PROVIDING END-TO-END QOS IN 4G NETWORKS

Rui Prior[1], Susana Sargento[2], Janusz Gozdecki[3], Rui Aguiar[2]
[1]LIACC, University of Porto, Portugal
[2]Institute of Telecommunications, University of Aveiro, Portugal
[3]Department of Telecommunications, AGH University of Science and Technology, Poland
rprior@ncc.up.pt, ssargento@det.ua.pt, gozdecki@agh.edu.pl, ruilaa@det.ua.pt

ABSTRACT
This paper proposes an end-to-end QoS solution for 4G IP-based networks, able to support all types of services, from legacy to adaptive multimedia, and able to handle user mobility, intra- and inter-domain, across different access technologies. The issues of session signaling and resource reservation for individual flows in the access, resource management in the core, and QoS control across domains are addressed in an integrated fashion. The proposed solution is scalable, based on DiffServ with layered resource control: resource management in the core is performed on a per-aggregate basis, whereas in the wireless link, where resources are scarce, per-flow QoS control is used.

KEY WORDS
End-to-end QoS, 4G networks, heterogeneous, mobility, intra/inter-technology

1. Introduction

Next generation wireless communication systems, usually referred to as 4G, will provide a wide range of services to the users. These services, ranging from legacy applications, such as data transfer, to voice and multimedia calls and advanced value-added services, must be supported across a great diversity of network access technologies and by operators targeting very different market segments. In order to satisfy user requirements, proper end-to-end QoS must be provided to the application flows. The requirements of seamless mobility of users and scalability further complicate the issue: the provision of seamless end-to-end QoS in such a demanding and heterogeneous scenario is still a major challenge in networks research.

In order to provide end-to-end QoS to the application flows, enough resources must be available along the entire flow path. In the most demanding scenario, where the mobile terminals communicating are attached to different access domains, this path comprises (1) the access networks of both terminals, (2) the core network of the access domains where the access networks belong, and (3) the inter-domain path, consisting of all the transit domains traversed by the flows. Several solutions for QoS support in 4G IP based networks exist in the literature [1], [2], [3]. These solutions assume that the core network is overprovisioned, and therefore, only access QoS support is required. Other approaches also take into account core issues (e.g. [4]), but do not provide a fully integrated QoS approach to IP-based communication for different types of applications and protocols, usually disregard adaptive applications, and do not take into account mobility issues.

In this paper we present an overall solution for end-to-end QoS support in a 4G architecture, addressing the issues of session signaling and resource reservation for individual flows in the access, resource management in the core, and QoS control across domains in an integrated fashion. A uniform approach for all types of services is proposed, and multiple QoS service models, according to the overall network configuration (defined by operator policies) are supported. This QoS approach is able to handle high user mobility, both intra- and inter-technology, and both intra- and inter-domain. Scalability is achieved through the use of a DiffServ framework [5] with different levels of resource control: resource management in the core is performed in a per-aggregate basis, using information provided by a monitoring platform, whereas in the wireless link, where (radio) resources are very limited, QoS control is performed per-flow.

The paper is organized as follows. Section 2 contains an overview of the proposed 4G architecture. Section 3 discusses the session signaling and resource reservation approach proposed for individual flows at the access networks. Resource reservation and QoS control at the core are discussed in section 4. Section 5 describes the approach for inter-domain QoS control. Finally, the main conclusions are presented in section 6.

2. Architecture Overview

The proposed architecture supports a wide range of services with seamless mobility of users across very heterogeneous networks. This heterogeneity stems not only from the diversity of access technologies which must be supported, but also from the need to be inclusive of operators with quite different dimensions, characteristics and business cases. The support of different access

technologies is required to allow the optimization of the coverage/performance/cost factor under very different utilization scenarios, which range from Local Area Networks (LAN) to Broadcast Diffusion Networks (such as WiFi, WiMax, UMTS and DVB).

The development and fast deployment of advanced communication services in such a heterogeneous environment requires the definition of a uniform architecture, capable of hiding the inherent complexity from those services. This uniformity is achieved by using IPv6 as a convergence layer that hides the specificities of the different access technologies from the applications and services. The native support of mobility in IPv6 is also of major importance in 4G communication systems. In order to provide completely seamless mobility, however, an extension based on the support for fast handovers [5] is applied to IPv6. These issues and their relation with QoS aspects have already been addressed in the literature (e.g. [1]).

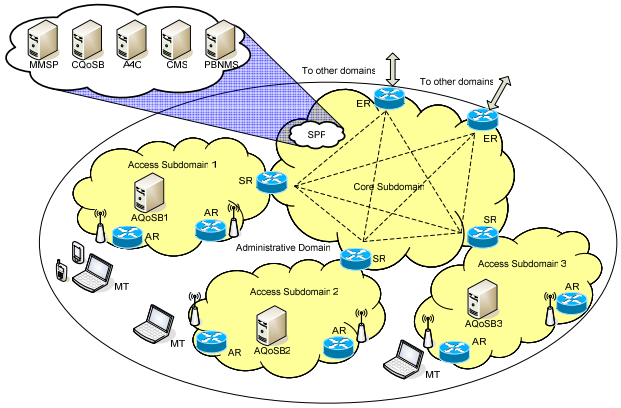

Figure 1 : Network architecture

Figure 1 shows the proposed architecture for the next generation network. Each administrative domain (AD) may contain several access networks (AN), each of them capable of supporting different access technologies, and a core subdomain providing interconnection between the access networks, via subdomain routers (SR), and to other administrative domains, via edge routers (ER). The architecture contains QoS elements in the AN, denoted by AN QoS Brokers, that control the admission of new flows and the handovers, and manage network resources, configuring the Access Routers (AR) accordingly, in a PDP-PEP (Policy Decision/Enforcement Point) relationship. An important feature of the AN QoS Brokers is the ability to optimize the usage of operator resources by load balancing users and sessions among the available networks (possibly with different access technologies) through the use of network-initiated handovers.

In order to provide QoS to all kinds of services, including legacy IP applications, novel functionalities are added to the ARs to mark and recognize individual flows, and to translate other QoS reservation mechanisms, such as the IntServ [7] Resource Reservation Protocol (RSVP) [8] into DSCP markings and QoS Broker requests. The entity performing these functions is termed ARM (Advanced Router Mechanisms) [9]. A QoS client module in the terminals, able to mark application packets for a QoS service and to issue requests to the broker, may also perform the resource requests, providing finer control over QoS to the user or the applications.

In the core network (CN), there is a Service Provisioning Platform (SPP) that provides the building blocks for creating services and applications on top of this network. The SPP contains a CN QoS Broker, responsible for resource management in the core, dealing with aggregates of flows traversing the core and inter-domain resources. Policies for resource management are defined by the PBNMS (Policy-Based Network Management System) and sent to the CN QoS Broker, where they are cached in a local repository for use. The Central Monitoring System (CMS) collects statistics and other network usage data from network monitoring entities, and configures these entities to perform both passive and active probing. The information collected and processed by the CMS is fed to the PBNMS and the QoS Brokers, which use it for proper network (re)configuration and resource management. A Multimedia Service Platform (MMSP), consisting of a broker and proxy servers, is responsible for the provision and control of multimedia services. It is also capable of mapping application level QoS configurations to network resource requirements and of performing QoS requests for the flows, as an alternative to the QoS client on the terminal or the ARM. This architecture, thus, has a large degree of flexibility in QoS signaling, enabling the use of a diversity of QoS access signaling scenarios that fulfill the needs of the different applications and business cases of different operators. Unification of the scenarios is achieved by the centralization of admission and handover control at the AN QoS Brokers.

An A4C (Authentication, Authorization, Accounting, Auditing and Charging) server is also present in each domain. In order to improve the network efficiency and scalability, AN QoS Brokers retrieve from the A4C a subset of the user profile when a user registers in the network. This subset, termed NVUP (Network View of the User Profile), contains information on the set of network level services (classes of service, bandwidth parameters) that may be provided to the user, reflecting the user's contract with the operator. Similarly, a Service View of the User Profile (SVUP), containing information on the higher level services available to the user (e.g., voice calls, video telephony, and the respective codecs), is retrieved by the MMSP to control multimedia services.

QoS support in the core is based on the DiffServ model, for scalability; in the access, IntServ-like per-flow reservations are used for better control. Though resource management is performed on an aggregate basis in the core and inter-domain segments of the path, information on the aggregates is propagated to the AN QoS Brokers,

where it is used for admission control in order to achieve end-to-end QoS. This combination of per-flow and per-aggregate processing in a two-layer hierarchy allows our architecture to provide fine-grained QoS control while keeping the scalability properties of per-aggregate core resource management, decoupled from per-session signaling. The next sections will detail the different pieces of the overall QoS approach and how they fit together.

3. Session Signaling

The ANQoSB is the central element that performs admission control for new flows and controls the handovers. For this purpose, the ANQoSB has detailed knowledge on the topology and resource usage of the AN, and is aided by metering information collected by the CMS. Although core and inter-domain resources are managed on an aggregate basis, communication between CN and AN QoSBs provides the latter with the necessary information to build maps containing the available resources to the different access networks in the same domain and to other administrative domains. Three tables are maintained by the ANQoSBs: one with information on resources of the AN, another with information on resources of the paths between ANs and between the AN and the edge routers, and a third one with alarm levels corresponding to the availability of resources in the inter-domain route, detailed in section 5. These tables, along with information on the set of network QoS services available to the user, contained in the NVUP, are used by the ANQoSB for admission control.

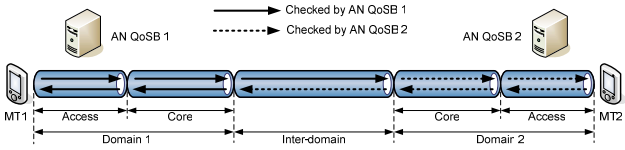

Figure 2 : Admission control in the inter-domain call scenario

In order to establish a reservation for a flow with fully end-to-end QoS, requests must be performed to the ANQoSBs of both endpoints of the flow. The admission control process will be different according to the relative location of the endpoints. When a mobile terminal, MT1, initiates a session to another one, MT2, there are 3 possibilities for their relative location: (1) they are connected to the same AN, (2) they are connected to different ANs in the same domain, or (3) they are connected to different administrative domains. In the first case, a single ANQoSB is involved, and resource checking is performed for the AN only, since communication is local. In the second case, ANQoSB1 checks for resources in the first access network, AN1, and the core, and ANQoSB2 checks for resources in AN2. In the third case, each ANQoSB checks for resources in the respective AN and in the core of the domain where they belong, and for transmission resources in the inter-domain path segment (see Figure 2), as will be detailed in section 5.

3.1 Session Initiation

As was previously mentioned, in order to support all the required applications and operator business cases, the network architecture is very flexible regarding the initiator of the QoS requests, which may be the MT, the ARM, the MMSP, or even an application server. In order to take advantage of this flexibility, different scenarios for the integration of the application setup and negotiation signalings and the network QoS signaling, necessary for the establishment of sessions with end-to-end QoS, were developed. A thorough description and analysis of all signaling scenarios is presented in [10].

Figure 3 illustrates a simplified example of a multimedia session initiation using SIP (Session Initiation Protocol) [11] in the scenario where terminals themselves issue QoS requests, for the case where the terminals are connected to different administrative domains. It is worth noting that, although illustrated with SIP, the scenario also works with different signaling protocols.

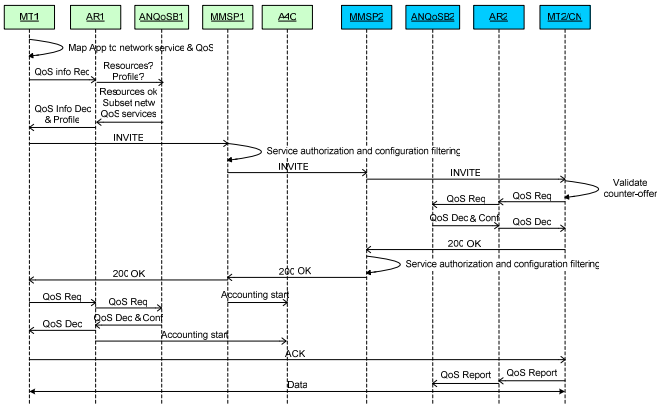

Figure 3 : Multimedia service setup with end-to-end QoS: MT issuing QoS requests

The calling terminal (MT1) begins by mapping the application requirements to network service and QoS requirements. It then sends a request to its QoS broker, ANQoSB1 (via a QoS attendant at the access router, AR1), with information on the required network and QoS parameters for the session. ANQoSB1 answers with information on the services that may be used according to the user profile and the current network status. This step prevents the terminal from trying to initiate services that cannot be supported by the access network or that the terminal is not allowed to use, in face of the user subscribed services. If allowed by ANQoSB1, MT1 sends an INVITE to MT2. When receiving this INVITE message with an initial offer of QoS configurations, MMSP1 performs service authorization, filtering out those not allowed by the SVUP, and forwards it to the MMSP of the callee (if MT2 was roaming, the message

would go first to its home MMSP). If the service is authorized, the INVITE is forwarded to the MT2. MT2 matches the QoS configurations in the INVITE to its own set, requests resources to ANQoSB2 and, accordingly, generates a counter-offer, included in the 200 OK. (The 180 Ringing message is omitted since it is not relevant in terms of QoS information). The counter-offer in the 200 OK is subject to authorization and filtering by MMSP2. Now that the location of MT2 is known, MT1 issues a request to ANQoSB1, selects the final configuration among those in the counter-offer, and sends an ACK containing this final configuration. The ACK triggers a QoS report to ANQoSB2, and accounting processes are initiated in the A4C allowing for transport- or service-based charging. Configuration of the access routers is triggered by the QoS requests at both sides.

3.2 Mobility

Mobility plays a central role in 4G networks, and the requirement for seamless handovers is probably the most demanding one in terms of timing. In a heterogeneous network, handovers may be performed across different access technologies; therefore, in this architecture they are performed at the access-agnostic layer 3. The handover process in our architecture is extended from the fast handover mechanism defined in [5], associated with the Candidate Access Router Discovery (CARD) protocol [12], used to propagate to the MT information on prospective networks for handover. In order to fully exploit the resources of each access technology, the capability of session renegotiation is provided by coordination of handover and application signaling.

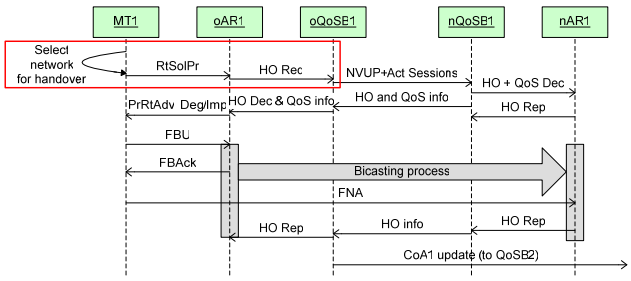

Figure 4: Fast handover process

Figure 4 illustrates a basic intra-domain, inter-AN handover process (procedures specific to inter-technology handovers are presented later in this section). In the case of a user-initiated handover, the terminal sends a Router Solicitation for Proxy (RtSolPr) message with an indication of the new network to perform handover, selected based on information provided by CARD. The old AR (oAR1) sends a handover request message to the old QoS Broker (oQoSB1), which pushes the NVUP, along with information on the set of active sessions, to the QoS Broker of the prospective network (nQoSB1). If nQoSB1 accepts the handover, the new AR is configured, and the decision is communicated to the oQoSB1 and then to the terminal by the Proxy Router Advertisement (PrRtAdv) message. The terminal then sends a Fast Binding Update (FBU) message confirming the handover. The FBU indicates that the terminal will move and triggers a bicasting process [5], where each packet sent to MT1 via the old network is duplicated at oAR1 and also sent via the new network. The Fast Neighbor Advertisement (FNA) message, sent by the terminal upon handover, tells the new AR1 that the handover is completed. Both QoSBs are informed of the fact, and the bicasting process stops, since the terminal may no longer receive information via the old network. Furthermore, oQoSB1 informs QoSB2 of MT1's new CoA, so that it can update filter configurations at AR2.

Network-initiated handovers are equally possible, providing a means to optimize operator resources. The differences between terminal- and network-initiated handovers are that in the latter the RtSolPr and HO Req (box in Figure 4) are absent and the PrRtAdv message contains an indication that the handover is mandatory.

The integration of handover and session renegotiation is achieved by means of the PrRtAdv message, which, if applicable, contains indication of the need to perform service degrading or the possibility of service improvement, used as a trigger for session renegotiation.

The most frequent handovers are intra-technology. Usually, no renegotiation is performed in these handovers, but in case of cell congestion some QoS degradation (reduction of reserved resources) may be required and, conversely, when leaving the congested cell, QoS may be improved again. With SIP, renegotiation for improvement is initiated by sending a re-INVITE together with the FBU. The renegotiation process is performed in parallel with the handover. Since the time to complete the handover is usually much shorter (in the order of 50-100 msec) than the renegotiation process (eventually, up to 1-2 sec), the handover completes and the activation of the improved QoS is performed in the new network (if not, the ACK that activates the changes is delayed until the handover is complete). Renegotiation for degrading is more demanding: if the handover is completed before renegotiation, the new network might be flooded with more traffic than it can handle, but the handover process cannot wait for the renegotiation to complete, since it needs to be fast due to the imminent loss of signal. This issue is solved by providing the network with some available bandwidth to cope with this traffic; also intelligent resource management can be used, temporarily supporting the overload for a short period if the user has an important profile, while temporarily reducing the available bandwidth of low priority users.

Although not as frequent, inter-technology handovers are also supported and, indeed, this is one major advantage of 4G networks, allowing features such as the automatic increase in the quality of a videoconference when arriving

at a 802.11 HotSpot, or the dropping of the video component of a multimedia call without dropping the call when leaving the HotSpot and keeping the call via a GPRS connection. In this case, the differences in QoS levels in the different networks are potentially very large. Service improvement poses no problems, and works similarly to the intra-technology case, but for service degrading, large differences in QoS levels prevent the new AN from temporarily supporting the overload. In this case, the solution is to increase the handover time. This approach is feasible in inter-technology handovers since the cell overlapping area is usually much larger, requiring only an adjustment of the signal strength thresholds that trigger the handover in order to give more time for the handovers.

A handover to a different AN implies a change in the core aggregate to the edge router (or to the AN of the correspondent node in intra-domain calls). Therefore, when receiving the message from oQoSB1 (Figure 4) with information on the NVUP and active sessions of the user, nQoSB1 needs to check for available resources in its AN and in the intra- and inter-domain path segments, ensuring that the new path has sufficient resources to accommodate the flows with the required QoS.

Inter-domain handovers are usually more complex, involving a new complete registration process and, therefore, the disruption of the active sessions. In order to avoid this, we resort to Context Transfer (CT) to install security information in the new domain, including the security associations, derived from those installed in the previous domain. CT is performed through the CNQoSBs, directly or via the A4Cs, depending on the existence of federation between the domains. QoS admission control is also required in the new domain. To decrease the handover time, the CT and QoS admission control processes are integrated. After this preparation phase, the decision concerning the handover is sent to the mobile terminal in the PrRtAdv message. The terminal activates the handover by sending the FBU message. The inter-domain handover process is, thus, similar to an intra-domain one, mostly differing by the inexistence of bicasting and by the use of CNQoSBs as proxies of the ANQoSBs.

4. Intra-Domain QoS Control

The intra-domain QoS control covers QoS resource management for an administrative domain from the user terminal to the edge router (ER). The main requirements for the intra-domain QoS architecture are: 1) scalability of the signalling within the administrative domain; 2) flexibility (easy to manage); 3) efficiency in the usage of network resources; 4) support for the mobility of users.
In this DiffServ environment, per-class aggregate resources are dynamically allocated, by the CNQoSB, based on actual network traffic, operator polices and other conditions. The monitoring subsystem plays an important role in this process, identifying aggregates to/from where resources should be reassigned.

The core resource management is based on:

1 - Policies received from the PBNMS – information containing the description of the different transport services and the network topology. The CNQoSB has a bilateral interface with PBNMS: it requests for policies at start-up time and receives unsolicited policy definition when policies are changed in PBNMS. The CNQoSB generates alarms to the PBNMS reporting, e.g., continuous resource over usage or ANQoSB fault.

2 - Measurements supplied by the CMS – CNQoSB detects if usage of a given link is above a certain threshold and can reallocate resources from less used links in order to increase the capacity of that link (first part of Figure 5).

3 - Requests from the ANQoSB – ANQoSB can directly ask the CNQoSB to change the amount of resources of a given link (second part of Figure 5).

Figure 5 depicts the resource management process in the core. The CNQoSBs reconfigure the bandwidth reserved for the aggregates on the basis of measurements and in response to requests sent by ANQoSBs. The CMS periodically sends the *Measurement Data* message with the monitoring results (the bandwidth occupied per class, the mean/maximum packet delay and loss in a class, etc). With this information, the CNQoSB has information on the congestion status of each class, and can reconfigure its routers (bandwidth per class, queue length, etc.) if required. In the case of core reconfiguration, the CNQoSB sends an *Agg Info* message to the ANQoSBs of the access networks affected by the reconfiguration to push an aggregate map update. Measurement information is usually used for long term reconfigurations, enforcing domain policies. Note that the CNQoSB can be provided with the measurement data on a periodic basis as well as on the requests sent to the CMS.

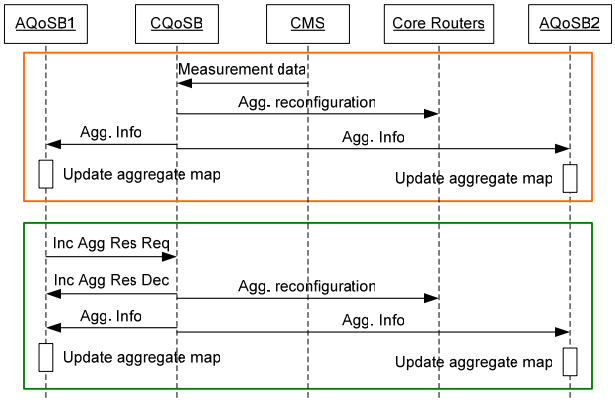

Figure 5: Resource management in the core network

Core reconfigurations may also be requested by an ANQoSB by sending an *Inc Agg Res Req* message to the CNQoSB when more bandwidth is required in a core aggregate to its access network. The CNQoSB answers this request and, if possible, reconfigures the routers and sends an *Agg Info* message to the ANQoSBs affected by the reconfiguration to update their aggregate maps.
The joint usage of these two mechanisms assures network flexibility while simultaneously minimizing the amount of signalling information exchanged in the connections between the CNQoSB and the CMS, and between the CN and AN QoSBs.

Considering that the core network is usually not the bottleneck in terms of bandwidth, core reconfigurations should be infrequent, and so should measurement information sent by the CMS. Note that, each core reconfiguration may imply sending resource map updates to all the ANQoSBs that have to refresh the information related to this reconfiguration. Therefore, the use of partial, on demand reconfigurations decreases the signalling load and improves the network efficiency.

5. Inter-Domain QoS Control

Though much attention has been paid to intra-domain QoS, much less has been done in the scope of inter-domain QoS control. While the solutions of over-provisioning or static DiffServ configurations are simple, they cannot provide any guarantees regarding end-to-end QoS. Additional mechanisms must, therefore, be used for inter-domain resource management. Our approach is based on inter-domain routing with QoS constraints. A solution for inter-domain QoS should be scalable and based on an evolution from the existing inter-domain routing, which the dynamic nature of QoS information should not compromise. Additionally, in order to gain acceptance, it should be simple and impose minimum requirements on intra-domain routing and QoS control.
Our approach to inter-domain QoS control is based on 3 main pieces: 1) a set of well-known traffic classes globally supported by all operators, 2) service level agreements (SLA) between adjacent domains, and 3) an inter-domain routing protocol propagating QoS information.

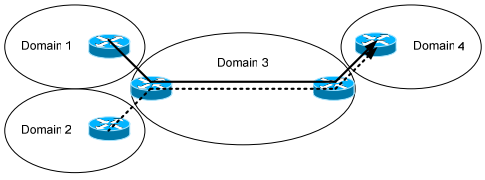

Figure 6 : Virtual trunk type service level specifications

The well-known classes are a small set of traffic class templates with particular characteristics that should be globally supported by all network operators (e.g., a conversational class for small-sized packets with very low delay and jitter). These classes have well-defined per-domain limits for the major QoS parameters (delay, jitter, percent loss), and limits for the complete path may be derived by combining the values of the traversed domains. Since they are merely templates, operators have to map the well-known classes into the specific classes implemented in their own domains.

The SLAs must contain Service Level Specifications (SLS) that specify a set of aggregates, each corresponding to a particular (ingress point, egress point, service class) triplet. These aggregates may be regarded as virtual trunks connecting, for each traffic class, two different domains across a third domain directly connected to both. Figure 6 illustrates the concept: an SLS between domain 1 and domain 3 specifies that X traffic may flow between domain 1 and domain 4 (continuous line) for a given traffic class; an SLS between domain 2 and domain 3 specifies that Y traffic may flow between domain 2 and domain 4 (dashed line) for the same traffic class. Aggregates are managed internally within each (transit) domain by the respective CNQoSB (ensuring that enough resources are assigned): there are no requests from CNQoSBs in different domains.

The inter-domain routing protocol needs to be capable of conveying QoS information. Currently, BGP (Border Gateway Protocol) [14] is the most common protocol for inter-domain routing. In IPv6 internets, version 4 of BGP with multiprotocol extensions, commonly referred to as BGP4+ [15], is used. In addition to propagating reachability information (routes) between peering domains (eBGP - External BGP), BGP is also used to share information on learned routes among the different edge routers of a given domain (iBGP - Internal BGP).
The standard BGP protocol does not carry QoS information: the usual metric for the path selection policy is the number of traversed Autonomous Systems (AS). This selection criterion may easily lead to the choice of sub-optimal paths: shortest path may be the one with lower bandwidth links or with more congestion. BGP, however, can be extended to support QoS routing, and proposals for doing so exist in the literature (e.g. [16]). In our proposed extension to BGP, QoS information is conveyed in the UPDATE messages by a newly defined Path Attribute, QoS_INFO, which is optional and transitive. This attribute contains the following information:

1. Allocated bandwidth for each of the well-known service classes (minimum along the path).

2. Congestion alarm level for each class (maximum along the path): 0–no congestion; 1–very light congestion; 2–medium congestion; 3–serious congestion.

These values are updated by the BGP-speaking routers at each transit domain, taking into account the virtual trunks for all classes between the domain to which the route is

advertised and the "next hop domain" for the route. Notice that these virtual trunks are shared among different AD-to-AD aggregates: in Figure 6, for example, all traffic transported from domain 1 to domain 4 via domain 3 and belonging to the same service class shares the red virtual trunk, independently of the origin and destination access domains.

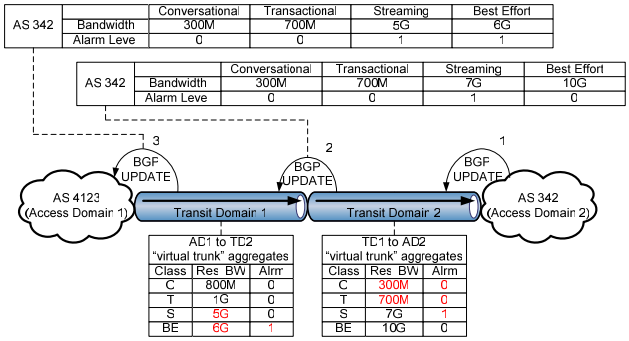

Figure 7 : Example of the propagation of QoS information

Figure 7 shows an example of the propagation of QoS information with this BGP extension. The virtual trunks are configured independently in the transit domains (TD). When TD2 receives the UPDATE from AD2 and the route is selected, it propagates the route to the upstream domain, TD1, with information on the reserved bandwidth for the virtual trunks for the different classes from TD1 to AD2. On receiving this UPDATE, TD1 decides if this route should be used; if it is, the route is propagated to AD1 combining the QoS information from the received UPDATE with that of local virtual trunks from AD1 to TD3. As can be seen, the path reserved bandwidth announced to TD1 is the minimum along the path, whereas the congestion alarm level is the maximum.

Multi-level congestion alarms reduce the fluctuations in the usage of the aggregates: with multi-level alarms, when level 1 (light congestion) is reached no domain should use that route to replace a previously established one, but domains that were already using the route should not switch to a different one unless a higher congestion level is reached, avoiding the synchronized route flapping problem. The policy for assigning an alarm level to a virtual trunk at a transit domain must be such that alarm level updates are infrequent.

The information on the reserved bandwidth is used to select the route with the larger amount of bottleneck bandwidth, while the alarm levels are used to eliminate congested routes from the set of possible choices. The other relevant QoS parameters (delay, jitter, percent losses) are not included, since they have per-domain upper limits for each of the well-known classes (upper limits for each route may be derived from the number of traversed domains). It is worth noting that multi-level alarms, when coupled with information on allocated resources, provide coarse-grained information on resource availability. Taking into account scalability reasons, we consider inter-domain routes between two ADs to be the same for all traffic classes. Each route then represents the best tradeoff for all service classes to the destination domain.

The information on inter-domain routes must be retrieved by CNQoSB in order to manage core resources (Figure 8); this task is performed by a BGP module installed in the CNQoSBs, which are, therefore, iBGP speakers. Conversely, bandwidth and alarm information on the aggregates between edge routers in the domain must be propagated to other domains. This information, stored at the BGP PIB (Policy Information Base) of the edge routers for use by the respective Decision Processes (the process which selects a given route to a destination AS among all the available ones), is configured and updated by the CNQoSB in a similar way to the other router parameters. If the route is selected, the edge routers propagate it to their peers in the upstream domain with the updated QoS_INFO attribute. As was previously mentioned, the CNQoSB propagates information on the inter-domain routes to the ANQoSBs, where it is used for admission control purposes.

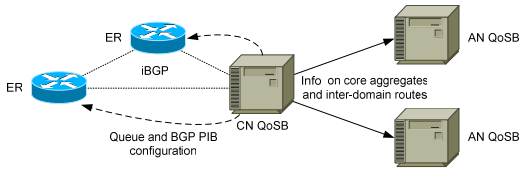

Figure 8 : Inter-Domain Communications

6. Conclusions

This paper presented an end-to-end QoS solution for 4G networks able to, in a scalable way, support differentiated QoS for a large diversity of users and services, supporting the large mobility of the users, both intra- and inter-technology, and both intra- and inter-domain, and assuring the QoS guarantees of legacy and demanding multimedia services. The proposed solution achieves end-to-end QoS by the close interaction between application and network signaling (session setup and mobility with QoS), and by the two layer approach to resource reservation in the access and core networks. Based on the characteristics of the solution, we conclude that it is able to support end-to-end QoS with low signaling load processing load in the resource management elements. This QoS solution is being simulated and its performance and scalability will be addressed in future work. A prototype implementation is undergoing in the scope of the Daidalos project, and will be used to prove the concepts here presented.

7. Acknowledgement

This work is in part supported by the EU FP 6 for Research and Development Daidalos (IST-2202-506997).

References:

[1] V. Marques et al., An IP-Based QoS Architecture for 4G Operator Scenarios, *IEEE Wireless Communications, 10*(3), 2003, 54-62.

[2] D. Wisely, E. Mitjana, Paving the Road to Systems Beyond 3G - The IST MIND Project, *Journal of Communications and Networks, 3*, 2002, 1042-1046.

[3] Joachim Hillebrand, et al., Quality-of-Service Signaling for Next-Generation IP-Based Mobile Networks. *IEEE Communications, 42*(6), 2004, 72-79.

[4] Koch, A QoS Architecture with Adaptive Resource Control – the AQUILA Approach, *Proc. QoFIS'2001*, Coimbra, Portugal, 2001.

[5] S. Blake (ed.) et al., An Architecture for Differentiated Services, *IETF RFC 2475*, Dec. 1998.

[6] R. Koodli (ed.), Fast Handovers for Mobile IPv6, *Internet Draft*, Oct. 2003.

[7] R. Braden et al., Integrated Services in the Internet Architecture: an Overview, *IETF RFC 1633*, Jun. 1994.

[8] R. Braden (ed.) et al., Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification, *IETF RFC 2205*, Sep. 1997.

[9] D. Gomes et al., A Transsignaling Strategy for QoS Support in Heterogeneous Networks, *Proc. ICT'2004*, Fortaleza, Brasil, 2004.

[10] R. Prior et al., Heterogeneous Signaling Framework for End-to-end QoS support in Next Generation Networks, *Proc. HICCS-38*, Hawaii, 2005.

[11] J. Rosenberg et al., SIP, Session Initiation Protocol, *IETF RFC 3261*, Jun. 2002.

[12] M Liebsch et al.: Candidate Access Router Discovery, *Internet Draft*, Dec. 2003.

[13] AQUILA Project (IST-1999-10077), Deliverable D1203, Final system specification, Feb. 2003.

[14] Y. Rekhter and T. Li, A Border Gateway Protocol e (BGP-4), *IETF RFC 1771*, Mar. 1995.

[15] T. Bates et al., Multiprotocol Extensions for BGP-4, *IETF RFC 2858*, Jun. 2000.

[16] G. Cristallo and C. Jacquenet, Providing Quality of Service Indication by the BGP-4 Protocol: the QOS_NLRI attribute, *Internet Draft*, Jun. 2003.