

A Home-Proxy Based Wireless Internet Framework in Supporting Mobility and Roaming of Real-Time Services

Jonathan CHAN[†], Björn LANDFELDT^{††*}, Ren LIU^{†††},
and Aruna SENEVIRATNE[†], *Nonmembers*

SUMMARY Due to the lack of support in mobility and quality of service, today's IP-based networks have some inherent limitations for delivering multimedia services in a mobile environment. In the past few years, these issues have been addressed in the research community and the resulting techniques are being standardised. However, these developments have been done in isolation and become incompatible with each other. In addition to these technical issues, the future infrastructure for charging and accounting mobile multimedia services is expected to be increasingly complicated. In this paper we present a home-proxy based framework which can facilitate the integration of mobility support and QoS management. Furthermore, it enables centralised accounting, which simplifies the cost recovery processes of roaming services. To prove the viability of our design, we have built and tested this framework on a DiffServ wide-area backbone using an MP3 streaming application.

key words: *wireless Internet, mobility management, mobile QoS, charging and accounting, real-time services*

1. Introduction

In the past few years, we have seen a rapid growth in cellular mobile telecommunications and Internet penetration. The natural evolution of these technologies is towards a wireless Internet, which will provide access not only to voice and short message services, but also to all information services from anywhere at anytime. Because of its popularity and the recent advances in wireless network technologies, the provision of wireless Internet services is becoming increasingly profitable. We can expect to have more small-scale Internet service providers (ISPs), offering regional services in large public areas such as railway stations and shopping centres.

Nevertheless, today's IP-based networks have some inherent limitations for realising the above vision. Firstly, the traditional Internet protocol suite is not capable of supporting mobility because Internet routing uses an IP address to identify the point of attachment

of a device. Secondly, the best effort nature of the Internet services, together with the volatility of the radio propagation characteristic have made it impossible for these networks to provide the quality of service (QoS) guarantees necessary to deliver any real-time service. Finally, traditional charging and accounting models are based on the assumption of a single service, best effort, and a static networking environment. Therefore, they have to be altered to suit the needs of wireless Internet services.

In the past few years, some of these issues have been addressed by the research community and the resulting techniques are being standardised. The IETF has adopted Mobile IP [1] to handle mobility in Internet environments. Moreover, there has been considerable progress in developing techniques for providing QoS guarantees in the Internet [2], [3]. In addition, the AAA and ROAMOPS working groups in the IETF are developing requirements for authentication, authorisation and accounting (AAA) as applied to network access. In today's Internet, AAA implementations such as RADIUS [4] and DIAMETER [5] have been deployed for roaming services.

However, until recently these developments have occurred largely in isolation. Consequently, the mobility management and QoS management techniques that are adopted currently are incompatible, and do not provide an integrated solution to the problem of supporting real-time services. Moreover, the charging and accounting methods for a roaming service are concerned mainly with Internet dial-up access. Hence, they do not fully address the issues associated with device mobility and multiple service classes in the Internet.

The objective of this paper is to describe a framework which provides an integrated solution to the above issues. The focus of our proposal is to use an application level proxy instead of a network agent on the home network. We call this a home proxy (HP). We will demonstrate in this paper that this technique can facilitate the integration of mobility support with QoS management mechanisms. Furthermore, it enables centralised accounting, which simplifies the cost recovery processes between administrative domains.

The rest of the paper is organised as follows. In Sect. 2, we give an overview of related work in these areas. In Sect. 3, we propose our home-proxy based

Manuscript received August 25, 2000.

[†]The authors are with the School of Electrical Engineering and Telecommunications, the University of New South Wales, Australia.

^{††}The author was with the School of Electrical Engineering and Telecommunications, the University of New South Wales, Australia.

^{†††}The author is with CSIRO Telecommunications and Industrial Physics, Australia.

*Presently, with the Ericsson Research, Sweden.

framework in details. In Sect. 4, we describe our experimental testbed that consists of a DiffServ wide-area backbone and a foreign administrative domain with wireless access capability. To prove the viability of our design, we present the performance of an MPEG layer 3 audio (MP3) streaming application under the influence of user mobility and network congestion in this framework. Finally, we provide a summary in Sect. 5.

2. Related Work and Challenges

The research community has recently started to address the problems associated with integration of the numerous techniques that have been designed for mobility and QoS management in Internet environments. This work is particularly important for supporting real-time services in wireless Internet, as they require more than the notion of simply forwarding packets to the mobile host (MH).

To understand the issues associated with integrating QoS support into mobility management frameworks, consider the provision of QoS capable services in the wireless Internet using Mobile IP, the current proposal for managing mobility in Internet environments.

A MH using Mobile IP can move between different subnets while maintaining on-going connections without changing its IP address. This transparency is achieved through the co-operation of two mobility agents. At the MH's home network, a home agent (HA) is responsible for capturing packets intended for the MH and tunnelling them to a foreign agent (FA) at the subnet the MH is visiting. As the MH moves away from its home network, it requires the MH to obtain a care-of-address (COA) from this foreign network and informing the HA. Through this address binding refresh process, packets can be tunneled correctly from the home network to the current subnet visited by the MH.

2.1 Issues with Mobile IP

Although Mobile IP is a simple and scalable solution for IP mobility, it suffers from performance and security problems, and has a number of drawbacks especially when serving users with high mobility and QoS expectations. Currently there are many enhancements being proposed to Mobile IP. They include:

2.1.1 Route Optimisation

Since all packets destined to a MH have to be routed through its HA, the chosen path can be significantly longer than the direct route. In a QoS supported mobile network, a longer path and delay will produce a higher probability of call dropping and service refusal.

The route optimisation extension of Mobile IP [6] suggests a means for corresponding hosts (CH) to cache

the location of a MH, so that their packets can be tunneled to the MH directly. An alternative approach is to introduce the concept of a location register, through which the MH can update its current location. Then the CH can query the location register to obtain the location of a mobile user before transmitting a packet [7]–[9]. However, in order to enable optimal routing, all the above proposals need to introduce mobility extensions in the CHs. Unfortunately, this is not available today nor will it be available in the near future.

2.1.2 Fast Handover and Location Updates

Mobile IP does not mandate fast handover control and real-time location updates. If the visited network is some distance away from the home network, the large delay in re-registration can cause many misrouted packets. Another disadvantage is that when inter-subnet handovers occur, the MH obtains a new COA and thus packets destined to the MH will be encapsulated with a different tunnel header. Even if we assume that network resources can be dynamically reallocated across a QoS capable Internet backbone during handovers, this rerouting process may take a long time when the HA and the FA are far apart.

The micro-mobility concept is a promising approach to manage efficiently user mobility within a single administrative region. Firstly, location updates within a region are handled locally, thus avoiding frequent re-registrations across the Internet to the HA. Secondly, by using a specific packet delivery scheme within the region, it prevents the costly re-establishment of end-to-end routing between the HA and the FA. This regional framework is normally coordinated by a domain gateway, which serves as the interchange entity for mobility management within a domain (micro-mobility) and mobility management across different domains (macro-mobility). So far, only Mobile IP has been considered as the solution for macro-mobility management. For micro-mobility management, however, many types of regional network architecture have been proposed and they can be divided into four categories: cascade of tunnelling (e.g. hierarchy of foreign agents (HFA) [10]); dynamic per-host routing (e.g. HAWAII [11], Cellular IP [12]); overlay routing via an alternative technology (e.g. IP over Mobile ATM [13]); and dual care-of-addresses assignment (e.g. TeleMIP [14]).

It is unclear which of the above is the most appropriate for micro-mobility management. For instance, management schemes like HFA are much closer to the original approach of Mobile IP. However, it can be argued that they are inefficient because packets need to be encapsulated and de-encapsulated multiple times while traversing the hierarchy. Other techniques such as Cellular IP, IP over Mobile ATM, and TeleMIP could be more suitable for IP-based mobile environments. How-

ever, the dynamics of re-routing depend on routers or ATM switches being mobility capable, and the dual assignment of care-of-addresses requires modifications in both FA and the Dynamic Host Configuration Protocol (DHCP). Currently the functionality that is necessary to provide these capabilities is still under development.

2.1.3 Tunnelling across QoS Domains

Most proposals assume that mobility across different administrative regions is handled by Mobile IP, which in the base specification uses tunnelling (IP encapsulation) to forward packets from the HA to the FA. However, this cannot be applied directly into a network that supports QoS. The incompatibility comes from the IP-in-IP encapsulation, because the insertion of a tunnel header offsets the packet payload, which prevents the fields in the transport and higher layers from being accessed normally.

Consider the Resource Reservation Protocol (RSVP) which is the QoS signalling protocol adopted by the IETF. When RSVP signalling messages enter a tunnel, they are encapsulated with a tunnel header that carries an “IP-in-IP Encapsulation” rather than a “Router-Alert” option. Consequently, RSVP-capable routers cannot recognise the packets, and resources are not reserved accordingly. Moreover, even if the required resources could be reserved, those intermediate RSVP routers will not be able to access port numbers correctly in order to distinguish data packets belonging to different flows. Therefore, it will not be possible to honour the per-flow state resource reservations.

To resolve these problems, a RSVP tunnelling algorithm has been proposed in [15]. However, this approach requires further complication in both signalling and encapsulation at the tunnel endpoints, and it increases considerably the overhead of transferring small packet payloads like voice data.

RSVP is proposed to be used with and tightly connected to the IntServ architecture, but it is also considered as the signalling protocol to be used with DiffServ [16]. In this case, the same shortcomings of QoS and tunnelling will apply to both IntServ and DiffServ environments. In the case of a DiffServ infrastructure without RSVP signalling, tunnelling poses fewer problems since the DiffServ codepoint (DSCP) can be copied forwards and backwards between the tunnel header and the original IP header when encapsulation and de-capsulation take place. However, in certain networking scenarios when path- or source-dependent services are desirable, multiple-field (MF) classification has to be invoked in the ingress and/or egress DiffServ routers [17]. Similarly to RSVP compliant routers without modifications, these DiffServ edge routers cannot access the higher layer information in the packet payload due to the extra location offset created by the tunnel header.

2.2 Issues with Charging and Accounting

Besides the technical aspects of integrating QoS support into IP mobility management, wireless Internet access also creates certain challenges for the current techniques of charging and accounting that are mainly for fixed-user access and best-effort data services. It is to be expected that the future infrastructure for charging and accounting will be increasingly complicated due to its support of roaming access and multiple service classes. These complications extend to the following areas:

2.2.1 Authentication and Reconciliation of Roaming Service

As mobile users roam among various ISPs (or administrative domains), they can access as many services as desired through the visited ISP. Inevitably, a share of trust must exist between the visited and the home administrative domains, such that the visited ISP will not charge the home domain excessively for the provided services. Similarly, the visited ISP needs to be sure that the home domain will honour the payment for resources that the customer has consumed.

The current approach to establishing this trust relationship is based on initial strong identity verification, credit history checks, and online authentication at the start of a session [18]. Furthermore, because of the liability for paying foreign ISPs for its roaming customers, the home ISP needs to process credit limit checks and fraud detection, and to verify conformance to usage policy and service level agreements. To achieve this, accounting records should be transferred within some time intervals to the home ISP for reconciliation [19]. Depending on the number of ISPs within a consortium, these direct business/trust relationships between ISPs can cause a scalability problem. Consequently, a third-party entity or “broker” [18] has been proposed which acts as a settlement agent, providing a common point of contact for charging and settlement services for various administrative domains.

2.2.2 Pricing and Charging for QoS

Currently the predominant form of Internet retail pricing is based on a flat-rate or per-time fee. Users in such charging schemes are forced to share a single service class that is not guaranteed and prone to variable delay. However, this situation should be greatly improved in the future when the Internet can deliver multiple service classes. In the future Internet, charging policies should be service-dependent, such that users of delay- and bandwidth-sensitive applications will have the option to pay a higher price for the better service quality.

Unfortunately, network services with multiple classes cannot be charged using simple pricing schemes such as flat-rate or per-time because users are likely to ask for the best service quality at all times. Therefore, many alternative or complementary pricing schemes [20], [21] have been proposed for future Internet services. They can be usage-based (in terms of volume of traffic), capacity-based (in terms of conformance to a service contract), auction-based (in terms of network congestion levels), or a combination of these. While these schemes have their attractive properties, their successful implementation relies on the availability of efficient traffic measurement and network monitoring algorithms. It is still an issue for further research to find a solution that is easy to implement by ISPs and simple enough for the average user to understand.

3. A Home-Proxy Based Wireless Internet Framework

Similarly to Mobile IP [1], our home-proxy based wireless Internet framework aims to provide mobility management in a IP-based mobile environment. Moreover, this architecture has been designed from the beginning to inter-operate with a QoS capable Internet backbone and to simplify the cost recovery processes of a roaming service.

Figure 1 is a brief comparison of the mobility management techniques used in our home-proxy based framework and Mobile IP. In both schemes, the CH is not required to be mobility-aware or capable of detecting changes in network conditions. This implies that when a connection is made to a MH, all packets destined to this MH are normally forwarded to its home network. Then the HP or HA intercepts these packets and redirects them to the current subnet of the MH. Unlike the HA of Mobile IP, our HP uses a split-connection approach to relay information to the MH. This approach requires the HP to accept the incoming connection on behalf of the MH, and then makes a separate connection to the MH. Once these two connections have been established, the HP connects them

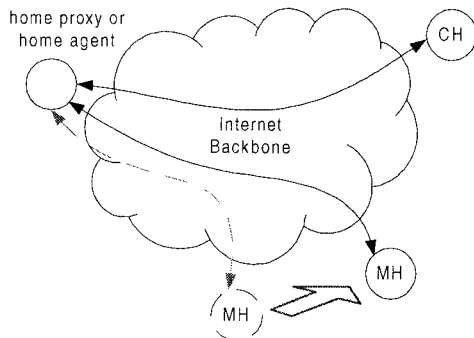


Fig. 1 Mobility management techniques in the home-proxy based framework and Mobile IP.

to form a complete path between the CH and the MH. When the MH moves to another location, it simply obtains a new COA from the new subnet and informs the HP. The HP tears down the old connection and makes a new connection to MH using its latest COA. When this new connection is established, the HP joins it with the incoming connection such that the path between the CH and the MH can be re-established. The details of this approach will be discussed in following sections.

3.1 Advantages of the Split-Connection Approach

This split-connection approach has two distinct advantages over the current method of redirection in Mobile IP. Firstly, it avoids the use of tunnelling between the MH's home network and its visiting network. As explained in the previous section, various incompatible issues can arise when passing tunnelled packets across IntServ or DiffServ domains. Our framework eliminates these problems by replacing tunnelling with a new connection when the MH moves from one visited subnet to another. This implies that the roaming traffic of a MH is now indistinguishable from other traffic generated by stationary hosts, and thereby mobility management can be gracefully integrated into any QoS management architecture.

Secondly, the split-connection approach allows the HP to initiate Internet services on behalf of the MH while it is away at a foreign ISP. That is, instead of directly requesting a service connection at the foreign network, the MH can send out its service requests via a control connection to its HP. The HP then makes two separate connections: one from the HP to the CH, and the other from the HP to the MH. Similar to the service establishment initiated by the CH, the HP merges the connections to form a complete path between the CH and the MH. By initiating service connections via the HP, the visited ISP and all intermediate service providers would see the roaming traffic of a MH as if it were originated from a fixed host in the home network. This arrangement has several advantages:

- The visited ISP is not directly dealing with a random roaming user, but only provides a local IP address and a control connection to each user for service requests. Therefore, the user's trust and accountability has been removed from the visited ISP, and complex authentication and reconciliation as mentioned in Sect. 2.2.1 may not be necessary.
- Although the subscribers are at a foreign ISP, the home network can easily verify their conformance to usage policy and service level agreements by monitoring and controlling the HP.
- A fair charging policy can be achieved for roaming services. Just like many charging schemes in today's telephone networks, the future Internet can charge the initiating party for the cost of a con-

nection. With such a policy, when a MH in a foreign domain requests for a connection, the home administrative domain would pay for the connections to the visited ISP and to the CH. In the case where the CH is initiating a connection to the MH, the CH would only pay for the connection to the MH's home network and the home administrative domain would pay for the cost of redirection from the HP to the visited ISP.

3.2 Realisation of Split-Connection Approach

The home-proxy based framework is a novel concept for facilitating the integration of mobility support with QoS management mechanisms, and at the same time simplifying the cost recovery processes of roaming services. Nevertheless, the use of proxy or middleware is not a totally new idea in supporting user mobility. For instance, MSOCKS [22] handles mobility management at the transport layer and introduces a proxy in-between the client and server. SLM [9] operates above TCP and switches TCP streams between the MH and CH. Moreover, SIP mobility support [8] extends the application layer protocol SIP to switch UDP traffic between the MH and CH. We believe that all the above schemes can be applied, with some modifications to support mobility management between the MH and HP. However, the IP stack has to be modified in MSOCKS, and mobility support in SIP has not yet been consolidated. Therefore, we have chosen to expand the SLM approach.

Figure 2 illustrates the basic function of SLM in setting up data sessions. If a data connection is initiated by the MH (Fig. 2(a)), a reflector module intercepts and redirects the connection to the SLM module at the MH. The SLM module then examines the connection and sends out a control message indicating the session characteristics and destination address to the

MH's home register (HR). After verifying the connection request, the HR authorises the SLM module at the HP to make two separate connections that have the requested characteristics: one from the HP to the CH, and the other from the HP via the SLM module to the SLM module of the MH. Similar procedures are invoked when the MH binds a socket to listen for an incoming connection (Fig. 2(b)). The reflector module indicates this socket binding process to the SLM module of the MH, which in turn informs the HR about its readiness to accept the incoming connection. After verifying the binding request, the HR instructs the SLM module at the HP to wait for incoming connections. When the CH initiates a data connection, the HP uses proxy-ARP or similar techniques to accept this connection on behalf of the MH. Then the HP consults the HR about the preferences and policies of this MH, and makes a separate connection, via its own SLM module, to the SLM module at the MH. Similar to its peer at the HP, the SLM module at the MH accepts the incoming connection for the MH, and makes a separate connection to the waiting application.

Like any mobility management framework, this split-connection approach requires proper location management support through which the MH can update its latest location and the connection destined for the MH can be redirected to its current location. However, before this, the MH must obtain a COA from the visited subnet. This COA can be the IP address of a mobility agent (like Mobile IP) or a unique IP address assigned by a DHCP server. Since our framework does not require tunnelling or a mobility agent at the visited subnet, the later approach is more appropriate.

Figure 3(a) is a simple illustration of the function of the HR to direct a connection to the MH. After obtaining a COA from the DHCP server, the SLM module at the MH is responsible for updating its latest location

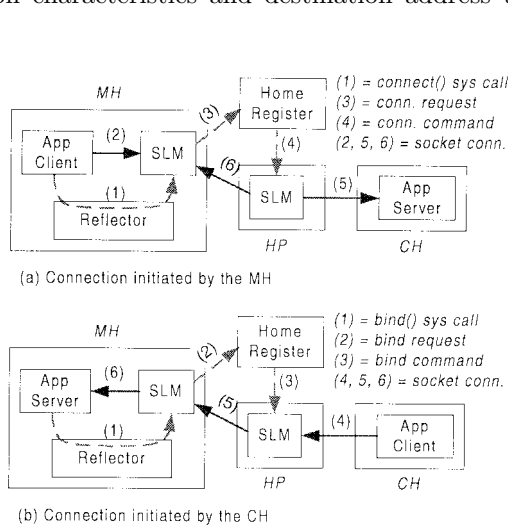


Fig. 2 Connection establishment via SLM.

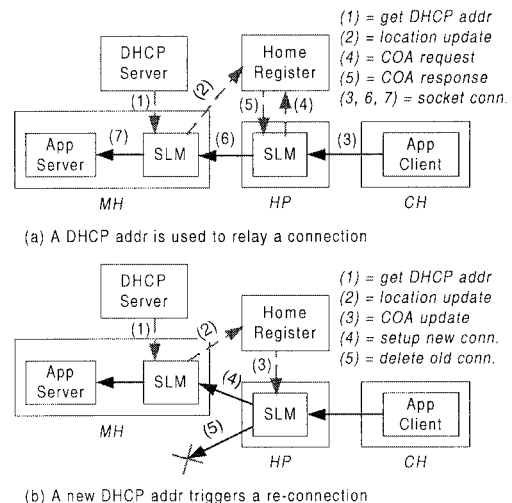


Fig. 3 Using DHCP address to establish and maintain a connection.

at the HR. When a connection destined for the MH arrives, the SLM at the HP needs to obtain the current COA of the MH and then establish a new connection to it.

After the establishment phase, the SLM module treats data connections identically regardless of whether the MH or CH establishes the connection. When the MH moves to another subnet (Fig. 3(b)), it obtains a new COA and informs the HR. Then its HR instructs the HP to tear down and re-establish the inter-SLM connection between the HP and MH. Since the connection state is retained by SLM modules and the original connections are shielded from the details of the network, the movement of the MH is transparent to the applications at both the MH and CH.

From the above description, it is evident that the realisation of a split-connection approach is not difficult. In fact, most of the basic ideas, such as the proxy-based mobility management (SLM and others), DHCP, and HR, have already been proposed before or are widely used in practice. Therefore, our main task has been to integrate them in a manageable and collaborative manner.

3.3 Support of Macro- and Micro-Mobility

In our framework, inter-SLM connections are torn-down and re-established as the MH moves from one subnet to another. Just like Mobile IP, this approach does not allow frequent handovers and fast location updates especially when the visited domain is some distance away from the home network. For the provision of mobile real-time services, the micro-mobility management technique could be used to supplement our split-connection approach that handles macro-mobility. However, unlike the integration of micro-mobility and Mobile IP in Sect. 2.1.2, our management of macro- and micro-mobility should be relatively simpler because the split-connection approach operates at a higher layer.

We assume that each regional administrative domain is equipped with one or more wireless Internet gateways (WIG) which serve as the interchange between conventional IP routing and micro-mobility routing. Therefore, it is important that all DHCP addresses given to MHs should direct their incoming traffic to pass through the WIG to which they are allocated. In principle, any kind of the regional network architecture can be used to realise micro-mobility management. However, we have chosen a dynamic per-host routing scheme similar to HAWAII [11] or Cellular IP [12] because of its advantages in routing data packets to a MH without tunnelling or address conversion. In this approach, packets are forwarded on a hop-by-hop basis from the WIG over a dynamically established path to the MH, whose IP address is unchanged as long as it stays at the visited domain. Since the IP address has no location significance within the access network, the

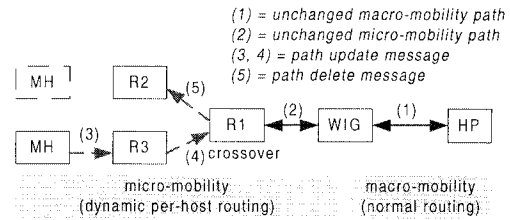


Fig. 4 Dynamic update of data path between the WIG and MH.

user's mobility would become transparent to the QoS enabled Internet backbone.

Per-host routing technique in our framework is used in the following manner. To dynamically maintain a routing path from the WIG to its current location, the MH needs to send out control messages to update the forwarding entries at each router along the path. Figure 4 illustrates the basic scenario for rerouting as the MH moves from one subnet to another within the domain. After receiving the path-update message, a router needs to modify its routing table accordingly for the downstream traffic to the MH and propagate this control message to an upstream router based on its default gateway setting. Then a similar path-update process is repeated at each router between the MH and the WIG until the control message encounters the WIG or a router keeping a forwarding entry for the old path. Then the WIG or this crossover router would send out a path-delete message to the downstream routers along the old path to eliminate any out-of-date forwarding entries. At this time, when packets destined for the MH arrive at the WIG, they are routed to the MH using the per-host forwarding entries newly established. To efficiently manage these dynamic routing paths, the per-host routing entry at each router should be refreshed periodically.

The deployment of our micro-mobility management requires the routers at the access network to be mobility-aware. Nevertheless, these routers are no different to normal IP routers except being able to process and propagate path-update or path-delete messages. Another minor modification is also required at the access network to maintain a unique IP address for each visiting MH. Instead of installing a DHCP server within every subnet, we need to set up DHCP relay agent at the wireless subnets and co-ordinate the assignments of COAs via a centralised DHCP server at each domain.

3.4 Support of Advance Macro- and Micro-Reservation

It is generally difficult to promise a specified level of QoS to a MH since there may not be enough resources in the part of the network that the mobile user is moving into. The approaches taken to alleviate this problem can be based on advance reservation. How-

ever, the current proposals such as [23]–[25] are mainly designed for the IntServ/RSVP architecture and are somewhat difficult to implement because of their significant changes to the QoS framework, such as modifications to all intermediate RSVP routers.

In order to reduce the complexity of resource reservation for MHs, it is assumed in our framework that due to inadequate network resources, the bottleneck is likely to be at the Internet backbone or at the wireless link. This implies that the network paths joining basestations and the WIG should be either well engineered or over provisioned. We believe that this is not difficult to achieve in practice because the network between the basestation and WIG is normally owned by the same administrative domain and has relatively large and cheap bandwidth. Based on this assumption, we can divide resource reservation into two levels: macro-reservation and micro-reservation [26], similar to the mobility management scheme we proposed earlier. The schematic of our advance reservation scheme is shown in Fig. 5. Our advance resource reservation scheme operates as follows:

3.4.1 Advance Macro-Reservation

The WIG plays an important role in our resource reservation scheme for mobile users. At the macro-level, it

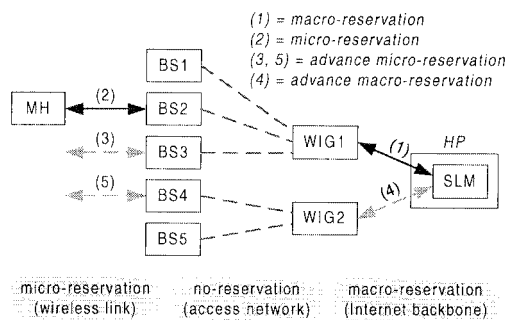


Fig. 5 Advance reservation at macro- and micro-levels.

behaves as the first or last hop node of the underlying QoS framework to complete the QoS path across the Internet backbone. For instance, it can act as an ingress or egress router for the DiffServ architecture, or as a RSVP proxy agent [23], [24] for the IntServ architecture. The benefit of our advance macro-reservation scheme is that it requires no further changes in the existing QoS framework, because the SLM module at the HP can simply establish multiple connections to different WIGs if their domains contain the predicted locations of future moves. But this is only achievable if the regional DHCP server is willing to release a COA in advance for the HP [26].

3.4.2 Advance Micro-Reservation

At the micro-level, the WIG allocates wireless resources at basestations in which the QoS path is expected to be initiated or terminated. In our framework, QoS support across the Internet backbone is managed by the process of macro-reservation. Therefore, we argue that by assigning wireless resources for a QoS path at multiple basestations, we are effectively extending the QoS support of this Internet connection in advance into multiple locations. Moreover, as the MH moves from one location to another, the number of cells participating in resource pre-allocation can be different. Nevertheless, regardless of the user location in the administrative domain, the same QoS path is used to carry traffic across the Internet backbone, and therefore no change is necessary to the macro-reservation.

3.5 The Full Picture of the Home-Proxy Based Framework

Figure 6 shows the overall architecture of our home-proxy based wireless Internet framework. The role of the HP is three-fold. Firstly, it shields the movement of the MH from the CH. Secondly, it allows graceful integration of macro-mobility and advance macro-reservation across the Internet backbone. Thirdly, it

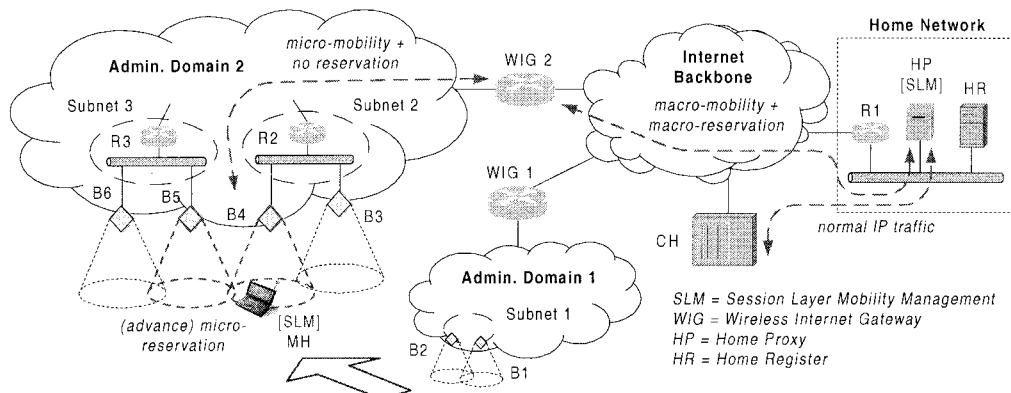


Fig. 6 An overall architecture of the home-proxy based framework using a dual-level approach of mobility management and advance resource reservation.

can initiate Internet services on behalf of the MH, which greatly simplifies the cost recovery processes of roaming services in foreign administrative domains.

The WIG is another key entity in this architecture. Firstly, it shields the regional movement of the MH from the HP. Secondly, it extends the QoS guarantees of macro-reservation to multiple locations within the domain. Although we make certain assumptions about the regional access networks (e.g. mobility-aware, rich in capacity), the overall framework imposes no strict requirement of their presence. The use of micro-mobility and micro-reservation is aimed at improving the performance of mobility and QoS supports without special modifications in the Internet backbone.

4. An Experimental Testbed

To prove the viability of our design, we have built and tested this framework in a realistic environment. Our experimental testbed consists of a home network and a foreign administrative domain located at the University of New South Wales and CSIRO Telecommunications and Industrial Physics respectively. Figure 7 shows the various types of network equipment at both premises. For ease of development, the HR, HP, WIG and mobility aware routers in this testbed were implemented in Java and C on a RedHat 6.2 Linux platform. Although most of the framework components are connected by 100 Mbps Ethernet links, we could only obtain from the NSW Regional Network Organisation a two Mbps ATM link between UNSW and CSIRO TIP due to the limited resources available. Also, the wireless interface has a relatively small bandwidth of 5.5 Mbps provided by the WaveLAN 802.11 card. These limitations on

network bandwidth agree with our earlier assumption that the bottleneck of network resources is likely to be at the Internet backbone or at the wireless links.

4.1 Simplifications on Implementation

Compared to the framework proposal in the previous section, we made several simplifications to our testbed. (1) We assume that there are only two types of service available: premium and best effort. (2) We consider network bandwidth as the only QoS parameter in the testbed, and assume there is some simple admission control mechanism for premium service in both the home network and the foreign administrative domain. This function is emulated by two simplified Bandwidth Broker (BB) modules at the HR and WIG, which keep track of the bandwidth consumed by premium service at the Internet backbone or the wireless interfaces. However, our BB module is easily replaced by a proper admission control mechanism once it becomes available in the DiffServ framework. (3) We have merged the functions of mobility aware router and basestation (e.g. R2, B3 and B4 in Fig. 6) into a new entity called the Wireless Subnet Router (WSR). For total separation of network traffic between subnets, the wireless interface at each WSR is configured to operate at a disjoint sub-band frequency. (4) The focus of our testbed is not on the link layer handover. Moreover, the delay in DHCP assignment and the detection of subnet arrival can depend heavily on the software implementation [27], [28]. To isolate these effects from the handover performance at network and session layers, we have implemented a mobility emulator at the MH that periodically generates DHCP parameters for use in macro- and micro-mobility. (5) A QoS capable MAC layer is essential to properly support multiple service classes across the wireless link. In our testbed implementation, however, the WSR can only control the downlink traffic by using the Linux Traffic Control (TC) software [29]. (6) Due to the small number of MHs and WSRs, we do not implement advance resource reservation. We only allocate network resources when the MH arrives at a new subnet. (7) Without losing generality, we have placed the CH at the home network to avoid any unexpected network congestion.

4.2 Testing Scenarios

The RTP application we used to conduct testing consisted of a MP3 player (FreeAmp) at the MH and a MP3 jukebox (Obsequiem) at the CH [30]. The Obsequiem server was slightly modified so that it can perform RTP streaming in unicast mode. When a mobile user has chosen a song, his/her request to the MP3 jukebox triggers a MH-initiated connection establishment via the SLM (see Fig. 2(a)). From this request, the Obsequiem server learns the home IP address of the mobile

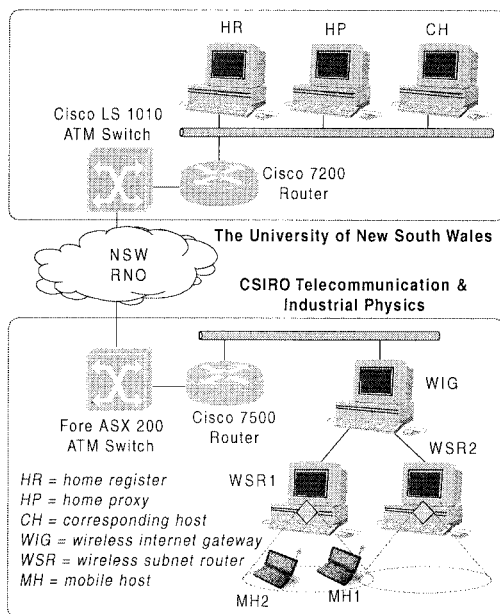


Fig. 7 The layout of our experimental testbed.

Table 1 Performance of handover in the experimental testbed.

	Macro-mobility Handover		Micro-mobility Handover	
	Best Effort	Premium Service	Best Effort	Premium Service
Time of Interruption (ms) [†]	159 ± 26	361 ± 418	126 ± 45	215 ± 132
Number of Packet Lost [†]	6.3 ± 0.9	19.4 ± 13.8	2.5 ± 0.7	9.7 ± 9.6

[†] Average and standard deviation values are calculated from 100 handover events.

user and the port number to be opened by FreeAmp. In the meantime, the FreeAmp application opens a UDP socket at the MH, and its HP is informed and prepares itself for the incoming traffic (see Fig. 2(b)). When the Obsequiem server sends the RTP data stream to the home network of the MH, the HP is ready to receive the RTP packets and relay them to the current location of the MH.

In this experimental testbed, we are particularly interested in the influence of user mobility and network congestion on the RTP streaming traffic. Although there is only one foreign administrative domain in the testbed, our mobility emulator can explicitly request a macro- or micro-mobility handover when the MH moves between the two WSRs. In addition, the mobile user can request his/her home network to upgrade or downgrade the current network service at any time. The BB module at the HR is capable of sending out a configuration script to the Cisco ingress router enabling it to mark the packets destined to the MH as either high priority or best-effort. To provide premium service, a high priority packet arriving at the ATM interface is sent via a CBR PVC to the foreign administrative domain. Similarly, when a high priority packet reaches the WSR hosting the mobile user, the TC software sends it to a high priority, low delay FIFO queue for delivery to the MH. In default best-effort case, the ingress router transfers the packet via a UBR PVC and the TC software at the WSR directs it to a low priority RED queue with 60 kbytes of buffer.

During the experiment, subnet handover was generated every 10 seconds at MH1. The handover event was repeated 100 times for each of the following conditions: macro-mobility with best effort service; macro-mobility with premium service; micro-mobility with best effort service; and micro-mobility with premium service. To test the effectiveness of the premium service class, we introduced heavy best effort traffic in the background, i.e., 94% link capacity from the CH to the WIG, and 91% link capacity from WSR1 to MH2. In the presence of background traffic, the round-trip time for “pinging” the WIG and CH from MH2 was around 250 ms and 300 ms respectively.

4.3 Performance of Mobility Management

We tested the MP3 application in the above scenarios and recorded the time of interruption and the number of RTP packet losses at MH1 during handover. Table 1 is a summary of the performance, and presents

each result in terms of average and standard deviation. Compared to macro-mobility management at the session layer, micro-mobility management at the network layer provides a better handover characteristics (i.e. shorter interruption and less packet loss). Particularly with our MP3 player when an input buffer size of 20 kbytes was set, the audio interruption was just a very short skip when there was no background traffic present (i.e. in the case of best effort in Table 1).

On the other hand, we noticed a significant worsening in the handover performance for both macro- and micro-mobility when heavy background traffic was present (i.e. in the case of premium service in Table 1). This deterioration is mainly due to the lack of multiple service class support in the wireless interface. Consequently, all mobility control protocols in the uplink direction have to compete with the downlink background traffic, and they experience a longer delay. Our view is supported by the earlier observation of poor connectivity to the WIG at MH2.

The worst handover characteristics occurred when MH1 was performing macro-handover under heavy background-traffic conditions. Beside the lack of multiple service class support in the wireless link, the macro-mobility control protocols have to pass through a congested Internet backbone. Moreover, it was found that the Cisco ingress router needs a certain time interval before it can activate a new marking scheme for packets destined for the new user location. During this transition period, there are three to five packets being marked as best effort although the user requires premium service class. As a result, these packets are either being dropped in the best effort queues, or they arrive out of sequence at the MH.

In a rough comparison with other prototypes, the handover performance of our testbed has a slightly low rating. For instance, considering a user space implementation of Mobile IP [31] with frequent location re-registration, the handover interruption time can be as low as 123 ms on average [32]. Another example is Cellular IP which would drop only one packet during a hard handover with a light load condition [12]. However, since most of the coding in our testbed uses the Java language, the performance values in Table 1 can be significantly improved if our modules were written in a more efficient programming language.

4.4 Performance of Resource Reservation

After a user had selected the premium service, we were

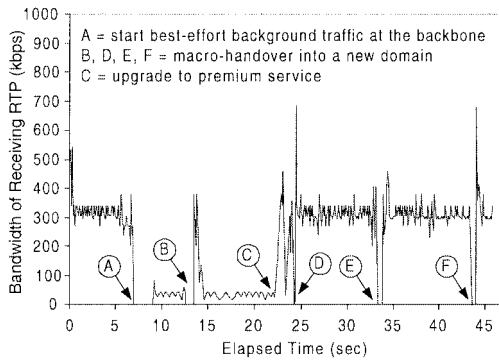


Fig. 8 RTP arrival rate of best effort and premium service classes under macro-mobility.

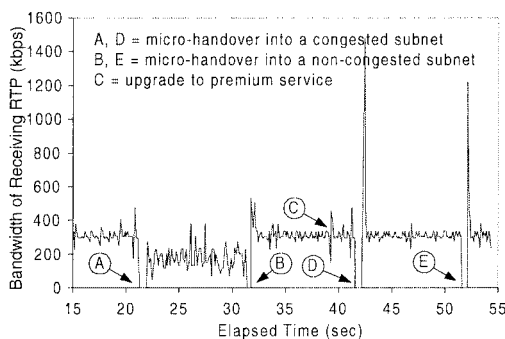


Fig. 9 RTP arrival rate of best effort and premium service classes under micro-mobility.

able to successfully maintain the MP3 streaming traffic despite heavy network congestion and user mobility. In an ideal seamless mobility environment, we would expect to receive our MP3 traffic at a constant bit rate of 320 kbps. In a practical environment, however, subnet and domain handovers would cause additional jitter in the RTP traffic. Moreover, since our SLM and WSR implementations do not retransmit UDP packets, we would expect some traffic interruption as mentioned in the previous section.

In order to observe the effectiveness of our Macro- and Micro-Reservation mechanism, we introduced background traffic separately at the Internet backbone and the wireless link, while the arrival rate of RTP packets was measured at MH1. Figure 8 illustrates the arrival rate of RTP packets when the MH was performing macro-handovers. Due to the heavily congested Internet backbone, only a small portion of the MP3 traffic managed to get through and arrive at the MH. However, once the user had upgraded to premium service, the RTP packets were received at the MH regardless of a congested backbone and frequent user mobility.

Figure 9 shows a similar result when MH1 moved in and out of some congested and non-congested wireless subnets. Although the MH was performing micro-handovers every 10 seconds, the arrival rate of RTP

traffic was maintained as long as premium service was selected. Moreover, it is notable that our micro-reservation scheme does not require any reconfiguration of the marking scheme at the ingress router of the home network, and therefore is transparent to the macro-reservation mechanism at the Internet backbone.

5. Conclusions

In this paper we have presented an overview of our home-proxy based wireless Internet framework, which supports a dual-level mobility and resource reservation approach to meet the stringent QoS requirements of real-time services.

At the macro-mobility level, we have proposed using Session Layer Mobility (SLM) management to handle user mobility. Compared with the current specification of Mobile IP, this approach appears to have several advantages:

- The home network can initiate Internet services on behalf of the MH while it is away at a foreign ISP. Therefore, the user's trust and accountability has been removed from the visited ISP, and complex authentication and reconciliation processes as mentioned in Sect. 2.2.1 may not be necessary.
- The tunnelling mechanism in Mobile IP is replaced by inter-SLM connections when the MH moves from one domain to another. This implies that the roaming traffic of a MH is now indistinguishable from other traffic generated by fixed hosts, and thereby mobility management can be fully decoupled from any QoS management architecture at the Internet backbone.
- As macro-mobility management is at the session layer, our framework can support both terminal mobility (e.g. subnet handover) and personal mobility (e.g. handover between different terminals or different kinds of network).
- In a mobile environment, session customisation (i.e. protocol conversion and/or content adaptation) is an important tool to ensure continuous network connectivity for mobile users, and the HP in our framework offers a convenient insertion point for these services.

It is worth mentioning that the additional roles of proxy in personal mobility and session customisation have been widely reported in the literature. Because of this, we decided not to have a repeat mentioning of these functions and their benefits in the paper. Interested readers are referred to [33] for more information.

For the provision of mobile real-time services, the micro-mobility management technique has been introduced to improve the performance of frequent handovers within an administrative domain. We have chosen a dynamic per-host routing scheme because of its advantages in routing data packets to a MH without

tunnelling or address conversion.

As our mobility management is designed to gracefully integrate into any QoS management architecture, the SLM module can perform advance resource reservation across the Internet backbone (Advance Macro-Reservation) without special requirements on intermediate routers and modifications of protocols.

With a simple assumption about the capacity of regional access networks, we have proposed that by assigning wireless resources at multiple basestations for a QoS path from HP to WIG, we are effectively extending the QoS support of this Internet connection in advance into multiple locations (Advance Micro-Reservation). Compared with other proposed schemes of advance reservation in [23]–[25], our dual-level reservation approach is simpler with little additional complication for the underlying QoS architecture.

To prove the viability of our ideas, a proof-of-concept testbed has been designed, built and tested. As our experiments illustrated, this framework can be readily realised because it involves only the integration of several existing facilities in a manageable and collaborative manner. The preliminary results are encouraging, and therefore we believe this framework can be used effectively to support mobility and QoS in the emerging wireless Internet environment.

Acknowledgement

The work of Jonathan Chan is funded through a CSIRO Postgraduate Scholarship and an Australian Postgraduate Award.

References

- [1] C. Perkins, "IP mobility support for IPv4, revised," IETF Internet Draft, Sept. 2000.
- [2] R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet architecture: An overview," IETF RFC 1633, June 1994.
- [3] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differential services," IETF RFC 2475, Dec. 1998.
- [4] C. Rigney, A. Rubens, W. Simpson, and S. Willens, "Remote authentication dial in user service (RADIUS)," IETF RFC 2138, April 1997.
- [5] P. Calhoun, A. Rubens, H. Akhtar, and E. Guttman, "DIAMETER base protocol," IETF Internet Draft, Sept. 2000.
- [6] C. Perkins and D. Johnson, "Route optimization in mobile IP," IETF Internet Draft, Feb. 2000.
- [7] R. Jain, T. Raleigh, C. Graff, M. Bereschinsky, and M. Patel, "Mobile Internet access and QoS guarantees using Mobile IP and RSVP with location registers," Proc. IEEE ICC'98, pp.1690–1695, June 1998.
- [8] E. Wedlund and H. Schulzrinne, "Mobility support using SIP," Proc. ACM WoWMo'99, pp.76–82, Aug. 1999.
- [9] B. Landfeldt, T. Larsson, Y. Ismailov, and A. Seneviratne, "SLM, a framework for session layer mobility management," Proc. IEEE ICCCN'99, pp.452–456, Oct. 1999.
- [10] E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP regional registration," IETF Internet Draft, July 2000.
- [11] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "IP micro-mobility support using HAWAII," IETF Internet Draft, July 2000.
- [12] A. Campbell, J. Gomez, S. Kim, A. Valk, C.-Y. Wan, and Z. Turnyi, "Design, implementation, and evaluation of Cellular IP," IEEE Pers. Commun. Mag., vol.7, no.4, pp.42–49, Aug. 2000.
- [13] A. Acharya, J. Li, F. Ansari, and D. Raychaudhuri, "Mobility support for IP over wireless ATM," IEEE Commun. Mag., vol.36, no.4, pp.84–88, April 1998.
- [14] S. Das, A. Misra, P. Agrawal, and S.K. Das, "TeleMIP: Telecommunications-enhanced mobile IP architecture for fast intradomain mobility," IEEE Pers. Commun. Mag., vol.7, no.4, pp.50–58, Aug. 2000.
- [15] A. Terzis, J. Krawczyk, J. Wroclawski, and L. Zhang, "RSVP operation over IP tunnels," IETF RFC 2746, Jan. 2000.
- [16] Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, "A framework for integrated services operation over DiffServ networks," IETF Internet Draft, May 2000.
- [17] D. Black, "Differentiated services and tunnels," IETF RFC 2983, Oct. 2000.
- [18] S. Glass, T. Hiller, S. Jacobs, and C. Perkins, "Mobile IP authentication, authorization, and accounting requirements," IETF RFC 2977, Oct. 2000.
- [19] J. Arkko, P. Calhoun, P. Patel, and G. Zorn, "DIAMETER accounting extension," IETF Internet Draft, Sept. 2000.
- [20] M. Borella, V. Upadhyay, and I. Sidhu, "Pricing framework for a differential services Internet," European Trans. Telecom., vol.10, no.3, pp.275–288, May 1999.
- [21] G. Fankhauser, B. Stiller, and B. Plattner, "Arrow: A flexible architecture for an accounting and charging infrastructure in the next generation Internet," Baltzer Netnomics, vol.1, no.2, pp.201–223, Feb. 1999.
- [22] D. Maltz and P. Bhagwat, "MSOCKS: An architecture for transport layer mobility," Proc. IEEE INFOCOM'98, March 1998.
- [23] A. Talukdar, B. Badrinath, and A. Acharya, "MRSVP: A resource reservation protocol for an integrated services network with mobile hosts," Technical Report, DCS-TR-337, Rutgers University, 1997.
- [24] I. Mahadevan and K. Sivalingam, "Architecture and experimental results for quality of service in mobile networks using RSVP and CBQ," ACM/WINET, vol.6, no.3, pp.221–234, June 2000.
- [25] A. Terzis, M. Srivastava, and L. Zhang, "A simple QoS signaling protocol for mobile hosts in the integrated services Internet," Proc. IEEE INFOCOM'99, pp.1011–1018, March 1999.
- [26] J. Chan, B. Landfeldt, A. Seneviratne, and P. Sookavathana, "Integrating mobility prediction and resource pre-allocation into a home-proxy based wireless Internet framework," Proc. IEEE ICON 2000, pp.18–23, Sept. 2000.
- [27] J.-O. Vatn and G. Maguire, Jr., "The effect of using co-located care-of addresses in macro handover latency," Fourteenth Nordic Tele-traffic Seminar (NTS 14), Aug. 1998.
- [28] A. Fladenmuller and R. De Silva, "The effect of mobile IP handoffs on the performance of TCP," ACM/Baltzer MONET, vol.4, no.2, 1999.
- [29] W. Almesberger, J. Hadi, and A. Kuznetsov, "Differentiated services on Linux," IETF Internet Draft, March 1999.
- [30] Information available at <http://obs.freeamp.org/>
- [31] M. Rodriguez, "An implementation of mobile IP under Linux," Information Available at http://www.hpl.hp.com/personal/Jean_Tourrilhes/MobileIP/index.html
- [32] D. Coudeville, "Comparison between two protocols for mo-

bility management in the Internet: Mobile IP and SLM," Master Thesis, UNSW/ENSICA (Ecole Nationale Supérieure des Ingenieurs des Constructions Aeronautiques), pp.95–98, Aug. 2000.

- [33] P. Maniatis, M. Roussopoulos, E. Swierk, K. Lai, G. Appenzeller, X. Zhao, and M. Baker, "The mobile people architecture," *ACM Mobile Computing and Communications Review*, vol.3, no.3, pp.36–42, July 1999.



Jonathan Chan received his Bachelor of Electrical Engineering with honours at the University of New South Wales in 1990. Then he has worked at Macau Telecom, Macau Electricity, Andersen Consulting (Aust.), and Dow Jones Telerate (Aust.). He is currently completing his Ph.D. at the same university. His current research interests are in QoS management and mobile computing.



Björn Landfeldt was born in Stockholm, Sweden. Started studying electrical engineering at the Royal Institute of Technology, Stockholm. He completed his Ph.D. in electrical engineering at the University of New South Wales in 2000. He is with Ericsson Research, Networks and Systems, Sweden and his areas of interest are mobile networking and QoS management.



Ren Liu received his Ph.D. from the University of Newcastle, Australia. He was a lecturer of Telecommunications in BUPT before coming to Australia. Dr. Liu joined Telecommunications Group of CSIRO Australia in 1995, where he is now a senior research scientist. His current research interests are in the areas of internetworking design, Diff-Serv, Internet traffic measurement and modelling.



Aruna Seneviratne completed his Ph.D. at University of Bath, UK, awarded in 1983. Since graduating he has held academic appointments in the UK and Australia. He currently holds the Mahanakorn Chair of Telecommunication at the University of New South Wales, and leads the Telecommunication Group. His current research interests are in QoS Management particularly in mobile data communication systems.