QoS and micromobility coupling integrated with UMTS

Luis Angel Galindo Sánchez
Network Technology
Telefónica Móviles España
Madrid, Spain
galindo_la@tsm.es

Pedro M. Ruiz
Dept. of Information and Communications Engineering
University of Murcia
Murcia, Spain
pedrom@dif.um.es

***Abstract*— QoS is one of the big challenges Internet protocol designers are facing during the last years. The complexity of the problem is making usual the overprovisioning of the network to avoid bottlenecks. However, in wireless networks in which the spectrum is very expensive such an approach does not seem economically viable and there is a need to offer QoS guarantees. In addition, the adoption of an All-IP network paradigm as the basis for the future wireless networks, is putting a great demand on QoS mechanisms being able to deal with the mobility of the nodes without compromising the QoS they provide. We present an integration of the RSVP protocol with the HMIP micromobility protocol to improve the QoS guarantees which are given in micromobility scenarios. We demonstrate through extensive simulations how this approach clearly outperforms both for TCP and UDP transport the traditional use of RSVP for these scenarios.**

Keywords.- QoS-coupling,, HMIP, Micromobility,UMTS

I. INTRODUCTION

Mobile networks are evolving from traditional circuit-switched architectures to an "All-IP" paradigm. These future mobile networks are witnessed as a high-bandwidth wired core IP network, complemented by a variety of heterogeneous wireless and wired technologies. Mobile terminals will be able to access bandwidth-demanding applications and advanced services while roaming across these different access technologies. This new paradigm poses strong challenges for mobility management and QoS protocols to operate in those networking scenarios.

As a consequence of this, the wireless research community is investigating new technologies to facilitate the wireless broadband access to IP-based networks such as the Internet. These technologies include GPRS, UMTS and emerging WLAN technologies among others.

New mobile users expect a flexible and technology-independent access to new and existing Internet services. Multimedia applications including audio and video streams often require strong QoS guarantees similar to those offered by circuit-switched networks. In order to provide the required QoS guarantees to specific applications on existing packet-switched networks, the network is required to incorporate specific protocols to support QoS. These mechanisms offer a reliable and controlled packet data transport which allows applications to receive the same service guarantees independently of the fixed or mobile nature of the network.

Network-layer QoS protocols such as RSVP[1] and DiffServ[2] were designed within the IETF for IP fixed networks. The use of these protocols in IP-based mobile networks, in which hosts can dynamically change their point of attachment to the fixed network, usually evidences limitations in terms of operation and performance. So, the real challenge is being able to maintain the service level requested by the applications even when the mobile device changes several times its position within the network during the same session, as it happens when using the Mobile IP [3] protocol. Other relevant problems which come up in these scenarios are the variability of the physical medium and the resource scarcity of wireless links. This often leads to QoS violations which are perceived by the applications as packet losses, increased delays and even service disruptions.

Mobility management and QoS support mechanisms have been usually researched as independent threads of work. This has ended up in uncoupled mechanisms producing sub-optimal solutions. The QoS coupling approach, which we describe in this paper, is based on the idea of defining tight interaction mechanisms to make QoS and mobility management protocols to work together towards enhanced QoS support in IP-based mobile networks. The idea is to make QoS mechanisms aware of the changes in the point of attachment of mobile terminals so that they can react accordingly. In this way, the re-establishment of a reservation just after the handover produces a minimal disruption in the application-perceived QoS guarantees.

The remainder of the paper is organized as follows: section II describes well-known QoS and micromobility protocols in IP-based networks. Section III describes our proposal for QoS coupling with the HMIP protocol. Our results in a simulated macro and micromobility UMTS scenario are presented in section IV. Finally, in section V we offer some conclusions and future work.

II. MICROMOBILITY AND QOS

The usage of micromobility protocols in mobile network scenarios improves mobility performance when handovers occur frequently, since it is not necessary to register continuously with the home agent. However, QoS provision remains without solution.

When a handover occurs, QoS is degraded if IntServ is being used, since it is necessary to reserve resources in the new route. However, resources in the route towards the new access point won´t be reserved until a refresh of the reserve state will be done. The mean refresh time used by the Soft State approach used by RSVP is not quick enough (about 30 s.) to avoid possible QoS-violations after the handover.

The concept of coupling consists of implementing a signaling mechanism that enables a protocol to inform other protocol of a special event happened. For example, coupling could be used for the QoS protocol (e.g. RSVP) to be informed by the mobility management protocol when a handover occurs. Thus, in the case of RSVP, it avoids the RSVP agent to wait until the next refresh period to reserve the resources over the new route.

In the case under study, we identify three coupling levels:

- No coupling. The QoS and mobility management protocols work separately without any interaction between them.
- Loose coupling. If an event happens in one protocol then it produces some actions in the other.
- Hard coupling. The information related to both protocols (QoS and micromobility) is transported as a whole, it could be as a protocol extension or as a new protocol managing QoS and micromobility. An example of this is INSIGNIA[4] protocol, which used in ad-hoc networks.

Efficiency, complexity and applicability are the criteria to choose between these coupling options. With no coupling, transparency and independence in the deployment and implementation are guaranteed, but there are no improvements in the performance and efficiency. On the other hand, hard coupling allows optimizing efficiency, but protocol independence is lost and it attacks some of the mainstays of Internet as they are its end-to-end nature and its layering design [5]. As a trade-off to improve the overall efficiency, the selected approach in our study is loose coupling, albeit it has the cost of having some dependency between QoS and mobility management protocols.

The QoS model used in the access network is Integrated Services (i.e. RSVP). Our study is focused in the access network, since in this part of the network, the handovers are managed by the micromobility protocol without the Home Agent (HA) taking part in this process. Thus, we are interested in evaluating the effects of the handovers in the QoS guarantees provided within the access network. We assume the use of the DiffServ approach within the Core Network (CN), with edge routers performing the mapping between the QoS requirements requested by IntServ to the different DiffServ classes available in the core network. This model is commonly known as ISSLL [6].

Based on this ISSLL model, we will use RSVP as the QoS reservation protocol since is supported in both IPv4 and IPv6 and it is designed to work not only with the actual but also future routing protocols, both unicast and multicast. In addition, as the micromobility protocol in the access network we will use Hierarchical Mobile IP (HMIP[7]), which is being developed in the IETF and has been demonstrated to show a good performance.

III. QOS COUPLING FOR HMIP

We propose to use a loose coupling between the QoS and micromobility management protocols. This integration aims at maintaining the QoS guarantees in a mobile environment by notifying the RSVP's local path repair (LPR) about handovers. In addition, with this approach the updates in the reservation only affect to the links to which the route changes, avoiding any extra overhead of processing QoS information end-to-end.

In the loose coupling option, the location changes of the MH, (i.e. the updates of the network routing information) trigger the generation of repair messages (RSVP PATH). This mechanism only repairs the part of the reservation lost, facilitating a quick re-establishment of the reservation. The signaling required for this re-establishment of the resource reservation must not be done until the new route is stable in the network. Thus, in any handover it is expected to have some fixed delay (the delay until the new route is re-established) which can not be reduced and highly depends on the underlying local mobility management protocol.

Combined with the use of the coupling, it is possible to apply other complementary techniques in order to further improve the QoS management during the handover:

- Prioritization of signaling packets. It consists on prioritizing the signaling traffic of the resource reservation protocol given the importance of delivering the signaling packets even when the link is overwhelmed.
- Prioritization of packets in course. It consists on prioritizing packets of a flow, which has lost the reservation as a consequence of a handover. When a handover occurs, packets will be routed through nodes that they have not the reservation established. These packets are called packets in handovers. To overcome that there are extensions to Integrated [8] and Differentiated Services [9] protocols that allow making the reservation before the handover occurs. These reservations are called "in advance". However, this approach requires a planned handover, which is not always possible.
- Context transfer protocols. IETF has created a working group [10] to investigate about a context transfer protocol, this is a protocol that takes care of transferring the QoS information (and other contextual information) from the old access router to

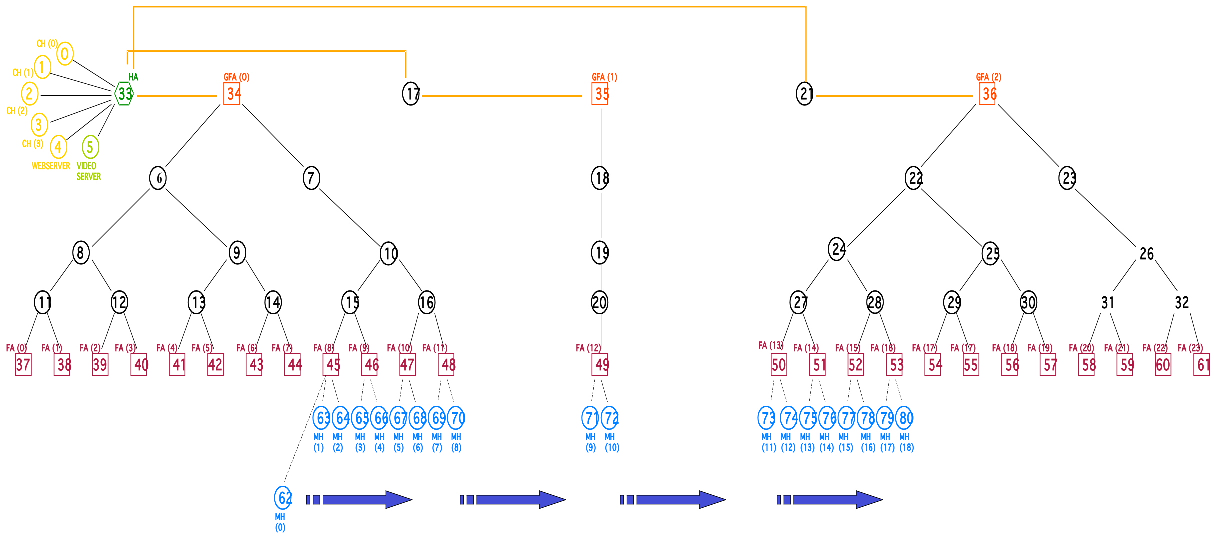

Figure 1. NS modelling of the business park

the new one. This aims, in the case of QoS context transfers, at speeding up the re-establishment time.

For the simulations in our study we use prioritization of the QoS signaling as a complementary technique to the coupling between the mobility management and the QoS protocols. In our studies, routers will have a fixed reservation of 30 kb/s for RSVP signaling traffic.

IV. RESULTS

A. Network scenario

We have assumed a network deployment in a business park. In this scenario there are companies buildings, some nearby and others at a medium far. The suggested density for these types of scenarios is variable, with high density of users inside the company buildings and low in the spaces between buildings. Usually, users do not change their location during the session. In this scenario, the user mobility is low (less than 10 km/h) if the user is walking. It possible to have a vehicle in order to communicate the distant buildings, but the speed is lower than 50 Km/h. It is important to note that although users are not moving constantly, it could occur that many users change from one base station to other simultaneously. In the companies, users are connected through WLAN or TDD. When the user moves toward other building in the park, a handover to a public WAN is done (GPRS/UMTS) and other handover occurs when the user goes into other company building, being that network WLAN or TDD too.

B. Scenario Modeling

Traffic types have been selected according to the four different classes defined in UMTS. Conversational and Streaming classes are very similar in characteristics. The same applies to the Interactive and Background ones. Thus, we have chosen two types of real traffic sources in order to characterize the UMTS classes:

- UDP traffic. It will be characterized according to the parameters of minimum delay and high error rate, representing a typical Conversational application. The small differences between Conversational and Streaming classes allow assimilating both cases, since the results of simulation would be very similar.
- TCP traffic. It will be characterized according to the parameters of high delay and minimum error rate, typical of the Interactive traffic. The similarities of this traffic with the Background traffic allow simulating both classes with TCP without loosing generality.

In order to simulate the UDP traffic, we have chosen videoconference as application. It has been used a file with traces of the ISABEL application got from a H.263 coded video file. The packet size is not uniform, since depends on image quality and the differences between adjoining images. The mean rate of this source is 275 kb/s.

For the TCP traffic, we have selected the FTP file transfer. The FTP arrivals can be modeled as Poisson distribution with a fixed rate per hour. A session is divided in two parts: “connections” and “spaces”. “Connections” represent either data transfer or commands and “spaces” represent the time from the end of a connection to the beginning of the following. “Space” times can be characterized by a lognormal or loglogistic distribution. The size of the connection packets follows a Pareto distribution.

The simulation packet size is 512 bytes and the maximum window size is 20 packets.

The NS-2[11] simulator allows to choose between several types of FTP transmitters and receptors. In our case, we have chosen Sack1 for transmitters and Sack1/DelACK for receptors. This choice enables to send TCPs with slow starting, eliminate the congestion, quick retransmissions, quick recovery, strategy of recovery faced with loses based on Sack and receivers that support ACK delay.

Finally, for interfering traffic we have used constant bit rate traffic (CBR) generated by and UDP agent. The transmission rate is chosen in order to saturate the links in the scenarios where it is needed. The interference UDP traffic is of 450 kb/s for each of interfering mobile nodes.

In the scenario under analysis (see figure 1), the mobile host MH(0) will move from a WLAN/TDD mesh network, which it could represent a company network, to an UMTS network, finally registering in other WLAN/TDD mesh network. In figure 1 it is shown the topology aspect of the network used. The node labeled with 0, CH(0), acts as remote node sending data flows to the different mobile hosts that are registered in the three networks (the UMTS network and the two companies networks). From the 19 mobile hosts, only the MH(0) will move through the three networks, starting in the company 1 (whose GFA is the node 34).

The node 33 (HA) performs the functions of a Mobile IP home agent, forwarding the packets destined for the MH to GFA (external agent that acts as a gateway in the hierarchy of external agents of HMIP). The links 33-34, 33-35 and 33-36 emulate the Internet path. The characteristics of the companies networks links were obtained by experimentation, whereas the UMTS link properties were obtained from tests in UMTS emulators at Ericsson Spain S.A.

The section GFA-GFA corresponds with the links situated between the nodes 33 to 17 and 33 to 21. The section GFA-W corresponds with the links that are placed between nodes 34, 35 and 36 (GFA) with the nodes 11 to 16, 20 and 27 to 30, respectively. The path W-FA corresponds with the links between the previous nodes and the FA nodes (that correspond with the base stations)

C. *Analysis of the results*

The planning of the different events of the simulation is as follows:

- The Mobile Host starts to receive traffic in t=5s. The rest of the nodes start to receive traffic (this is the interfering traffic).
- The simulation starts in t=0s and ends in t=300s.
- The handovers start in t=10s and occur each 40 seconds (t=10,50, 90,130, 170, 210 …) thus this handovers are done with micromobility until t=130, where a handover from the network of the first company to the UMTS network occurs. In t=170, a new handover to the network of the second company is done and from this moment micromobility handovers each 40 seconds will occur. These handovers between companies networks and the UMTS network are macromobility handovers and are managed with Mobile IP.
- The reservation for the data flows which is done in the access network is 300kb/s

For the QoS classes of *Interactive* and *Background*, we have used an FTP source as we described in the previous section. In order to evaluate the pros and cons of our coupling proposal for this type of traffic, we have analyzed the evolution of the TCP window size. This parameter is extremely representative of the sending rate that the application is able to maintain for the different cases under study. Figure 2 shows the evolution of the TCP window size without interference traffic, when there is interference traffic but no reservation, with reservation without coupling and with both reservation and coupling.

As we can see in figure 2, the best result is obtained without interference traffic. This is the ideal case of our study. As it can clearly be observed, the worst case corresponds with the situation of interference traffic without reservation. The two intermediate cases ordered by better results are with reservation and coupling and with reservation without coupling.

When there is no interference traffic, the ideal case, the TCP window is growing as ACK packets are received from the end point. As we can see in each of the handovers, there are packet losses even in this ideal case and, as a consequence, the window size decreases. When there is interference traffic without reserve the TCP window does not grow due to the continuous losses produced by link saturation.

The most interesting case is the comparison between the cases in which there is a reservation with and without coupling. In the case without coupling, it is noticed that the results with bandwidth reservation improve those without reservation. However, we can realize that the window size is lower than the one in the ideal case and the one in the coupling case. The reason for this is that when there is no

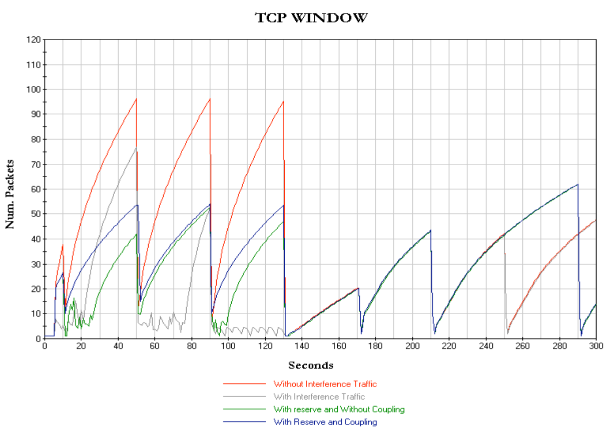

Figure 2. Evolution of the TCP window size

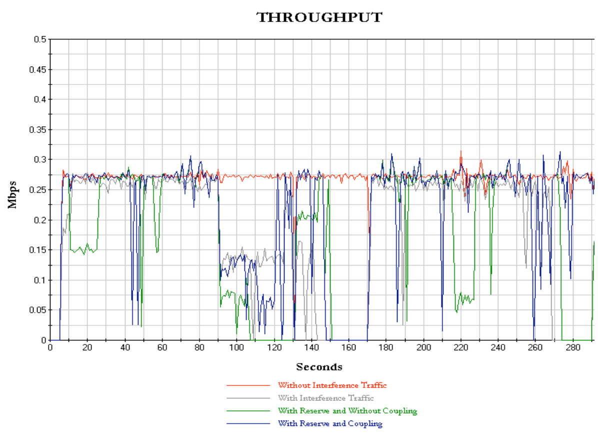

Figure 3. Comparison of videoconference bandwidth

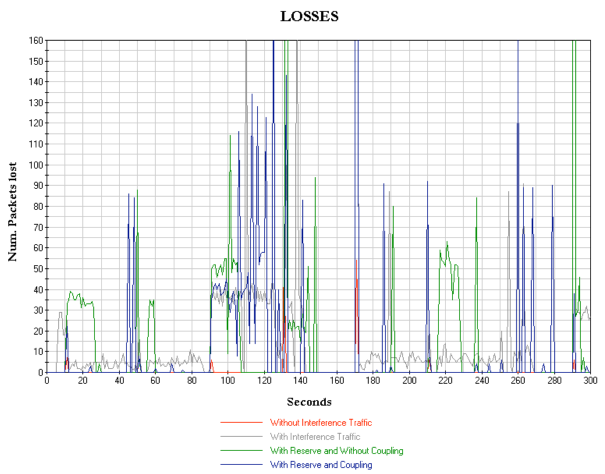

Figure 4. Evolution of videoconference losses

coupling, after the handover there is a period of time until the LPR process starts during the reserve is lost and packets are lost. When we use coupling, as soon as the handover is completed, the reservation is re-established making the window grow immediately.

As we mentioned in the previous section, in order to simulate UDP traffic we have used a file coded in H.263. For the inherent behavior of H.263, data rate is variable in relation with the difference between the picture to be coded and the previous one. The results of the bandwidth received by the MH are shown in figure 3.

As it can be noticed in the figure, the ideal case happens when there is no interference traffic. The variable rate of the codec produces some variation in the instantaneous bandwidth, which it is negligible compared with the medium bandwidth value that is around 275 kb/s.

The most interesting results come from the comparison between the cases with reservation and coupling and those without coupling. In the first case, during the periods without handovers the effective bandwidth is approached enormously to the ideal case. However, just after a handover occurs, reservations are lost and they are not re-established until the LPR process is started. When we use coupling, it can be observed that just after the handover, the reservation recovery is forced and the reservation is re-established almost immediately. Thus, the case with coupling shows a better behavior and the results approach to those of the ideal case.

The difference between using coupling or not is clearly noticed seeing the packet losses that occur after the handovers. As we observe in figure 4, after the handover, the use of reservations without coupling takes several seconds more to reestablish the QoS reservation.

V. CONCLUSIONS AND FUTURE LINES

We have analyzed different ways of cooperation between QoS provision mechanisms and mobility management protocols. We have shown through simulations of the loose-coupling approach between HMIP and RSVP in a scenario that emulates a hybrid deployment of UMTS and a micromobility environment in a company network, that this approach offers an important improvement in the applications performance for different UMTS traffic classes based both on TCP and UDP. The main advantage of the coupling is that, although the handover can not be speeded up, the reservations are reinstalled as soon as the new path is stable. In order to improve still more the reestablishment of the reserve, we have used as a complementary mechanism the prioritization of the signaling packets that has shown to be an effective technique.

As future work, we are studying the dependency between coupling and the mobility management protocols. In particular, we would like to compare this viewpoint for different micromobility protocols such us BCMP [12].

REFERENCES

[1] Braden, R., et. Al. "Resource ReSerVation Protocol (RSVP) -- Version 1, Functional Specification". Internet Engineering Task Force, Request for Comments (RFC) 2205, September 1997.

[2] Braden, R., Clark, D. Shenker, S., "Integrated Services in the Internet Architecture: an Overview ". Internet Engineering Task Force, Request for Comments (RFC) 1633, June 1994.

[3] Perkins, C., "IP Mobility Support". Internet Engineering Task Force, Request for Comments (RFC) 2002, October 1996.

[4] Lee, S.B., Gahng-Seop, A., Zhang, X., and A.T. Campbell, "INSIGNIA: An IP-Based Quality of service Framework for Mobile Ad Hoc Networks", Journal of Parallel and Distributed Computing (Academic Press), Special issue on Wireless and Mobile Computing and Communications, Vol. 60 No.4. p.374-406, April 2000.

[5] B. Carpenter. Architectural Principles of the Internet, RFC 1958, IETF, www.ietf.org, June 1996. rfc1958.txt.

[6] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Worclawski, E. Felstaine. A Framework for Integrated Services Operation over Diffserv Networks. RFC 2998, IETF, www.ietf.org, November 2000. rfc2998.txt.

[7] Soliman, C. Castelluccia, K. ElMalki, L. Bellier, "Hierarchical MIPv6 mobility management". Internet Engineering Task Force Internet Draft (work in progress), July 2001, (draft-ietf-mobileip-hmipv6-05.txt).

[8] A. Tulkdar, B. Badrinath, A. Acharya. MRSVP: A Resource Reservation Protocol for an Integrated Services Packet network with Mobile Hosts. ACTS Mobile Summit'98, June 98.

[9] J-C Chen. QoS Architecture Based on Differentiated Services for Next Generation Wireless IP Networks. Internet draft, IETF, www.ietf.org. July 2000. draft-itsumo-wireless-diffserv-00.txt

[10] IETF Seamoby Working Group. Context Transfer, Handoff Candidate Discovery, and Dormant Mode Host Alerting (seamoby), http://www.ietf.org/html.charters/seamoby-charter.html, 2002

[11] The NIST Network Simulator version 2, http://www.isi.edu/nsnam/ns, 2002.

[12] K. Ceszei, N. Georganopoulos, Z. Turanyi, A. Valko, "Evaluation of the BRAIN Candidate Mobility Management Protocol", Proceeding IST Summit '01, September 2002