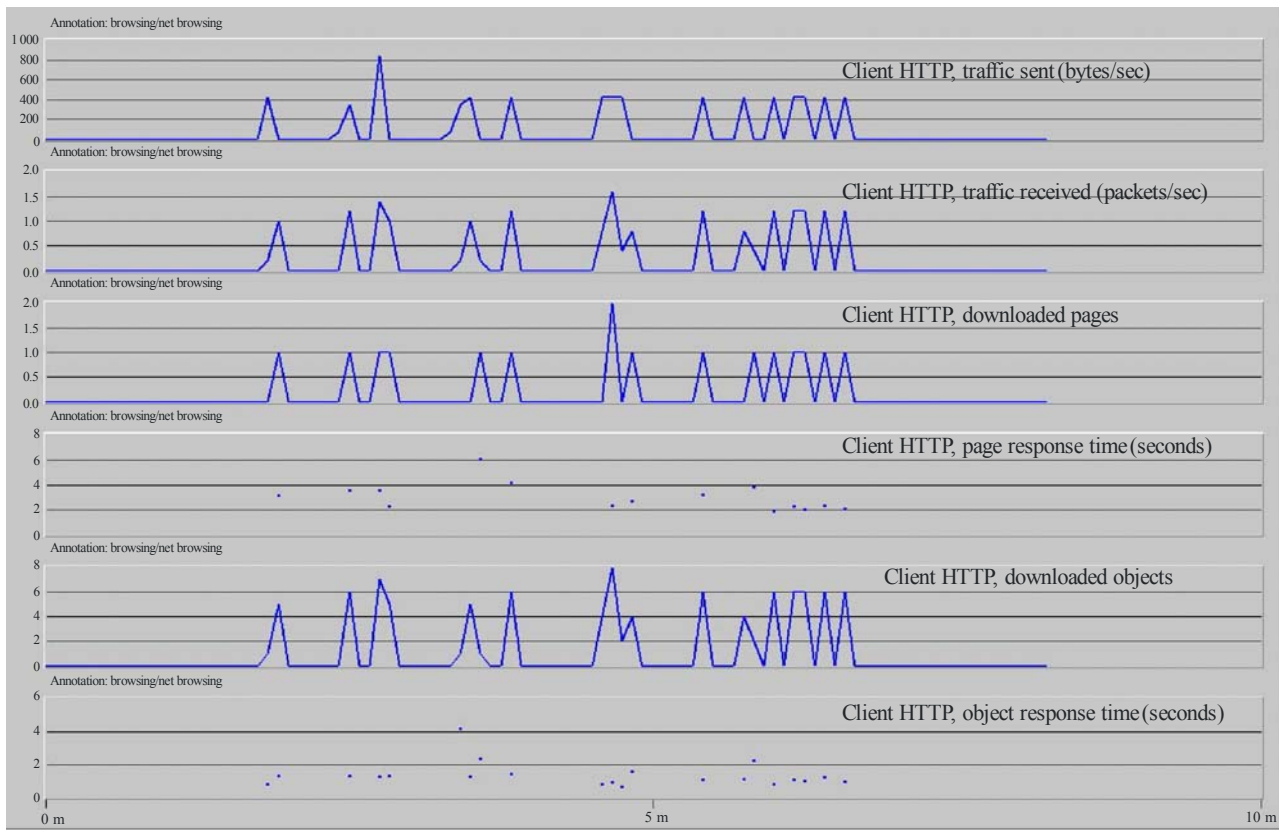
S.1897-19

3.4.6.5 FTP Traffic

Figure 20 shows the best effort class of FTP traffic sent in packets per second and received packets per second and response times.

FIGURE 20

FTP traffic between client 4 and server 4

S.1897-20

The relative performances of the TOS classes for video, voice and HTTP is shown in Table 3. Traffic sent with the highest TOS marking (4) (uses CRA) is less affected by a rain fade event with the lowest delay. The lowest priority traffic (TOS 1) experiences considerable delay and packet drops occur during the rain fade. This priority is probably unsuited for high bandwidth applications using TCP.

TABLE 3

Relative performance of the TOS classes

Relative TOS Perfomance						
	Video		**Voice**		**HTTP**	
	FWD	REV	FWD	REV	Page	Object
TOS 1	0.4619	0.4458	0.5036	0.4319	3.927	1.404
TOS 2	0.4433	0.3823	0.5024	0.4341	3.312	1.338
TOS 3	0.3911	0.3625	0.4399	0.4265	2.984	1.262
TOS 4	0.3713	0.2937	0.4071	0.3484	-----	-----

NOTE – All results are expressed in seconds.

3.4.7   Summary

Cross-layer design based simulation framework is used to assess the impact of rain fade and mitigation due to DVB-S2 on the multimedia applications. Simulations demonstrated the effect of fading on applications, e.g. video conferencing, VoIP, HTTP, Data Base Query and FTP in terms of QoS parameters. The QoS parameters included bandwidth allocation, end-to-end delay, delay variation and Jitter.

4   Cross-layer based QoS for VoIP in hybrid satellite-WiMAX networks

4.1   Introduction

This section presents results of the cross-layer optimization of real-time traffic transmission over hybrid networks using DVB-RCS/S2 satellite and WiMAX shown in Fig. 4. This simulation is focused on experiments of Voice over IP (VoIP) transmission reducing total transmission delay and jitter. An end-user-centric scenario with codec rate adaptation is employed. Two possible cross-layer codec rate adaptations have been analyzed: a) bank of codec is available at both transmission ends and a codec switch is performed driven by the RTCP (Real Time Control Protocol) reports; b) use of adaptive wideband codecs (inherently cross-layer across the networks). The cross-layer design mechanisms demonstrate potential advantage of reducing total transmission delay and jitter and increases significantly the total capacity even in GEO satellite environments.

**4.2   Cross-layer VoIP rate adaptation**

This contribution proposes cross-layer distributed rate control approach for VoIP flows relying on the Real-time Transport Control Protocol (RTCP) reports. A bank of codecs is available at both transmission ends and a codec switch is driven by the RTCP reports. Figure 21 shows the network protocol architecture and the required information flows.

4.2.1   RTCP driven approach

Real Time Protocol (RTP) receivers provide reception quality feedback using RTCP report packets which may take one of two forms depending upon whether or not the receiver is also a sender. The only difference between the sender report (SR) and receiver report (RR) forms, besides the packet type code, is that the sender report includes a 20-byte sender information section for use by active senders. The SR is issued if a site has sent any data packets during the interval since issuing the last report or the previous one, otherwise the RR is issued.

The information about jitter, delay and packet loss is extracted from these reports at each end of the transmission and sent up to the application layer. Jitter is measured by RTCP and included in the RR messages sent by the receiver. As this value is measured in sampling units, in order to convert to time units, one must divide by the sampling rate of the media codec.

FIGURE 21

RTCP-driven architecture
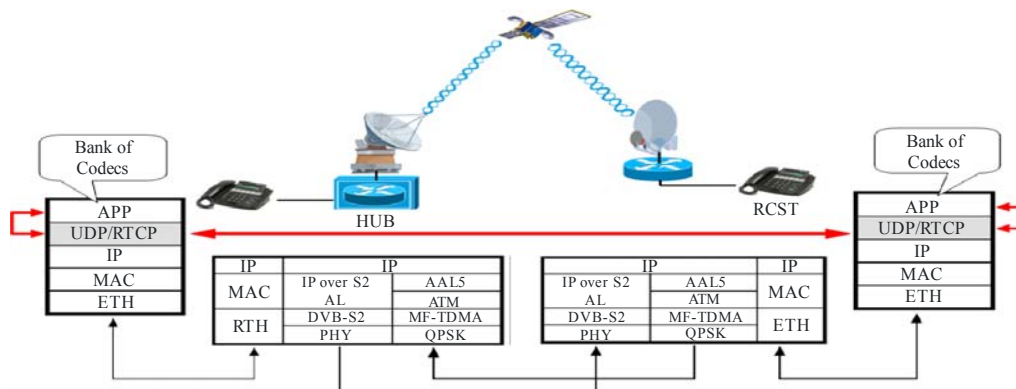
S.1897-21

TABLE 4

Codecs considered for the RTCP-driven scenario

ITU-T Codec	Modulation Type	Coding bit rate (kbps)	Reason to be selected
G.711	Companded PCM	64	Narrowband, most commonly used, "Tool quality"
G.729	CS-ACELP*	8	Narrowband, most commonly used after G.711
G.722.1	Transform Coding	24 / 32	Wideband

* Conjugate-Structure Algebraic Code Excited Linear Prediction.

Delay between two peers can be calculated with the difference among three times, such as, Delay since Last Sender Report (DLSR), Time of Last Sender Report (TLSR) fields in RTCP receiver report packets and report receiving timestamp. Inter-arrival delay jitter and packet loss rate are obtained from the inter-arrival jitter field and cumulative number of packet lost fields in RTCP receiver report packet, respectively. Inter-destination delay jitter can be calculated with the delay values received from all the other multicast group members.

4.2.1.1 Bank of codecs

A bank of codecs is assumed at each transmission end, which is a realistic assumption for many of currently available VoIP software packages. The application can switch to a different codec according to the information extracted from the RTCP reports. In particular, switching to a lower (higher) bit-rate codec takes place whenever the delay or jitter value reported by the RTCP is above (below) than the required ones. Table 4 shows the codecs and their key technical characteristics.

4.2.2 Adaptive multirate wideband (AMR-WB) approach

Second cross-layer mechanism is based on the RTCP reports interpreted along with the signalling extracted from the Adaptive MultiRate Wideband (AMR-WB) packets.

4.2.2.1   AMR-WB codec

AMR-WB coding algorithms are based on an Algebraic Code Excitation Linear Prediction (ACELP) technology, consisting of 9 speech rates. This same technology has been utilized in various speech codec standards, such as GSM enhanced full rate (GSM-EFR) (3GPP TS 06.51) and narrowband GSM-AMR (3GPP TS 26.071). The main novelty in AMR-WB is the sub-band structure, which enables significant savings in complexity and memory consumption. The audio band is split into two frequency bands so that the internal sampling frequency of the core is 12.8 kHz, having an audio bandwidth of 50-6 400 Hz. Separate processing is performed for the frequency range from 6 400 to 7 000 Hz: more bits can be allocated to the perceptually important lower band. At low-bit rate operation the higher band is synthesized based on the lower-band characteristics, while at the highest bit rate additional bits are reserved for coding the high-band signal. Another advantage of the sub-band structure is the fact that with 12.8 kHz sampling, the 20 ms frame contains 256 samples, thus enabling efficient bit-level operations and quantization schemes (e.g., for ACELP algorithms).

It should be emphasized that the utilization of wideband speech in an IP network does not introduce additional system complexity compared to narrowband. Within the IP network the operation is by definition transcoder free, since compressed speech is transmitted in IP packets end-to-end.

4.2.2.2   Cross-layer rate adaptation – CMR

The cross-layer signalling data of the AMR-WB packet called Codec Mode Request (CMR) for the VoIP flow rate adaptation is used for simulations. The CMR indicates the desired coding mode to the other end. The CMR must be then computed (to be written in the IP voice payload) and extracted from the IP payload received to select the appropriate codec for transmission. The CMR is computed based on quality measurements, which is assumed to be based on both the link state and the RTCP reports. Two major differences between the adaptive RTCP-based scenario and the AMR-WB-based scenario are: 1) the frequency of the reports sent by the RTCP may not be synchronized with the speed of adaptation required by the PHY and 2) the highest bit rate of the AMR-WB codecs is less than half of the Recommendation ITU-T G.711 and hence, the same system load allows for a higher number of VoIP connections.

Note that the inbound cross-layer signalling of both AMR-WB and RTCP allows for a fully distributed cross-layer VoIP bit rate adaptation.

4.3   Satellite-WiMAX performance model

4.3.1   Delay model

The delay budget model for each subnetwork is given by:

$$T_{tot} = T_{codec} + T_{network1} + T_{network2} + T_{tplayout} \qquad (1)$$

$$T_{network\ i} = T^i{}_{MAC} + T^i{}_{trans} + T^i{}_{prop}, \ i{=}1\ (DVB\text{-}S2/RCS),\ 2\ (WiMAX) \qquad (2)$$

T_{codec} is the delay introduced by the codec, which is as follows:

$$T_{codec} = T_{fr} + T_{la} + T_{proc} + T_{pack} \qquad (3)$$

where T_{pack} is the delay introduced when encapsulating more than one voice packet per IP packet, T_{fr} is the framing delay, T_{la} is the look-ahead delay (for prediction purposes) and T_{proc} is the processing delay. $T^i{}_{MAC}$ is the delay introduced in the MAC queuing and scheduling by network i. $T^i{}_{trans}$, the transmission time in network i. $T^i{}_{prop}$, the propagation time in network i, a GEO satellite Network. And finally, $T_{playout}$, playout delay to smooth out the jitter.

4.4 Performance results

4.4.1 Satellite subnetwork: Aggregate rate adaptation

This section presents simulation results for the satellite subnetwork. The aggregated expedited forward (EF) type of traffic in the forward link, which is controlled at MAC layer to guarantee a maximum delay of 270 ms (i.e. 20 ms is excess of the propagation delay) is simulated. A fully loaded system corresponds to an EF load of 10 Mbit/s with a bank of codecs and 3.5 Mbit/s using the AMR-WB codec (due to the lower maximum bit rate). Two thresholds one for decreasing and another for increasing codes, forming a hysteresis range are set to avoid quick fluctuations between VoIP speech rates when channel performs fast fading.

Figure 22 left shows the RTCP-driven adaptive scenario that counteracts a rain fade event of 12 dB. Note that the aggregated rate adaptation during the rain events is approximately, five minutes long. In both cases the system provides the guaranteed delay since channel attenuation is not noticeable by the EF traffic. It can also be observed that this latter case outperforms the RTCP-driven rate adaptation due to the following reasons:

1)   *cross-layer scenario does not add signalling overhead (RTCP reports) on the system;*

2)   *AMR-WB codec has inherent higher voice quality (Wide Band);*

3)   *AMR-WB highest bit rate is lower than the Recommendation ITU-T G.711 bit rate and therefore the system can admit a higher number of connections;*

4)   *like in the RTCP-based adaptive baseline, VoIP adaptation can be seen as a system load control.*

Both models allow not only providing guaranteed delay, but also allow the system having enough resources for servicing all users. A 30 MHz bandwidth channel (Rs) for the DVB-S2 is considered with 23 ACM modes (η_m being the spectral efficiency of ModCod m). Therefore, the maximum capacity of a RCST in users ($N_{u,max}$) when all users are using the highest VoIP speech rate is achieved as follows.

Each speech payload packet has a length $R_{VoIP} \cdot t_{pckt}$, where R_{VoIP} is indicated in Table 4, and t_{pckt} is the packet duration, 20 ms for all codes. After adding all header (RTP/UDP/IP of 12/8/20 bytes respectively), the total number of bytes of VoIP (n_{VoIP}) stream is: $n^i{}_{VoIP} = n^i{}_{RTP} + h_{RTP} + h_{UDP} + h_{IP}$, where i is the speech code used. On the other hand, the number of bits per frame in the DVB-S2 standard is $T_f \cdot \eta_m \cdot R_s$. Where T_f is the duration of N transmitted Base Band Frames (BBFRAMEs).

Therefore, the number of users that can be served are:

$$N_u = \left\lfloor \frac{T_f \cdot \eta_m \cdot R_s}{n^i{}_{VoIP}} \right\rfloor \tag{4}$$

FIGURE 22

100 DVB-RCS terminals aggregated expedited forward throughput for a channel attenuation
of 12 dB and RTCP-driven rate control (left) and adaptive AMR-WB-based (right).
Note that AMR-WB system load is lower due to lower bit rates for the best quality
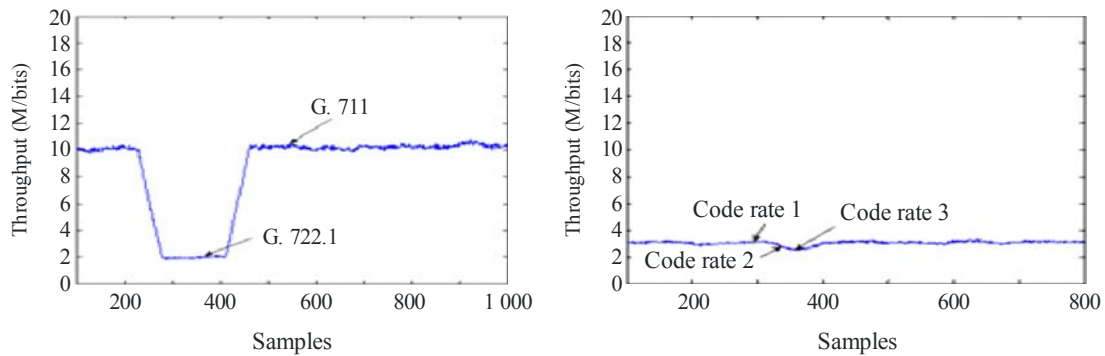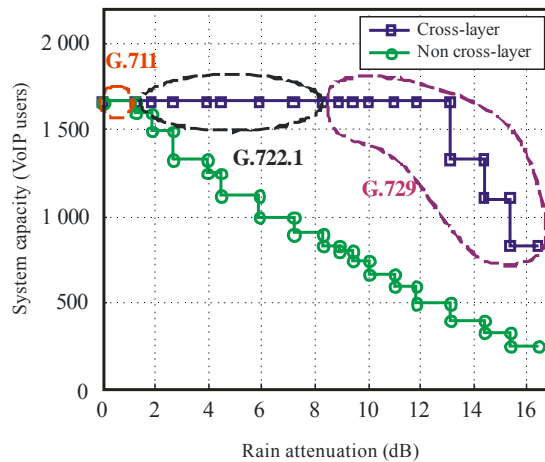

S.1897-22

A comparison between the cross-layer and non-cross-layer approaches is presented in Fig. 23. Maximum capacity is achieved under clear sky conditions (0 dB). The staircase form is due to the modulation and coding adaptation to the channel attenuation. Spectral efficiency gets lower as rain attenuation increases and thus decreasing capacity. It can be observed that cross-layer design allows maintaining the number of users for attenuations lower than 13 dB, and even for an attenuation of 15 dB, the capacity is not critically reduced. However if cross-layer is not applied, capacity drastically decreases, e.g. 50% of capacity is lost when attenuation reaches 8 dB.
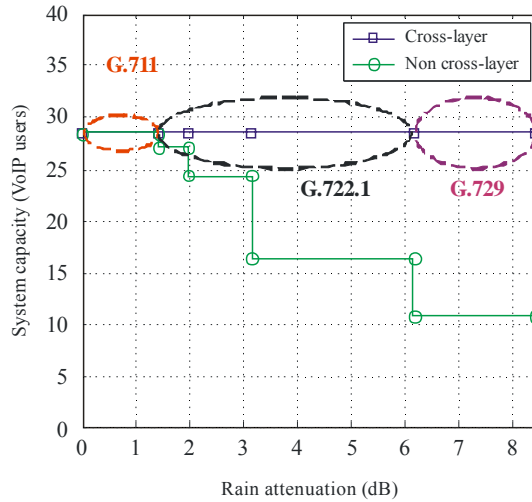
FIGURE 23

VoIP capacity for the satellite subnetwork as a function of
channel conditions Forward link (DVB-S2)

S.1897-23

FIGURE 24

**VoIP capacity for the satellite subnetwork as a function of
channel conditions Return link (DVB-RCS)**

S.1897-24

In Figure 24, a similar procedure is carried out for computing capacity on the return channel (DVB-RCS). In this case, a 1 MHz channel and 26.5 ms of frame duration (T_f) are assumed. It can be observed that cross-layer speech coder adaptation avoids a reduction in the number of VoIP users.

4.4.2 Terrestrial subnetwork performance

This section presents subnetwork performance simulation results. A rate adaptation at end user level, between BS and UE to maintain the capacity offered by the WiMAX subnetwork is considered. The capacity performance results with and without cross-layer design approach are provided.

FIGURE 25

**WiMAX VoIP capacity performance for cross and non-cross-layer,
depending on channel conditions**
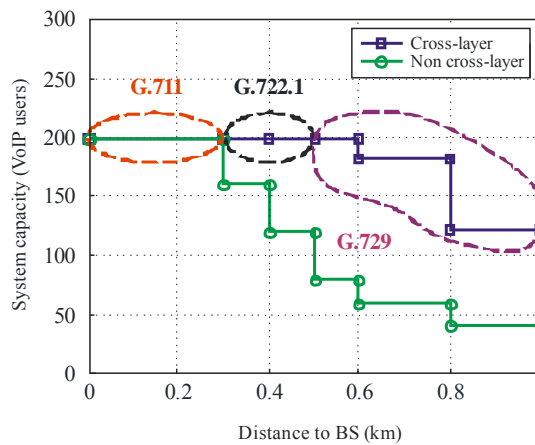
S.1897-25

As in the satellite subnetwork, an adaptive physical layer composed by 6 ModCods is assumed in the terrestrial segment to counteract channel fading. Channel Quality Information (CQI) of user is available in the ACM scheduler. Thus, both physical layer and VoIP are assumed to be able to adapt to medium-long term variations due to propagation losses (not to short-term fading). A downlink of 2 MHz channel bandwidth for RCST and its associated BS is considered. Figure 25 shows the system capacity for the terrestrial subnetwork, showing cross-layer VoIP capacity gain against a non-adaptive design. WiMAX coverage is assumed to be 2 km. Considering all users are in the border of the cell, capacity using adaptive approach is more than 60%, while in the other case is lower than 20%.

4.5   Summary

This Annex presents a cross-layer design of VoIP transmission over a satellite network using DVB-S2/RCS protocol and WiMAX system. Two approaches based on rate control and distributed concepts have been proposed and compared. It is shown with cross-layer design, that VoIP capacity can be controlled through rate adaptation in order to combat capacity decrease due to bad channel conditions in both the subnetworks: 1) aggregate VoIP rate adaptation to maintain the capacity that is offered by the satellite to the terrestrial network; 2) end-user rate adaptation to maintain the capacity that is offered by the WiMAX subnetwork. It can be concluded that cross-layer design approach yields the capacity improvement of 70% for satellite segment and 43% for the terrestrial network.

Annex 3

Cross-layer design for satellite link using TCP as transport protocol

1   Scope

GEO satellite links have a unique combination of characteristics that affect the throughput of TCP-based traffic. The critical elements are both the large Round-Trip Time (RTT) and the high Packet Error Rate (PER). Recommendation ITU-R S.1711 provides the TCP enhancements defined for networks with large Bandwidth-Delay Product (BDP) environment. This Annex provides the TCP performance as a result of interactions with lower layer protocols and, in particular, modulation and coding levels. This simulation considers different TCP versions, i.e. New Reno, SACK, Hybla and Westwood+.

2   Introduction

TCP versions such as NewReno, SACK, Westwood+, Hybla, Veno, Compound, Scalable and Binary Increase Congestion Control (BIC) are considered. A possible improvement to cope with multiple losses in a window of data is represented by the TCP NewReno that modifies the fast recovery phase of TCP Reno. TCP NewReno can recover from multiple lost packets (without waiting for the RTO expiration) by using the 'partial ACKs' received during the fast recovery phase.

In Selective ACKnowledgment (SACK), the receiver informs the sender on the successfully received packets, so that the sender can re-transmit just the actually lost packets. SACK can be implemented with both fast recovery and fast retransmit algorithms of the TCP NewReno version.

TCP Westwood and Westwood+ version sets *cwnd* and the *slow start threshold* (*ssthresh*) after a loss episode on the basis of an end-to-end bandwidth estimate, B_{we}, made before the loss event was detected. The *ssthresh* is made equal to $B_{we} \times$ RTT. Hence, TCP Westwood avoids a conservative *cwnd* reduction, thus allowing a faster recovery phase. A similar modification is made when RTO expires: *ssthresh* = $B_{we} \times$ RTT and *cwnd* is reset to its initial value. The bandwidth B_{we} of a connection is continuously estimated by considering the amount of data sent and the ACK interarrival time.

TCP Hybla proposes a modification to the *cwnd* update algorithm on the receipt of an ACK in order to accelerate the *cwnd* increase in both slow start and congestion avoidance phases for connections with large RTT values. Hybla foresees the adoption of SACK in order to recover quickly from losses due to both channel errors and the more aggressive injection of data in the network.

TCP Veno is an end-to-end congestion control scheme that can improve TCP performance. Veno enhances the Reno congestion control algorithm by using the estimated state of a connection based on Vegas. Instead of using Vegas' estimate on the network condition to prevent packet loss proactively, one can use it to determine whether packet losses are due to network congestion or the wireless channel. Specifically, if a packet loss is detected while the estimate indicates that the network is not congested, then the loss is due to the radio channel, so that *cwnd* is reduced less aggressively, preventing unnecessary throughput degradation.

Scalable TCP (S-TCP) modifies the congestion control algorithm of TCP. S-TCP "scales" well with the congestion window value and, hence, with BDP. S-TCP is based on a Multiplicative Increase Multiplicative Decrease (MIMD) algorithm, according to which the congestion window is increased by a factor α upon an ACK reception and decreased by a factor β upon a packet loss event. The simulation experiments assumed $\alpha = 0.01$ and $\beta = 0.125$.

In the Compound TCP (C-TCP), a new state variable namely, *dwnd* (*delay window*) is introduced, which controls the delay-based component in C-TCP. The conventional *cwnd* remains untouched to control the loss-based component in C-TCP. The C-TCP sending window now is controlled by both *cwnd* and *dwnd*.

The BIC TCP is a TCP variant for high-speed long-distance networks. When a packet loss occurs, BIC reduces its window by a multiplicative factor. The window size just before the reduction is set to the maximum and the window size just after the reduction is set to the minimum. BIC performs a binary search using these two parameters by jumping to the midpoint between the maximum and the minimum.

3 Reference network architecture

Figure 26 shows a network architecture featuring a GEO bent-pipe satellite considered for simulation experiments. A client (TCP receiver) is connected to a router-earth station which is connected via satellite to another router-earth station linked to a remote server (TCP sender) from which files are downloaded. A return channel via satellite is used to send both lower layer signalling (i.e. link quality estimation for modulation and coding adaptation) and transport layer ACKs.

FIGURE 26

Reference network architecture
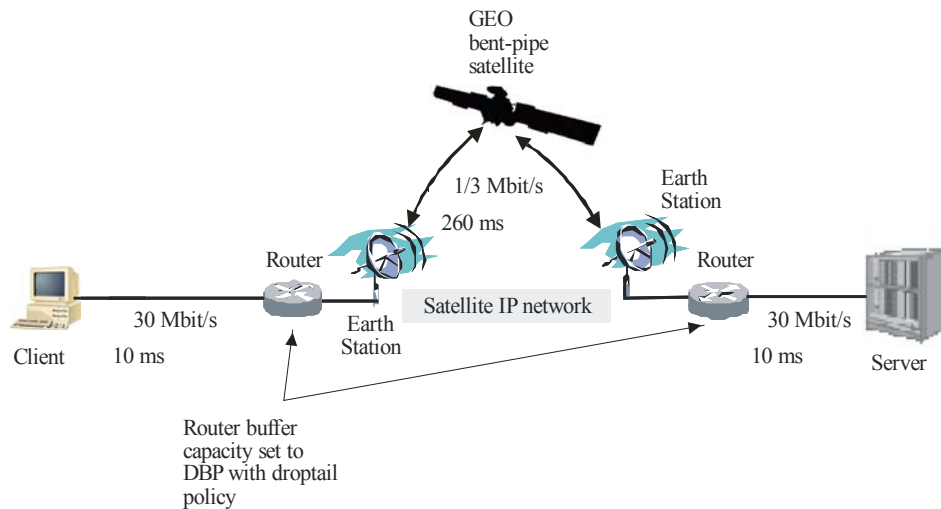
The GEO satellite link at Ka band with a bandwidth of 2 MHz for both uplink and downlink is considered. Two Modulation and Coding (MODCOD) options are considered at PHY, referring to a DVB-S2-like system adaptation. The selection of the MODCOD at PHY determines the information bit-rate available to convey IP datagrams.

3.1 Simulation parameters

The earth station-to-earth station one-way propagation delay is about 260 ms. The terrestrial links from the earth station and the client and from the other earth station to the server are at 30 Mbit/s with a (one-way) propagation delay of 10 ms. Hence, the satellite link is the system bottleneck and the propagation delay contribution to RTT is equal to 560 ms.

3.2 Assumptions

Some of the assumptions include: (*i*) earth stations with a Line-of-Sight (LoS) path to the GEO satellite; (*ii*) a memory-less channel with uncorrelated losses (fixed users); (*iii*) attenuation fluctuation only due to slow troposphere events (long-term variations of the received signal strength due to cloud attenuation and rain fades) so that the channel can be considered of the AWGN type; (*iv*) residual packet losses after the decoding process are uncorrelated and occurring according to a given PER at the transport level; (*v*) channel variations are very slow compared to the delay of the feedback signal to inform the sending earth station to modify its transmission mode; (*vi*) no Automatic ReQuest repeat (ARQ) technique is used at layer 2, since the adoption of an ARQ scheme would entail a long delay to recover packet losses that could cause an RTO expiration at the TCP level.

The TCP goodput performance for the transmitted packets has been addressed. An FTP application (persistent TCP connection) that produces TCP traffic according to an ACK-clocked model is assumed. ACM is used for the satellite link to provide an acceptable quality in different radio channel conditions. The transmission mode (i.e. MODCOD) adaptation is performed by the sending earth station on the basis of the channel quality measure made by the receiving earth station. Two modulations, i.e. BPSK and QPSK and a convolutional encoder/Viterbi decoder, the standard NASA 1/2 rate convolutional code with constraint length 7 and derived punctured code with rate 3/4 have been assumed. Two transmission modes (MODCODs) operated at parity of transmission bandwidth (2 MHz) and power considered are:

Mode #1: BPSK with rate 1/2 convolutional encoder, and resulting information bit-rate of 1 Mbit/s;

Mode #2: QPSK with rate 3/4 convolutional encoder, and resulting information bit-rate of 3 Mbit/s.

An Ethernet packet of 1 500 bytes has been adopted. Therefore, the BDP product is 560 000 bits (about 47 packets) for mode #1 and 1 680 000 bits (140 packets) for mode #2. Each router has a buffer with a capacity of packets equal to BDP for mode #2 (i.e. the maximum BDP value between mode #1 and mode #2). The queues adopt a "droptail" policy to manage the congestion, i.e. arriving packets are lost according to the buffer capacity.

4   Performance results

The simulation experiments employed the ns-2 2.29 version. Simulations have been performed for static radio channel conditions, i.e. assuming links with fixed E_b/N_0 all over the simulation length. To define an optimal E_b/N_0 threshold for the selection between mode #1 and mode #2, two criteria have been compared:

Scheme *a*: Select the transmission mode on the basis of a PER threshold value (i.e. *non-cross-layer, classical approach* for current air interfaces with ACM);

Scheme *b*: Select the transmission mode with the aim of maximizing the goodput** performance at the transport layer (*implicit cross-layer design, top-down method*).

The most convenient switching point, in terms of E_b/N_0, depends on a suitable trade-off between large bandwidth versus higher PER, for good TCP performance. Such trade-off depends on the adopted TCP version.

Referring to the classical scheme *a*, Fig. 27 shows the TCP goodput as a function of the E_b/N_0 value considering a PER switching threshold equal to either 10^{-1} (upper graph) or 10^{-4} (lower graph). In both cases, such selection criterion is inefficient. In fact in the upper graph the switching point is too anticipated for some TCP versions, resulting in a drop of performance when changing from transmission mode #1 to transmission mode #2. Whereas, in the lower graph, the switching point is too delayed so that the system is constrained to use transmission mode #1 when mode #2 would be more convenient. It is not an efficient utilization of resources. The results in Fig. 27 prove that an appropriate selection criterion between mode #1 and mode #2 should not be based on PER performance, but rather on the transport layer performance and be directly related to E_b/N_0 (cross-layer approach). That is why scheme b is selected. The results are shown in Fig. 28 for TCP NewReno, SACK, Westwood+, Hybla and in Fig. 29 for TCP Veno, Compound-TCP, Scalable-TCP and BIC-TCP. From these results, it is evident that the selection driven by the TCP goodput behaviour permits to achieve a better efficiency and, better QoS performance perceived at the application layer. It is therefore evident that the optimal selection between mode #1 and mode #2 depends on the TCP version adopted. Finally, the results show that among all the TCP versions compared, TCP Hybla achieves the best performance followed by Westwood+.

** Goodput is the application layer throughput, i.e. number of useful information bits, delivered by the network per unit of time, excluding protocol overhead as well as retransmitted packets.
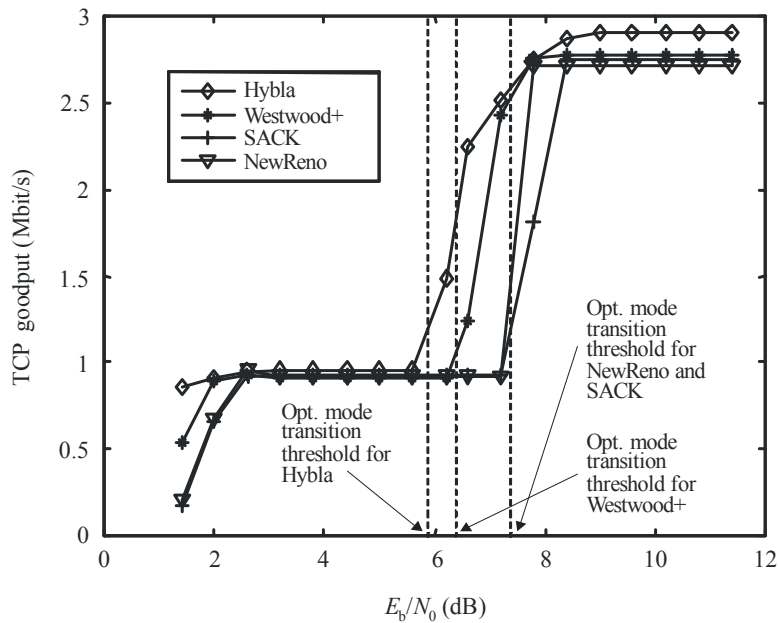
FIGURE 27

Simulation results for the different TCP versions as a function of the E_b/N_0 (reference value for BPSK) in the case of scheme *a*; the maximum 95% confidence interval amplitude for all the curves is ±7%
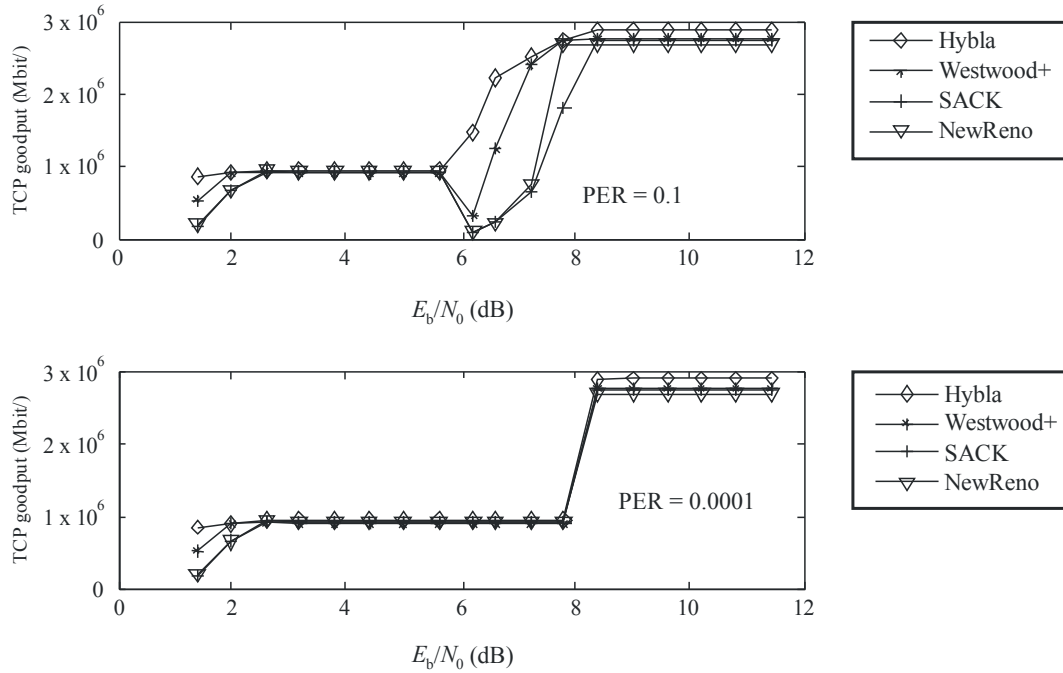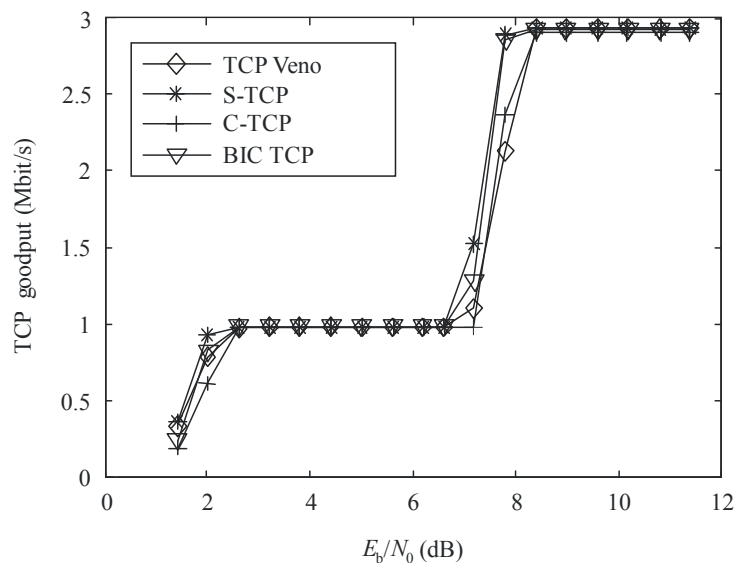

S.1897-27

FIGURE 28

Simulation results for the different TCP versions (NewReno, SACK, Westwood+, Hybla) as a function of the E_b/N_0 (reference value for BPSK) with scheme *b*; the maximum 95% confidence interval amplitude for all the curves is ±7%

S.1897-28

FIGURE 29

**Simulation results for the different TCP versions (Veno, Compound, Scalable and BIC)
as a function of the E_b/N_0 (reference value for BPSK) with scheme b; the maximum 95%
confidence interval amplitude for all the curves is ±7%**

S.1897-29

5 Summary

Cross-layer design is critical to improve the throughput performance and the QoS provisioning in IP-based satellite communication networks. This simulation experiments considered interactions between mode adaptivity at layer 1 and TCP goodput behaviour at layer 4 in the presence of a satellite error-prone channel. Simulation results demonstrate that physical layer mode selection should be driven by TCP protocol version and the related goodput performance. These results demonstrate the advantages for coordination or cross-layer interactions between PHY and TCP behaviour. PHY switching thresholds among transmission modes should be determined on the basis of the higher layer performance for efficient utilization of satellite radio resources.

6 Conclusions

The various simulations experiments results presented in this Annex provide guidance on the design of hybrid satellite-terrestrial networks depending on the network architecture.

The following results are of particular significance in cross-layer based design of such hybrid networks:

– Performance results for hybrid satellite – WiFi networks, show the influence of fading over QoS parameters such as bandwidth allocation, delay and delay variations for video conferencing, VoIP, HTTP and TCP/UDP applications (see § 3).

– The highest priority traffic like streaming multimedia or interactive voice traffic was unaffected by the rain fade (see § 2).

– For video conferencing rain fade caused a peak delay of about 0.859 seconds (see § 3).

– The end-to-end delay for VoIP reached 1.5 seconds for (best efforts) during rain fade.

– The packet delay variation has a peak close to 0.08 seconds squared, during rain fade and declined to 0.025 seconds by the end (see § 3).

–   The jitter had a peak of 0.0045 seconds during rain fade and had a range of 0.0025 to 0.002 at other times.

–   The HTTP page response time ranges from 2-6 seconds and object response time varies from 1-4 seconds during rain fade for best effort traffic class (see § 3).

–   For VoIP transmissions over satellite-WiMAX networks, cross-layer design allows maintaining the number of users for attenuation lower than 13 dB, and even for an attenuation of 15 dB. The capacity is not critically reduced. However, without cross-layer design 50% capacity is lost for attenuation of 8 dB (see § 4).

–   Cross-layer design for hybrid satellite-WiMAX networks yields capacity improvement of 70% for satellite segment and 43% for the terrestrial segment (see § 4).

–   Coordination between physical and transport layer is required and transmission mode switching at physical layer should be determined based on the higher layer performance for efficient utilization of satellite resources (see § 5).