Architecture for Providing QoS in an IP-based Mobile Network

Louise Burness[1], Eleanor Hepworth[2], Alberto López[3], Jukka Manner[4]

[1]BTexact Technologies, United Kingdom
[2]Siemens/Roke Manor Research, United Kingdom
[3]Univeristy of Murcia, Spain
[4]University of Helsinki, Finland

Abstract: *This paper describes a baseline architecture for the provisioning of QoS for mobile terminals in the all-IP radio access network defined by the BRAIN project. It describes the motivation for defining the baseline architecture, the problems it is addressing and the challenges faced in its definition. The paper then proposes a series of optional extensions that can be deployed to enhance the performance of QoS mechanisms within the baseline architecture.*

I. Introduction

Research in post-3G system architectures is extending native IP transport all the way to the mobile user, offering true wireless Internet access. Simultaneously, an increase in the diversity of the applications supported by IP networks, from web browsing to bandwidth-intensive multimedia applications, has identified the need for providing differential and guaranteed Quality of Service (QoS) across IP networks. While QoS provisioning in fixed network environments has been studied for some time [Blake98][Bernet00][Braden94][Braden97][Wroc97], the challenging problem of providing QoS in mobile networks is only just starting to be considered [SeaMoby][MobileIP].

The provisioning and maintenance of QoS in the mobile environment poses some unique problems. In addition to the challenges of maintaining fixed network level QoS, handovers between access routers, frequent changes of IP address, and contention for resources between mobile users force QoS mechanisms to deal with re-negotiation and increases the complexity of QoS provisioning. Basic QoS support is required since, typically, the access network is the principal bottleneck of resources, which are limited when compared to the resource availability in the core network. An access network without any QoS support will not be able to adequately support a wide range of services.

Mobile access network deployment is very diverse, with the choice of technologies and network topology dependent on the operator and the services they wish to supply to the users. As a result, a simple and flexible QoS architecture is required that both guarantees inter-operability with external networks, and allows the operator to customise their networks according to their requirements. The architecture should limit the complexity required in mobile nodes, such that only a core subset of QoS techniques need to be supported, which the access network can build on internally where required.

The IST BRAIN [BRAIN00] research project was aimed at developing a mobile wireless IP-based access network. The project considered problems from the application layer down to the air interface. This paper is focused on the network layer, and the provisioning of QoS in mobile environments with maximum flexibility, efficiency and high service levels.

The sections that follow discuss requirements for a mobile-aware network layer QoS architecture and describe the problems faced when trying to develop such architectures. A simple baseline architecture to provide basic QoS functionality within the access network is outlined, and a series of optional extensions that can improve performance aspects of this architecture are then proposed.

II. Why QoS is difficult in Mobile Networks

Unlike fixed network environments, the mobile node can potentially change its point of attachment to the network many times during a session. As well as the disruption during the handover itself, this necessarily leads to changed routing within the access network and possibly other changes in the end-to-end session. In addition, there is at least one hop over a wireless link, which has less predictable properties than a wired link. To enable QoS support in this environment, a number of key interactions need to be considered. These are summarised in Figure 1.

Interaction with Mobility

When the mobile node changes its point of attachment to the network, the path across the access network, and potentially the external network, will change. Depending on the mobility mechanism, the mobile's IP address may change as well. For active application sessions on the mobile, the network should negotiate QoS along the new route as part of the handover procedure. The QoS contract negotiated with the network should be maintained where possible, or re-negotiated if a lower service level that is still suitable for the application can be achieved.

Some mobility mechanisms use tunnels during handover to forward packets between the old and new access routers for the mobile device. This helps eliminate packet loss, but the packets must be allocated an appropriate QoS to ensure that the packets reach the mobile within the confines of the QoS contract. This should be achieved without needlessly allocating network resources, which may be relatively limited in the access network environment, especially at the very edge of the network.

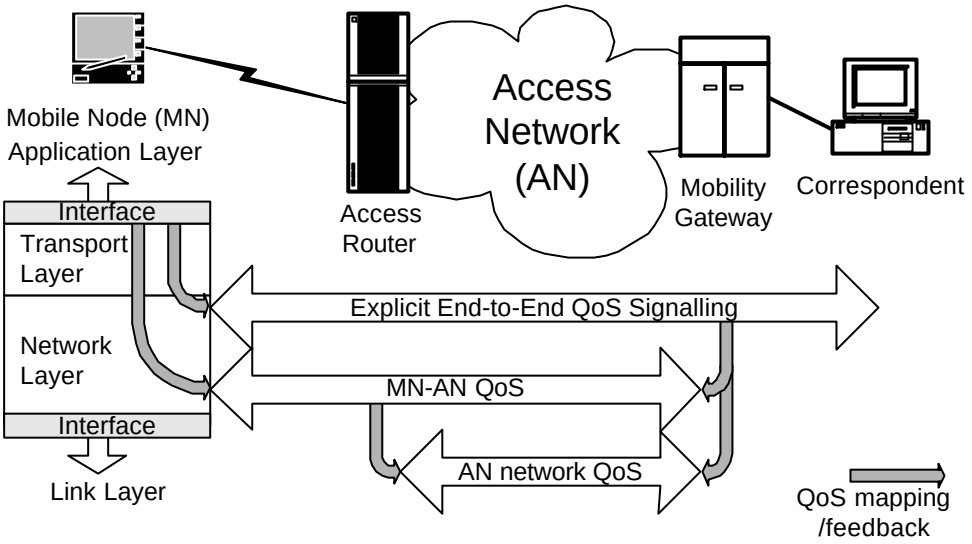

Figure 1: Summary of Interactions

Interaction with Applications and QoS Re-negotiation

The network layer QoS mechanism should provide asynchronous feedback to the application layer about the resource availability in the network. For adaptive applications, such as those that conform to BRENTA [BRAIN00], this feedback allows them to adjust to network conditions and request re-negotiation of network layer QoS contracts if required. Applications may also use implicit transport layer QoS information, for example TCP acknowledgements, to decide whether the QoS provided by the network is adequate. Feedback to applications may be more frequent in mobile compared to fixed environment because the paths across the network are more dynamic, as is resource availability in the current cell, either because other users have moved into the cell, or because the user has moved into a neighbouring cell.

Interaction with Link Layers

On-going developments in wireless link layers, e.g. HIPERLAN/2, have introduced sophisticated capabilities such as QoS support. The network level QoS mechanism must be able to determine the link layer QoS capabilities and to provide the link layer with enough information to co-ordinate resources on the air interface. This information could include link layer significant information such as error rate or transmit delay. The link layer should provide asynchronous feedback regarding the ongoing availability of resources at the air interface, which must be taken into account when requesting resources from the network and providing feedback to applications. The scheduling carried out at each layer must also be co-ordinated to ensure that packets are sent with acceptable delay.

Interaction with External Networks

The access network QoS architecture should interoperate with existing QoS mechanisms used in external networks. There can be no assumption of explicit end-to-end QoS signalling in the current Internet environment, where end-to-end QoS architectures are only partially deployed. The BRAIN

QoS architecture must interwork with external QoS techniques, if present, but not rely on them for the provisioning of QoS within the access network.

III. Baseline BRAIN QoS Architecture

The baseline QoS architecture is required to ensure inter-operability of access network QoS mechanisms with external networks, and to provide a simple, yet flexible means to provision QoS. It also defines and limits the functionality required within the mobile terminal. In order to allow the adaptation of the architecture to suit particular deployment scenarios or operator requirements, the architecture must be easily extensible without affecting the inter-operability of the solution.

The architecture must support the QoS functionality found in fixed networks i.e. admission control, traffic shaping, QoS resource negotiation, application feedback and QoS re-negotiation, but must also include the mobile specific functions described above. Additionally, inline with the BRAIN access network goals, the architecture must be scalable to many, potentially millions, of users and resilient with no single point of failure.

The IETF Integrated Services over Specific Link Layers (ISSLL) working group [Bernet00] has specified a framework for sending RSVP-controlled traffic through a Differentiated Services network, and it is this framework that the baseline architecture uses as its foundation. The framework is a powerful approach to providing per-application resource allocation using IntServ and RSVP [Wroc97] and the aggregation of flows across the network using DiffServ [Blake98]. Per-flow resource information is kept at the network edge nodes, which perform admission control and traffic shaping according the agreed QoS contract. The RSVP messages are exchanged end-to-end and the edge nodes use the information within these messages, and any policy information available within the network, to allocate a suitable DiffServ codepoint (DSCP) for a particular traffic flow. Figure 2 shows a possible deployment scenario of the ISSLL framework within the BRAIN access network (BAN).

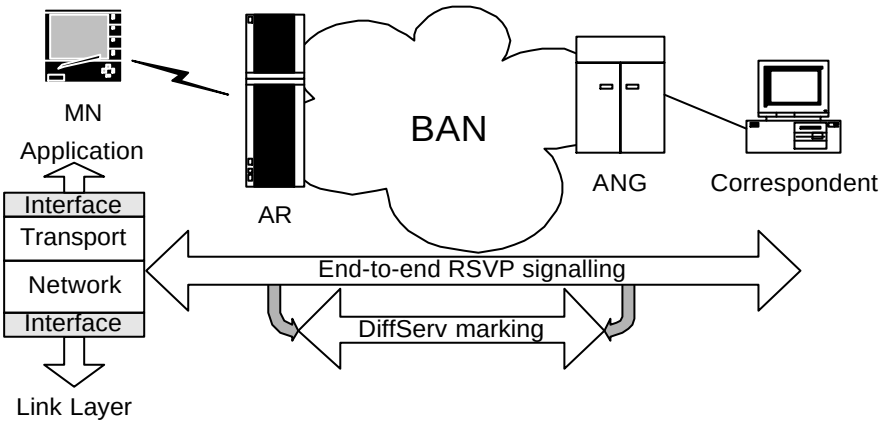

Figure 2: ISSLL deployment in BRAIN

The access router (AR) and the access network gateway (ANG) [Manner01b] are the network edge nodes that interpret the RSVP signalling. As an example, assume that the mobile node wishes to start a voice call with a correspondent node in an external network. Application layer signalling e.g. SIP [Handley99] is exchanged between the mobile and the correspondent nodes and the application passes a request for resources to the network layer. An RSVP message is constructed using this QoS information, which is sent to the correspondent. The access routers and mobility gateways interpret the end-to-end signalling messages, and map the reservations to suitable DSCPs. Once data begins to flow along the reservation, the edge nodes use information in the packet headers to classify and mark the packet with the appropriate DSCP. Apart from interactions with the link layer, which are yet to be specified fully, the operation of the baseline architecture conforms to the standard ISSLL framework.

The reasons for choosing the ISSLL framework as the baseline architecture are as follows:

- Scalability: the architecture minimises the amount of per-flow information that needs to be maintained by the network, confining it to the edge nodes.

- Resilience: there is no single point of failure within the architecture. Unless centralised admission control architectures are deployed, the state information distributed between multiple nodes.
- Extensibility: the use of standard protocols end-to-end and between applications and the network layer means that generic IP applications can be supported. Any mobility specific functionality required within the access network can be made transparent to the higher layer protocols.
- Functionality: the architecture provides mechanisms for QoS negotiation, re-negotiation, admission control and traffic shaping.
- Mobility: using DiffServ within the access network provides basic support for mobility because DiffServ uses per-hop behaviour, which, unlike IntServ, is not affected by routing changes.
- Ease of Deployment: the ISSLL architecture is simple and uses well defined QoS mechanisms that should be supported in existing network equipment.

IV. ISSLL and Mobility/Wireless Interactions

The ISSLL framework was developed for the fixed network environment, so there are aspects of its operation that could be optimised for the wireless and mobile environments. The following section provides an overview of these areas.

The mobile node and access routers may use a wide variety of air interface technologies, some with lower capacities than others e.g. HIPERLAN/2 versus GPRS. Therefore, it is desirable to minimise the signalling across this interface. The baseline architecture uses soft-state RSVP signalling [Braden97] for the communication of the mobile's QoS requirements to the network. The overhead of transmitting periodic refresh messages across the air interface is unsuitable in the GPRS scenario. This implies that modifications to the RSVP signalling model would improve performance in wireless networks.

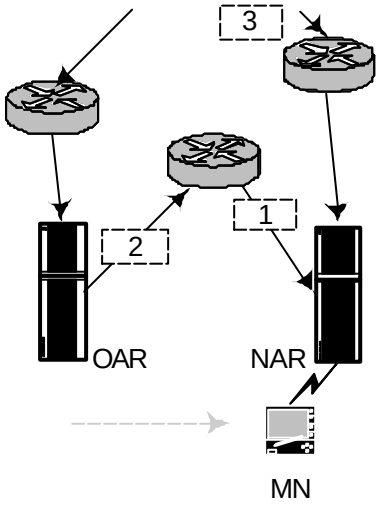

Figure 3: Tunnelling packets during handover

Another issue concerns the minimal support for seamless handover provided by the baseline architecture. According to the BRAIN handover framework [Keszei01], during handover packets are forwarded from the old access router (OAR) to the new access router (NAR) until the routing information about the mobile's new location is installed in the network. For delay sensitive applications e.g. voice, these packets may arrive with unacceptable delay and jitter characteristics. This is demonstrated in Figure 3, where the tunnelled packets 1 and 2 must reach the mobile before packet 3, which is travelling directly to the new access router. A mechanism is required that can improve QoS guarantees during handover.

Admission control and traffic shaping solely at the edge nodes means that the strength of service guarantees is limited. In access networks with no QoS routing or traffic engineering capabilities, traffic may be admitted into the network because overall enough resources are available, but hot spots within the network are not taken into account. The use of certain mobility mechanisms e.g. Cellular IP [Cambell00], introduce complexity when trying to support multiple paths across the network because of the way routing to mobiles is handled with a single explicit per-host route. This implies that a mobility mechanism that can support traffic engineering and a way to re-install reservations quickly after handover is required.

Another consideration is asynchronous feedback to adaptive applications. RSVP-aware nodes provide some level of feedback via the refresh mechanism, but these nodes may only be present at the edge of the network. Since we are concerned with resource availability along dynamic paths within the access network, a more general feedback mechanism is required for DS marked traffic. Also, fixed network QoS parameters do not fully characterise desirable QoS behaviour in mobile and wireless environments. This might lead to unsuitable error correction techniques being used across the wireless interfaces, or unnecessary QoS error reporting. This can result in QoS violations being reported every handover even though the application is able to cope with these transient violations.

Additionally, the baseline architecture assumes that end-to-end RSVP signalling is available, which is not feasible in the current Internet environment. Since it is likely that the access network will have

limited bandwidth compared to core networks, it is essential that the access network QoS mechanism does not rely on the presence of this signalling, but can allocate resources independently.

V. Extensions and Implementations

To allow operators to address some of the issues outlined above, extensions to the baseline architecture have been proposed, some of which are outlined in the following section. Further extensions can be found in [BRAIN00]. These extensions are optional and do not affect the inter-operability of the access network with external networks.

The strength of QoS guarantees is improved by providing admission control within the access network. QoS signalling messages are sent along the data path, and each router determines whether it has sufficient resources to support the new traffic flow. These messages can either be end-to-end RSVP signalling messages, or access network specific signalling messages that have been optimised for the mobile environment. An example of such an optimisation would be the coupling of the QoS and the mobility mechanism. The coupling can be loose, where QoS signalling is triggered by mobility events, or tight, where the QoS and routing information are signalled using the same messages. This ensures that the state information is installed along the new path as soon as possible after handover and simplifies the implementation (compared to having two interacting protocols). Simulation work to study the feasibility of coupling can be found in [Lopez01]. It also localises the signalling required to update the reservation to the area affected by the change in routing updates. Additionally, due to the lack of QoS support from correspondent nodes, mobile nodes need a way to reserve at least the local access network resources, especially across the wireless link. Modifications to the RSVP protocol have been defined that allow initial resource reservations and re-reservations due to terminal mobility to be carried out locally within the access network [Manner01a].

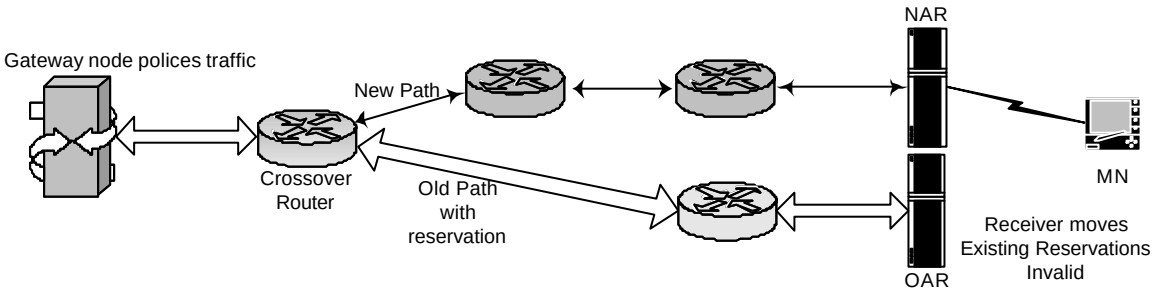

Figure 4: DiffServ handover markings

On handover occurs, there are two areas in the network where reservations for the traffic flow may be unavailable: between the old and new access routers, and between the crossover router and the new access router. This is shown in Figure 4. DiffServ markings are used to trigger special handling of traffic while the network repairs the data path and installs the QoS reservation. This codepoint is reserved for high priority handover traffic to ensure that this traffic is expedited without disrupting other traffic and maintains the QoS for handover traffic without requiring short-lived reservations between access routers. The crossover router must identify that the packet belongs to a session with no reservation. As the gateway-policing node has admitted the packet, the router assumes handover is in progress and remarks the packet with the special codepoint.

To address issues concerning interaction with link layers and the limitations of the IntServ parameters, QoS parameters used within the access network can be enhanced to include mobility-related parameters and wireless specific information. The mobility-related parameters propagate into the access network where they are useful during handover. For example, an application can specify time periods for QoS measurements such that QoS violations are not triggered unnecessarily during handover. The wireless specific information can include loss profiles and bit error rates that are suitable for the application. This is used by the link layer to determine suitable error correction techniques etc. to apply to the traffic flow. This information need not be link layer specific, but can be indicated by generic parameters such as acceptable IP packet loss rate. For example, voice traffic has

quite stringent packet loss requirements but at the same time requires packets to arrive within certain jitter bounds. Any additional delay introduced by the retransmission of packets may be unacceptable. The mobility-enhanced parameters are localised to the access network, and not exchanged end-to-end.

VI. Conclusions and Further Work

In this paper we discussed the issues associated with provisioning QoS in mobile networks and presented the baseline BRAIN QoS architecture. A series of optional extensions were presented that can be deployed within the baseline architecture to enhance the operation of QoS mechanisms.

The proposed extensions improve some aspects of the baseline architecture, but work is still required in this area to address some of the outstanding issues. Further work is needed in the areas of simplifying and enhancing QoS parameters, providing adequate feedback to applications and the performance of QoS mechanisms during handover. Work is also ongoing in the IETF, most notably within the SeaMoby [SeaMoby] working group, to address the issue of providing seamless handover with support for the exchange of QoS information. Enhancements to existing QoS signalling protocols have also been discussed within the IETF.

Follow up work will also be carried out in the IST project MIND to trial the QoS architecture with various extensions and to extend the architecture to more generic access network architectures including ad hoc networks.

Acknowledgements

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 – R&D, INRIA, King's College London, Nokia Corporation, NTT DoCoMo, Sony International (Europe) GmbH, and T-Nova Deutsche Telekom Innovationsgesellschaft mbH.

References

[Blake98] Blake, S., Black, D., Carlson, M., Davies, E., Wang,, Z., Weiss, W., "An Architecture for Differentiated Services". *IETF, RFC 2475*, Dec. 1998.

[Bernet00] Bernet, Y. et al, "A Framework for Integrated Services Operation Over Diffserv Networks". *IETF, RFC 2998*, November 2000.

[BRAIN00] Broadband Radio Access for IP-based Networks, IST-1999-10050, http://www.ist-brain.org

[Braden94] Braden, R., Clark, D., Shenker, S., "Integrated Services in the Internet Architecture: an Overview", *IETF, RFC 1633*, Jume 1994.

[Braden97] Braden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S., "Resource ReSerVation Protocol (RSVP) -- Version 1, Functional Specification". *IETF, RFC 2205*, September 1997.

[Cambell00] Cambell, A. T., Kim, S., Gomez, J.,Wan, C-Y., Turanyi, Z., Valko, A., "Cellular IP", *IETF, <draft-ietf-mobileip-cellularip-00.txt> (work in progress)*, January 2000.

[Handley99] Handley, M., Schulzrinne, H., Schooler, E., Rosenberg, J., "SIP: Session Initiation Protocol", *IETF, RFC 2543*, March 1999

[Keszei01] Keszei, C., Georganopoulos, N., Turanyi, Z., Valko, A., "Evaluation of the BRAIN Candidate Mobility Management Protocol", *IST Mobile Communication Summit*, September 2001.

[Lopez01] López, A., Velayos, H., Manner, J., Villaseñor, N., "Reservation Based QoS Provision for Mobile Environments", *Proc. 1st IEEE Workshop on Services and Applications in the Wireless Public Infrastructure, Evry (Paris) France*, July 2001 (To appear).

[Manner01a] Manner, J., Raatikainen, K., "Extended Quality of Service for Mobile Networks", *9th IEEE/IFIP Workshop on Quality of Service, Karlsruhe Germany*, June 2001 (Published in the Springer LLCS Series).

[Manner01b] Manner, J., Kojo, M., Suihko, T., Eardley, P., Wisely, D., Hancock, R., Georganopoulos, N., "Mobility Related Terminology", *IETF, < draft-manner-seamoby-terms-01.txt> (work in progress)*, March 2001

[MobileIP] The IP Routing for Wireless Mobile Hosts Working Group, http://www.ietf.org/html.charters/mobileip-charter.html.

[SeaMoby] The Context and Micro-Mobility Routing Working Group, http://www.ietf.org/html.charters/seamoby-charter.html.

[Wroc097] Wroclawski, J., "The use of RSVP with IETF Integrated Services", *IETF, RFC 2210*, September 1997.