Requirements and Current Work of QoS for Mobile IP and Wireless Internet

Yucel Balkanci, University of Texas at Arlington

Abstract

This paper presents the requirements to provide certain levels of QoS (Quality of Service) for mobility, followed by different approaches that are motivated to solve the problem. All the approaches are based on DiffServ (Differentiated Services), which is superior to IntServ if QoS for mobility and wireless internet is considered. The Hierarchical QoS Architecture Approach is based on controlling several parameters of a wireless LAN cell: limited distance to ensure the same bandwidth for all hosts, the constrained rate of traffic sources to limit the use of the channel in function of the required QoS and the limited number of active hosts to keep the load low enough to provide the requested QoS. The second approach is the Moby Dick project which is still in progress today. Main objectives of this project are to facilitate the development of seamless access to existing and emerging IP-based applications, to propose an architecture for wireless Internet access by developing new mechanisms for seamless hand-over, QoS support after and during hand-over, AAA, and charging, to facilitate new business opportunities for operators, manufacturers, services providers, and content providers for wireless, access, and backbone technology and services, and to contribute actively to standardization bodies, such as Internet Engineering Task Force and Internet Research Task Force.

I. Introduction

For telecommunication and wireless internet industries there is one main ambitious paradigm for developers: to provide the ability for users and customers to access all kinds of information like data, voice, video, etc anywhere in minimum amount of time. We are still far from catching this dream since internet means the whole world and the whole world means a combination of different technologies and architectures trying to work together to provide the requested services with the desired QoS from one end to the other one. For cellular networks this objective is closer to be fulfilled at least for voice connectivity. Till the near past, cellular devices were only being used for voice connections through the cellular networks. As we consider today’s demands and expectations we can easily say that the two industries, the internet and cellular networks, have to converge in another form, let’s say in the form of mobile internet. This means that cellular networks have to work together with all other kinds of wired and wireless networks that are part of the internet.

To solve the QoS problem for mobile IP, four steps need to considered and understood well. First step is to find out the requirements. Second is to evaluate today’s QoS solutions against these requirements. Third is to decide to make additional improvements for the current solutions. The last step is to evaluate the results and consider new solutions. In the following section the requirements of QoS for mobile IP is explained in detail. Requirements section is followed by two approaches, The Hierarchical QoS Architecture Approach and The Moby Dick Project. Finally, the paper ends with the conclusion part.

II. Requirements of QoS for mobile IP

Mobile IP is used to route packets to a mobile device as this mobile node changes its connection spot and moves to another one. As this mobile node changes its location there should be a mechanism that provides the same or expected QoS to the mobile node through the new link.

The problem scenario is as follows: a mobile node may change its access router after a handover process. As a result the path for the packets being sent or received from CN to mobile node might change. The current session's packets might start using a new care-of-address. Further, the handover may take place between to subnets and these two subnets might have totally different administrative control and functions.

We can divide the requirements into four categories: (1) Performance requirements (2) Miscellaneous requirements (3) Interoperability requirements and (4) Standard requirements.

Performance Requirements:

Interruption needs to be minimized during the handover process. During the handover, the middle node on the new path might have not received the QoS requirements and as a result the packets through this node might be treated by the default QoS instructions. This kind of QoS interruptions must be minimized. The number of packets with the default QoS treatments can be counted to calculate the amount of interruption.

Consider the following example: For RSVP (Resource reservation protocol) when the route changes the resource reservations have to be changed as well. Soft state is used to deal with this dynamic situation. A soft state is a set of state information at a router that expires unless regularly refreshed. If a route for a session is changed then a soft state will expire and new resource reservations will invoke the appropriate soft states on the new routers along the route. During the handover of RSVP, OPWA (one Pass with Advertisement) model of reservation causes the latency of about one round trip time between CN and MN. This means that in the intermediate nodes packets might receive default QoS treatment for one round-trip time. This kind of latency might be acceptable at the initialization state of the session but it is not acceptable in the middle of the session.

Re-establishment of the QoS treatment needs to be localized at the part of the route where changes took place after the handover. This will minimize the delay for the connection and overhead on the network. Sometimes it is not possible to keep this re-establishment local like in the case of handover between different administrative domains.

The previous QoS state along the old path needs to be released. The QoS mechanism needs to support a way (explicit or timer-based) to release old states. For example for RSVP it is timer-based. The soft state will expire unless it is refreshed periodically. It needs to be noted that, during the handover process, it is not always possible for the MN to send a tear down message explicitly along the old path since the connection with the old access router will be lost.

Interoperability Requirements:

There are two sub-categories: (1) Interoperability with mobility protocols and (2) Interoperability with heterogeneous packet paths as regards QoS paradigms.

Several mobility protocols are already defined or may be defined in future in IETF. Localized mobility management, fast handover, context transfer are several examples. The QoS mechanism should take advantage of these mobility protocols for the optimized operation.

After the handover takes place, there might exist a different QoS paradigm along the new path. The QoS mechanism for Mobile IP should make the necessary forwarding treatment along the packet paths that are consist of different QoS paradigms. Assume that a MN is handed-over from DiffServ network and the second one is at the edge of an IntServ network.

The new access network would expect to exchange the RSVP messages so that proper QoS treatment for the stream can be done in the new network. QoS mechanism should have provisions to handle these kinds of heterogeneous network combinations.

Miscellaneous Requirements:
After handover the MN might connect from one access router to another one. There could be multiple paths over which MN's packet may propagate. Several examples of these paths are: triangle route via Home Agent, route-optimized path between the MN and its CN, temporary tunnel between old and new access routers, reverse tunnel from the new access router (Foreign Agent) to Home Agent, etc. A QoS mechanism should support QoS along different paths. On the other hand a QoS mechanism may not be able to support all of these paths at the same time.
Huge number of devices with Mobile IP will be connected to the internet via wireless links. This means that QoS for Mobile IP may provide information to the wireless link layers for them to support the required QoS. This information should be meaningful both ways: to applications so that they can choose the IP service of certain QoS characteristics and to wireless link QoS managers so that they provide this information to lower level mechanisms.

Standard Requirements:
Even though it is not possible to set quantitative targets, QoS mechanism for Mobile IP should satisfy certain standards like security, scalability, conservation of wireless bandwidth, low processing overhead on mobile terminals, robustness against failures and so on.

III. Facts about IEEE 802.11B

If we consider the characteristics of a wireless LAN environment, we can see that it is not easy to provide QoS over these wireless channels. We need to analyze the 802.11 MAC layer under performance related issues. The more active hosts there are the more overhead 802.11 MAC layer raises.
Assume that there is a single host transferring data over 802.11B with a bit rate of 11Mb/s.

$T_{single} = t_{pr} + t_{tr} + SIFS + ACK + DIFS$

Where t_{pr} is the preamble time (144 µs), t_{tr} is the frame transmission time (size/bit rate), SIFS (Short Inter Frame Space) is 10 µs, ACK (ACK Transmission Time) is 210 µs, and DIFS (Distributed Inter Frame Space) is 30 µs.

$r = t_{tr} / T_{single} = 1.11ms / 1.51ms = 0.735$

This means that a single node transmitting data over an 11 Mb/s radio channel can only use a bandwidth of 8.08 Mb/s. In fact there will be more nodes using the same channel at the same time. This means that there might me collisions and the efficiency might go down and down. If we want to provide the expected QoS, we need to limit the number of nodes using the same 802.11 link. As a simple example scenario: Assume there is no restriction of how many nodes can connect and transmit/ receive data over the same link. Assume that at the time T only two Mobile nodes are connected. At the time $T + \Delta t$ 20 more nodes connect through the same link with a request of streaming video. It will definitely be harder to provide the Sufficient QoS for every host getting connected to the same channel.

Another problem with the 802.11 is the distance between the nodes that the transmission takes place. If because of the distance, one of the nodes can not use high bandwidth capacity then this will affect the other mobile nodes that can use higher bandwidth. This means that if we want to provide the QoS efficiently we need to restrict the usage of the link to the nodes that can all use high or same level of bandwidth.

Finally, traffic sources should be constrained by configuring traffic shapers in hosts to obtain the target QoS.

IV. Hierarchical QoS Architecture Approach

Differentiated Services

The architecture is composed of two main parts: the core and access networks. Packets are forwarded by the core routers and the access network connects the edge router to the mobile node. Typically, s DiffServ domain would be under the control of one administrative entity. The services provided across a DiffServ are defined in a service level agreement (SLA). If the destination is beyond the customer's DiffServ domain, then the DiffServ domain will attempt to forward the packets through the other domains, requesting the most appropriate service to match the requested service.

As part of the DiffServ standardization effort, specific types of PHBs (Per-Hop Behavior) need to be defined, which can be associated with specific differentiated services. Currently there are three categories including the best effort (BE) issued.

Expedited Forwarding PHB

RFC 2598 defines the expedited forwarding (EF) PHB as a mechanism that can be used to support what is referred to as a premium service. This kind of service is a low-loss, low-delay, low-jitter, assured bandwidth one. By its nature, internet involves queues at each node, or router, where packets are buffered and queued. As a result guaranteeing the EF PHB is not easy. EF packets have higher priority, so the interior nodes must treat the incoming traffic in such a way that queuing effects do not appear. In other words, the requirement on interior nodes is that the aggregate's maximum arrival rate must be less than the aggregate's minimum departure rate.

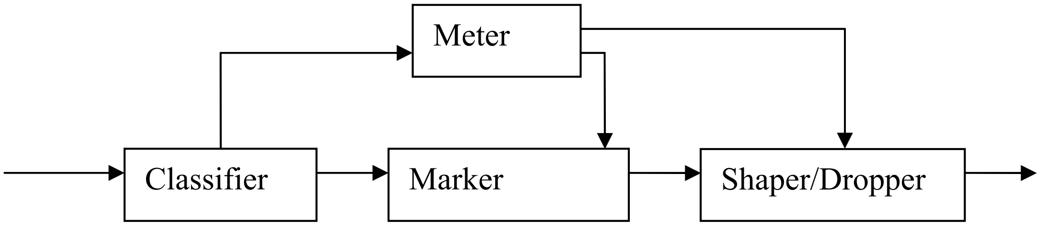
Assured Forwarding PHB

The assured forwarding (AF) PHB is designed to provide a service superior to best-effort but one that does not require the reservation of resources within an internet and does not require the use of detailed discrimination among flows from different users. AF traffic needs a minimum amount of bandwidth, but if available this amount can be increased. Users are offered the choice of a number of classes of service for their traffic.

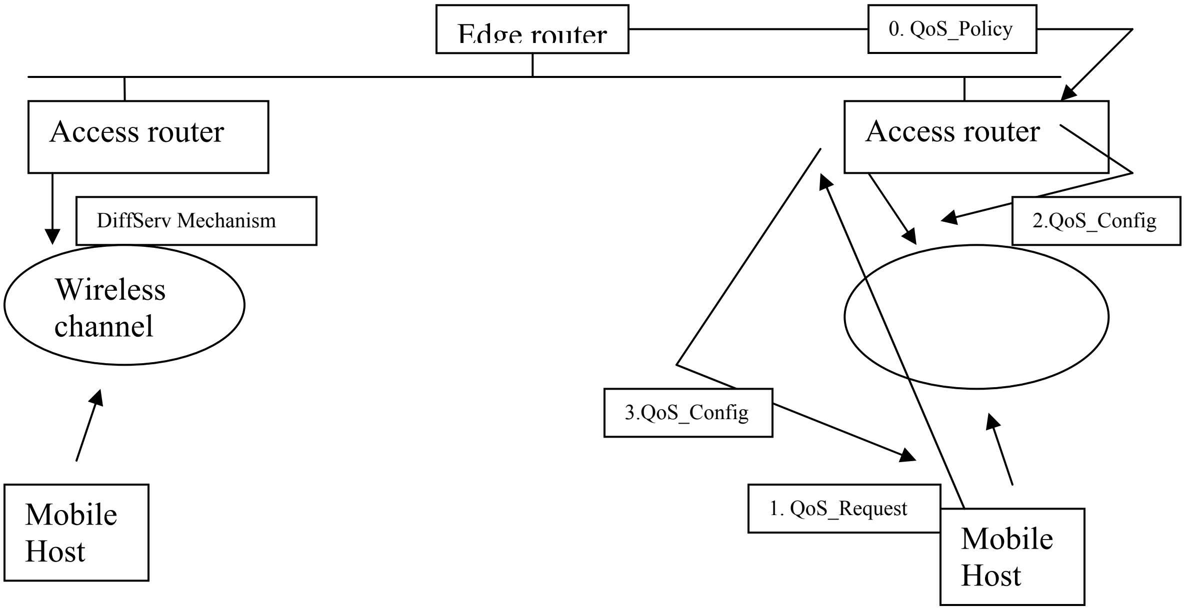
Best Effort PHB

This kind of traffic is not guaranteed any kind of QoS. If the resources are available the packets will be treated with a better quality of service.

At the edge router the packets are classified and marked with a DSCP (Differentiated Services Codepoint). There are five traffic conditioning functions: Classifier separates submitted packets into different classes. This is the foundation of the DiffServ. A Meter measures submitted traffic for conformance to a profile. The meter determines whether a given packet stream class is within or exceeds the service level guaranteed for that class. Marker might be used to mask the packets that exceed the traffic. Shaper polices traffic by delaying packets as necessary so that the packet stream in a given class does not exceed the traffic rate specified in the profile for that class. Dropper drops the packets when the rate of packets of a given class exceeds the amount specified in the profile.



The above functions give us the ability to extend the DiffServ model to the wireless environment so that we can provide IP level QoS to mobile nodes. The main difference between using the DiffServ model in wireless and wired environment is for the wireless environment the conditions may change rapidly and unpredictably as more mobile nodes try to connect to the system or as more applications are launched. The manager of QoS has keep track of all the QoS requests for all the mobile hosts, and it has to update this information periodically to adapt the rapid changes. There has to be an admission control that can prevent a host to connect to the system if there are sufficient resources in a cell. The QoS management has to be tightly coupled with mobility management so that the overall performance perceived by mobile hosts be acceptable for each QoS class.



The above figure shows the architecture. As it is seen each wireless cell is connected to an access router. These access routers are connected to edge routers through wired over-provisioned LAN. All mobile hosts and access routers are fully functional by DiffServ mechanism, so that traffic sources are controlled. The system is hierarchical because there are two time scales and two levels of management: intra-cell and inter-cell management. The first level of management (intra-cell) is local to one cell and performed by the access router. Mobile hosts inform the access router and it configures the QoS mechanism. The second level (inter-cell) management concerns a set of cells connected to an edge router. At this level, conditions does not change that fast as it is in first level of control. This global management is done by the edge router as it fixes long term policies for access routers.

Local QoS Management

The available bandwidth of the link depends on the number of the active hosts in a cell and on the aggregated traffic of each class. So the access router has to keep track of all the hosts and their QoS expectations.

Bandwidth allocation is done according to the class of QoS (EF, AF or BE). To perform bandwidth allocation, the proportion of the useful bandwidth in function of the traffic load and the number of hosts in the 802.11b wireless LAN are measured. The access router provides the requested QoS if it is possible to do so. This is the result of the soft state principle. The QoS management module in the mobile host configures the output rate of the EF and AF/BE classes. All allocations are time-sensitive. When a mobile host stops sending packets, its allocation is canceled after the interval.

Global QoS Management

The edge router is considered to be the global QoS manager for manipulating the cells. An edge router sets the policies to be followed by the access routers. This approach proposes five rules for reservation: (1) reserve a given bandwidth in all cells, (2) reserve a given bandwidth in the cells on a given path, (3) reserve a given bandwidth in the cells on a frequent mobility path, (4) reserve a given bandwidth in the neighbor cells, and (5) reserve a given bandwidth in one cell.

The forth rule refers to EF and the fifth rule refers to AF.

QoS signaling

Between all elements (mobile hosts, access routers, edge routers) of this architecture there is an exchange of information to manage QoS as well as mobility. The IPv6 stack has two modules: QoS and mobility management,, and DiffServ mechanism. Cooperation between the management modules is done by signaling protocol that either uses data packets for communication (in-band signaling) or generates ICMP control packets (out-of-band signaling). Signaling protocol format is as follows:

Command	parameters
QoS_Request	bandwidth in 64 Kb/s units
QoS_Config	EF rate, AF/BE rate, AF weight
QoS_Policy	bandwidth in 64 Kb/s units, traffic class, source address, policy type
HO_Request	target AR, current QoS allocations
HO_Ack	source AR, host route
HO_Deny	source AR

In band signaling consists of inserting commands into data packets transmitted between a mobile host and an access router. If there is no data traffic, ICMPv6 that contains signaling commands is used. This solution is also used to send signaling commands to a remote entity, which is the case of the hand-off request.

V. Moby Dick - Mobility and Differentiated Services in a Future IP Network

In order to continue to evolve 3rd Generation mobile and wireless infrastructure towards the Internet - targeting IST 2000 IV 5.2 "Terrestrial Wireless System and Networks", the project Moby Dick will define, implement, and evaluate an IPv6-based mobility-enabled end-to-end QoS architecture starting from the current IETF's QoS models, Mobile-IPv6, and AAA framework.

Key Issues:

- Definition of a common architecture integrating QoS, IPv6 mobility, and AAA (out of the separate architectural approaches for each component currently provided by the IETF) with respect to wireless issues.
- Implementation and evaluation of an IPv6-based end-to-end technological approach to fulfil the requirements of present and future mobile communication services.
- Implementation and evaluation of QoS models (e.g. Differentiated Services) in highly dynamic and heterogeneous network topologies (understanding of QoS models is normally restricted to relatively static environments).
- Definition of a suitable charging concept which would enable permanent mobile IP based services on a large scale (a strong requirement related to AAA, but currently not a topic within the IETF).
- Trans-European trial to test the implementation by using SOKRATES-ERASMUS exchange students as test-users.
- Actively participate in IRTF (Internet Research Task Force) AAAArch (Authentication, Authorization, and Accounting Architecture) working group, and monitor in particular ETSI, 3GPP (3rd Generation Partnership Project), MWIF (Mobile Wireless Internet Forum), IEEE (in particular 802.11).
- Follow and actively influence ongoing relevant IETF standardization activities in particular in the working groups

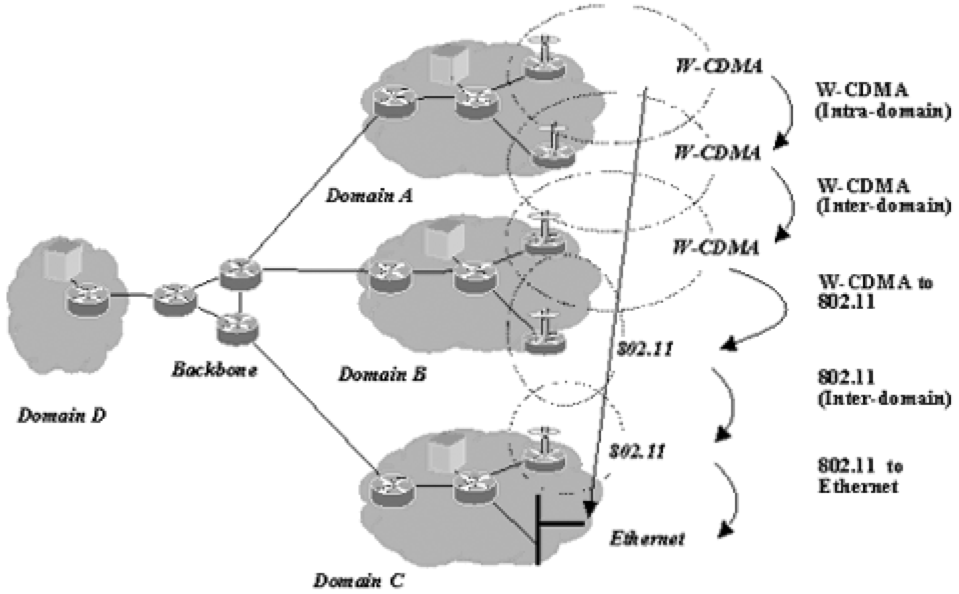


Figure: Network Structure and Components

Enabling IP QoS in Mobile Environments

QoS Constrains in Mobile Environments

As an example, a VoIP (e.g.) call should be able to proceed as a Mobile Terminal (MT) roams between a fixed network, a wireless LAN, and an UMTS service provider. Multiple and different handovers have to be established in this type of environment: between technologies (vertical handover) and between cells of the same technology (horizontal handover). These handovers are expected to be seamless, that is, undetected by the user. This imposes serious problems: these VoIP calls require the reservation of certain network resources, the assurance of certain QoS parameters, and often the confirmation of administrative information. This means that in some of those handovers, a couple of QoS reservation mechanisms and mobility functions must be performed, which are ultimately supervised by AAAC *(Authentication, Authorization, Accounting and Charging)* procedures (and/or servers). To make it transparent to the user, these handovers have to be done in a short period, without disturbance in the established communications. The network design and architecture is crucial in this environment. The success of such networks will be dependent of maters like the decision of which functions have to be performed, in which order, in which entities and the placement of control elements on the network, among others.

Relying on an IP network brings specific problems. In fact, the IP transport can be overlaid into a wireless network in multiple ways. Issues as QoS assurances in mobile environments can be structured in five layers of complexity: I) physical; ii) cell; iii) network, iv) management, and v) administration.

- The physical layer problems are associated with the fundamental connection capabilities. If any of the transmission links (wired, wireless) places any limits to QoS (e.g. delay assurances), then this will impact the overall end-to-end (e2e) service.

- The cell problems are motivated by the flow handover across neighbouring cells. QoS parameters have to be satisfied at the time of handover between similar cells. During the handover period, many packets can be in transit, and these will have to be properly treated in order for the end to end QoS to be kept.
- The network problems are due to changes in the IP network of attachment. This can be described as the classical mobile IP problem, increased with QoS constraints, and strong timing limitations for the handover phase (e.g. the MH is connected to a wired network and swaps to the UMTS network).
- Management problems are concerned with network management, core network configuration in order to assure the required QoS levels, fast on-demand reconfiguration, hierarchical network control, and inter-domain contract establishment.
- Administration is here used in the context of accounting, logging, and charging users and ISPs for resource usage. Although this may seem not to be related with QoS problems, charging models may be closely coupled with the QoS level provided. In this case, issues as handover can be dependent on the costs of specific QoS levels, and the MT can decide to change this level according with advertised connection costs, accepting lower quality connections for cost considerations.

The issue of handover illustrates the network changes in heterogeneous mobile environments. An analysis on the issues of QoS provision in a mobility environment has to consider these multiple context switches (CS) also, as these will place worst case requirements. Context changes in the connection can be caused by:

CS I. mobile terminal changing wireless cells – a MT crosses the border between two wireless cells (either UMTS or WLAN). This should ideally not lead to more AAAC traffic.

CS II. Mobile terminal changing IP Autonomous Domain (AD) – a MT becomes connected to a different AD (its wireless provider can have its network partitioned in different AD for management considerations, e.g.). This will require AAAC traffic.

CS III. Mobile terminal changing wireless provider – a MT switches wireless provider. This will require AAAC traffic, as different providers correspond to different AD also, in our network model.

CS IV. Mobile terminal changing transport technology (either wireless or wired) – a MT switches from wired to wireless connections or between different wireless technologies (this does not necessarily imply changes in the AD, e.g. moving from the corporation LAN to the WLAN). This may, or may not, require AAAC traffic.

When QoS issues are considered in these CS, the handover problem becomes much more complex.

Proposed Management Architecture

Regardless of the lower layers approach, Management and Administrative issues have to be handled during handover, and these issues will impose further delay requirements to the handover process. Figure below presents a simplified architecture optimized to QoS provisioning in mobile environments, considering all other network modules that must exist in such an environment.

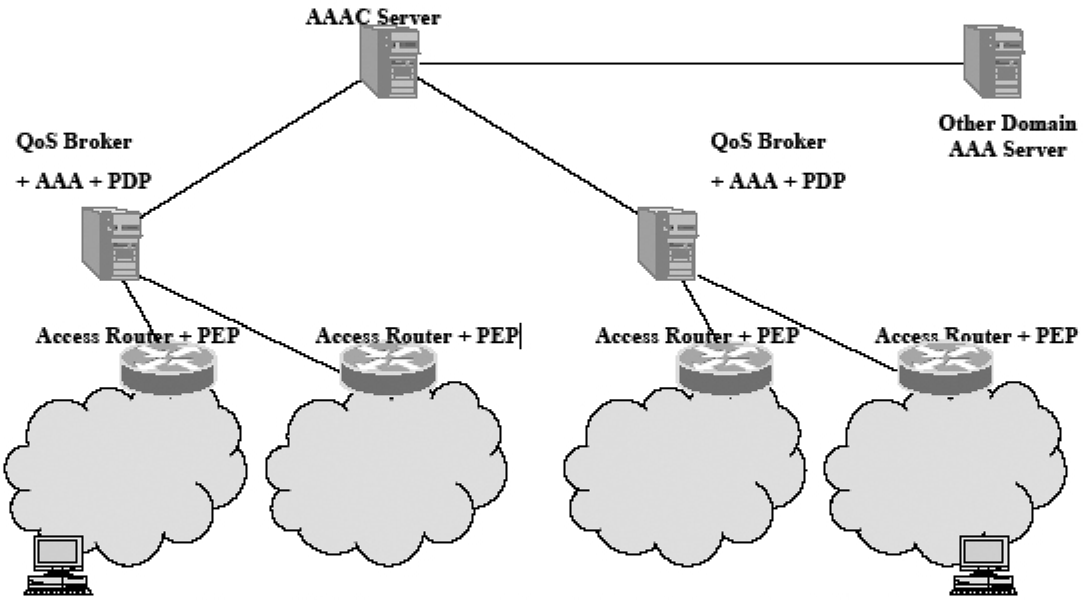


Figure 2. Management and Administrative Structure

This architecture is based on administrative boundary principles (each administrative domain has its own independent set of management entities) and on the concatenation of independent QoS-enabled connections in order to achieve end-to-end services.

In this scenario, an AAAC entity is the responsible to keep all information related with the user (the user profile) and to supply an adequate subset of that profile to proper network management system (NMS) entities. These entities, after user authentication and authorization, are the responsibles for the set-up of the user connections (with the demanded QoS) and for its usage. These NMS entities cover a QoS Broker, some basic AAA control functions, Policy Decision Points (PDPs) and Policy Enforcement Points (PEPs). They are thus responsible for: connection setup, verification if the user has permissions for certain resource reservations, enforcing that the user does not exceeds the amount of resources that it is paying for, and informing the AAAC server of the user usage of resources (for charging and accounting purposes).

The control messaging flow will be the following. The MT connects to a certain layer 3 network, gets a CoA (care-of-address) and initiates the authentication process, by sending an authentication message to the Access Router (AR). The AR will forward this request to the “QoS+AAA+PDP” entities, which will contact the AAAC server for that user profile. If that user belongs to that domain, the AAAC server sends the profile to the proper QoS+AAA+PDP entities, otherwise it will contact the MT’s home AAAC server to get its profile. After the relevant aspects of this profile are stored in the QoS+AAA+PDP entities, it will send the MT a message confirming its authentication, and inform the PEP of the user permissions. After this point, the MT may ask authorization to use any service that its profile allows, and the QoS entity will provide such connections. If sufficient network conditions are not available at that moment, there will be a negotiation between the MT and the QoS entity that either is successful or not. If it is, the MT may start using the resources that were provided, and the AR will send to the NMS information on that usage. If the user limits are exceeded, the PDP informs the MT of that and simultaneously asks the PEP to cut the resources. Periodically, or after MT relocation or on call completion, the QoS+AAA+PDP will inform the AAAC server of the MT total usage in that period, for accounting and charging purposes. If the MT changes to another layer 3 network, either that network is served by the same QoS+AAA+PDP entities, and it just has to demand for a new connection set-up (because that MT's profile is already stored at the QoS+AAA+PDP) or the full process with contact of that MT's AAAC home server is done.

VI. Conclusion

There are lots of problems to be solved when we consider mobility across different technologies (both wired and wireless). Especially, handover itself is a crucial problem. When a mobile terminal with a specific SLA (service level agreement) moves from one domain to another, there are no guaranties that the new domain will have enough resources to comply with the SLA because of the shared nature of the access network. In addition, when we consider that this new domain may belong to another Service Provider, the cost for the same SLA may be higher and the customer may not want to continue with the same SLA (that is, he may want to reduce the SLA to maintain the cost at an acceptable level).The decision of if and when to do a handover, and to which domain/cell, may be taken by the client or by the network - and may depend on many different factors. Hierarchical structuring of networks, allowing independent solutions at different complexity layers will be a fundamental aspect for successful networks. In the future the increasing number of wireless local area networks will be the basis for the high performance Wireless Mobile Internet.

In this paper we talked about two projects and approaches: The Hierarchical QoS Architecture and The Moby Dick project. In Moby Dick project, it is focused on the mobility control on IP layer. This leads to the fact, that all wireless cells will have their own router. Changing wireless cell will imply the (re)establishment of resources in the network, under the supervision of the QoS Broker. In an alternative approach, the mobility will be handled by the physical layer as much as possible (e.g. until the layer is able to sustain the "advertised" QoS parameters to the QoS Broker), and changes in IP will be invoked only when required. The second choice will lead to lower network overhead and faster handovers. The specific choice between these approaches (and in the case of the second, exactly how much aggregation to perform) will depend on the core IP-network behavior, and this approach can be performed in different ways at different networks.

References:

[1] J. Antonio García-Macías et al. 2001. Quality of Service and Mobility for the Wireless Internet, *Proceedings of the first workshop on Wireless Mobile Internet, Rome, Italy*: 34-42

[2] Moby Dick Project, http://www.ist-mobydick.org/

[3] IETF NSIS Working Group 3 April, 2003, Requirements of a QoS Solution for Mobile IP

[4] Victor Marques, Rui Aguiar, Francisco Fontes, Jürgen Jähnert, Hans Einsiedler, Enabling IP QoS in Mobile Environments
IST Mobile Communications Summit 2001, Barcelona, Spain, September 9-12, 2001, pp. 300-305

[5] Victor Marques (PT Invocao), Rui Aguiar (Instituto de Telecomunicacoes,University of Aveiro), Jürgen Jähnert (University of Stuttgart), Karl Jonas (GMD Research Centre for

Information Technologies), Marco Liebsch (NEC Network Laboratories Europe), Hans Einsiedler (T-Nova Berkom), Francisco Fontes (PT Invocao),
An Heterogeneous Mobile IP QoS-aware Network
Quarta Conferencia sobre Redes de Computadores - CRC'01, Covilha, Portugal, November 29-30, 2001.

[6] Vlora Rexhepi 2000, Wireless Internet QoS
Thesis towards Master in Technological Design (MTD) diploma at the University of Twente, Department of Computer Science and Electrical Engineering
[7] Nortel Networks 2001, Benefits of Quality of Service (QoS) in 3G wireless Internet
http://www.nortelnetworks.com

[8] Rajeev Koodli, Cedric Westphal and Meghana Sahasrabude, Nokia Research Center Mountain View, CA, QoS and Mobility slides.

[9] Jens Zander, Center for Wireless Systems Royal Institute of Technology Stockholm,Sweden
Who wants to pay for Wireless QoS ? slides