****

ATM_ Forum/ 97-0341

****

SOURCE: Andrew T. Campbell, Raymond R.-F. Liao
Department of Electrical Engineering and
Center for Telecommunications Research
Columbia University, New York, NY 10027-6699, USA
e-mail: {campbell, liao}@ctr.columbia.edu
web: comet.ctr.columbia.edu/wireless

Yasuro Shobatake
Research and Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan
e-mail: masahata@csl.rdc.toshiba.co.jp

****

TITLE: Delivering Scalable Flows using QoS Controlled Handoff

****

DISTRIBUTION: WATM

****

DATE: April 28 - May 2, 1997 Chicago, Illinois, USA

****

ABSTRACT: In this contribution we argue that large-scale mobility requirements, limited radio resources, and fluctuating network conditions make the delivery of hard QOS guarantees in the wireless domain very difficult. To address this challenge, we are implementing a QOS-aware software mobile middleware platform called *mobiware* which operates between the application and radio ATM link layers, and runs seamlessly on mobile devices, base stations and mobile-capable ATM switches. At the heart of the mobiware platform lies a QOS controlled handoff algorithm which exploits the inherent scalability of audio and video flows during handoff. In this contribution we present an overview of the design, implementation and evaluation of a beta version of the QOS controlled handoff algorithm.

****

NOTICE: This contribution has been prepared to assist the ATM Forum. This document is offered to the Forum as a basis for discussion and is not a binding proposal on Toshiba nor Columbia University. Toshiba and Columbia University specifically reserve the right to add to, amend or modify the statements contained herein.

****

1 Introduction

Currently, the ATM Forum's Wireless ATM (WATM) working group is in the process of developing a set of new specifications for ATM mobility [1]. The work encompasses the definition of a new 25 Mbps air-interface, suitable location management and a fast handoff scheme. To date, most of the working group contributions have focussed on extensions to existing signalling specifications for handoff [2][3][4][5][6] and location management [3]. An open issue which has yet to be addressed by the group is the type of Quality of Service (QOS) that can be sustained over the air-interface during handoff and periods of persistent QOS fluctuation. Implicit in the term "QOS controlled handoff" is the notion that audio and video

flows can be represented as scalable flows and adapted during handoff to meet fluctuating network conditions. The QOS controlled handoff algorithm interacts with media scaling algorithms [10] to adapt audio and video flows to match available bandwidth as a mobile devices handoff between access points. Novel aspects of the mobiware QOS controlled handoff algorithm include the use mobile soft-state and hard-state to represent mobile flows, aggregation techniques for bundling and transporting mobile flows to and from mobile devices, and re-routing and QOS re-negotiation anchor points which limit the impact of small-scale mobility on the wireline network.

In this contribution we describe our model for handoff, the associated algorithms and the experimental setup used to implement signalling support for handoff. We also provide an indication of the initial performance of the middleware implementation of the QOS controlled handoff algorithm.

2 QOS Controlled Handoff Model

An important contribution of our approach is the consideration of two normally conflicting design goals: supporting the dynamic, transient and adaptive nature of device mobility and mobile flows while limiting any potential disturbance incurred by QOS re-negotiation during handoff. A number of goals motivate the design of the QOS controlled handoff scheme. These goals are based on certain assumptions concerning the nature of multimedia flows, providing QOS guarantees in the wireless domains and handoff dropping policy. For full detail of the motivation behind our handoff algorithm see [8].

To achieve these design goals we introduce three novel architectural concepts:

- *mobile soft-state,* models the dynamics of mobility and service adaptation through the periodic re-routing and QOS re-negotiation of flows as a mobile device roams. Mobile soft-state is established between a per mobile *routing anchor point (RAP)* and the mobile device. Hard-state exists between the RAP and the wireline network. In this context the RAP provides an interface between the hard-state and soft-state portions of flows;
- *logical anchor points,* limit the impact of small-scale mobility of a mobile device on the wireline network. Per mobile RAP and *QOS re-negotiation anchor points (QRP)* localize the periodic re-routing and QOS re-negotiation during handoff, respectively. In addition the RAP decouple handoff re-routing from resource allocation in our system; and
- *connection groups (CG),* provide a common routing representation for all virtual paths and virtual circuits *to* and *from* mobile devices. Connection groups are maintained between a per mobile RAP and mobile devices. The RAP allows collective control of all mobile flows associated with a mobile device during handoff. By aggregating flows through a single reference point (i.e., RAP) mobiware supports a fast and efficient handoff algorithm.

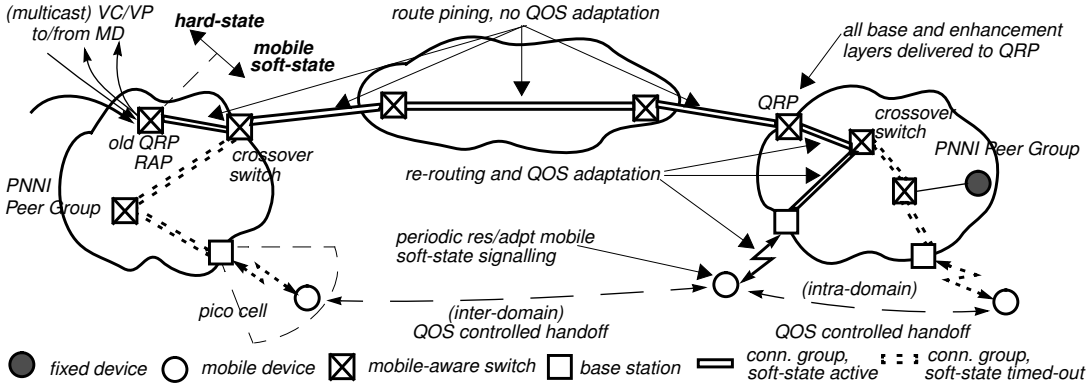

Figure1: QOS Controlled Handoff Model

QOS controlled handoff supports multicast flows where a mobile device can be both a source and receiver. The handoff scheme uses a soft-handoff approach for flows which terminate at a mobile devices and hard-handoff for flows which originated at mobile devices. This approach helps reduce the complexity of managing flows during handoff while supporting QOS needs. The QOS controlled handoff scheme is classified as a forward handoff type. As illustrated in Figure 2 the crossover switch is optimally chosen to be along the path between the mobile device and the RAP. Therefore, the QOS controlled handoff scheme can support a number of different flavors of handoff discussed in the working group.

As illustrated in Figure 1, the concepts of mobile soft-state, connection group and logical anchor points (viz. RAP and QRP) are modeled in the wireline and wireless domain providing QOS controlled mobility. The mobiware platform models the wireless portion of the ATM network as being divided into pico-cells each served by a base station connected to a wireline ATM network. Base stations are cell relays which translate the ATM cell headers from radio ATM format to that used by standard ATM. Each base station supports signalling, QOS control and adaptation management of flows based on the semantics of an adaptive network service. The existing wireline ATM network provides connectivity between base stations. We organize the wireless network into domains. A domain consists of a set of base stations which are under the management of mobile-aware ATM switches. A domain corresponds to a logical partition of the wireline network and the physical location of the base stations in hierarchies for scalable routing (in ATM Forum PNNI routing, these domains would be peer groups).

The QOS controlled handoff can be categorized as forward with partial re-routing which combines soft-handoff for receivers and hard-handoff for senders. It is our view that forward handoff simplifies the signalling and using soft handoff at receiver side allows smooth transition between the flow at the old access point to the flow at the new location. Partial re-routing handoff allows better bandwidth usage especially in case of inter-(routing) domain handoff case. Supporting connection grouping makes handling the complexity of handoff multiple flows to a mobile device relatively simple [8], reduces COS processing and handoff latency.

3 Handoff Algorithms

Mobiware models [7] mobile flows through the semantics of an adaptive network service and QOS controlled mobility. The semantics of the adaptive network service dictate that base layers undergo full end-to-end admission testing and support hard end-to-end QOS guarantees. In contrast, enhancement layers are admitted without any such test but compete for residual capacity between the QRP and the mobile device. Mobiware assumes that network resources are scarce in the WATM domain and abundantly available in the wireline ATM domain. The semantics of QOS controlled mobility assume that as mobile devices migrate between cells then mobile flows are scaled using media scaling [10] to match the available bandwidth in the target cell. To meet these QOS requirements, mobiware models mobile flows through a combination of mobile soft-state, logical anchor points and connection groups.

In the following we provide an overview of the QOS controlled handoff concepts and algorithms; for full details see [8].

3.1 Mobile Soft-State

A significant contribution of the QOS controlled handoff scheme is the use of mobile soft-state. Mobile soft-state is suited to support the dynamics of mobile flows which operate in mobile and QOS fluctuating environments. Mobile soft-state combines self-removable routing state and periodic QOS adaptation between mobile device and the RAPs. Mobile soft-state is designed to model the dynamics of small-scale and large-scale mobility in WATM environments. As illustrated in Figure 2, mobile devices periodically send *reservation (resv) messages* toward the RAP in the wireline network. These *resv* messages carry the mobile devices desired QOS requirement and are interpreted by the base station and fixed network infrastructure according to the rules of the mobiware adaptive network service. As illustrated in Figure 2, resources are allocated to a mobile device over a particular wireless/wireline route for the duration of the mobile soft-state *timeout*. This is achieved when a QRP responds to a *resv* message by sending an *adaptation (adpt) message* back toward the mobile devices with an indication of the allocated

resources. Soft-state provides a receiver-driven QOS adaptation mechanism for the mobile device to periodically probe the wireless network for better quality. This approach is particularly suitable when individual receivers of a multicast flow have differing QOS needs. The mobiware QOS model also operates on very fast traffic control time scales where traffic variations are countered by buffering and scheduling algorithms with resources allocated by a measurement based admission control algorithm.

3.2 Logical Anchor Points

A RAP and QRP are defined for each mobile device. The major goal of mobile-to-QRP signalling is for QOS adaptation using the *resv/adpt* messages pair. The QRP is located in the same local domain as the mobile devices to minimize the impact that small scale mobility has on the wireline portion of the network. The RAP serves as a gateway for a mobile device to the outside network as all the flows to and from this mobile device are routed through the associated RAP and aggregate to and from the mobile device (see section 3.3 on Connection Groups). The RAP's major function is soft-state management, which includes anchoring the connection group, interfacing soft/hard states, consolidating *resv* and generating *adpt* messages.

The QRPs are fixed for intra-domain handoffs and relocated for inter-domain handoffs. In general, the location of RAP remains fixed. The RAP is initially collocated with the QRP, and is fixed for the duration of the handoff to preserve the mobile soft-state connection group path as a basis for future handoff. During handoff the location of the crossover switch has to be chosen between the RAP and the mobile device. By extending the soft-state connection group path the handoff re-routing algorithm can select the optimal location of the crossover switch, as presented in Figure 1 which illustrates a zigzag handoff which is common in the pico-cellular environment.

As illustrated in Figure 1, originally the RAP and QRP are collocated in the same domain. After the first handoff, the QRP is located to the new domain while the RAP stays in the same location to maintain the existing connection group. The mobile device then migrates back to the old cell. The crossover switch is optimally chosen to be near the new location of the mobile device. The mechanism for locating QRP and RAP is simplistic. During the first connection setup, with the knowledge of the pre-assigned location of the QRP (queried from PNNI peer group leader), the soft-state is set up between the mobile and the QRP/RAP. During handoff, a *handoffSetup* message will also set up the soft-state between the new QRP and RAP

The RAP and QRP handle small-scale mobility and inter-domain handoff re-routing. However, suboptimal routing can still occur. This develops when a mobile device migrates further away from its initial RAP location during multiple inter-domain handoffs. To resolve this mobiware executes a path optimization algorithm. The path optimization algorithm is conservative and executes on longer time scales than handoff. The reasoning behind this conservative approach is that the path optimization requires a complete rebuild of all the flows to and from the mobile devices. This will, by its nature, be quite disruptive to all flows in operation. Hence the need to limit total rebuilds to those times when it is absolutely necessary.

3.3 Connection Group

QOS controlled handoff uses the concept of a *connection group (CG),* which is analogous to virtual paths in ATM, to aggregate all flows between the mobile devices and the RAPs. Mobiware extends the connection bundling concept first introduced in [5] [7] from a VP to a higher level of aggregation allowing VC/VPs with different QOS requirements to be grouped together and transported. The connection group concept promotes the idea of supporting individual VC and VP with different QOS requirements at ATM switches and base stations.

The connection group is maintained by the *resv/adpt* mechanism. The *resv* message probes the path between the mobile device to the RAP to detect any individual VC/VP setup or timed-out/torn-down. On the return path, the *adpt* message updates connection group mapping tables in all the intermediate switches and the base station. During handoff, a new portion of connection group is set up using the multicast operation *addBranch*. After handoff, the removal of the old branch from a connection group tree is managed automatically through the soft-state operations used in the mobile environment.

4 Handoff Walk-through

In this section we provide a step-wise walk-through of the QOS controlled handoff algorithm. Each phase of the handoff is considered in turn. These phases comprise the *handoff initiation*, *path setup* and *path teardown* phases. Figure 2 presents each phase and it respective constituent parts.

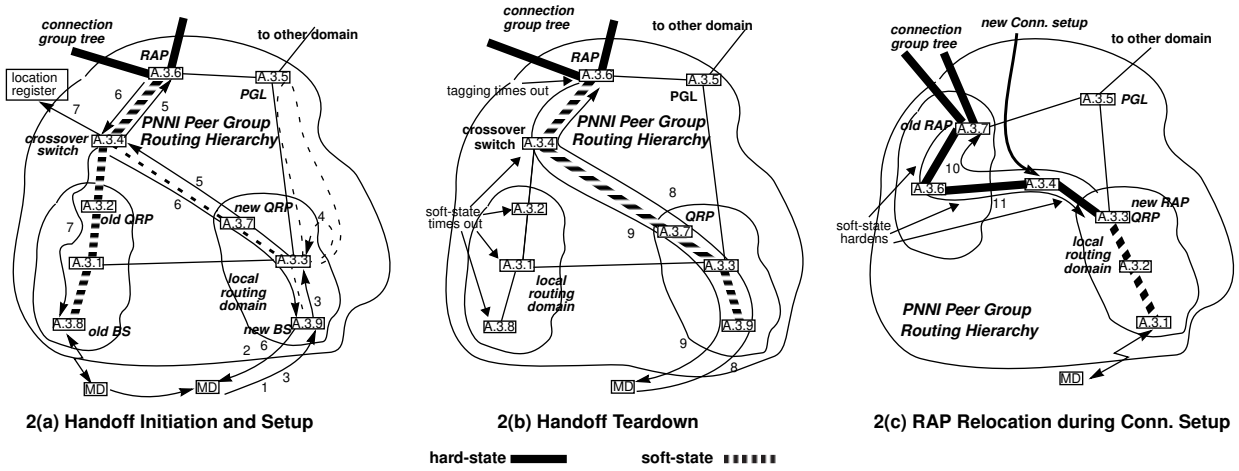

Figure 2: Handoff Walk-through

4.1 Handoff Initiation Phase

The first phase of QOS controlled handoff determines whether a new base station can provide a stronger signal at the desired level of QOS. A QOS monitoring algorithm resident at mobile device monitors the beacon messages which are periodically broadcast by all neighboring base station. In addition to indicating the signal strength of neighboring cells, beacon messages indicate the residual capacity currently available at the neighboring base stations. By periodically monitoring beacons from all neighboring base stations the mobile device is able to determine link qualities and occupancy of adjacent base stations and use this QOS state information as a basis to initiate handoff after a suitable dwell time.

The next step in handoff is the establishment of a new signalling channel between the mobile device and the new base station. A mobile device issues a *(1) signalRequest* to the new base station over a dedicated *meta signalling* channel using the base station address found in the beacon message. This results in the creation of a signalling channel when the base station responds with a *(2) signalResponse*. The signalling channel carries all signalling and *resv, adpt* messages between the mobile and wireline network.

4.2 Path Setup Phase

Once the signalling channel has been successfully created the mobile device initiates a forward handoff to the new base station. It does so by issuing a (3) *handoffSetup* message which includes connection group route state information and the desired QOS required. Mobile devices express their desired QOS in terms of the semantics of the adaptive network service which are mapped into connection group QOS requirements in terms of connection group base layer requirements and enhancement layer requirements, respectively. Admission control located at the new base station first determines whether sufficient resources are available to support the minimum QOS of the requested handoff.

The base station's parent mobile switch uses connection group routing information in combination with PNNI routing to (4)*locateCrossover* within the range of the old mobile soft-state path and establish the shortest branch between the new base station and the crossover switch on the existing connection group tree with the desired QOS. The PNNI peer group leader is interrogated in the case where the mobile device roams outside the current peer group domain of the old base station. This suggests that

the existing crossover point would provide a suboptimal routing point. An optimal re-routing point could to be established by end-to-end hierarchical routing algorithm.

The mobile switch then issues a (5)*handoffSetup* message to the newly identified crossover switch. The QOS values in *handoffSetup* message can be modified by the distributed handoff algorithm based on the availability of resources at the traversed nodes as the *handoffSetup* messages is forwarded toward the RAP. The returned *(6)handoffAck* message commits switch and base station resources for the new connection group branch. The semantics of the adaptive network service provides hard guarantees to the base layers and admits enhancements layers based on the availability of residual resources. Existing mobiles' reservations *above* the base layer may be altered to allow a new mobile device to enter a cell. The base station drops the handoff call (6)*handoffDrop* if insufficient residual capacity is available to meet the group connection base layer resource requirements of the new mobile device.

The handoff algorithm interacts with the adaptive transport *media scaling agents* at either the base station or the mobile-aware switches during connection group setup. In the case of media scaling, the level of media scaling is dependent on the current utilization of the wireless link, application specific desired QOS and the semantics of the adaptive network service. Media scaling may result in the "scaling-down" of mobile delivered quality when a mobile device enters a pico-cell and "scaling-up" when they leave. QOS filters support the delivery of different combinations of layers to particular mobile devices based on the available resources, which is essential for multicast to heterogeneous receivers. For full details on mobiware media scaling see [10].

Handoff registers new mobile devices with the domain location management which in turn allocates a new proxy ATM address as mobile devices roam into cells within a new PNNI Peer Group. A (7) *locationUpdate* message is used to register the mobile devices at the home location in this case and update the cache register at the old base station.

4.3 Path Teardown

QOS controlled handoff is based the notion of soft-handoff for roaming receivers and hard-handoff for roaming senders. In the case of soft-handoff, the mobile device simultaneously receives flows from the old and new base stations. Once the crossover switch sends a *handoffAck* message to the mobile device it begins to forward the new flow to the mobile devices. This results in duplicate cells arriving at the mobile device from both the old and new base stations. The crossover switch uses cell tagging to preserve ATM cell level sequence integrity at the mobile device during handoff. This is achieved by inserting an end-to-end OAM cell for both flows for every K data cells.

After tagging has commenced mobile devices determine suitable synchronization points in the old and new flows and initiate flow switching from the old to the new flow. After flow switching the old flow is rendered redundant. Old flows continue to arrive at the mobile devices as long as the route between the old base station and mobile switching point is active; that is, old flows are switched through to the mobile while the mobile soft-state is still installed and has not timed-out. Mobile devices do not, however, have to process old flows. One advantage of this mobile soft-state tear-down approach is to allow path re-use if a mobile device comes back before the path states time-out. This is equivalent to pre-reserving network resource following the tracking of a mobile device, which can speed up handoff in a pico-cell environment.

The mobile device at the new cell starts its QOS adaptation and mobile soft-state management procedure by sending (8)*resv* message toward its QRP and RAP. The resv messages probes for higher quality enhancement layers and consolidates connection group to VP/VC mapping between the mobile device and QRP point. The resv message is also used to consolidate the connection group mapping between the QRP and RAP point. Note that there is no QOS adaptation between these points. The full complement of connection group base and enhancement layers are delivered to the QRP. QOS adaptation is active from the QRPs to the mobile devices. The RAP responds to a *resv* by sending an adpt messages back toward the mobile. This message confirms the connection group mapping, refreshes the mobile soft-state between the RAP and QRP. In addition, the adpt message also informs the mobile device of the resources available to it over the next interval and scales the available resources for the mobile device. Media scaling filters the flow at the base station or at the QRP.

After flow switching the mobile device refrains from sending any further periodic *resv* messages to the old base station and closes the signalling channel to the old base station. Once the mobile soft-state timer expires the old branch of the connection group tree between the old mobile switching point and the old base station is timed-out and removed. Media scaling is once again invoked at the old base station to determine if deallocated resources can be utilized by the any existing mobile devices at old base station. If the *resv* and *adpt* message pairs of these mobile devices confirm the residual bandwidth availability, the media scaling agent scales up the corresponding media flows.

4.4 RAP Relocation Optimization

As illustrated in 4c mobiware avoids potential sub-optimal routing to the RAP for new calls. A RAP relocation feature allows the routing algorithm to route a new connection to a merge point between the current RAP and QRP point as oppose directly to the RAP. RAP relocation optimization is illustrated in 4c. This procedure does not require any re-routing of existing flows from the old RAP location to the new one. All that is required is that the RAP is relocated and the soft-state between the new and old RAP points in hardened (10) and (11).

5 Experimental Setup and Evaluation

The mobiware testbed (illustrated on Figure 3) consists of 4 ATM switches (viz. ATML Virata, Fore ASX100/ASX200s, NEC Model 5, Scorpio Stinger) and 4 base stations. The base stations include multihomed 200 MHz Pentium and wavePOINTs with ATM 25 Mbps wireline access to the wireline ATM network. The air-interface is based on 2 Mbps WaveLAN air-interfaces to a number of mobile devices based on Pentium PCs and notebooks running MPEG decoders. The data path supports the transmission and delivery of ATM cells end-to-end. The PCs run Linux, Windows/NT and **xbind** (based on CORBA technology). Mobiware platform runs on mobile devices, base stations and the ASX100 ATM, ATML and NEC mobile capable switches supporting the QOS controlled handoff algorithm in all network nodes.

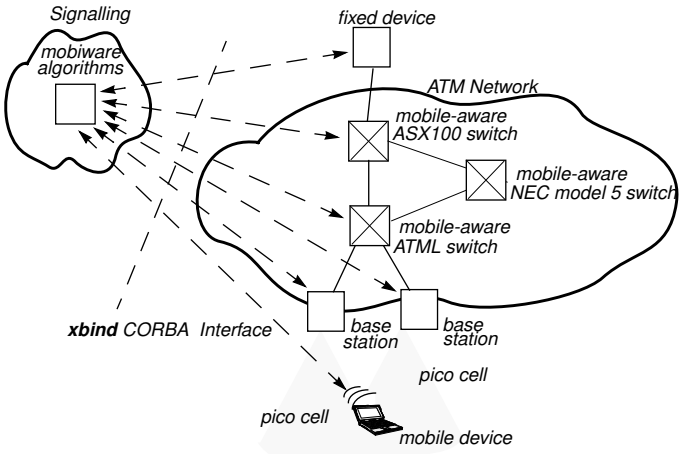

Figure 3: Mobiware Testbed

Currently we have implemented the handoff protocol as part of the mobiware middleware platform augmenting our existing switch software to support mobile-capable operations required for QOS controlled handoff. This includes supports mobile soft-state, connection groups and RAPs and QRPs. In addition we have implemented two types of media scaling algorithms that run at the access points and scale audio and video flows to the available bandwidth. For full details of these media scaling algorithms and performance of these algorithms see [10]

In the following we present some initial performance evaluation of the handoff signalliing system to support QOS controlled handoff using the experimental test bed described in Figure 3.

Table 1 shows the measurements associated with QOS controlled handoff signalling:

- *handoff delay*: this is the delay measured at the mobile device between issuing a *handoffSetup* and receiving *handoffSetupAck.* This includes establishing the cross over switch, extending the existing connection group from the cross over switch to the mobile device through the new base station;
- *soft-state refresh delay*: this is time taken to send a *resv* message from the mobile device to the RAP point(which for the measurements below is situated at the ASX 100 switch) and to receive an *adpt* message back from the RAP;
- *per node soft state teardown delay*: this is the time taken from the time out of the soft state for a connection to clearing the switch table; and
- *hard state teardown delay:* for comparison with soft state tear down we implemented a hop by hop hard state teardown; this is the delay measured at the mobile device between issuing a *teardownConnectionGroup* to receiving a *teardownConnectionGroup*Ack .

TABLE 1. Signalling Message Delay Measurement

measured delay	handoff	soft-state refresh	per-node soft-state teardown	hard state teardown
Mean (msec)	153	18	12	132
95 Percentile Confidence Interval	(125, 181)	(16, 20)	(10, 15)	(129, 158)

The initial results of the middleware implementation look promising. A caveat to these measurements is that they are hot of the press and will be refined once we understand more about delays in the system. One point is that these results show that the overhead due to soft-state maintenance doesn't significantly increase the overall signalling latency.

6 Conclusion

In this paper we have argued that large-scale mobility requirements, limited radio resources, and fluctuation network conditions make the provision of hard QOS guarantees in the wireless domain rather difficult. Currently we are implementing a QOS-aware middleware platform called mobiware and its QOS controlled handoff algorithm which exploits the inherent scalability of audio and video flows. Novel aspects of our QOS controlled handoff include the use of soft-state and hard-state to represent mobile flows, aggregation of flows to and from mobile devices and RAP and QRP which limit the impact of small-scale mobility on the wireline network.

References

[1] Raychaudhuri, D., (NEC USA), Dellaverson, L., (Motorola), Umehira, M., (NTT Wireless Systems), Mikkonen, J., (Nokia Mobile Phones), Phipps, T., (Symbionics), Porter, J., (Olivetti Research), Lind, C., (Telia Research) and Suzuki, H., (NEC C&C Research), "Scope and Work Plan for Proposed Wireless ATM Working Group", ATM Forum Technical Committee, ATM Forum/96-0530/PLEN, April, 1996.

[2] Rauhala, K.,"Requirements of Mobile ATM Handover", ATM Forum/96-989, August 19-23, 1996

[3] Acharya, A., et al,"Primitives for Location Management and Handoff in Mobile ATM Networks" ATM Forum/96-1121, August 19-23, 1996.

[4] Veerarghavan, Karol M. and Eng K. Y., "A Combined Handoff Scheme for Mobile ATM Networks, ATM Forum/96-1700, December 2-6, 1996

[5] Porter, J. and Gilmurray D., *Tunnelled Signalling for the Support of Mobile ATM*, ATM Forum Technical Committee, ATM Forum/96-1699, Dec., 1996.

[6] Deane, J., *WATM Interim Mobility Protocol (WIMP) Protocol*, ATM Forum Contribution 96-1118, August 1996.

[7] Campbell, A., *Mobiware: QOS-aware Middleware for Mobile Multimedia Communications,* Proc. IFIP 7th Intl. Conf. on High Performance Networking, White Plains, New York, April 1997.

[8] Campbell, A.T., Liao, R-F. and Shobatake Y., "*Supporting QOS Controlled Handoff in Mobiware*", 6th Winlab Workshop on 3rd Generation Wireless Information Networks, New Brunswick, March 20-21, 1997.

[9] comet.ctr.columbia.edu/wireless: Wireless Media Systems Project at Columbia University.

[10] Balachandran A., Campbell A.T. and Kounavis M., "*Active Filters: Delivering Scaled Media to Mobile Devices*", Proc. 7th International Workshop on Network and Operating System Support for Digital Audio and Video, St Louis, May 1997.

[11] **xbind** Project at Columbia University, URL: http://comet.ctr.columbia.edu/xbind/