Satellite and Next Generation Networks: proposal for QoS architectures

Stéphane Combes
Alcatel Space, 26 Av. Champollion, 31037 Toulouse cedex, France
stephane.combes@space.alcatel.fr
Tel: (33) 53435 6938/ Fax: (33) 53435 5560

Olivier Alphand, Pascal Berthou, Thierry Gayraud
LAAS-CNRS
7, avenue du Colonel Roche
31077 Toulouse cedex 4, France

Abstract

Broadband satellite systems will need specific functions and interfaces in order to fit seamlessly into an end-to-end Next Generation Network (NGN) Quality of Service (QoS) architecture. This paper will define NGN overall architecture, showing how the coupling between Application/session layers and Network/transport layers can achieve end-to-end QoS support. Possible integration scenarios of satellite access into such architectures will be depicted, with a special emphasis on transparent and regenerative satellite architectures based on DVB-RCS [14] type of access.

Detailed architectures proposed by two recent research projects, IST Satellite Broadband Multimedia System for IPv6 (SATIP6) [21] and ESA Integrated Resources and QoS Management in DVB-RCS networks (QoSforRCS) [11] [7], will be presented. On the one hand, we will develop a “QoS assured” architecture proposed in the frame of QoSforRCS: it provides dynamic Service Level Agreement (SLA) management and strict QoS control. Admission Control (AC) function is located in the satellite Hub which allows fine control and billing by the Service Providers. On the other hand, a “QoS enabled” architecture proposed in the frame of SATIP6 is presented: it allows simple QoS differentiation between applications based on configuration by the end user or the subscriber’s network administrator. This architecture may include AC function and is designed for Virtual Private Network (VPN) deployment.

1. INTRODUCTION

1.1 WHAT IS NGN?

The following definition is an excerpt of ETSI’s NGN-Starter Group conclusions (see [15]):

« The term NGN is commonly used to give a name to the changes to the service provision infrastructures that have already started in the telecom and IT industry. As such it is not a term that can be precisely defined but is rather an umbrella term to describe developments following PSTN/ISDN/GSM phase 2+ era.

One of the main characteristics of NGN is the uncoupling of services and networks, allowing them to be offered separately and to evolve independently. Therefore in the NGN architectures proposed there is a clear separation between the functions for the services and functions for the transport. An open interface is provided between both. NGN allows the provisioning of both existing and new services independently of the network and the access type used.

NGN will have to provide the capabilities (infrastructure, protocols, etc.) to make the creation, deployment and management of all kinds of services (known or not yet known) possible. This comprises services using all kinds of media (audio, visual, audiovisual), with all kinds of encoding schemes and data services, Conversational, Unicast, Multicast and Broadcast, Messaging, simple data transfer services, Real time and Non-Real time, delay sensitive and delay tolerant services. Services with different bandwidth demands from a few kbit/s to hundreds of Mbit/s, guaranteed or not. Within the NGN there is an increased emphasis on service customisation by the Service Providers whereby some of them will offer their customers the
possibility to customise their own services. NGN will comprise service related APIs (Application Programming Interfaces) in order to support the creation, provisioning and management of services.

In NGN the functional entities controlling policy, sessions, media, resources, service delivery, security, etc. may be distributed over the infrastructure, including both existing and new networks. When they are physically distributed they communicate over open interfaces. That is why the NGN architectures proposed in standards bodies and fora consist of layers and planes and show a lot of reference points. New protocols are being standardised to provide the communication between these functional entities. Interworking between NGN and existing networks such as PSTN, ISDN and GSM is provided by means of Gateways.

NGN will support both existing and "NGN aware" End Terminal Devices. Hence terminals connected to NGN will include analogue telephone sets, fax machines, ISDN sets, cellular mobile phones, GPRS terminal devices, SIP terminals, Ethernet phones through PCs, digital set top boxes, cable modems...

Specific issues are certainly the migration of voice services to the NGN infrastructure, Quality of Service related to real time voice services (bandwidth guarantees, delay guarantees, packet loss guarantees, etc) as well as Security. NGN should provide the security mechanisms to protect the exchange of sensitive information over its infrastructure, to protect against the fraudulent use of the services provided by the Service Providers and to protect its own infrastructure from outside attacks.

One of the main consequences of these principles is the clear split between Access Network and Core Network. The latter has to be able to propose multimedia services to end-users regardless the type of Access Network used to get the connection: UMTS, ADSL… or Satellite Access Network.

Another general requirement is the Personal Mobility. End-users can get a connection anywhere in the network and always retrieve the same service environment. Therefore, two different networks are always considered: The Home Network is the network the end-user has subscribed to. Services are always provided by the Home Network; thus, the service environment retrieved by the end-user is always the same, and it is independent of the physical connection point. The Visited Network is the network providing the connection point (and therefore generally owning the access network) to allow to reach the home network, and the access to the core transport network.

1.2 NGN QoS AND SATELLITE ACCESS

Near real-time data communications (e.g. WWW browsing) has already demonstrated that the Internet, and the underlying generation of packet network technologies, which operate on the basis of ‘best effort’, provide insufficient QoS support (e.g. World Wide Wait). In the NGN, packet network technologies will have to support real-time services (voice, multimedia…) and will a fortiori require improved QoS support. These networks will thus have to include advanced QoS management capabilities such as:

- QoS aware network equipment (Intserv [5], DiffServ [3])
- Admission control mechanisms
- Interaction between call signaling, 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 other access networks: cable access [25] or 3GPP [1]. Their absence in the satellite field may lead to slow down the uptaking of NGN and their associated services by satellite.

Security threats such as denial or theft of service have to be taken into account in the design of such systems. In such NGN, the revenue stream will mainly come from these QoS enhanced services. Network users can thus go to extreme lengths to steal from these pricey services. The addition of security technology is necessary to protect the system from such threats. Reasonable costing technology has to be employed in order to force potential fraudulent users to spend an unreasonable amount of money to circumvent them. In addition, the provisioning of voice services over a satellite NGN has to be as secure as the PSTN networks.
are today. In order to provide the same level of privacy and resistance to denial of service attacks than
dedicated lines, the satellite system will have to apply cryptography-based security mechanisms.

The present paper discusses two complementary QoS architectures which aim at promoting the integration of
satellite systems into QoS enhanced NGNs. Section 2 and 3 introduce the overall NGN QoS architecture,
highlighting the interactions between the QoS specific functions, such as Admission Control, and the
network resources management processes. Target satellite network architectures, one based on traditional
transparent payloads, the other one on advanced regenerative ones, are described in Section 4. Then Section
5 and 6 respectively detail the two envisaged satellite network QoS architectures.

2. FRAMEWORK FOR NGN END-TO-END QoS CONTROL

This section details a proposal for a generic NGN QoS architecture, spanning multiple network domains and
various networking layers that can be involved. It takes into account latest works in progress in
standardization bodies or fora such as ITU-T, ETSI TIPHON [10], UMTS, DSL Forum or IETF.

2.1 General QoS Model

2.1.1 The QoS layers

The general model for QoS handling is based on a four layers splitting (see Figure 2), mainly characterized
by the scope of the handled entity. The model itself is general enough to describe the role of each layer and
the relations with the next upper and lower layers, but without precluding any actual configuration:

1. On top of the model, the Application Layer owns the logic of the service to be provided to the end-user
(e.g. music distribution, conference handling or personal routing). When dealing with multimedia sessions,
the application layer defines the QoS associated to the service, in a transport technology independent way.

2. The Session Layer handles the signaling procedures which includes session establishments between end-
users, access to other end-points, like conference bridge servers, and the negotiation of session
characteristics. Besides the end-users, the network elements providing session signaling capabilities are also
considered in this layer, like e.g. Session Initiation Protocol (SIP) [30] proxies. The entities addressed by the
session layer are the end-users. They can, for instance, be known by their Uniform Resource Identifier (URI)
which can be considered as a logical identifier. The URI of an end-user is expressed using the email-like
form user@domain. Such signaling allows to describe:

- The user end-points of the session
- QoS parameters (codec, frames per packet, frame size, jitter buffer delay, FEC, mean delay variation,
  packet loss)
- Other service-related parameters

3. The Connection Control Layer handles the bearer set-up for each media component, including admission
control and resource reservation. Bearers are considered end-to-end, or at least edge-to-edge inside a network
domain.

4. The Transport Layer is responsible for conveying the data packets from a source to a destination, using
the physical network. Therefore, the hop-by-hop mechanisms to traverse routers are considered here, that is,
the actual technology used in the transport network, as e.g. DiffServ, IntServ over DiffServ [2] or Multi-
Protocol Label Switching (MPLS) [29].

2.1.2 The QoS process

The contractual relation between the Service Provider and the Service Subscriber specifies which levels of
QoS the Subscriber accepts to pay for. Thus, the service subscription has a main component based on the
QoS Subscription, corresponding to the commitment of the Service Provider with regards to QoS. The QoS
Subscription could be embedded in a global service package, therefore not directly seen by the Subscriber as
an explicit parameter.
The layered general model for QoS handling then allows to negotiate the QoS to be used for the session set-up in the following way:

- Firstly, the QoS has to be negotiated end-to-end at the Application/Session layer. This implies the verification of the user rights in the subscriber profile in both the originating and terminating sides. For a given session request, the set of QoS possibilities that could be used for the session set-up have to be selected and negotiated, taking into account different sources of QoS information, as e.g., the end-user request, the subscriber profile, the application request or the known network capabilities. This QoS negotiation involves the home network and the visited network, for both the originating end-point and the terminating end-point.

- Once the session characteristics including QoS have been negotiated and agreed, the Connection Control layer has to verify the corresponding bearers against the operator policies in the visited network, to decide if the bearers needed to implement the media component can be admitted or not. The main policy is Admission Control, that is, the verification that the requested resources are actually available in the network domain, and therefore the reservation of those resources.

- Finally, the Transport layer is requested to set-up the bearers in the transport network, and applying the actual technology used in that network. Next, it is up to the Transport layer to activate the corresponding mechanisms in the transport network according to the operator policies: RSVP, DiffServ or MPLS.

If the bearer set-up fails (because of policy refusal or unavailable QoS, for instance), then the Session Layer has to decide if the session can still proceed or not. The session characteristics can be re-negotiated at any moment by any of the end-users: add new media, remove media or modify the characteristics of a given media.

QoS is considered to be a part of each media component description, and therefore to be part of the session description. This means that QoS aspects can be also negotiated or re-negotiated, at any moment, as any other element of the session description. When the QoS part of a session description is re-negotiated, the same QoS process is run again including the four layers, for all the media components of the multimedia session. The media components which are not involved by this re-negotiation will obviously not go through a new admission control and/or reservation procedures.

2.1.3 The QoS network

Whatever the considered standardization reference, NGNs always differentiate two kinds of network elements, whether they belong to the Application or the Transport plane, as shown on Figure 1.

The application plane elements provide services to users. Service is requested through user/call signaling protocols, such as ITU-T multimedia signaling protocols (H225 [22] + H245 [23]), IETF SIP combined with SDP [16] or ITU-T Gateway Control Protocol (H248 [24]).

The transport plane provides a packet oriented transport service and the desired QoS. Such QoS is requested thanks to QoS signaling protocols, such as the IETF Resource Reservation Protocol (RSVP [4]) and Common Open Policy Service protocol (COPS [10]). QoS signaling can be exchanged with user endpoints and/or the application plane and across different network domains.

2.2 QoS in the Application/Session layer

The terminology used to describe the QoS architecture is aligned with the one being standardized at the 3GPP [1]. The Application layer is provided by the final Application itself, which can even own a dedicated subscriber profile. The Session layer is provided by the Call Session Control Function (CSCF).

The negotiated session characteristics are carried in a SDP description which is conveyed by the SIP session signaling protocol and which contains for each media component of the multimedia session:

- The type of the media component: voice, audio, video, application, data.
- The list of codecs and encoding formats, if relevant (e.g., voice, video).
The Traffic Information (bandwidth, burstiness, packet size, etc.) and QoS Sensitivity information (end-to-end delay, maximum jitter, packet loss, etc.) can be either implicitly derived from the medium type and the codec information (if possible), or explicitly specified in the media description.

The Application always stands in the Home Network, it provides the effective application logic to the end-user. The Application is by definition a non-standardized element (there is indeed a very strong willingness in the standardization bodies like 3GPP and IETF to definitely avoid service standardization).

The Proxy Call Session Control Function (P-CSCF) function is mainly due to the inter-domain roaming configuration, where the visited domain is responsible for taking care of bearer resources requested by the roaming end-user or by the Serving-CSCF (on demand from an application). It uses the media component description in SDP carried by e.g. SIP or Real-Time Streaming Protocol (RTSP [32]).

The Serving Call Session Control Function (S-CSCF), located in the Home Network, provides access to the home service environment. All (SIP) session signaling is routed via the Home Network, while the (multimedia) data streams (and eventual resource reservation signaling) usually follow a more straightforward way.

2.3 QOS IN THE CONNECTION/TRANSPORT LAYER

When the Session negotiation has ended, the P-CSCF associated to each one of the end-users (that is, the originating end-user and the terminating end-user) receives the final result of the negotiation process (a Session Description containing an agreed SDP coded description of each media component, and therefore including traffic and QoS information).

At this point, the end-users and the concerned applications have agreed on which media components to use with which QoS level. The next step is to verify if the bearers implementing the media components can be admitted or not, and to set-up those bearers.

Figure 3 shows the QoS architecture in the connection control and transport layers:

- The P-CSCF provides the negotiated session description to the Policy Control Function (PCF). The PCF function translates, for each media component, the media description into a set of (IP) QoS parameters. The translated information is then used by the PCF for application of the operator's policies, including resource reservation and connection set-up.

- The Admission Control (AC) function is requested by the PCF function to verify the actual availability of requested resources in the core transport network.

- The Access Gate (AG) is located in the transport data path and performs per micro-flow processing (policing, monitoring statistics, etc.) The AG performs all the inter-working between the Access Network specific procedures and the Core Network. The AG acts as a Policy Enforcement Point (see section 3.1) and enforces the QoS policy decisions taken by the PCF function.

The AG manages access to QoS services through “Gates”. Each QoS session is associated with a gate entity, which corresponds to a packet classifier and policer. It ensures that only IP flows that have been granted access to QoS services by the AC can pass through. “Opening” a gate is done selectively for each flow (or each call in case of a telephony session). In a two-phase resource commitment (“QoS assured” mode where a session completes only if all the required resources are available and assigned to the session), gates should only be opened when the destination user picks up.

The interface between PCF and AG is standardized by the 3GPP and is COPS based (see section 3.1). For consistency reasons, it is expected that the interface between P-CSCF and PCF will also be COPS based. The interface between PCF and AC is supposed to be proprietary.

In some access networks, the resource management problem (reservation, allocation) is solved using an explicit protocol between the end-user and the AG. In other access networks, there is no protocol at this level (e.g. in case of ADSL... or satellite access); therefore, a specific entity has to be considered in the access network to solve at least the following questions:
- Identification of the end-points (end-users), and especially to solve the question of getting the actual IP address of the end-user.

- Admission control on the access network, probably including the end-user access link.

- Actual resource allocation when user access link is dynamically shared, as it is the case with satellite access.

This new functional entity is called the *Access Resource Control (ARC)* function in the present document, as shown on Figure 3. Both the ARC and the AC functions are requested by the PCF function on a session by session basis. In case of SIP sessions, this approach allows to perform only once the translation between SDP parameters into an IP flow description at the PCF, and next ask:

- the ARC entity to get resources in the access network
- and the AC entity to get resources in the core network.

### 2.4 GENERAL ARCHITECTURE

Figure 4 summarizes the overall architecture in one single picture, showing the relations between all the involved elements in a simplified mono-domain case. It shows that, at the end, the application itself can control the subscribed QoS to be delivered to the end user for a given session.

Thus, the Admission Control function performs admission control of micro-flows needed to implement the media component of a multimedia session (e.g. audio, video or application sharing), independently of how physical resources are organized, implemented and managed in the Transport Domain.

### 3. ADMISSION CONTROL AND LINK WITH RESOURCES MANAGEMENT

This section introduces the Policy Based Admission Control framework, as defined by the IETF, and proposes different interaction modes between AC and Resource Management functions.

#### 3.1 POLICY BASED ADMISSION CONTROL

IETF working groups have developed explicit signaling of the QoS requirements (IntServ model) from the user. Inside each QoS-enabled router, the admission control component only takes into account the requestor’s resource reservation request and the available capacity to determine to whether accept QoS requests or not. However, the IntServ mechanisms do not include some important aspects of admission control. Indeed network managers and service providers must be able to monitor, control and enforce the use of network resources and services based on policies. These policies can be derived from criteria such as the identity of users and the type of applications, traffic/bandwidth requirements, security considerations and time-of-day/week. Similarly, DiffServ mechanisms also need to take into account policies that involve various criteria such as customer identity, ingress points and so on.

Another point is about end-to-end QoS for which we face the following alternatives:

- IntServ allows it as long as all routers are IntServ enabled
- DiffServ defines a Per-Hop Behavior (PHB), which can vary from router to router or at least from administrative domain to administrative domain
- There is no end-to-end QoS !

The IETF policy model [34] is a general framework for providing policy-based control over admission control decisions. The two main architectural elements for policy control are the *PEP (Policy Enforcement Point)* and the *PDP (Policy Decision Point)*. PEP is a component at a network node and PDP is a remote entity that may reside at a policy server. The PEP represents the component that always runs on the policy aware node (e.g. the Edge Router). It is the point where policy decisions are actually enforced. Policy decisions are primarily taken by the PDP. The PDP itself may use additional
mechanisms and protocols to achieve additional functionalities such as user authentication, accounting, or policy information storage.

According to these definitions, in the NGN architecture: PCF acts as a PDP and AG acts as a PEP.

The following interaction models between PDP, PEP and AC functions can be envisaged:

- **Pull model:** When a PEP receives a notification or a message that requires a policy decision, it then formulates a request for a policy decision to the PDP. This is the “Pull” model: network elements “pull” policy decisions from policy servers. The standardized protocol between PDP and PEP is COPS. How COPS is triggered when PEPs receive RSVP is described in [18]. This model is also known as “Outsourcing” mode in COPS terminology. In this scenario, the PDP (PCF) waits for a policy decision request coming from the PEP (AG). Such a request is generated when the PEP function at the AG is activated by a QoS request. This is a likely scenario when RSVP or a dedicated Access Network protocol is used.

- **Push model:** Inversely, PDPs may “push” policy decisions to PEPs without having been notified first by the PEPs. This is the case when a PDP directly receives QoS signaling messages, either from an end-point or from a Call/Application Server. This is the “Push” model or “Provisioning” mode in COPS terminology which is described in [6]. In this scenario, the PDP (PCF) gives to the PEP (AG) all the information needed to perform the bearer reservation. The bearer reservation can be triggered by the user (e.g. RSVP or Access Network specific protocol) or by the AG itself (e.g. Access Network specific protocol). The PCF had previously received all needed information to perform its policy decision from the P-CSCF (e.g. SDP session description). It is then up to the PCF to trigger the AC on a micro-flow basis (only for micro-flows requiring QoS). After bearer set-up, the PCF (or the AG) will check the consistency between the QoS authorized by the policy decision and the QoS requested at the transport layer (if a user initiated QoS protocol such as RSVP was used at this level).

This is the most likely scenario for Applications not supporting a QoS signaling protocol like RSVP.

- **Scenario not using a Network Session layer:** In this scenario, end-points do not use any network session signaling with P-CSCF but we suppose that micro-flows still need to be set-up with specific QoS characteristics. Admission control is therefore needed, but only at the Connection/Transport layer. It is therefore assumed that user end-points know each other and have exchanged addressing and media component types information by some means: direct SIP or H.323 exchanges between end-points (as in the Internet today), information retrieved from an Application Server.

For instance, if the Application Server is a File Transfer Protocol (FTP) server, the peer agreement could be the handshake during the opening of a FTP session. In this scenario, RSVP is the protocol likely to be used between the end-user and the AG. Then the Pull model applies between the AG and the PCF.

### 3.2 LINK WITH RESOURCE MANAGEMENT

Whatever the interaction model between policy elements is, the interaction with the actual Resource Management functions, booking or allocating bandwidth or other network resources, can be performed according to one of the 3 following modes:

- **Best effort:** This mode is definitively not suited for applications requiring stringent QoS and will not be addressed in the developed scenarios.

- **QoS-Enabled:** A “QoS-Enabled” session allows the endpoints to complete the session establishment either with or without the desired resources. Such session will use dedicated resources if available, and use a best-effort connection as an alternative if resources cannot be dedicated. In cases where resources are not available, the originating and/or terminating user end-point might check with the user to obtain guidance on whether the session should complete or not.

This mode of operation will be considered in the QoS architecture developed in section 6.
- **QoS-Assured**: A “QoS-Assured” session will complete only if all the required resources are available and assigned to the session. A provider may choose to block a call when adequate resources for the call are not available. For voice service, public policy demands that the phone system provide adequate quality at least in certain cases: e.g., for emergency communications during times of disasters. Call blocking enables a provider to meet such requirements. This mode should be preferred for real-time services especially for a full telephony service and will be assumed in the developed scenarios.

This mode of operation will be considered in the QoS architecture developed in section 5.

4. **SATELLITE NETWORK ARCHITECTURE**

4.1 **TRANSPARENT SATELLITE SYSTEM**

In the following, we will take the example of a Satellite Access Network (SAN) with a star topology, based on the DVB-RCS standard and supported by a transparent bent-pipe conventional satellite. In such a transparent architecture the SAN acts as an interface between the subscriber networks and the ground (backbone) networks towards the Service Providers (SP); the Satellite Terminals (ST) and the Gateways (GW) are the SAN components interfacing with the subscriber’s networks and the ground networks, respectively. The architecture precludes direct (one-hop) mesh communications between STs.

For the purpose of this paper it is assumed that the SAN has a single GW access point to several SPs and that the SAN is managed by a single operator, denoted as Interactive Network Access Provider (INAP). It is also assumed that the GW can communicate with STs in several beams.

The satellite transparency requires that the ST receiver (and GW transmitter) is MPEG-2/DVB based, and that the ST transmitter (and GW receiver) is DVB-RCS based. A multi-beam architecture would require that the GW supports several forward links (in TDM mode) and several return link carrier groups (in MF-TDMA mode). For the return link the ATM profile is taken as baseline.

The satellite transparent architecture enables to provide Internet/Intranet access services with Internet/Intranet servers that can be reached through the GW. Access to corporate networks (VPN, LAN interconnection) is also possible.

The satellite service subscriber’s network can be of one of the following types:
- Residential (with a single or several subscribers per subscriber’s network)
- SOHO/SME (with a single or several subscribers per subscriber’s network)
- Residential + teleworking (with a single or several subscribers per subscriber’s network)
- Corporate (with a single subscriber).

Several subscribers per subscriber’s network correspond to the multi-dwelling case.

Subscriber’s networks include a variety of Customer Premises Equipment (CPE) including PCs, TV sets, IP Phones (standalone telephone with Ethernet interface) or Softphones (Software applications for telephony residing on user’s PC). In multi-dwelling situations, the ST can be owned by an Access Provider (telecommunication operator) and denoted Point-of-Presence (POP). Service distribution to users behind the POP can rely on various technologies (Ethernet, DSL, WiFi).

Figure 5 shows a high-level view of the SAN together with the attached external networks. Please note that:

- SP networks can be public, as it is the case with Internet SP (ISP) or various Application Service Providers (ASP), such as Multicast SP or Internet Telephony SP.
- SP networks can be private, as it is the case with VPN.
- The network interconnecting the GW with SP networks can consist of a variety of interconnected backbone networks.
The SAN and the SP networks are interconnected at Edge Routers (ERs), which can be located in the GW itself, if there is a direct connectivity between the GW and the SP networks. ISPs offer Internet access (i.e. IP address) and Internet services (e.g. WWW, e-mail), whereas ASPs offer application services such as telephony, video-telephony, video-on-demand or gaming. ASP network can include a Multimedia Call Server (MMCS), handling multimedia session control. This MMCS can be interfaced with 3rd party applications using SIP or Open Service Access (OSA)/Parlay interface [26].

It is supposed that residential users will have public IP addresses. Professional users can have public or private IP addresses. When they have private IP addresses, they are not able to use services provided by ISP or ASP unless they have a NAT function or use tunnels.

The usual operational dependencies between SP (ISP and ASP) and INAP are as follows:

- Service Level Agreement (SLA), which is a formal negotiated agreement between SP and INAP, typically includes a specification of QoS and security guarantees offered by INAP to SP.
- ASPs will not need to assign and manage IP addresses or provide AAA services at access level; this will be done by the INAP. The ASPs are provided with the knowledge of the binding between Uniform Resource Identifiers (URI, in the username@domainname form) and the IP addresses, for the duration of service delivery.
- ASPs will typically provide AAA services at application level (i.e. on URI) independent of the access (i.e. PVC and IP address) or even of the access technology used (i.e. DSL, cable, satellite).
- ISP’s user subscription is typically a flat fee that covers both INAP and ISP. ASP subscription will typically be a flat fee combined with a per call/per service invocation based fee.

The GW subsystems shown in Figure 5 perform the following functions:

- FLSS: The Forward Link Sub-System hosts the transmission and MAC functions that are related to the operation of the forward link. It mainly includes DVB modulator and IP to MPEG gateway.
- RLSS: The Return Link Sub-System hosts all the functions required at the GW to operate the return link. It handles MAC functions (synchronization, power control, scheduling and connection control) and multi-carrier demodulation.
- TISS: The Terrestrial Interface Sub-System provides the interface between the FLSS/RLSS subsystems and the terrestrial networks. Depending on the options, it includes either a simple ER or a Broadband Access Server (BAS). The BAS offers routing functions, subscriber management functions and can also offer ATM and IP QoS functions.
- TFSS: The Timing and Frequency Sub-System is responsible for generating stable clock reference for the GW synchronization purpose.
- GMSS: The Gateway Management Sub-System includes all the management functions for the GW equipment and the STs.
- ACSS: The Access Control Sub-System handles STs authentication and admission. It allocates connection identifiers and configures the bandwidth management function of the FLSS (Forward QoS) and RLSS (Return QoS) according to SLAs stored in the GMSS. It also manages “dynamic SLAs”, through the interaction with Session Servers; it is thus able to dynamically re-configure the traffic parameters in FLSS and RLSS.

4.2 REGENERATIVE SATELLITE SYSTEM

In a regenerative satellite system, a ST is a DVB-RCS terminal upgraded to work with future satellites with regenerative payloads. The baseline architecture includes three main segments shown on Figure 6:

- The space segment, which is composed of one satellite with On-Board Processor (OBP) and its associated control infrastructure. The OBP in this paper can be of packet or circuit type.
- The terminal segment, which is composed of STs which can be connected to an external network (in this case, they are denoted Gateways – GW, as in the transparent DVB-RCS model) or to the customer premises equipment (“user” STs), either directly or through a LAN. The ST part of a GW is denoted GW-ST.

- The control and management segment, which includes a Network Management Center (NMC) for non-real time high-level configuration operations and a Network Control Center (NCC) for real-time controls. In the system considered, NMC and NCC functions are located behind the same ST, denoted NCC-ST. Those three elements compose the Management Station (MS).

Compared with the star system, the NCC includes a simplified RLSS (including MF-TDMA scheduler & ST controller, and forward signaling handler) and the ACSS (Connection Manager, ST Management system). It includes additional functions to manage the RCS signaling generated at the OBP.

The NMC is very much similar to the GMSS (Network and GW Management) but includes additional functions to manage the OBP.

The GW-ST includes a modified FLSS part (forward link is actually MF-TDMA according to DVB-RCS but generating several carriers in parallel using stacked modems) and a modified RLSS part (return link is actually TDM downlink received from the OBP).

5. QOS ASSURED ARCHITECTURE

In this section we develop typical application scenarios based on a hub-centric architecture allowing to offer “QoS-Assured” sessions.

The reference QoS management architecture is based on DiffServ IP QoS framework and dynamic bandwidth-on-demand management (see [7]). We suppose that ATM based link layer is used as return link transport mode. Permanent Virtual Channels (PVC) are offered between each ST and the GW.

This architecture makes use of a central ARC equipment co-located with the GW. The ARC is responsible for the resource management in the SAN and is able to dynamically request modifications of satellite access bandwidth for users who engage into starting or modifying an application session. It is a stand-alone product interfacing on one side with the SPs’ equipment (Web portal, AAA server, Softswitch) and on the other side with the GW subsytems. By coupling the service and transport planes, such equipment is the true enabler of NGN QoS, as shown through the typical application scenarios described below: dynamic SLA through the "Turbo button" feature and conversational services.

5.1 TURBO-BUTTON APPLICATION

This is an example of a time-limited Bandwidth-on-Demand (BoD) service. When a user establishes an Internet access session (e.g. via PPP [33]), the uplink and downlink bandwidth parameters are set to its Internet SP (ISP) defaults or to subscriber specific values indicated through RADIUS [28] to the ARC.

When the end user wants some extra bandwidth during a session (e.g. to download or upload big files, or to get better ping times when playing games), he/she can access the ISP portal and use the ISP turbo button service to boost bandwidth temporarily (probably on a time credit basis, e.g. the user is allowed 1 hour of Turbo per month).

The sequence of events for this application scenario takes place as shown on Figure 7:

1. The user invokes a bandwidth boost through the Web browser. This can be done explicitly through a click on a “Turbo button” on the ISP Web page, or implicitly through the use of some specific Web services.

2. A SOAP [35] interface is triggered from the ISP portal, including the policy to be applied. The ARC, knowing the allowed policies per ISP, can accept the request or not.

3. If the parameters are accepted by the ARC, it can re-configure the subscriber’s bandwidth parameters in the BAS and the ACSS.
4. ACSS re-configures the FLSS (if QoS functions are implemented in the FLSS) and RLSS.

5.2 QoS ENHANCED APPLICATION

This application scenario is typical of a session-based BoD. It shows the establishment of a conversational real-time session, such as VoIP. DHCP [9] is used as an auto-configuration protocol for user’s host and SIP server addresses.

This sequence of events for this application scenario takes place as shown on Figure 8:

1. Two default PVCs are established at ST log-on: a BE PVC and a QoS PVC. Their traffic characteristics are set to a default minimum. The BE PVC can be managed as described in the previous scenario.

2. When the IP phone is turned on, it looks for a DHCP server.

3. TISS (BAS) acts as a DHCP relay [27] towards the INAP’s DHCP server. The BAS provides the PVC to IP address binding via DHCP option 82. Note that in this setup no access level authentication and accounting are supported; the authentication and accounting will be handled by the ASPs at application/service level (e.g. authentication during SIP register, with Call Detail Records (CDRs) generated by the MMCS).

4. Through an API based interface, the DHCP server sends the PVC to IP address binding to the ARC.

5. The end user establishes a multimedia session, using, for example, SIP or HTTP/RTSP.

6. In case of SIP based conversational services, the MMCS uses a COPS interface towards the ARC in order to request for transport resources. In case of HTTP/RTSP based streaming services (e.g. VoD), a SOAP interface is used to issue the request; this request specifies the involved IP micro-flows and the required throughput per micro-flow.

7. The ARC finds out which QoS PVC corresponds to the specified micro-flows and performs admission control. If the request is accepted by the ARC, it can re-configure the subscriber’s bandwidth parameters in BAS and ACSS.

8. ACSS re-configures the FLSS (if QoS functions are implemented in the FLSS) and RLSS.

6. QoS ENABLED ARCHITECTURE

In this section we develop typical application scenarios based on a ST-centric architecture allowing to offer “QoS-Enabled” sessions.

6.1 QoS SERVER DESCRIPTION

As in the previous scenario, the IP DiffServ architecture is supported by the SAN (see [19] and [20]). The ST is supposed to be the entry point of the SAN DiffServ domain and therefore implements DiffServ border router functions such as IP flows classification based on multi-field criteria. In order that a data packet received on the ST router entry point takes the advantage of the services associated to a class, some form of signaling mechanism has to be used. Two such mechanisms appear practical:

- **In-band signaling**, carried within the IP packet header, using DSCP field. This is the standard DiffServ case where applications are able to drive packet marking themselves.

- **Out-of-band signaling**, usually before any effective data transfer. This is the standard IntServ case, or IntServ over DiffServ, where a dedicated protocol is used in order to specify the needed resources or forwarding class for a given IP stream. In that case, a stream identifier is needed to allow the router to identify and process the packets accordingly. This identifier can be the 4-uple source and destination addresses and ports in IPv4 (multi-field classifying) or the FlowId field in IPv6.
The second mechanism, close to a reservation procedure, has been chosen for the ST implementation in the SATIP6 project. The stream identification is done by an entity called « QoS Server » running on the ST. When an application has to use one of the available services, it has first to send a 5-uple {source address, source port, destination address, destination port, service identifier} to the QoS Server to identify the stream and the required QoS. The application is responsible of this reservation removal; however a watchdog is able to remove any too old reservations.

This solution can be easily linked to an AC mechanism which is essential in such an architecture. Flow granularity and state maintenance in the ST do not matter here. Indeed ST is a subscriber’s access router and therefore the number of considered streams is not very high.

Nevertheless, this kind of QoS reservation is seldom supported by today’s applications (e.g. they usually do not support RSVP). Next parts will describe two different QoS proxy mechanisms, which are able to perform this task on behalf of the application.

6.2 A QOS AGENT FOR NON QOS-AWARE APPLICATIONS

6.2.1 Providing QoS for every kind of application

Very few applications implemented today are aware of the QoS provided by the underlying networks. As the applications are rarely able to define their own requirements, a user-oriented solution allowing any application to take benefits from network services has been defined. The solution proposed in the following is called « QoS Agent ».

Implemented on the user terminal, the « QoS Agent » waits for application streams, maps them statically or dynamically to the chosen QoS level according to the user’s choice and, through interaction with the QoS Server, the ST is configured in order to take the request into account.

The satellite system architecture and the ST functionality of streams classification and forwarding related to a chosen policy are shown in Figure 9. The user is then able to remotely configure its ST through the QoS Agent and QoS Server, in order that the requested services may be available to any of the applications, during the time they are used.

6.2.2 Functional description

The QoS Agent has been designed in order to provide signaling of QoS requests even if the user’s applications do not support any kind of QoS interface.

Except the user terminal (CPE), no satellite network entity (ST included) is able to know precisely which applications are run on the user terminal. Actually, when a packet is sent in the network, every reference to the sending application is hidden by successive encapsulations. Apart from the applications using fixed and well known port numbers, it is not possible for an external entity to identify the sending application without systematically extracting the application header of all the packets in order to know which QoS to request.

The QoS Agent is a daemon running on the CPE, checking any new outgoing application connection. A graphical interface (see Figure 10) shows at any time network connections status and the applications that use these connections. This module allows the user to assign one of the available QoS services and to remotely configure the multi-field classifier in the ST, for only one or for all the streams the application uses. Alternatively, a network administrator might configure the application to QoS class mapping without the user’s intervention.

Each time a service is selected by the user for a given application, the QoS Agent sends to the QoS Server the concerned connections’ list with the reference of the associated service using a dedicated transactional protocol. If resources are not available, the QoS Agent is immediately informed. Using the received information, the QoS Server is able to tag and redirect these packets coming from the user terminal towards the appropriate requested IP service.
Integrated to QoS Agent, a « monitoring » module shows the user terminal outgoing traffic (see Figure 11). The user is then able to monitor the improvement provided by the implementation of the selected service.

Please note that this system only works for outgoing connections, i.e. with DVB-RCS return link QoS management. The user is charged for the traffic sent, and not for the traffic received. The forward link QoS management is not discussed here.

Nevertheless, if the traffic is bi-directional and if the considered scenario involves two users in the SAN, as the receiver wants the received stream to be upgraded, then it has to negotiate with the sender. The sender will then select a better and more expensive service. The mechanism described in section 6.3 better fits this kind of applications.

6.2.3 User Cases

Two concrete user cases will now be described more precisely to illustrate how the QoS Agent is used.

6.2.3.1 Expedited Forwarding service selection for a VoIP application

In the following we take the example of a company with multiple subsidiaries. This company wants to use VoIP between the different sites in order to save money. But the Expedited Forwarding (EF) DiffServ service [8] appears really expensive, as it is charged much more than Assured Forwarding (AF) [17] or Best Effort (BE).

The QoS Agents activated by default on all the company computers are configured by the system administrator in order to map all the VoIP communications (related to applications like GnomeMeeting or Microsoft Messenger) to AF. During communication set-up, the caller and callee QoS Agent configure their respective ST in order to map the right service to the exchanged flows.

If, during the communication, AF quality of service suddenly no more meets the requirements of an acceptable conversation, the users are able to request an enhanced service level through the QoS Agent. In order to do so, the graphical interface allows the user to select the EF service for the VoIP application.

At the end of the communication, as soon as the network connections are closed, the QoS Agent notifies the ST that there is no use in maintaining this QoS mapping any longer. Charging for EF traffic will instantaneously stop.

6.2.3.2 Implementation of a Web server using an Assured Forwarding service

In this second user case, we consider the same entity that wants to implement an Intranet in order to make easier information flow exchanges within the company. It has been decided to map all the flows going out from the Web server onto the AF service.

In order to do so, network administrators used the QoS Agent located on the Web server to configure the ST in order to map the server’s outgoing flows to AF service. A wildcard mechanism implemented in the QoS Agent allows to define automatically, for the httpd application (the web server), an association like: « all the flows sent from the web server port 80 use AF service ». The 5-uple sent to the ST’s QoS Server is therefore {web server IP address, 80, *, *, AF}.

6.3 QoS SIGNALING USING A SIP PROXY

The same QoS Server function in the ST could also directly be triggered by application servers, such as SIP proxies, when located at the user’s premises. It would play the same role as the MMCS in the “QoS assured” architecture developed in section 5.2, but the QoS selection being performed on the ST side. It merely allows QoS class selection which is suitable for a “QoS enabled” mode but it could be extended in order to support AC, possibly with an additional interaction between the ST and the Hub in order to reserve additional resources.
6.3.1 An enhanced SIP proxy

The entity in charge of QoS reservation is a standard SIP proxy with enhanced functionalities. The main functionality of a SIP proxy is to route the SIP messages from a SIP entity to another SIP entity and so on to the final destination. If we want the QoS reservation to remain transparent to the application, the enhanced SIP proxy must intercept session descriptions from SIP messages exchanged between caller and callee, translate the parameters of each media in the session in QoS characteristics and manages QoS reservations with the QoS Server on behalf of the application.

To be able to match our requirements, this proxy will be « stateful » in order to keep session state and all SIP requests and answers have to go through the proxy using Record-Route field in SIP INVITE message header. Moreover, the selected transport protocol is UDP, which is more suited for satellite link than TCP. Indeed the Round-Trip Time of the satellite link (around 500 ms) makes TCP handshake a lengthy process, especially for short-lived flows like signaling ones.

The functionality to be added to the standard SIP proxy is given below:

- a SDP analyzer so that the proxy is able to understand the session descriptions defined in SDP format included in SIP messages.
- A media table updated at SIP session establishment. Each media is defined by a 4-uple (IP source address, IP destination address, source port, destination port), the kind of media (audio, video…) and the Real-Time Protocol (RTP) profile. The media negotiated by caller and callee are gathered using the Call-ID, which is a unique identifier of the related SIP session.
- A SDP/DiffServ mapping from the kind of media (and the codec) to the DiffServ service (and the corresponding peak rate), essential to the QoS module.
- A QoS module taking into account the resource reservation corresponding to one media in a SIP session. It manages QoS reservations with the QoS Server.

6.3.2 QoS-Aware Proxy location

The QoS reservation related to the traffic on satellite uplink always takes place in the concerned ST.

A QoS-Aware SIP proxy is deployed in each user LAN interconnected by the SAN. This distributed architecture is well suited for an access or mesh topology based on a regenerative satellite because it answers the following two concerns:

- scalability concerning flow QoS management in user LAN;
- session establishment delays: the number of round trips of session and QoS signaling on the satellite link are minimized.

These two aims can only be achieved provided the central PCF delegates some of its policy decision responsibilities to the different STs. Thus, at the ST logon, its QoS Server will be supplemented, with local PDP functionalities which enable it to provision resources to local entities without requiring the PCF acknowledgment which lasts at least one satellite Round-Trip Time. The two main benefits are more interactive session establishments and a significant reduction of the signaling load on the satellite access network segment. However this resource pre-allocation can also lead to underutilization of satellite resources. In order to achieve a relevant trade-off between resources utilization and session establishment interactivity, when the local aggregated reservations exceed a defined threshold (for instance of bandwidth), the local PDP will handle all the additional local requests under the outsourcing model by routing them again individually towards the PCF. This distributed architecture will therefore improve both interactivity and scalability in both access and mesh topologies.

6.3.3 User Cases

Figure 12 shows a typical session establishment concerning two users starting a videoconference. Each QoS-Aware SIP proxy captures and translates the SDP description of each media (audio and video) and then sends resource reservations to its related QoS Server.
7. CONCLUSION

The integration of a satellite access network within a Next Generation Network or QoS enhanced network is a complex matter, which was tackled in both ESA QoSforRCS and IST SATIP6 projects. Those two projects helped derive two complementary approaches, starting from the same generic QoS management model, which is based on a layered QoS model, with processes at various network protocol layers, from the transport layer up to the application layer.

On the one hand, a Gateway centric (and therefore Service Provider centric) architecture, allowing “QoS-Assured” sessions, was described. This one allows the same level of guarantee than traditional IntServ approaches but without requiring applications to use RSVP or other QoS specific protocols. The processes and protocols have then been particularized for two example scenarios: the Turbo-button scenario (for dynamic SLA management, triggered by the user itself) and the enhanced QoS scenario (for conversational services). This architecture can be thought more appropriate for public Service Provider’s networks.

On the other hand, a Satellite Terminal centric (and therefore user centric) architecture, allowing “QoS-Enabled” sessions, was described. This one allows to use the traditional DiffServ approach without requiring applications to be able to perform QoS marking on IP packets they generate and without complex application proxies in the Satellite Terminal itself. A QoS Agent, residing on the user’s PC or server, allows to dynamically configure the Satellite Terminal classification function based on rules made up by the user or its network administrator. An extension of this approach with an application specific proxy on the user’s network has also been shown. This architecture can be thought more appropriate for VPN.

Although not all applications will need complex resource management functions or per-flow admission control, these new developments may pose severe scalability concerns. Future networks’ tendencies need to be watched closely and important trade-off’s still have to be made: this is the concern of current technical activities started within the ESA sponsored ARTES program (DOMINO-2) and standardization groups (ETSI Broadband Satellite Multimedia group [13] and SatLabs [31]).

ACKNOWLEDGEMENTS

The authors would like to thank all the SATIP6 and ESA QoSforRCS project members for their contribution to the above topic throughout the past two years.

ACRONYMS

3GPP 3rd Generation Partnership Project
AAA Authentication, Authorization, Accounting
AC Admission Control
ACSS Access Control SubSystem
ADSL Asymmetric Digital Subscriber Line
AF Assured Forwarding
AG Access Gate
API Application Programming Interface
ARC Access Resource Control
ASP Application SP
ATM Asynchronous Transfer Mode
BAS Broadband Access Server
BE Best Effort
BoD Bandwidth on Demand
CDR Call Detail Record
COPS Common Open Policy Service
CPE Customer Premises Equipment
CSCF  Call Signaling Control Function  
DHCP  Dynamic Host Configuration Protocol  
DSCP  DiffServ Code Point  
DVB  Digital Video Broadcast  
EF  Expedited Forwarding  
ER  Edge Router  
ESA  European Space Agency  
ETSI  European Telecommunications Standards Institute  
FEC  Forward Error Correction  
FLSS  Forward Link SubSystem  
FTP  File Transfer Protocol  
GMSS  GW Management SubSystem  
GSM  Global System for Mobile  
GW  Gateway  
HTTP  HyperText Transfer Protocol  
IETF  Internet Engineering Task Force  
INAP  Interactive Network Access Provider  
IP  Internet Protocol  
ISDN  Integrated Services Digital Network  
ISP  Internet ST  
IST  Information Science and Technology  
IT  Information Technology  
ITU  International Telecommunications Union  
LAN  Local Area Network  
MF-TDMA  Multi-Frequency Time Division Multiple Access  
MMCS  MultiMediA Call Server  
MPEG  Motion Picture Expert Group  
MPLS  Multi-Protocol Label Switching  
MS  Management Station  
NAT  Network Address Translation  
NCC  Network Control Centre  
NGN  Next Generation Network  
OBP  On-Board Processor  
OSA  Open Service Access  
PCF  Policy Control Function  
PDP  Policy Decision Point  
PEP  Policy Enforcement Point  
PHB  Per-Hop Behavior  
PSTN  Public Switched Telephone Network  
PVC  Permanent Virtual Channel  
QoS  Quality of Service  
RCS  Return Channel for Satellite  
RLSS  Return Link SubSystem  
RSVP  Resource ReSerVation Protocol  
RTP  Real-Time Protocol  
RTSP  Real-Time Streaming Protocol  
SAN  Satellite Access Network  
SDP  Session Description Protocol  
SIP  Session Initiation Protocol  
SLA  Service Level Agreement  
SOAP  Simple Object Access Protocol
REFERENCES
1. 3GPP: TS 23.228 v5.5.0 (2002-06). 3rd Generation Partnership Project – Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2 (Release 5), 3GPP TS 23.228 v5.5.0, June 2002.
2. Y. Bernet et al., A framework for integrated services operation over Diffserv networks, IETF RFC 2998, November 2000.
22. ITU-T H.225: Call signalling protocols and media stream packetization for packet-based multimedia communication systems.
Figure 1 - Application & Transport planes
Figure 2 - Layered General QoS model
Figure 3 - QoS in the Connection/Transport Layers
Figure 4 - Global QoS Architecture
Figure 5 - Transparent satellite system topology
Figure 6 - Regenerative satellite system topology
Figure 7 - Turbo-button application
Figure 8 - QoS enhanced application
Figure 9 - QoS management by ST
Figure 10 - QoS Agent GUI
Figure 11 - Applications throughput display
Figure 12 - SIP session establishment
Packet Flow
Application Plane
Service Domain 1
Transport Domain 1
Transport Domain 2
Transport Domain 3
Service Domain 1
QoS Signalling
Call Signalling
Packet Flow
QoS Signalling
Packet Flow
QoS Signalling
Packet Flow
QoS Signalling
Packet Flow
1: BE and QoS PVCs established at ST log-on

2: DHCP session establishment

3: DHCP server

4: API

5: SIP

6: COPS

7: SNMP/COPS

8: re-configure

MMCS

ARC

DHCP

server

Gateway

TSS

MMCS

ASP Network

IP Backbone

ER
<table>
<thead>
<tr>
<th>Pres No</th>
<th>Application</th>
<th>IP Version</th>
<th>Protocol</th>
<th>QoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2952</td>
<td>eggcups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3005</td>
<td>gameRoom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3210</td>
<td>soh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3219</td>
<td>minos-dock:32008&lt;-&gt;mucha,bass:22</td>
<td></td>
<td>TCP</td>
<td>1</td>
</tr>
<tr>
<td>3322</td>
<td>testQoSserver</td>
<td></td>
<td>TCP</td>
<td>4</td>
</tr>
<tr>
<td>3306</td>
<td>mozilla-bin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3308</td>
<td>minos-dock:32023&lt;-&gt;spiderman:69</td>
<td></td>
<td>TCP</td>
<td>0</td>
</tr>
<tr>
<td>3308</td>
<td>minos-dock:32030&lt;-&gt;spiderman:69</td>
<td></td>
<td>TCP</td>
<td>0</td>
</tr>
</tbody>
</table>

**Presselected Service**

<table>
<thead>
<tr>
<th>AF_qos</th>
</tr>
</thead>
</table>

**Monitoring**

<table>
<thead>
<tr>
<th>Monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Rate:</td>
</tr>
</tbody>
</table>

**Clear** | **Validate**