A Study on QoS Provision for IP-Based Radio Access Networks[1]

Alberto López[1], Jukka Manner[2], Andrej Mihailovic[3], Hector Velayos[4], Eleanor Hepworth[5], and Youssef Khouaja[6]

[1] Department of Communications and Information Engineering, University of Murcia, Facultad de Informática, Campus Universitario de Espinardo - 30071 Murcia, Spain
alberto@dif.um.es

[2] Department of Computer Science, University of Helsinki, P.O. Box 26, FIN-00014, Helsinki, Finland
jmanner@cs.Helsinki.FI

[3] Center for Telecommunications Research, King's College London, Strand, London, WC2R 2LS, UK
andrej.mihailovic@kcl.ac.uk

[4] Agora Systems S.A., c/ Aravaca 12 3ºB - 28040 Madrid, Spain
hvelayos@agora-systems.com

[5] Siemens/Roke Manor Research, Roke Manor, Romsey, SO51 0ZN, UK
eleanor.hepworth@roke.co.uk

[5] France Télécom R&D/DMR/DDH, 4 rue du Clos Courtel - BP 59 35512 Cesson Sévigné, France
youssef.khouaja@rd.francetelecom.fr

Abstract. The fast adoption of IP-based communications for hand-held devices equipped with wireless interfaces is creating new challenges for the Internet evolution. Users expect flexible access to Internet based services, including not only traditional data services but also multimedia applications. This generates a new challenge for QoS provision, as it will have to deal with fast mobility of terminals being independent of the technology of the access network. Various QoS architectures have been defined, but none provides full support for guaranteed service levels for mobile hosts. This paper discusses the problems related to providing QoS to mobile hosts and identifies the existing solutions and future work needed.

1 Introduction

The emerging wireless access networks and third generation cellular systems constitute the enabling technology for "always-on" personal devices. IP protocols, tradi-

[1] This work has been performed in the framework of the IST project IST-1999-10050 BRAIN, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues from Siemens AG, British Telecommunications PLC, Agora Systems S.A., Ericsson Radio Systems AB, France Télécom - CNET, INRIA, King's College London, Nokia Corporation, NTT DoCoMo, Sony International (Europe) GmbH, and T-Nova Deutsche Telekom Innovations-gesellschaft mbH.

S. Palazzo (Ed.): IWDC 2001, LNCS 2170, pp. 139–157, 2001.

tionally developed by the Internet Engineering Task Force (IETF), have mainly been designed for fixed networks. Their behaviour and performance are often affected when deployed over wireless networks.

The telcom world has created various systems for enabling wireless access to the Internet. Systems such as the General Packet Radio Service (GPRS), Enhanced Data Rate for GSM Evolution (EDGE), Universal Mobile Telecommunications System (UMTS) and International Mobile Telecommunications (IMT-2000) are able to carry IP packets using a packet switching network parallel to the voice network. These architectures use proprietary protocols for traffic management, routing, authorisation or accounting, to enumerate some, and are governed by licenses and expensive system costs.

From the QoS point of view, the problems with mobility in a wireless access network and mobility-related routing schemes are related to providing the requested service even if the mobile node changes its point of attachment to the network. Handovers between access points, change of IP-addresses, and mechanisms for the intra-domain micro mobility mechanisms may create situations where the service assured to the mobile node cannot be provided, and a violation of the assured QoS may occur. A QoS violation may result from excess delays during handovers, packet losses, or even total denial of service. In the case where the user only requested differentiation according to a relative priority to flows, a short QoS violation may fit within acceptable limits. If the flows were allocated explicit resources, the new network access point and route from the domain edge should provide the same resources.

Several research projects within the academic community, e.g. INSIGNIA [Lee00], and in the industrial community, e.g. ITSUMO [Chen00], have sought to combine mobility with guaranteed QoS. In the BRAIN project [BRAI00], we are envisioning an all IP network, where seamless access to Internet based services is provided to users. By using IETF protocols, we are designing a system that would be able to deliver high-bandwidth real-time multimedia independent of the wireless access network or the wireless technology used to connect the user to Internet. This implies the need for IP mobility support and also end-to-end QoS enabled transport. The provision of QoS guarantees over heterogeneous wireless networks is a challenging issue; especially because over-provisioning is not always possible and the performance of the wireless link is highly variable. We focus our architecture on wireless LAN networks, since these provide high bandwidths but may also create frequent handoffs due to fast moving users - this type of architecture is most demanding in view of mobility management and QoS.

2 QoS and Mobility Background

This sections presents QoS and mobility architectures relevant to the further discussion. We have not covered all existing architectures in or study but at least those considered most important or promising in order to understand completely all the issues concerning QoS and mobility interactions.

In the following discussion, the term mobile node (MN) is used to refer to a mobile host or mobile router. If mobile host (MH) is used, the term mobile router does not apply, and vice versa.

Regarding ***QoS*** we have considered IETF-presented architectures for providing different levels of services to IP flows, although much work has been done within the academic community and the telcom industry; for example INSIGNIA and ITSUMO are mature proposals for providing QoS to data flows. INSIGNIA has its own in-band signalling mechanism and ITSUMO is based on the DiffServ framework.

The IETF architectures can be classified into three types according to their fundamental operation; the Integrated Services framework [Wroc97] and the Resource Reservation Protocol (RSVP [BZB+97]) provides explicit reservations end-to-end; the Differentiated Services architecture (DiffServ, [BBC+98], [BBGS01]) offers hop-by-hop differentiated treatment of packets. There are a number of 'work in progress' efforts, which are directed towards these aggregated control models. These include aggregation of RSVP [BILD00], the RSVP DCLASS Object [Be00] to allow DSCPs to be carried in RSVP message objects, and the operation of Integrated Services over Differentiated Services networks ([Bern00], [WC00]) proposed by the Integrated Services over Specific Link Layer (ISSLL) Working group. On the application level the Real-Time Transport Protocol (RTP, [SCFJ96]) provides mechanisms for flow adaptation and control above the transport layer.

For ***Mobility Management*** we have based or study on an analytical method we call the Evaluation Framework [EMS00], which has been adopted for facilitating detailed analysis and comparative evaluation of mobility protocols. This framework facilitates the selection of the most promising candidates for mobility management and introduce a categorisation for distinguishing protocols and their associated purposes. This analysis is closely related to QoS development, since both mobility and QoS protocols are expected to have awareness of certain, if not all, of their functionality.

For the interaction study we have considered several mobility architectures present today. On the macro-mobility side Mobile IP [Perk00] is the current standard for supporting macroscopic mobility in IP networks and its Ipv6 counterpart, Mobile IP support in IPv6 [JP00], based on the experiences gained from the development of Mobile IP support in IPv4, and the opportunities provided by the new features of the IP version 6 protocol.

For the support of regional mobility we identified two major categories: *Proxy-Agent Architectures (PAA)* which extend the idea of Mobile IP into a hierarchy of Mobility Agents and *Localized Enhanced-Routing Schemes (LERS)* which introduce a new, dynamic Layer 3 routing protocol in a 'localised' area.

In the first group (PAA) examples include the initial Hierarchical Mobile IP [Perk97] and its alternatives, which place and interconnect Mobility Agents more efficiently: Mobile IP Regional Registration [GJP01], Transparent Hierarchical Mobility Agents (THEMA) [MHW+99] and Fast Handoff in Mobile IPv4 [El01]. The new Mobile IP version 6 [JP00] has had some optional extensions by applying a hierarchical model where a border router acts as a proxy Home Agent for the Mobile Nodes. They include "Hierarchical MIPv6 mobility management" [SCEB01] and "Mobile IPv6 Regional Registrations [MP01].

In the second group (LERS) there are several distinctive approaches: *Per host forwarding schemes* where soft-state host-specific forwarding entries are installed for each MN (HAWAII [RLT+99], Cellular IP [CGK+00], Cellular Ipv6 [SGCW00]); *Multicast-based schemes* which make use of multicast protocols for supporting point-to-multipoint connections (dense mode multicast-based [SBK95][MB97][TPL99] and the recent sparse-mode multicast-based [MSA00]); and *MANET-based schemes* adapted for mobile ad-hoc networks (MER-TORA [OTC00] [OT01]).

Figure 1 shows some of the many IP mobility protocols, which category they fall into and very roughly how they relate to each other.

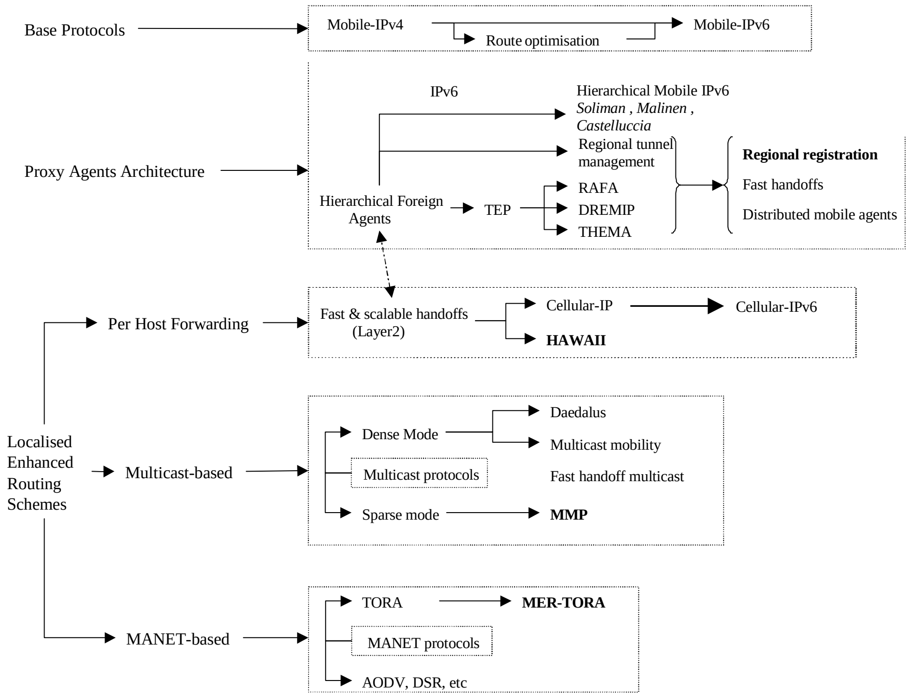

Fig. 1. Classification of mobility protocols

We will pay special attention to Handover Management, as it is considered one of the most important features of the mobility protocols when considering the interaction with QoS protocol because of the likely re-negotiation of QoS parameters. Handover refers in general to support for terminal mobility wherever the mobile node changes its point of attachment to the network.

We can identify several handover types: A **Layer-2 handover** happens if the network layer is not involved in the handover, **intra-access network**[2] **handover** when the new point of attachment is in the same access network, **inter-access network**

[2] Access Network (AN): An IP network, which includes one or more ARs and gateways.

handover when the new access router is in a different access network. **Horizontal** or **vertical** handover are said to happen if the old and the new access router[3] use the same or different wireless interface (technology) respectively.

We can also distinguish three different phases in a handover: the **Initiation Phase**, when the need for a handover (and its initiation) is recognized , the **Decision Phase**, when the best target access router is identified and the corresponding handover is triggered, based on measurements on neighbouring radio transmitters and eventual network policy information, and the **Execution Phase**, when the mobile node has been detached from the old access router and attached to the new one.

In a **planned** handover, contrary to an **unplanned** handover, some signalling messages can be sent before the mobile node is connected to the new access router, e.g. building a temporary tunnel from the old access router to the new access router.

Specific actions may be performed depending on the handover phase. For example, the events may initiate upstream buffering or advance registration procedures at the mobile node. These mechanisms characterize furthermore the handover type: **smooth** handover is a handover with minimum packet loss, **fast** handover allows minimum packet delays and **seamless** handover that is a smooth and fast handover.

3 Interaction of Mobility and QoS

This section discusses the problems related to guaranteeing service levels to mobile nodes. We classify the problem areas into three groups, namely topology related problems (3.1), and macro (3.2) and micro mobility (3.3) related issues. Solutions to these problems are presented in Section 4.

3.1 Depth of Handovers

We can identify several types of handover situations, which create different amounts of control signalling between different entities; handovers within the same Access Router (AR), between ARs and between access networks. The same physical handover can create different logical handover situations to different MN flows if the flows use different network gateways. Figure 2 shows a sample network topology to illustrate the levels of handovers while a MN moves within and between two networks.

The different levels of handovers create variable load of signalling in the access network. Also, if the QoS architecture has a signalling mechanism, such as RSVP, it adds to the need to signal in certain handover situations.

If the AR node does not change during a handover, the handover control only needs to handle radio resources since the routing paths do not change.

If the AR changes but the gateway stays the same due to similar routing, the handover affects the radio resource availability and the access network resources. In addi-

[3] Access Router (AR): An IP router between an Access Network and one or more access links.

tion, the new AR may need to check for admission control at the same time. All RSVP-reservations need to be refreshed.

If the gateway changes, either within the same access network or when the MN changes networks, flows may experience a drop in their QoS until the QoS signalling has updated the nodes on the paths. The time interval during which the MN is not receiving the subscribed QoS needs to be minimized.

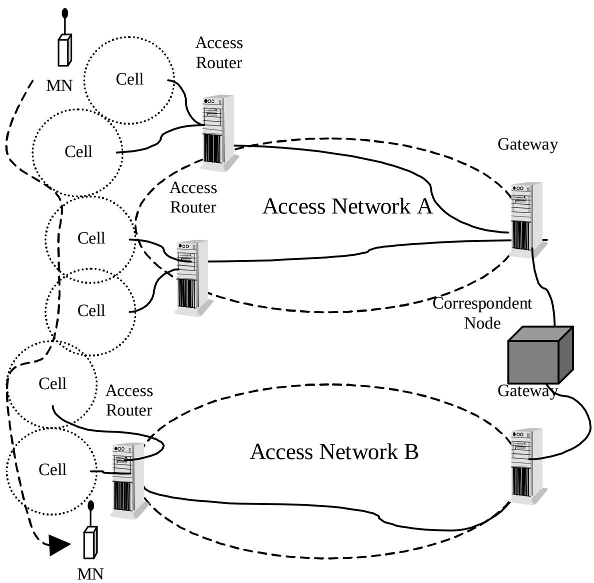

Fig. 2. Example network topology regarding different handover scenarios

3.2 Macro Mobility Issues

The first macro-mobility problem arises from the *triangular routing* phenomenon. Packets from the MN usually follow a direct path to the CNs, packets from the CNs are re-routed via the MN's home network to its point of attachment in a foreign network, from where they are forwarded to the MN's current location. Several QoS architectures operate best when packets follow the same route in the forward and reverse direction. Triangular routing can affect the service level guarantees of these schemes.

It is possible to tunnel the upstream flow to follow the downstream using *Reverse Tunnelling* [Mont01]. However, routers in the tunnel may not be able to recognize some encapsulated parameters of the QoS protocols apart from IP addresses. For example, if RSVP packets use the *Router Alert option* to indicate to routers on the path that they require special handling, when RSVP messages are encapsulated with an outer IP header, the Router Alert becomes invisible. Although solutions to this have been proposed e.g. RSVP extensions to mobile hosts [AA97], they still add complexity to the operation of QoS protocols on mobile environments.

Other main concern for QoS when the host is moving is the time needed to re-establish the routes, and hence, the time needed to re-configure resource management required to provided QoS in the new location. Even in Route Optimisation, transmission of binding updates directly to CNs result in a large update latency and disruption during handover. This effect is greatly increased if MN and HA or CN are separated by many hops in a wide area network. Data in transit may be lost until the handover completes and a new route to the MN is fully established. Route Optimisation (as a protocol specification) however includes Smooth Handoff support using Previous Foreign Agent Notification extension, which can be used to avoid the described disruption.

There are other problems related to signalling load and address management. Highly mobile MNs create frequent notifications to the home agent, which can consume a significant portion of wireless link resources. Since the current Mobile IP standard requires the mobile to change the care-of address (either FA or co-located) at every subnet transition, it is more complex to reserve network resources on an end-to-end path between the CN and the mobile. For example, if RSVP is used, new reservations over the entire data path must be set up whenever the care-of address changes. The impact on the latency for re-establishment of the new routes is critical for QoS assurances.

Mobile IPv6

Mobile IPv6 makes use of the new features provided by IPv6 protocol. They help to solve most of the problems discussed above which arise with the use of Mobile IP in IPv4 networks. For example *Route Optimisation* is included in the protocol, and there are mechanisms for movement detection that allow a better performance during handover. The *Routing Header* avoids the use of encapsulation, reducing overhead and facilitating, for example, QoS provision.

Although the Mobile IPv6 solution meets the goals of operational transparency and handover support, it is not optimised for managing seamless mobility in large cellular networks. Large numbers of location update messages are very likely to occur, and the latency involved in communicating these update messages to remote nodes make it unsuitable for supporting real-time applications on the Internet. These problems indicate the need for a new, more scalable architecture with support for uninterrupted operation of real-time applications.

3.3 Micro mobility Issues

The domain internal micro mobility schemes may use different tunnelling mechanisms, multicast or adaptable routing algorithms. The domain internal movement of MNs affects different QoS architectures in different way. IntServ stores a state in each router; thus a moving mobile triggers local repair of routing and resource reservation within the network. DiffServ on the other hand has no signalling mechanism, which means that no state needs to be updated within the network, but the offered service level may vary. At least the following design decisions of a micro mobility protocol need to be considered when combining mobility and QoS architectures within a network:

- the use of tunnelling hides the original packet information and hinders Multi-Field classification,
- changing the MN care-of-address during the lifetime of a connection,
- multicasting packets to several access routers consumes resources,
- having a fixed route to the outer network (always through the same gateway) is less scalable,
- adaptability and techniques (speed and reliability) to changing routing paths,
- having an optimal routing path from the gateway to the access router and
- support for QoS routing.

Multicast approaches can have ill effects on the resource availability, for example, because the multicast group can vary very dynamically. The required resources for assured packet forwarding might change rapidly inside the domain, triggering different QoS-related control signalling and resource reservations.

The use of tunnelling can affect the forwarding of QoS-sensitive flows since the original IP-packet is encapsulated within another IP-packet. However, as long as the tunnel end-points are capable of provisioning resources for the tunnelled traffic flows, the agreed QoS level need not be violated. Tunnelling has the advantage that multiple traffic flows can be aggregated onto a single reservation, and there is inherent support for QoS routing. Micro-mobility schemes that rely on explicit per-host forwarding information do not have such simple support for QoS routing, because there is only one possible route per host. Both IntServ and DiffServ have been extended to cope with tunnelling ([TKWZ00], [Bla00]) and the changes to the IP-address ([MH00]). Some coupling of the macro and micro mobility protocols and the QoS architecture may still be needed to ensure an effective total architecture.

4 Solutions

This section identifies various schemes for providing parts of an all-inclusive support of QoS-aware mobility. A full support of mobile terminals with QoS requirements can be accomplished by a combination of these schemes.

4.1 Strict Shaping at Network Edges

Network operators already intercept each packet arriving from an external network and decide whether the packet can be allowed into the core network. This admission control is performed by a node called the firewall and is based on IP addresses and port numbers e.g. identifying applications. Firewalls are typically deployed for security reasons and usually scan both incoming and outgoing packets.

The firewall operation can be modified by using different rules for performing the admission control. Instead of just preventing known security problems, the edge nodes would use defined bandwidth and QoS policies on a per-flow basis for controlling the traffic admitted into the network. Both the access routers and the gateways perform the admission control, the former for flows originating from mobile nodes and the latter for flows emerging from external networks.

When a previously unknown packet arrives, the edge node will check for the Service Level Agreement (SLA) and policies stored for the particular MN being contacted. A central bandwidth broker is in charge of the policy management, and once it receives a request from an edge node, it checks its databases for the proper forwarding rules and returns them to the edge node. Adjusting the load created by best-effort traffic is vital.

This method can be used to adjust the load admitted into each service class, if the network is operating with aggregate service classes, and not per-flow, as with RSVP. This can decrease the network load and thus allow for smoother handovers, especially if the traffic belonging to the best-effort class is not consuming all leftover capacity. Therefore, there is enough bandwidth left to support moving terminals.

The access routers should not need to make the primary policing decisions when the arriving load exceeds the capacity of the forward link. If we allow downlink traffic to flood the access network, mobility management schemes are affected. A bandwidth broker could be used to co-ordinate the access network resources and configure the gateways to drop excess traffic.

4.2 Coupling of Micro-mobility and QoS

In order to improve the behaviour of reservation-based QoS, as defined in the Integrated Services architecture [BCS94], in the dynamic micro-mobile environment, the QoS and micro-mobility mechanisms can be coupled to ensure that reservations are installed as soon as possible after a mobility event such as handover. Reservations are installed using a QoS signalling protocol, the most widely adopted of which is RSVP, which will be used in the following discussions as an example of an out-of-band soft state mechanism. In this study we present three levels of coupling over three different micro-mobility architectures: *proxy agent architectures* [CB00][GJP01][MS00b][MP01], *MANET-based scheme*s [OTC00] and *per-host forwarding schemes* [SGCW00][RLT+99][KMTV00]. The three scales of coupling presented for consideration are described on the following sections.

4.2.1 De-coupled

In the de-coupled option, the QoS and micro-mobility mechanisms operate independently of each other and the QoS implementation is not dependent on a particular mobility mechanism. Changes in network topology are handled by the soft-state nature of the reservations.

After a mobility event, the QoS for the traffic stream will be disrupted while until a new reservation is installed via refresh messages between the node where the old route and new route intersect, known as the crossover router (figure 3), to the new access router (NAR). The reservation between the crossover router and the old access router (OAR) cannot be explicitly removed, and must be left to timeout, which is not the most efficient use of network resources. This will occur every time the MN moves AR, which may be many times during one RSVP session, and can lead to poor overall QoS for an application.

These problems are common to all micro-mobility schemes.

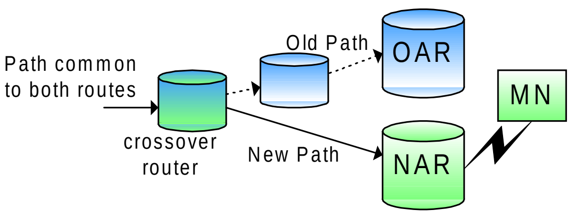

Fig. 3. Concept of a crossover router

4.2.2 Loosely Coupled

The loosely coupled approach uses mobility events to trigger the generation of RSVP messages, which distribute the QoS information along new paths across the network. The RSVP messages can be triggered as soon as the new routing information has been installed in the network. This mechanism is the Local Path Repair option, and is outlined in the RSVP specification [BZB+97] and has the effect of minimising the disruption to the application's traffic streams because there is a potentially shorter delay between handover and reservation set up. It also avoids the problem of trying to install a reservation across the network before the routing update information has been propagated. The latency for installing the reservation can also be reduced by localising the installation to the area of the network affected by the change in topology, i.e. between the crossover router and the NAR. The areas of the network affected by the topology change can have reservations installed across them almost immediately, instead of having to wait for the update to travel end-to-end, or for the correspondent node to generate a refresh message for reservations to the MN. In the case where the QoS must be re-negotiated, however, end-to-end signalling is required. The old reservation should be explicitly removed, freeing up unused resources immediately.

However, the loosely coupled approach requires additional complexity within the inter-mediate network nodes to support the interception and generation of RSVP messages when the router is acting as the crossover node. Another disadvantage is that bursts of RSVP signalling messages are generated after handover to install multiple reservations. This does not happen in the de-coupled case, because the reservation signalling messages are generated when refresh timers expire, not by the same triggering event.

In the **proxy agent architectures** the loosely coupled approach ensures that the reservation is not installed until the registration information generated by the MN has propagated across the network. In **MANET based schemes** and the **per-host forwarding schemes,** the loosely coupled ensures that the new routing information has been distributed into the network before attempting to install the reservation. The reservation is installed in the network as soon as the route to the MN is stable without having to wait until the next timeout to send QoS messages.

4.2.3 Closely Coupled

The closely coupled uses the same signalling mechanism to propagate the mobility and QoS information, either as an extension to the QoS/MM signalling protocol or via a unique QoS-routing protocol. This approach minimizes the disruption to traffic streams after handover by ensuring that the reservation is in place as soon as possible after handover by installing routing and QoS information simultaneously in a localised area. It also provides a means to install multiple reservations using one signalling message. The reservation along the old path can also be explicitly removed.

In the **proxy agent architectures**, support for the opaque transport of QoS information in the registration messages is provided, and is interpreted by the mobility agents. This allows the MN to choose a mobility agent based on the available resources and provides a degree of traffic engineering within the network. In the **MANET-based** and **per-host forwarding schemes**, the messages that install the host-specific routing information in the network also transparently carry opaque QoS information. The reservations are installed at the same time as the routing information, minimizing the disruption to the traffic flows.

4.2.4 Comparison of Approaches

Coupling reservations with micro-mobility mechanisms allow reservation set up delays to be minimised and packet loss reduced. Reservations along the new path can be installed faster because QoS messages can be generated as soon as the new route is established, reducing the disruption to the data flows. Also scalability and overhead are improved because a minor number of update messages are sent or they are localised to only the affected areas of the network. Moreover, it ensures that the request for a QoS reservation only occurs when there are valid routes to the MN in the network.

The closely coupled approach requires support from particular micro-mobility mechanisms so that the opaque QoS information can be conveyed across the network. This has the consequence that the QoS implementation will be specific to a particular micro-mobility mechanism, and extensions to the micro-mobility protocol may be needed to support the required functionality. However, the closely coupled approach maintains consistency between the reservation and the routing information within the network, and can reduce the amount of signalling required to set-up multiple reservations.

The choice between whether to use the loosely coupled approach or the closely coupled approach is a trade-off between a QoS solution that is tied to a micro-mobility protocol and the performance advantage close coupling provides. The closely coupled approach potentially provides improvements in performance and efficiency, but at the expense of additional complexity and loss of independence from the underlying micro-mobility mechanism.

4.3 Advance Reservations

The mobile host may experience wide variations of quality of service due to mobility. When a mobile host performs a handover, the AR in the new cell must take responsibility for allocating sufficient resources in the cell to maintain the QoS requested (if any) by the node. If sufficient resources are not allocated, the QoS needs may not be met, which in turn may result in premature termination of connections.

It is clear that when a node requests some QoS it is requesting it for the entire connection time, regardless of whether it is suffering handoffs or not. The currently proposed reservation protocol in the Internet, RSVP, implements so-called *immediate reservations*, which are requested and granted just when the resources are actually needed. This method is not adequate to make guaranteed reservations for mobile hosts. To obtain mobility independent service guarantees a mobile host needs to make *advance resource reservations* at the multiple locations it may possibly visit during the lifetime of the connection.

There are a number of proposals for advanced reservations in the Internet Community that can be classified into two groups, depending on the techniques they use:

- *Admission control priority*
- *Explicit advanced reservation signalling*

Those groups are not necessarily distinct, as both approaches could be used together. Admission control strategies are transparent to the mechanism using explicit advanced reservations, other than when a request is rejected.

4.3.1 Admission Control Priority

It is widely accepted that a wireless network must give higher priority to a handover connection request than to new connection requests. Terminating an established connection from a node that has just arrived to the cell is less desirable than rejecting a new connection request. *Admission control priority based mechanisms* rely on this topic to provide priorities on the admission control to handover requests without significantly affecting new connection requests.

The basic idea of these admission control strategies is to reserve resources in each cell to deal with future handover requests. The key here is to effectively calculate the amount of bandwidth to be reserved based on the *effective bandwidth* [EM93] of all active connections in a cell and the effective bandwidth of a new connection request.

There are a number of different strategies to do this:

- ***Fixed strategy***: One simple strategy is to reserve a fixed percentage of the AR's capacity for handover connections. If this percentage is high, adequate capacity will most likely be available to maintain the QoS needs of handover connections, but at the expense of rejecting new connections.
- ***Static Strategy***: the threshold values are based on the effective bandwidths of the connection requests. There is a fraction of bandwidth reserved for each of the possibly traffic class. This fraction may be calculated from historic traffic information available to the AR.
- ***Dynamic Strategy***: each AR dynamically adapts the capacity reserved for dealing with handover requests based on connections in the neighbouring cells. This will enable the AR to approximately reserve the actual amount of resources needed for handover requests and thereby accept more new connection requests as compared to in a fixed scheme. Such dynamic strategies are proposed and evaluated in [NS96] and [YL97].
- ***Advanced Dynamic Strategy***: this strategy assumes an analytical model where handover requests may differ in the amount of resources they need to meet their QoS requirements, and therefore it is more suitable for multimedia applications. A proposal for this strategy is described in [RSAK99].

This kind of admission control strategy can be used on statistically access control as the one performed on non hard guaranteed QoS provision, such as some DiffServ PHBs or Controlled Load on IntServ model. It is not enough for hard guarantees in all paths followed by a mobile node.

4.3.2 Explicit Advanced Signalling

Admission Control strategies are not enough to accommodate both mobile hosts that can tolerate variations in QoS and also those that want mobility independent service guarantees in the same network. To obtain good service guarantees in a mobile environment, the mobile host makes resource reservations at all the locations it may visit during the lifetime of the connection. These are known as *advanced reservations*.

There are a number of different approaches for advanced reservation in the literature. We present here two of the most relevant for supporting Integrated Services (MRSVP [TBA98]) and other for supporting Differentiated Services (ITSUMO approach [Chen00]).

MRSVP

Mobile RSVP introduces three service classes to which a mobile user may subscribe: ***Mobility Independent Guarantees*** (MIG) in which a mobile user will receive guaranteed service, ***Mobility Independent Predictive*** (MIP) in which the service received is predictive and ***Mobility Dependent Predictive*** (MDP) in which the service is predictive with high probability.

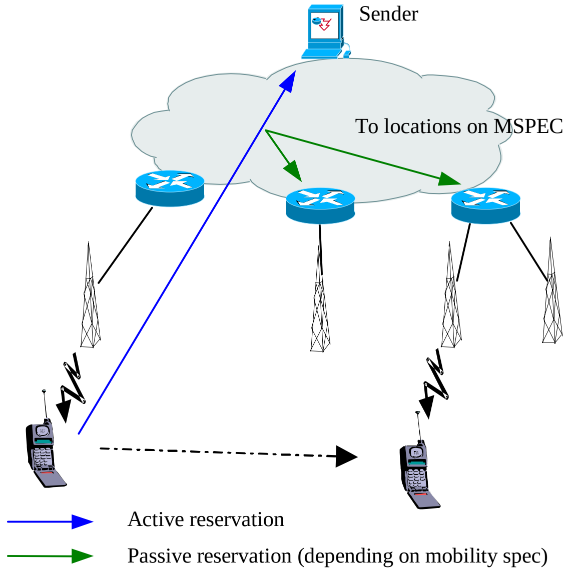

Fig. 4. MRSVP advanced reservations.

MRSVP allows the mobile node to make advance resource reservation along the data flow paths to and from the locations it may visit during the lifetime of the connection. These are specified in the Mobility Specification (MSPEC) as shown in figure 4. The advance determination of the set of locations to be visited by a mobile node is an important research problem, although several mechanisms have been proposed to approximately determine them by the network.

Two types of reservations are supported in MRSVP: active and passive. A mobile sender makes an ***active*** reservation from its current location and it makes ***passive*** reservations from the other locations in its MSPEC. To improve the utilization of the links, bandwidth of passive reservations of a flow can be used by other flows requiring weaker QoS guarantees or best effort service. However, when a passive reservation becomes active (i.e. when the flow of the mobile node who made the passive reservation moves into that link), these flows may be affected.

ITSUMO Approach

The ITSUMO approach has a different philosophy on advanced reservations. Although the mobile node itself has to explicitly request a reservation and specify a mobility profile, the advanced reservation is 'made' by Global QoS Server (GQS) on its behalf. Based on the local information and the mobility pattern maybe negotiated in the SLS, the QGS envisions how much bandwidth should be reserved in each QLN (QoS Local Node). The QGS then updates periodically the QLNs likely to be visited by MN. Rather than actively reserving resources in each of the access points, this scheme it is likely that either a passive reservation (utilized for best effort traffic) or an "handover guard band" could be used.

The clear difference with the previous approach is that advanced reservation in MRSVP has to be signalled by the mobile node explicitly to every station according to its mobility pattern. This mobility pattern is known and processed by it. In the ITSUMO approach this information is updated periodically by the QGS, according to the mobility pattern informed by the MN but processed on the QGS. So it could be said that MN relies the explicit advanced reservation in the QGS (figure 5).

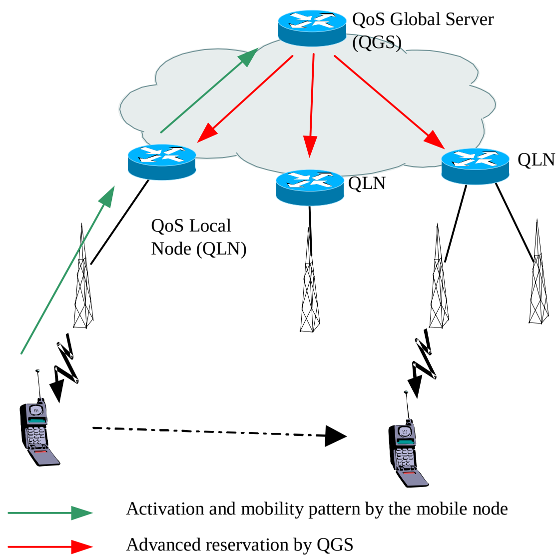

Fig. 5. ITSUMO advanced reservations

4.4 Pre-handover Negotiations

Pre-handover negotiations associate the change to a new cell to the actual resource availability in the new cell, as opposed to advance reservation schemes. When the network or the mobile node deems that a handover should occur, the access router can request some indication of resource availability from neighbouring access routers.

This needs support from the convergence layer between the IP-layer and the link layer. The link layer would need to communicate the overall resource availability of an access point in order to let the IP-layer to make a decision about a possible handover. Also an indication of a forthcoming handover is needed.

Initially, context transfer would enhance handovers between access routers, allowing access routers to communicate directly or through the MN, the QoS and other contexts of a moving MN. A further refinement to the scheme would allow both access router and gateways to communicate the mobile's context during a handover. This would allow to reduce the time during which the mobile has no specific resources allocated to it

4.5 Solutions in Third Generation Mobile Communication Systems

The currently evolving design of the third generation mobile communication systems (3G systems) aims to provide real-time multimedia services in wide area cellular networks [Walk99]. These systems will include a packet switched backbone (PSB) to carry data traffic in the form of IP datagrams, in addition to the traditional circuit switching for voice calls. As the standardization of 3G systems evolves, more and more IETF protocols are incorporated into the architecture. UMTS Release 2000 considers the PSB as an IP backbone using the same protocols as IP fixed networks, while the Radio Access Network (RAN) will use proprietary protocols. For the IP-based data transmission, this RAN is seen as a link layer.

Mobility management and the provision of QoS in 3G systems are still different from IP based fixed networks. Three types of mobility are considered in 3G systems: terminal, personal and service mobility. Service mobility provides the same set of services regardless of the current point of attachment to the 3G network. Personal mobility allows users to receive their personalized service independent of their location in the network. Terminal mobility across different operators is a key requirement in 3G systems. To this end, the support of Mobile IP is being considered with some proposed extensions [Das00]. In essence, the Internet Gateway Serving Node (IGSN) will act as Foreign Agent supporting macro mobility, while the movements of the terminal inside the Universal Terrestrial Radio Access (UTRA) are not visible outside the 3G network. The provision of QoS in 3G systems will incorporate two new features with respect to 2G systems and their evolutions: support for user/application negotiation of UMTS bearer characteristics and standardized mapping from UMTS bearer services to core network QoS mechanisms.

5 Conclusion

In this paper we discussed problems related to mobility and QoS. We deduced that the main problem in this field is following the movement of the mobile host fast enough to minimize the disruption caused to the QoS received by the application traffic flows. Also the depth of the handover signalling and the related QoS control affect the service outcome.

In Section 4 we studied solutions for the interoperability of mobility and QoS. We presented several schemes that provide parts of a total solution to mobile QoS. We discussed performing strict flow shaping at the network edge, coupling of micro-mobility and QoS protocols, advanced reservations, pre-handover negotiations and context transfer, and the 3G approaches.

It has become apparent that even though there exists several good partial solutions, we still need adaptive applications. Handovers, for example, still cause some disturbance to data streams. RTP can provide to this adaptability. The whole notion of end-to-end QoS still seems very distant. It is possible to provide adequate service to mobile hosts in a private access network, but when the corresponding node is behind some wider public network, keeping the promised QoS becomes harder.

A new IETF Working Group, Seamoby, is aiming to provide seamless mobility across access routers and even domains. The work of this group will hopefully lead to better mobility support, especially for the problematic multimedia streams. Part of the work done is on context transfer issues.

References

[AA97] Awduche, D. O., Agu, E., "Mobile Extensions to RSVP", Proceedings of ICCCN'97, Las Vegas, NV, Sep. 1997, p.132-6.

[BBC+98] Blake, S., Black, D., Carlson, M., Davies, E., Wang,, Z., Weiss, W., "An Architecture for Differentiated Services". Internet Engineering Task Force, Request for Comments (RFC) RFC 2475, Dec. 1998.

[BBGS01] Bernet, Y., Blake, S., Grossman, D., Smith, A., "An Informal Management Model for DiffServ Routers". Internet Draft (work in progress), February 2001 (draft-ietf-diffserv-model-06.txt).

[BCS94] Braden, R., Clark, D. Shenker, S., "Integrated Services in the Internet Architecture: an Overview ". Internet Engineering Task Force, Request for Comments (RFC) 1633, June1994.

[Be00] Bernet, Y., "Format of the RSVP DCLASS Object". Internet Engineering Task Force, Request for Comments (RFC) 2996, November 2000.

[Bern00] Bernet, Y. et al, "A Framework For Integrated Services Operation Over Diffserv Networks". Internet Engineering Task Force, Request for Comments (RFC) 2998, November 2000.

[BILD01] Baker, F., Ituralde, C., Le Faucheur, F., Davie, B., "RSVP Reservation Aggregation". Internet Draft (work in progress), April 2001 (draft-ietf-issll-rsvp-aggr-04.txt).

[Bla00] Black, D. "Differentiated Services and Tunnels". Internet Engineering Task Force, Request for Comments (RFC) 2983, October 2000.

[BRAI00] Broadband Radio Access for IP-based Networks, IST-1999-10050, http://www.ist-brain.org.

[BZB+97] Braden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S., "Resource ReSerVation Protocol (RSVP) -- Version 1, Functional Specification". Internet Engineering Task Force, Request for Comments (RFC) 2205, September 1997.

[CB00] Castelluccia, C., Bellier, L., "Hierarchical Mobile IPv6", Internet Draft (work in progress), July 2000 (draft-castelluccia-mobileip-hmipv6-00.txt).

[Chen00] J.C. Chen, A. McAuley, A. Caro, S. Baba, Y. Ohba, P. Ramanathan. A QoS Architecture for Future Wireless IP Networks. In Twelfth IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2000), Las Vegas, NV, November 2000)

[CGK+00] Campbell, A., Gomez, J., Kim, S., Valko, A., Chieh-Yih Wan, Turanyi, Z., "Design, implementation, and evaluation of cellular IP". IEEE Personal Communications, Vol. 7, August 2000, pp. 42-49.

[Das00] Das, S., Misra, A., Agrawal P., Das S.K., "TeleMIP: Telecommunications-Enhanced Mobile IP Architecture for Fast Intradomain Mobility", IEEE Personnal Communications, August 2000

[El01] El-Malki, K., et.al. "Low Latency Handoff in Mobile IPv4". Internet Draft (work in progress), May 2001 (draft-ietf-mobileip-lowlatency-handoffs-v4-01.txt).

[EM93] Elwalid, A. I., and Mitra, D. "Effective bandwidth of general markovian traffic sources and admisssion control of high-speed networks," IEEE/ACM Transactions on Networking, vol. 1, no. 3, pp. 329-343, June 1993.

[EMS00] Eardley, P., Mihailovic, A., Suihko, T., "A Framework for the Evaluation of IP Mobility Protocols", In Pproceedings of PIMRC 2000, London, UK, September 2000.

[GJP01] Gustafsson, E., Jonsson, A., Perkins, C,. "Mobile IP Regional Registration", Internet Draft (work in progress), March 2001 (draft-ietf-mobileip-reg-tunnel-04).

[JP00] Johnson, D., Perkins, C., "Mobility Support in IPv6", Internet Draft (work in progress), November 2000 (draft-ietf-mobileip-ipv6-13.txt).

[KMTV00] Keszei, C., Manner, J., Turányi, Z., Valko, A., "Mobility Management and QoS in BRAIN Access Networks". BRAIN International Workshop, King's College, London, November 2000.

[Lee00] Lee et al., "INSIGNIA: An IP-based quality of service framework for mobile adhoc networks". Journal of parallel and distributed computing, Vol 60 no 4, pp 374-406, April 2000.

[MB97] Mysore, J., Bharghavan, V., "A New Multicasting-based Architecture for Internet Host Mobility", Proceeding of ACM Mobicom, September 1997.

[MH00] Metzler, J., Hauth, S., "An end-to-end usage of the IPv6 flow label". Internet Draft (work in progress), November 2000 (draft-metzler-ipv6-flowlabel-00.txt).

[MHW+99] McCann, P., Hiller, T., Wang, J., Casati, A., Perkins, C., Calhoun, P., "Transparent Hierarchical Mobility Agents (THEMA)", Internet Draft (work in progress), March 1999 (draft-mccann-thema-00.txt).

[Mont01] Montenegro, G., "Reverse Tunnelling for Mobile IP". Internet Engineering Task Force, Request for Comments (RFC) 3024, January 2001.

[MP01] Malinen, J., Perkins, C., "Mobile IPv6 Regional Registrations", Internet Draft (work in progress), March 2001 (draft-malinen-mobileip-regreg6-01.txt).

[MS00] El Malki, K., Soliman, H., "Fast Handoffs in Mobile IPv4", Internet Draft (work in progress), September 2000 (draft-elmalki-mobileip-fast-handoffs-03.txt).

[MS00b] El Malki, K., Soliman, H., "Hierarchical Mobile IPv4/v6 and Fast Handoffs". Internet Draft (work in progress), March 2000, (draft-elmalki-soliman-hmipv4v6-00.txt)

[MSA00] Mihailovic, A., Shabeer, M., Aghvami, A.H., "Multicast for Mobility Protocol (MMP) for emerging Internet networks", In Proceedings of PIMRC2000, London, UK, September 2000.

[NS96] Naghshineh, M. and Schwartz, M. "Distributed call admission control in mobile/wireless networks," IEEE Journal on Selected Areas in Communications, vol. 14, pp. 711-717, May 1996.

[OT01] O'neil, A., Tsirtsis, G., "Edge mobility architecture - routeing and hand-off". British Telecom Technology Journal, Vol. 19, no 1. January 2001. (Available at: http://www.bt.com/ bttj/vol19no1/oneill/oneill.pdf)

[OTC00] O'Neill, G., Corson, S., "Homogeneous Edge Mobility for Heterogeneous Access Technologies", Proceedings of the IPCN 2000, Paris, France, May 2000.

[Per97] Perkins, C., “Mobile IP”, IEEE Communication Magazine, May, 1997.

[Perk00] Perkins, C., “IP Mobility Support for IPv4, revised”, Internet Engineering Task Force, Request for Comments (RFC) 2002, September 2000.

[RLT+99] Ramjee, R., La Porta, T., Thuel, S., Varadhan, K., Wang, S.Y., “HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks”. Proceedings of the Seventh International Conference on Network Protocols (ICNP) 1999, pp. 283 – 292.

[RSAK99] Ramanathan, P., Sivalingam, K. M., Agrawal, P., Kishore, S., “Resource allocation during handoff through dynamic schemes for mobile multimedia wireless networks”. Proceedings of INFOCOM, March 1999

[SBK95] Seshan, S., Balakrishnan H., Katz, R. H., “Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience”, ACM/Baltzer Journal on Wireless Networks, 1995.

[SCEB01] Soliman, H., Castelluccia, C., El-Malki, K., Bellier, L., "Hierarchical MIPv6 mobility management". Internet Draft (work in progress), February 2001 (draft-ietf-mobileip-hmipv6-03.txt).

[SCFJ96] Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V., “RTP: A Transport Protocol for Real-Time Applications”. Internet Engineering Task Force, Request for Comments (RFC) 1889, January 1996.

[SGCW00] Shelby, Z.D, Gatzounas, D., Campbell, A. Wan, C., “Cellular IPv6”. Internet Draft, (work in progress), November 2000 (draft-shelby-seamoby-cellularipv6-00).

[TBA98] Talukdar, A., Badrinath,, B. Acharya, A., “MRSVP: A Resource Reservation Protocol for an Integrated Services Packet Network with Mobile Hosts”, In Proceedings of ACTS Mobile Summit’98, June 1998.

[TKWZ00] Terzis, A., Krawczyck, J., Wroclawski, J., Zhang, L., “RSVP Operation Over IP Tunnels”. Internet Engineering Task Force, Request for Comments (RFC) 2746, January 2000.

[TPL99] Tan, C., Pink, S., Lye, K., “A Fast Handoff Scheme for Wireless Networks”, In Proceedings of the Second ACM International Workshop on Wireless Mobile Multimedia, ACM, August 1999.

[Walk99] Bernhard H. Walke, “Mobile Radio Networks, Networking and protocols”, chapter 5, “Third Generation Cellular”, Ed. JohnWiley , England, Feb. 2000.

[WC00] Wroclawski, J. Charny, A., “Integrated Service Mappings for Differentiated Services Networks”. Internet Draft (work in progress), February 2001 (draft-ietf-issll-ds-map-01.txt). (expired but rewritten as new draft)

[Wroc97] Wroclawski, J., “ The Use of RSVP with IETF Integrated Services”. Internet Engineering Task Force, Request for Comments (RFC) 2210, September 1997.

[YL97] Yu, O. T. W., and Leung, V. C. M., “Adaptive resource allocation for prioritised call admission over an ATM-based wireless PCN,” IEEE Journal on Selected Areas in Communications, vol. 15, pp. 1208-1225, Sept. 19