4G Mobile Systems: Multimedia Content Transmission

Antonio Cuevas Casado
Jose Ignacio Moreno
Departamento Ingeniería Telemática, Universidad Carlos III de Madrid, Madrid, Spain

Abstract
Multimedia and streaming applications are getting more and more important in 3G networks and will be an important part of the future 4G networks. In this article, after introducing 3G and 4G concepts and architectures, multimedia support in these network technologies is described with strong focus on Quality of Service.

INTRODUCTION

Role for Multimedia in Mobile Systems

As Internet is taking an ever greater role in all kinds of communications (mobile, fixed, data, voice—VoIP) operators fear losing their central role in the telecommunications business value chain and becoming mere “bit pipes.” Operators want to find new business opportunities. Creating and delivering new services besides the classical SMS or voice call services is of paramount importance for mobile operators. Multimedia applications are one of these new services. As an example we can mention the success of mobile IP TV in Korea.[1]

Brief Outline of 3G Technology

Telecommunications have been affected by two major trends: the success of the Internet and the success of mobile telephony. GSM has been universally accepted as the mobile telephony solution. We are now facing an evolution from GSM, the second generation (2G) (mobile) networks, to third generation 3G. An intermediary step, already commercially available, is 2.5G in which a packet network (GPRS) has been deployed in parallel to a traditional telephone circuit network (GSM). 3G aims to “merge” Internet and mobile networks and provide larger bandwidth. But, contrary to GSM, there is not a universal 3G solution. Two groups, 3G partnership project (3GPP)[2] and 3GPP2[3], are defining the architectures of 3G networks. 3GPP2 is backed by partners from the USA and from some Asia-Pacific countries. Its solution is often termed CDMA2000. 3GPP has more acceptance in Europe and also in Asia-Pacific countries; its solution is named UMTS. UMTS and CDMA2000 aim to be similar, but mainly differ on the radio access link. We will concentrate in this entry on UMTS as it is the most accepted 3G solution.

The general UMTS architecture, as currently specified, is presented in Fig. 1. The architecture is split into three domains: user equipment domain, access network domain and the core network domain. The access network domain and the core network domain build the infrastructure domain. The architecture also includes interoperability with other kind of networks.

• User equipment (UE) is the equipment used by the user to access UMTS services. User equipment has a radio interface to the infrastructure domain.

• The access network domain (AN) consists of the physical entities that manage the resources of the access network and provides the user with a mechanism to access the core network domain.

• The core network domain (CN) consists of the physical entities that provide support for the network features and telecommunication services. These entities include functionalities such as the management of user location information, control of network features and services, the transfer (switching and transmission) mechanisms for signaling and for user generated information. The gateway GPRS support node (GGSN) is the first point of interconnection of other networks—especially the Internet- with UMTS.

Need for Fourth Generation Systems

IP-based communication is gaining more and more importance. As we just saw, it is a key part of UMTS architecture. UMTS provides an IP-packet service using tunneling mechanisms, but still employs all the mechanisms of 2G networks: this way, all the UMTS network is counted as “1 hop” at the IP layer. Adding new access network technologies involves a great amount of translation procedures from the former specific mechanisms to UMTS. On the other hand, on 4G networks scenarios, IP is used to bind all
different technologies, deploying technology-independent protocols for quality of service (QoS), mobility and almost all other aspects. The 4G architecture should be able, thus, to embrace almost any wireless (or even wired) access technology available. Instead of following the 3G philosophy of bringing the concept of packet switching into the existing connection-oriented network environments, several voices argue that aiming directly to 4G networks may cause them to become a reality much sooner than forecasted. As such, the research efforts being done in this field are huge.

**Introduction to QoS**

Several aspects influence the QoS that the user will enjoy when employing multimedia applications over telecommunications networks. These include, among others, codecs chosen and the performance of the network. This section presents briefly technical issues at network level; the performance to be achieved is imposed by human perception constraints, which are described in the next section.

As it is perfectly well known, IP networks were not designed to provide any QoS warranties. But the applications traditionally using IP as their communication technology (web, e-mail, etc.) could perfectly cope with its lack of warranties. However, telephony, interactive and many multimedia applications require transport networks with very stringent QoS requirements. The ways to endow IP networks with QoS mechanisms have been extensively studied. The most accepted solutions are IETF’s [5] integrated services (IntServ) [3] and differentiated services (DiffServ) [6]. Both endow the routers with QoS mechanisms such as classifiers, priority queuing and shaping (Fig. 2).

The differences lie in the level of detail in the classifiers and in the state maintained. The IntServ model provides end-to-end QoS guarantees by reserving per-flow resources (normally using the resource reservation protocol (RSVP) [3]) in all the nodes along the path. While this architecture provides excellent QoS guarantees, it has scalability problems in the network core because of per-flow state maintenance and per-flow operation in routers. DiffServ, on the other hand, requires no per-flow control on the core and, consequently, routers do not maintain any per-flow state and operation. As a result, DiffServ is relatively scalable in the forwarding/data plane but offers no strict QoS guarantees. The criterion to classify the packets in the core routers is the DiffServ code point (DSCP) field in the packet header. Several approaches also exist combining IntServ and DiffServ: e.g., IntServ in the “access part” of the network and DiffServ in the core. Of course, solutions based on other paradigms also exist and are even complementary to these ones.

![Fig. 1 UMTS architecture.](image1)

![Fig. 2 IP QoS: mechanisms for packets exiting a router interface.](image2)
Importance of QoS in Multimedia Transmission

In this section we shall briefly review the parameters that characterize subjectively user-perceived QoS. Generally speaking, multimedia applications are among the most demanding applications in terms of QoS. These parameters dictate the performance any communication system willing to support multimedia applications must achieve.

Many organizations, from international standardization bodies to equipment vendors or telecommunication operators, have studied user-perceived QoS. As an example, we refered the results provided by the ITU in its G.1010 recommendation (Table 1).

Recent works are in the direction of psychoacoustics; they try, for instance, to measure the effort a person must make to understand a conversation as a function of the QoS of the transport technology. The goal is to infer the user-perceived QoS without directly asking him about his feeling.

MULTIMEDIA IN 3G AND 4G

Evolution of Mobile Technology

Telecommunications have been affected by two major trends: success of the Internet and success of mobile telephony. GSM has been universally accepted as the mobile telephony solution. We are now witnessing an evolution from GSM to 3G. 3G aims to “merge” Internet and mobile networks and provide larger bandwidth. Two groups (3GPP and 3GPP2) are defining the architectures of 3G networks. However, both solutions are very similar and are striving to integrate mobile telephony and Internet.

The first release of 3GPP’s architecture already included IP networks integration. In the successive releases, IP networks play an increasing role (Fig. 3).

Release 99 was the initial standard, the evolution with respect to 2G (GSM) and 2.5G networks was addition of a packet-switched domain. It featured a new access technology with larger BW: CDMA. The next release, Release 4, saw almost no changes in the packet-switched domain. However, in the circuit switched domain, there was a separation of transport and control. In Release 4 the circuit switched domain may also be IP-based. In Release 5 the access technology was enhanced: high-speed downlink packet access (HDSPA). Besides, IP multimedia subsystem (IMS) was introduced as a means of controlling and integrating IP multimedia services. IMS follows a new business model: “the semi-walled garden.” In this model, the network provider gains a new role (besides the bit pipe or, eventually, service provider): it becomes a service broker. In Release 6, the interworking with Wi-Fi was specified and the IP “service platform” enhanced. Those two aspects are being further developed in Release 7.

Multimedia Applications in 3G and 4G

The range of applications under the “Multimedia” term is very wide. We can have video and audio calls; we can consider streaming events such as soccer matches or video on demand; we can also think about on-line video games. The QoS requirements they demand from the networks (see section “Importance of QoS in Multimedia Transmission”) are very diverse. The applications characteristics are also very varied. For instance, a video call between two users does not benefit from broadcasting but the live streaming of a football match will profit from broadcast technologies—as radio or shared wire (like Ethernet)—and broadcast/multicast transmissions (as multicast groups or broadcast channels in UMTS systems). Also, the transmission “directionality” of the applications changes: some streaming applications like broadcasting sport events are usually one-way, without the need of a return channel, while interactive applications (like on-line gaming or audio calls) need two-way communications. Some applications may need a service composition (enabled by the network operators’ service platform), e.g., a “Tele-meeting” application with audio, video, slide show, file sharing and editing, simultaneous audio-conversations translation, etc. Besides, we have also to face right-management issues: i.e., controlling what users can do with the received content. Disallowing the storage of received content is the easiest form of content protection. Digital rights management (DRM) is a hot and widely researched topic including techniques such as encryption and conditional access based on usage rules to protect and manage access to multimedia data. Content providers are reluctant to deliver premium content without DRM mechanisms in place to prevent illegal copying of valuable multimedia content such as music and movies. Since this delivery is often done using streaming techniques, DRM is of special relevance for streaming. Finally, we stress on a fundamental difference among streaming applications, this is whether they broadcast “live” content or recorded content (i.e., video on demand). In the first case, the initial delay accepted by the user is under a dozen of seconds, while in the later case, it can be up to a minute or even more. This implies how big the buffer in the receiver can be and thus, how the service is resistant to changing network conditions. Besides, live transmissions of events such as sports matches are broadcasted to large audiences and each video on demand transmission targets smaller publics.

Streaming Multimedia Content over 3G Technology

3GPP concluded the first version of the UMTS streaming service in March 2001. The work gave birth to the 3GPPS (packet-switched streaming) standard. The UMTS streaming service integrates simultaneously playing video,
Table 1  Audio video applications QoS requirements and characterization done by ITU.

<table>
<thead>
<tr>
<th>Medium</th>
<th>Application</th>
<th>Degree of symmetry</th>
<th>Customer demand</th>
<th>Data rate</th>
<th>Key performance parameters and target values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>Conversational voice</td>
<td>Two-way</td>
<td>High</td>
<td>4–13 Kbps</td>
<td>One-way delay &lt;150 ms preferred, &lt;400 ms limit</td>
</tr>
<tr>
<td>Audio</td>
<td>Voice messaging</td>
<td>Primarily one-way</td>
<td>High</td>
<td>4–13 Kbps</td>
<td>&lt;1 sec for playback, &lt;2 sec for record</td>
</tr>
<tr>
<td>Audio</td>
<td>High-quality streaming audio</td>
<td>Primarily one-way</td>
<td>Low</td>
<td>32–128 Kbps</td>
<td>&lt;10 sec</td>
</tr>
<tr>
<td>Video</td>
<td>Videophone</td>
<td>Two-way</td>
<td>Low</td>
<td>32–384 Kbps</td>
<td>&lt;150 ms preferred, &lt;400 ms limit</td>
</tr>
<tr>
<td>Video</td>
<td>One-way</td>
<td>One-way</td>
<td>Low</td>
<td>32–384 Kbps</td>
<td>&lt;10 sec</td>
</tr>
<tr>
<td>Data</td>
<td>Telemetry-two-way control</td>
<td>Two-way</td>
<td>Low</td>
<td>&lt;28.8 Kbps</td>
<td>&lt;250 ms</td>
</tr>
</tbody>
</table>
audio, images, and formatted text into mobile multimedia applications. The 3GPP standard specifies both protocols and codecs.

Let us first analyze the protocols involved: real-time streaming protocol (RTSP) and session description protocol (SDP) are used for session setup and control, while the real-time transfer protocol (RTP) is employed for transporting real-time media such as video, speech, and audio. To transport static media such as images and text, HTTP is used. Those are standardized protocols widely accepted in the Internet world. Another protocol, not so widely deployed, is also used: synchronized multimedia integration language (SMIL). SMIL is an XML-based presentation language. Roughly speaking, SMIL is HTML with additional notions of time and temporal behavior. Thus, it can describe a media screen and control the placement of media elements in space and time. SMIL is the “glue” that combines the different elements (audio, video, images, text, etc.) to create an interactive multimedia presentation.

3G solutions code the streams flows with either standard formats or with 3G-tailored ones. RealMedia or 3GP are among the first. They are more efficient than many of the formats used nowadays in PCs and broadband connections (such as asymmetric-digital subscriber line (A-DSL)) but still are very common and supported by many commercial terminals. 3G tailored formats are adapted more to the radio access medium since they can change their quality varying the number of layers sent: the better the quality, the more the bandwidth needed. Those adaptations can cope with the variations suffered by the radio link. Some of them are MPEG-4, RealMedia SureStream, or media resource centre (MRC).

In the 3GPP design, the UMTS network transmits the data and passes it to the application from L2, the lower 3GPP link layer. The application demultiplexes the data and distributes it to the corresponding video and audio decoders. 3GPP did not define further the architecture and several possible instantiations can be made. In the rest of this section we present one possible solution, focusing on the service logic support dealing with aspects such as accounting or file formats. Future sections explore the (radio) transmissions mechanisms stressing on QoS aspects.

Many are the possibilities to design the architecture backing the streaming service logic. Service uses IP as transport protocol. As such, within UMTS networks, it lies within the packet-switched domain, “beyond” the GGSN. The elements generally present in such architectures are displayed in Table 2. Solutions with a display of nodes very similar to the one presented in the table are commercially deployed and running in Spain.

Fig. 4 depicts a possible interaction between the nodes presented in Table 2. Solutions commercially deployed in Spain follow a very similar interaction.

The first part of the service delivery takes place from messages 1 to 10. The user accesses the wireless application protocol (WAP) portal, requests and pays for his chosen content. The information about his payment and the selected content are stored in the database. The user is informed about the successful operation in message 10.

The second part takes place from messages 11 to 18. The user interacts with the AAA (authentication, authorization, and accounting) server which authenticates the user (taking information from the lightweight directory access protocol (LDAP) server) and authorizes the user checking whether he has paid (querying the database). Finally the AAA server “grabs” the content from the streaming server and forwards the streaming flows to the user’s terminal.

| Table 2 3G streaming architecture: service-logic-related modules. |
|------------------|------------------------------------------------------------------|
| WAP server       | WAP front end, providing an interface to the user to access the streaming service |
| Payment server   | Controls that the user has agreed to pay for the service (and that he has paid for it) |
| Mediator         | Mediator between the Billing Server and the streaming service data base. It registers in this data base that the user has agreed to pay (and has paid) for a certain content |
| Database         | Stores the information about the available streaming contents and also about the users who have paid and what content they have paid for |
| Streaming server | Streams the contents to the users |
| AAA server       | Authenticates and authorizes the users. Authentication is done with the help of the LDAP server. Authorization is done querying the data base to know whether the user has paid or not. It also generates call data records for later billing and/or auditing |
| LDAP server      | Stores the MS-ISDN to IP address association. The MS-ISDN “identifies” a user; it is static. IP addresses are dynamically assigned. This data can be gathered from the HLR/HSS |
Current trend is to integrate the streaming service architectures with the IMS. Changes will affect mainly the AAA server which may have direct interaction with the home subscriber server (HSS) instead of interfacing with LDAP or payment servers.

Challenges Faced by 3G Networks in Streaming Multimedia Content

3G systems were still designed under the “telephony business model” where the network operator controls the whole service provision. Note that, nevertheless, Internet connections are a fundamental part of 3G. If the streaming service is controlled by the network operator many issues, such as QoS, are readily solved. The challenges to face are the lack of flexibility of this approach and, eventually, the scarce radio access link capacity.

3G systems employ only one access technology, CDMA, whose BW may be enough for streaming into mobile phones. But if the device used has larger capacities (e.g., display quality) the required BW may surpass the available one. This makes the range of 3G target devices narrower. Besides, in CDMA technology due to the “cell breathing” phenomenon, the available BW varies as a function of the number of users. If the application requires low BW (like voice call or low quality streaming) this has no negative consequences but for high-quality streaming and application requiring broad BW, this poses multiple constraints. Determining application needs and negotiating them is of paramount importance as we will see in section “QoS Issues and Challenges over 3G.”

QoS Issues and Challenges over 3G

Though UMTS networks integrate IP-based services and achieve interoperability with the Internet, they still follow the classical telecommunications business model: the network operator controls all the aspects of the service provisioning. This allows facing the problem of QoS in an integrated way. Indeed, UMTS achieves end-to-end QoS with assured warranties. In order to achieve a trade-off between complexity and resolution, four traffic classes are defined in UMTS: conversational (low-delay interactivity between users), streaming, interactive (high-delay interactivity between users), and background.

Section “Importance of QoS in Multimedia Transmission,” pointed the importance of providing QoS to telecommunication services so that users can enjoy them. We shall see next, how UMTS provides QoS to these four kinds of applications.

It is the UMTS bearer service that provides the requested QoS through the use of different QoS classes. The UMTS bearer service consists of two parts, the radio access bearer (RAB) service and the core network bearer service. The RAB service is implemented by a radio bearer service and an Iu-bearer service (Fig. 5).

It is clear that, given a traffic class between final users, mapping of QoS attributes along bearer services will be necessary.

However, mapping from upper to lower layers will be in most situations an operator’s choice, owing to the fact that it should be able to determine whether to accept or
deny—according to the resources available—a request. Let us outline the 3 possible mappings:

- Mapping from application attributes to UMTS bearer service attributes: this is an implementation issue.
- Mapping from UMTS bearer service attributes to RAB service attributes: the value of the attributes will normally not be the same as the corresponding value for the RAB. In the lower layer the requirements will be more restrictive because of the interaction with other bearer services. It is also needed to define both the service data unit (SDU) format information and the source statistics descriptor.
- Mapping from UMTS bearer service attributes to CN bearer service attributes: this is operator’s choice.

**QoS Negotiations in 3G**

QoS setup in 3G’s UMTS network relies on the packet data protocol (PDP) context activation procedure. Here, we will take a look at some general management strategies and focus on PDP context request and activation procedures and which architectural entities are involved.

Fig. 6 shows a mobile station (MS) initiated PDP context activation procedure. The procedure is as follows: the MS sends an Activate PDP Context Request (1) to the SGSN, which initiates the RAB Setup (3)—this procedure will be seen later. Once the bearer has been established, the SGSN sends a Create PDP Context Request to the GGSN (5), who creates the new entry in its PDP context table, which allows the GGSN to route PDUs between the SGSN and the external PDP network. If QoS received from the SGSN is incompatible with the PDP context being activated, then the GGSN rejects the Create PDP Context Request message. The compatible QoS profiles are configured by the GGSN operator. Last, (7) the SGSN returns an Activate PDP Context Accept with the negotiated QoS and other parameters not involved in QoS.

For each PDP address a different QoS profile may be requested. For example, some PDP addresses may be associated with e-mail, tolerating lengthy response times. If a QoS requirement is beyond the capabilities we can renegotiate the QoS profile as close as possible to the requested QoS profile. The MS either accepts the negotiated QoS profile or deactivates the PDP context.

The purpose of Step 3 of the previous procedure is to enable establishment of new RABs (radio access bearers) for a given MS and/or modification and/or release of
already established RABs. The same messages are used for the three above mentioned actions and it is only the content carried by the messages that is different. Next we describe the RAB Setup (Fig. 7).

The procedure is initiated by SGSN, sending a Request message (1) to the RNC to establish, modify, or release one or several RABs. After the corresponding action (2), the RNC returns a Response to the SGSN. If the SGSN receives a Response with a cause indicating that the requested QoS profile(s) cannot be provided (e.g., Requested Maximum Bit Rate not Available), then the SGSN may send a new Request. The number of re-attempts, if any, as well as how the new QoS profile(s) values are determined, is implementation dependent.

The network can also initiate a PDP context activation, namely when the terminal is going to receive “unsolicited data” e.g., in the case of the callee in a conversation service.

**Shortcomings of 3G in Streaming Multimedia**

As we have already pointed out, 3G, although open to the Internet, follows still the telephony business model. This solves many issues, like QoS, but 3G networks are thus far less “flexible” than Internet. Supporting several types of devices and, mainly, several types of technologies is much more difficult than in Internet.

Besides, many streaming applications and many situations (e.g., live sport matches broadcast) profit from broadcast technology and transmission. In 3G, the radio access technology is broadcast but not the transmission: the transmission is divided in several unicast channels. There are some broadcast channels that all the phones are prepared to use but their BW is low since they are, normally, used for signaling. Changes can be made to create more broadcast channels or to use the unicast channels as broadcast. Besides, the core 3G network should, ideally, also support broadcast and multicast. Those needed changes are big and they are not likely to be done, waiting for the arrival of 4G networks.

A Glimpse of 4G Solution for Streaming Multimedia

3G still uses a circuit switched architecture and over it an IP overlay is created. 4G uses IP natively and thus adding any new access technology is far easier than in 3G. 4G networks will likely employ IETF open defined standards for all the aspects, from mobility to streaming. As such, streaming solutions in 4G will be very similar to the currently existing ones in the Internet. However, aspects unique to 4G are to be considered such as seamless mobility and handovers and QoS that has to be supported over heterogeneous access and multiple devices. Besides, networks operators will create service platforms used to bundle services including streaming under the “semi walled garden” business model.

**4G SYSTEMS**

**Strategy for Transmitting Multimedia over 4G Technology**

As we said in the Introduction, 4G systems will, essentially, be an Internet-like network with extra features such as mobility. We pointed that in such networks the telecommunication operator must strive to be more than a bit pipe, it must offer new services or be a service broker, following the “semi-walled garden” business model. Multimedia is to become one of these “killer services.” Transmitting multimedia over 4G should follow similar strategies as when dealing with Internet but taking into account 4G particularities, such as mobility, broadcast transmissions, heterogeneous access networks, several types of devices employed for several types of applications and the need for service bundling.[13]

To address these particularities, the key point for the 4G network operator is to offer enabler services such as QoS (see section “QoS Negotiations in 4G”), AAA, and mobility, for all kind of applications (multimedia or not). Other issues are more specific to multimedia and to some types of multimedia applications. As an example, let us discuss about “Robust Header Compression.” For applications requiring very low delay and transmitting relatively low volume of data (e.g., audio calls) the payload may be very small. But the overhead in IPv6 networks (IP version 6 is likely to be used in 4G) is very high if we add features such as mobility. As an example, for GSM codecs of 13.2 Kbps and with a codification delay of 80 ms, the payload in the IP frames is 33 bytes while the overhead is 84 bytes, including RTP and UDP headers and IPv6 basic header, IPv6 routing header and IPv6 home address destination option. To avoid this very poor payload/overhead ratio, techniques such as Robust Header Compression are used to minimize the overhead. Many multimedia applications, such as live sport
events, can greatly benefit from the broadcast technologies and transmission being designed for 4G networks. 4G will employ many broadcast technologies, such as radio links or “shared wired” connections, such as Ethernet. Over these technologies it is very easy to build broadcast transmission (e.g., the multicast groups in Ethernet). Besides, multicasting techniques are ready to be employed in the IP core network: the IP Multicast Backbone.[14] Again we stress that different multimedia applications (streaming, video-phone calls etc.) impose different constraints to the networks. In this entry we will concentrate on the streaming multimedia applications and the mechanisms of 4G networks to handle them.

Besides dealing with transmission-technology related issues, let us briefly focus on the service platform. Its design is still unclear; however, it seems accepted that the semi-walled business model will be followed, the network operator doing the service brokerage and service bundling, charging the users, and diverting the money (retaining a percentage) to service provider partners. The 3GPP designed IMS is a good starting point for designing this service platform.

Types of Multimedia Supported by 4G

Since 4G are Internet-like open networks all kinds of (multimedia) applications will be supported in 4G. Besides, in 4G new radio access technologies with more BW will appear and, moreover, since 4G is based on IP, any access technology, wired or wireless, will be used. Thus, the BW constraints should be looser than in 3G or, even, do not exist, if the network is properly managed. So the range of multimedia applications should be very wide.

Among all the possible multimedia applications, some of them, like video and audio calls may be given “special treatment” since they are the core business of telecom operators. This may imply that the operators, besides supplying their infrastructure to transport bytes, will offer themselves (or via provider partners) the means to assist peers in establishing and controlling this video or audio calls (i.e., like the IMS) so that the service quality is enhanced. Besides, if the operators build rich service platforms, the service bundling and composition may give birth to very rich multimedia applications and attract users.

The constraints that 4G networks (or any other network) will have to respect to allow good QoS for multimedia applications are determined by user perception, as we saw in section “Importance of QoS in Multimedia Transmission.” One of the few and first studies done in a prototype 4G network can be found in Serrano et al.[15]

Streaming Techniques and Processes over 4G

Streaming in 4G is similar to streaming in the Internet. 4G particularities have to be taken into account, both for the service platform and for 4G networking features.

Concerning the service platform and service logic, these can be quite similar to the one presented for 3G in section “Streaming Multimedia Content over 3G Technology.” There should be a AAA server to control users. While in 3G this AAA is an add-on to the HSS, in 4G, all the user control will be the responsibility of this AAA server. According to most 4G works, this AAA server will follow the IRTF AAA architecture and, as such, be a diameter server.[16] The service platform should also take into account QoS as we will see in section “QoS Negotiations in 4G.”

The 4G specific aspects related to network topics are mainly two: first, ability to use any access technology and, second, that some of these technologies have broadcasting characteristics.

With regard to the heterogeneity in access networks, in 3G, the radio link performance also varies, but these performance changes are nil compared to the ones we can have in 4G: In 4G the same one terminal can attach subsequently (or even simultaneously) to several access technologies with different performances. This is indeed one of the biggest differences to take into account when designing the streaming solution.

Some streaming techniques can adapt to the varying access network conditions, for instance, as it is done in 3G, use “layered codecs” (see section “Streaming Multimedia Content over 3G Technology”). Besides, the needed resources for a stream may vary due to the burstiness nature of video and audio codecs. The varying BW required suggests to provide streams with a transport service whose QoS is not based on static reservations but, rather, on traffic prioritization and aggregates, such as DiffServ. But we have more options: for instance, when the terminal employs a large BW access, it requests data at higher rates, enlarges terminal’s playback buffer and stores the “excess data” in the buffer; if the terminal moves again to a low-BW access technology it can profit from the buffered data. Borras-Chia[17] studies the performance gain when correctly buffering the media. Several are the protocols to cope with the above presented situations. They target the downlink (base station to terminal) direction over low-BW links, such as wireless access. They are based on pre-fetching; W-JSQ[18] is one of them. W-JSQ pre-fetches parts of the ongoing audio/video stream into buffers in the wireless terminals. Channel probing is used to judiciously utilize the transmission capacities of the wireless links, which typically experience location-dependent, time-varying, and bursty errors. Roughly speaking, the access router schedules—perhaps helped by the QoS Broker, an entity we shall study next—the packet for the terminal that has the smallest pre-fetched reserve and is currently experiencing good transmission conditions on its wireless link. The pre-fetched reserves help the terminal to continue playout during periods of adverse transmission conditions on the wireless links (and also when bursty high action scenes are played out).
Another important issue when streaming non-live content in 4G is the availability of broadcast technologies. Taking advantage of this and of the "pyramid broadcasting technique," we allow the server BW to grow only logarithmically with the ratio of the content length to the start-up latency. On pyramid broadcasting systems, the server partitions the media file into \( n \) segments of length \( x_2^1, x_2^3, x_2^3, \ldots, x_2^{n-1} \). The server repeatedly broadcasts (in a multicast group for each segment) as in a carousel and, even if the segments have different lengths, the transmission rate is the same for all the segments (Fig. 8).

The client downloads the segments in order and starts playing the content when it receives the beginning of the first segment \((n = 0)\). The worst-case start-up latency is therefore the time the client waits for the server to loop through the entire first segment to reach the beginning.

**QoS Negotiations in 4G**

4G networks employ the IPv6 protocol as a "native" convergence layer. The QoS mechanisms offered by networks are expected to be very similar to the current ones designed by the IETF for IP networks: IntServ and/or DiffServ, previously explained in the section "Introduction to QoS." The big challenge is to integrate them with other aspects present in 4G networks, such as mobility including intertechnology mobility between access networks with very heterogeneous QoS characteristics. Pricing issues and coping with different users' profiles and policies between the participants in the multimedia sessions are also of paramount importance when dealing with QoS in 4G networks. Besides, QoS involves many aspects, other than QoS, at the network level, and 4G networks should aim to integrate them all, for instance, content adaptation or L2 QoS. Still we suppose we want to follow the Internet paradigm of complete separation between application and transport which turns into invalid monolithic solutions such as the ones present in 3G networks. We can figure out how big the endeavor of providing QoS in 4G networks is. Solutions such as IMS can be a starting point to face this issue. The QoS negotiation is one of the key pieces in this challenge which we will present and look into in this section.

Concerning IP-level QoS and in the IntServ framework, the RSVP is used to reserve resources in the routers along the path between the sender and the receiver(s). RSVP also allows freeing these resources when they are no longer needed. Normally these reservations are to be policed and it is common to have an entity termed BW broker (or, also, QoS Broker) that takes the policy decision and communicates it to the routers. This entity will be later studied in this same section. The primary messages used by RSVP are the "Path" message, which originates from the traffic sender, and the "Reservation" message, which originates from the traffic receiver(s). RSVP can also explicitly shut down the QoS sessions using RSVP teardown messages. Teardown messages can be initiated by an application in an end system (sender or receiver) or a router as the result of state timeout. A RSVP enabled router may consult the QoS Broker [using the common open policy service (COPS) protocol] about the decision to take on receiving RSVP Path, Reservation or teardown messages (Fig. 9). The decision taken by the QoS Broker is conveyed in a COPS message and then enforced by the router.

In the other IP-QoS solution, DiffServ, no resource reservation mechanism is defined. Routers should allocate enough resources for the high priority DSCPs, while the lower ones or the "classical" best effort (BE) traffic (DSCP 0) may use spare resources. DiffServ networks require access control in the edge routers, so that only authorized users can inject the core packets with high priority DSCPs. Depending on the type of edge routers, this access control can take place in different levels of detail. Since there is no "reservation protocol," in DiffServ, access routers must "detect" when admission control should take place (e.g., new traffic) and when the resources should be freed (e.g., traffic of certain kind remains inactive for too long).

4G networks are much more heterogeneous than 3G and thus a fine QoS control of the data transport service is needed. This control is done by the QoS Broker we presented in the above paragraphs. In a simplified way, the QoS Broker manages and monitors the network resources in one particular domain of operation. It also monitors the edges for incoming and outgoing resource reservations/utilization. The information thereby acquired is used in conjunction with information coming from other systems—management, user’s AAA etc.—to take admission control decisions and re-configurations and to convey them to the routers.

These are the classical QoS broker operation scenarios. Interaction with other elements of 4G networks brings richer possibilities and scenarios. For instance, the interaction with AAA assures providing each user his contracted QoS as stated in his user profile. Besides, the QoS Broker is responsible of assuring a constant QoS level even
if the user is moving through different access networks with different load states and heterogeneous characteristics. This includes managing the resources in the optimal way and taking into account the different users’ profiles and priorities. However, even if a very fine QoS control is performed, the required level of resources may not be able to be attained. In such cases, the QoS broker may ask to direct the flows to content adaptation nodes so that they are accommodated to the available resources. One way to achieve this re-direction is to ask the originators to do so. This can be done if the QoS Broker is able to interact with the flow originators or with proxies that interact with them; the QoS broker may even request the participants to reduce the QoS of their session. This will be seen next.

The biggest leap forward is to build an interface between the QoS broker and the service providers. This interface can also be used with proxies—not delivering any service but taking part in the application-level session being setup between other entities, for instance, SIP proxies assisting peers with SIP user agents in setting up sessions. This design is one of the 3GPP’s IMS’s key features, where the QoS broker is termed PDF (policy decision function). The idea is to have two levels: one handling QoS at application layer (this is done by the SIP proxies or the service providers), the other handling QoS at network layer (this is done by the routers). The QoS broker is able to communicate with both (Fig. 10).

The QoS Broker is, for the service quality (QoS), like an intermediary between the QoS defined at the application-level and its actual enforcement at the network level (routers). In IMS the QoS Broker (it is termed PDF) is only a decision maker while in 4G it is in charge of managing the QoS of the network. Thanks to this, the interaction between application and transport level QoS can be richer than in IMS. We can support, for instance, scenarios where the QoS Broker detects the link BW is decreasing (e.g., because the user moved to lower BW access network) and tells the SIP proxies to ask the users to change codecs to reduce the transmitted BW. In UMTS, the actual QoS reservation at network level is triggered by the terminals activating a PDP context, just in the same way we saw in section “Qos Negotiations in 3G.” In 4G networks the PDF context activation would be replaced by RSVP or by the DiffServ access control mechanisms we just saw. Another option is that the QoS Broker could setup the whole QoS in the routers just after receiving a request from the SIP proxies or the service providers. Salsano and Veltri\cite{21} compares both options (QoS Broker or terminal setting up network level QoS).

![MSC for QoS negotiation following the RFC2749 [20] model (resource reservation case).](image1)

![The QoS broker: bridging QoS negotiation and routers' configuration in 4G networks.](image2)
Advantages of Using 4G Technology

We can forecast that 4G networks’ future success is assured. The reasons we can give for such a success are diverse. First, it will spare costs to the network operators since, instead of managing two networks, the telephone and data network/Internet, the operators will need to handle only one. Second, we have to consider the enormous user acceptance of Internet applications. Besides, due to the openness of the Internet it is very easy to build more and better services upon it and to offer them (via any access technology) to the customers who will then find new opportunities to spend money, thus increasing the operators’ revenue. The business models such that the operators can profit from this bunch of new services are also under development. The users will not only like the integrated access to applications and the service bundles created by the networks operators, but also will enjoy these services on the basis of any-time any-where, always best connected and using any device.

All along this section, we saw the benefits of the 4G openness, making it possible to employ any solution developed for the Internet such as multicasting and “pyramid broadcasting and carrouselling” for video on demand.

Challenges in Streaming Multimedia

Content in 4G

The business model adopted in 4G is a fundamental challenge for the operators. There are also big technical challenges, for instance when trying to integrate different functionalities like QoS and mobility. We will review in this section the challenges faced by streaming applications in 4G.

The heterogeneity and variability of network access and the diversity of terminals, employing any terminal on any network and for any purpose is a major challenge to designers. It is fundamental and a big endeavor to stream the content in the best possible quality despite these constraints. As we saw, several are the techniques to achieve this, ranging from content adaptation to pre-fetching.

Shortcomings of 4G

4G architectures are still under discussion. They will profit from all the existing Internet technologies. Many of these technologies are already available but the real challenge is to integrate them. Not only have we to integrate the technologies but also the diverse entities participating in the service under coherent telecommunications business models. In previous network generations (2G and 3G) the technologies were integrated and, normally, there was one single entity in the business value chain so the telecommunication service is provided coherently. This led to less flexibility to create services than in 4G but, on the other side, for instance, the provision of QoS is done coherently among all levels, application and transport.

Besides, since 3G networks were designed initially to support voice and mobility, which was not the case for Internet, some aspects were provided more efficiently in 3G than in 4G, for instance the overhead for voice application is 30% less in 3G than in 4G.

FUTURE TRENDS

What Other Services Should 4G Strive to Support

Owing to the open nature of 4G networks, the services can be provided without operator control. For the operator not to become a mere bit pipe, it must build service platforms and offer over them many services (either by itself or via service provider partners). For these services to have great user acceptance, they must offer more than those services provided over non network-operator-partnered platforms. The success of these services is crucial so that the “semi walled garden” business model succeeds and the operators find a relevant place in the telecommunication market.

Offering “better services” ranges from delivering better video quality to more valuable or tailored services, for instance, considering user’s context and location. It also includes service bundling or providing a platform with single sign-on and unified billing to existing client-server or peer-to-peer applications. Service bundling and service provider–network operator partnerships can be built in many levels. For instance, we can think of a big company, offering its employees services like advanced on-line agendas with capabilities such as reserving company’s meeting rooms. Outside the company’s premises the partnered network operator provides the connectivity to the employees. The network operator can also offer this electronic agenda service (itself or via a partnered service provider) to all its customers but, perhaps, in a more basic way.

Services should be rich but, still, the user interface and tariffs must be kept simple and understandable, which is often difficult to achieve.

Further Research Opportunities

The research efforts in 4G networks are huge since, as we explained, they may become a reality sooner than forecasted. The time frame may become short but the work to do is hard, so the scientific community is paying lots of attention to 4G networks.

First, we need to adapt the IP-based 4G networks to support the services offered today in mobile and fixed telephone networks, with the same or better performance. This is far from simple owing to the very different technical natures of the telephone and the IP Internet networks. The architectures for providing those services are still under discussion but they are generally based on IETF stan-
standards and protocols, for instance, for mobility in Mobile IPv6(MIPv6) and for seamless micro-mobility (mobility within small—dozens or hundreds of meters—areas) in hierarchical MIP. The biggest challenge is to integrate the different functionalities achieving trade-offs so that the design of one solutions does not affect the performance of the others. This issue is well known and addressed by the research community. Many aspects are “ready;” the challenge lies now in integrating them. Issues such as mobility, fast handovers, QoS, and AAA must inter-work in the best possible way and prototypes already exist.[22]

The design of the transport network with possibilities to interact with service providers is, as we said, an important research aspect. But the design of services platforms with pervasive characteristics is also an important aspect to the research community. The best ways for the network operator–service provider interaction are being debated, with a strong bias on using the AAA infrastructure to do so. Integrating sensors, context, and different user “virtual identities” to build context-aware pervasive services and respecting user’s privacy (with techniques such as anonymization) is a field open to many research directions.

Last but not least, the definition of business models and redefinition of concepts like “network operator” is another extremely wide research opportunity. We can ask what is a network? Where are the limits of the network operator domain? Imagine, for instance, an airport with a huge internal public network (e.g., Ethernet plugs in VIP lounges and Wi-Fi all over). The airport may not want to manage the network and, as such, it may make it dependent and controlled by a commercial network operator. But, still, the airport wants all its travellers (irrespective of their telecommunications company) to profit from the Wi-Fi access. Now we can understand the questions previously raised.

**CONCLUSION**

Rich multimedia applications and the control of their delivery, not only in the transport aspect but also in the aspect of controlling the service and the users will be a fundamental task of the telecommunication operators. As such, there is no wonder about the strong interest for this kind of applications.

We analyzed the ways these applications are delivered over existing 3G networks and the possibilities to do so in the future 4G architectures. 3G solutions, influenced by the “monolithic” approach of the traditional telecommunication industry are good in terms of delivering an integrated QoS. However, they are far less flexible and open than 4G solutions and, as such, are poorer when talking about service bundling and creating rich multimedia applications. 4G has to integrate several aspects to assure QoS. The solutions seem to exist, the major problem laying in the fact that several business entities need to co-ordinate to deliver an end-to-end QoS. While this is a drawback of 4G, as we saw, the advantages of 4G are clear. Not only are they more flexible, they also support multiple terminals and networks (including new radio technologies with more BW) and all the advanced techniques designed for the Internet, such as “carrouselling” for video on demand, can be applied to 4G.

4G seems to be an ideal network to support multimedia, but research needs to be carried out (and this is intensively done) to get the most out of its multiple possibilities.

**SUMMARY**

Multimedia and streaming applications are getting more and more important in 3G networks and will be an important part of the future 4G networks. In 3G several commercial solutions exist already and the provision of the service is granted in a semi-monolithic way. While this is good for issues such as QoS, it lacks flexibility to create new services. Besides, the broadcasting support is not yet available. 4G networks are much more flexible, new services can be built and they can adopt better new technologies like the ones designed for the Internet.

**REFERENCES**

2. 3G Generation Partnership Project Homepage, available at http://www.3gpp.org/.
3. 3G Generation Partnership Project 2 Homepage, available at http://www.3gpp2.org/.
15. Serrano, P.; Bernardos, C.J.; Soto, I.; Moreno, J.I. Medida y análisis del tráfico multimedia en redes móviles de cuarta generación; Telecom I+D: Madrid, Spain, 2002.

BIBLIOGRAPHY