The Challenges of Provisioning Real-Time Services in Wireless Internet [†]

Jonathan Chan, and Prof. Aruna Seneviratne, University of New South Wales, Björn Landfeldt, Ericsson Research, Sweden

The issues involved in supporting real-time wireless Internet services go beyond the problem of forwarding packets to the correct destination of a mobile host. These issues include handover latency, availability of network resources after moving, and a proper charging and accounting infrastructure. Through a survey of current research in these areas, we discuss various concerns of integrating mobility, Quality of Service, and service charging techniques. Finally, we identify some open issues for future study.

Jonathon Chan

Aruna Seneviratne

Björn Landfeldt

INTRODUCTION

Over the last few years, we have seen a rapid growth in cellular mobile telecommunications and Internet penetration. In addition, these technologies have provided a diverse range of new services. Mobile telecommunications systems have enabled the transmission of voice and short messages from virtually anywhere at any time, and the Internet has provided access to information from a wide variety of locations such as office, home and Internet Cafes.

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, including real-time services, from anywhere at anytime. However, the convergence of these technologies to provide ubiquitous access has been hampered by three major factors:

- Firstly, the traditional Internet protocol suite does not support mobility [1].
- Secondly, the bandwidth of the wireless access networks available so far has been sufficient to support only the most rudimentary services such as e-mail and voice.
- Finally, the best effort nature of the Internet services, together with the volatility of the radio propagation characteristics have made it impossible for these systems to provide the quality of service (QoS) guarantees necessary to support any real-time services.

Some of the above mentioned issues have been addressed by the research community and the techniques are being standardised. The IETF has adopted Mobile IP [2] to handle mobility in Internet environments. There have also been tremendous advances in radio technologies that have enabled mobile systems to start offering data services – to a limited extent now [3],[4], and at much higher data rates in the next generation systems [5]. Moreover, there has been considerable progress in developing techniques for providing QoS guarantees in Internet environments [6],[7], and in wireless access networks[8]. These developments, until recently, have been largely done in isolation. Therefore, the availability of mobility management techniques and QoS management techniques are incompatible. Besides these technical issues, the current techniques of charging and accounting are mainly for fixed-user access and best-effort data services, and hence they may not be appropriate to the mobile multimedia environment.

[†] To appear in Telecommunications Journal of Australia, vol. 50 no. 3, Spring 2000, pp. 37 – 48.

The research community has just started to address the problems associated with integration of the numerous techniques that have been designed for mobility and QoS management in Internet environments. The objective of this paper is to describe challenges of integrating mobility management, QoS management and charging issues in the emerging wireless Internet to enable the provision of real-time services.

The organisation of the rest of the paper is as follows. We first give an overview of some techniques used in the Internet environments for managing mobility and QoS. In the next two sections, we provide details of various issues associated with integrating Internet QoS and IP mobility management techniques. We then describe the current development on resolving the charging and accounting on mobile multimedia services. Finally, we provide a conclusion in the last section.

BASIC ALGORITHMS OF IP MOBILITY AND QoS MANAGEMENT

The inherent limitations for supporting mobility and real-time mobile services in today's IP-based networks are well known [9], [10]. Therefore, the Internet Engineering Task Force (IETF) and the research community have been developing mechanisms that would make traditional IP capable of supporting mobility as well as mechanisms that will make the Internet more QoS aware.

Support of IP Mobility

The IP mobility support mechanisms proposed to date can be categorised into schemes that provide either personal or device mobility.

Personal mobility is the ability of users to access network services from any terminal in any location. These schemes enable the network to identify end users, as their access location and method of access change. The Session Initiation Protocol (SIP) [11] provides the framework for personal mobility in Internet environments without changes to the standard IP protocol.

Device mobility refers to the ability of the network to provide support for seamless roaming to mobile devices as their users move. To support device mobility, the IP forwarding mechanism, which is based on IP addresses with implicit location information, either needs to be changed or the addressing scheme has be modified. There are at least three alternatives available.

- **IP Encapsulation (Tunnelling)** — This process involves encapsulating a packet, as data, inside another packet. Because a topologically correct IP address is embedded inside the header of the new packet, i.e. the tunnel header, this encapsulated packet can follow the standard IP routing mechanisms and reach the IP subnet serving a mobile user. This method has been widely studied and has been adopted as the mobility management mechanism in the IETF Mobile IP [2].

- **Loose Source Routing —** As an option in the packet header, loose source routing enables the sender to specify a list of IP addresses that the packet must traverse. The source generates a list of addresses of the intermediate routers of which it wishes the packet to traverse, with the last entry being the current address of the mobile host. Contrasting to the normal formation of an IP packet, where the destination field is assigned with the address of the mobile host, this field is assigned by the source with the first entry of the address list. When the first intermediate router receives this packet, it fetches the next entry of the address list and places it in the destination field of the packet. This process is repeated until the packet reaches to the mobile host. This function has been carefully integrated in IP version 6 using a routing extension header which avoids current problems in IP version 4 with regards to security and performance [12].

- **Dynamic Host Routing** — The destination IP address is used as a mobile host identifier only, removing its association with the current location of the mobile host. Packets are forwarded on a hop-by-hop basis from a gateway over special dynamically established paths to a mobile host. The forwarding entries at each router along the path are refreshed periodically using update messages sent from the mobile host. This category of packet forwarding has been proposed lately in some micro-mobility architectures, which we will discuss in the later sections of this paper.

The main difference between these three routing schemes is the way location information is placed. In the case of tunnelling, it is embedded within the packet payload; with loose source routing, it is provided in the packet header; and in dynamic host routing, it is maintained in the forwarding table of each intermediate router.

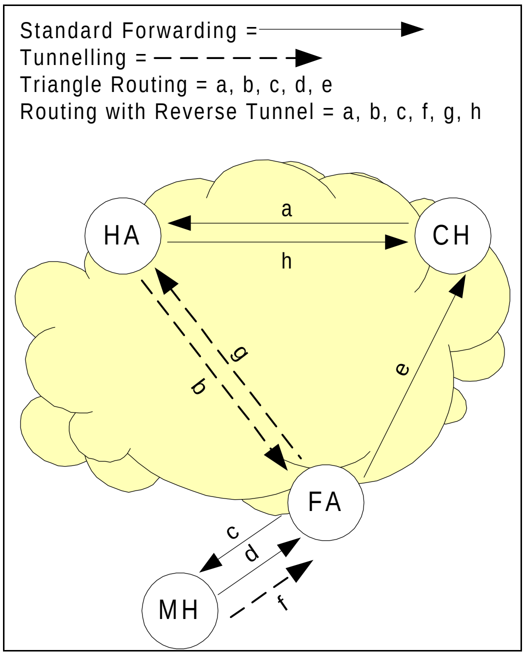


Fig. 1 *– Rerouting of IP packets in Mobile IP*

A commonly accepted realisation of IP Mobility is the IETF Mobile IP [2]. In Mobile IP, a fixed host (corresponding host – CH) which wants to communicate with another host is unaware whether the other host is mobile or fixed. This transparency is provided by using two network agents, one located at the mobile host's home network (home agent – HA) and the other located on the visited network (foreign agent – FA). The packets destined to a mobile host are always forwarded to its home network. In the case that this mobile host is currently away, the HA captures the packets and tunnels them to the FA on the visited network. The FA receives the packets, decapsulates them, and forwards them to the mobile host.

To facilitate the tunnelling, when a mobile host moves to a foreign network, it registers with the FA on that network and obtains a care of address (COA). Then it informs its HA of the current location via the FA. Mobile IP specifies the mechanisms for mobile hosts to discover foreign agents, and to deal with authentication and security.

Packets from a mobile host can be delivered in two different ways depending on the level of security in the foreign network (see Fig. 1). If routing is independent of source address within the foreign network, the mobile host can directly send packets to the corresponding host. This asymmetry of routing between the corresponding and mobile hosts is known as "triangle routing." On the other hand, if source-filtering routers are installed in this network, they will drop all packets originating from the mobile host because its source address is not topologically correct. One possible solution to this problem is to establish a reverse tunnel from the mobile host to its HA so that all tunnelled packets bear a correct source address [13]. When these packets arrive at the HA, they will be decapsulated and forwarded to the corresponding host.

Support of Quality of Service

Concurrently with the investigation of IP mobility, the IETF together with the Internet community have been developing mechanisms for providing QoS guarantees in the Internet environment.

- **Integrated Services (IntServ)** — The IntServ architecture [6] defines a flow as a stream of packets sharing the same source address, destination address and port number. To provide the necessary support, the IntServ routers are required to classify, regulate, and schedule data packets based on their flow-specific state. Since the amount of per-flow state and the handling of refresh messages increase proportionally with the number of flows at each router, the IntServ model does not scale well [10]. Resource Reservation Protocol (RSVP) was the first protocol adopted by the Internet community for realising IntServ. It is a signalling protocol which enables the advertising of available resources (PATH messages), and in the reverse direction, it allows the reservation of resources (RESV messages). RSVP is currently under revision for controlling bandwidth allocation within DiffServ domains [14], or setting up Label-Switched-Path inside MPLS networks [15].

- **Differentiated Services (DiffServ)** — Unlike the IntServ model, the DiffServ architecture [7] classifies individual flows only at the network ingress routers. The packets are then verified against certain admission control criteria, marked with one or more DiffServ Code Points (DSCP) and passed into the core network. The core routers examine their DSCP and apply some standard Per-hop Behaviours (PHBs) to forward each packet through the network. Since it aggregates flows with similar QoS requirements rather than dealing with each flow, the DiffServ model is more scalable than IntServ. However because of this aggregated traffic control, it is difficult to provide throughput guarantees for individual flows. Moreover, it is still unclear how to dynamically manage the allocation and utilisation of resources within and between the DiffServ domains. These issues are currently being addressed using Bandwidth Broker (BB) [16] and Multiprotocol Label Switching (MPLS) architectures [17].

Since these QoS frameworks have their own strengths and weaknesses, it is likely that they will co-exist in the future Internet as compatible and complementary technologies for providing end-to-end scalable QoS support. For instance, using RSVP to allocate bandwidth at the edges of the network and at peering points, while DiffServ is deployed in the backbone [10].

INCORPORATION OF INTERNET QOS INTO IP MOBILITY

To understand the issues associated with integrating a QoS management framework with a mobility framework, we consider the provision of QoS managed services in a wireless Internet using Mobile IP, the current proposal for supporting mobility in Internet environments.

Since Mobile IP was designed without the consideration of QoS, the framework has gone through various extensions to incorporate such requirements. The proposed efforts can be categorised into four areas.

Route Optimisation

Since all packets corresponding to a mobile host have to be routed through its HA in the forward and possibly in the backward direction, the chosen path can be significantly longer than the direct route. This indirect routing not only places an unnecessary burden on the network resources but also introduces longer delays for packet delivery. These shortcomings degrade services even further in a QoS supported mobile network, as longer path and delay mean higher probability of call dropping and service refusal.

To rectify this problem, the extension of route optimisation in Mobile IP [18] provides a means for corresponding hosts to cache the actual location of a mobile host so that their packets can be tunnelled to the mobile host directly. IP version 6 also integrates this idea into its base specification, but replaces the tunnelling mechanism with loose source routing that incurs less delivery overheads [19].

Another approach to this problem is to introduce the concept of a location server, which has been used successfully in cellular systems and recently considered for wireless ATM [20]. Through this location server, the mobile host can update its current location, and the corresponding host can then query the latest location of a mobile user before transmitting a packet. Since the corresponding host knows the actual location of a mobile host, both triangle routing and tunnelling can be avoided. The associated call setup and signalling protocols can be implemented by either changing the Mobile IP protocol (like MIP-LR [21]) or setting up user agents at the application-layers (like the SIP with mobility support [22] or Session Layer Mobility management [23]).

It is note worthy that in order to enable optimal routing, all the above proposals have to introduce mobility awareness in the corresponding hosts. This requires either modifications in the IP protocol stack (binding cache followed by tunnelling or loose source routing), or the addition of location server and signalling protocols.

Tunnelling Across QOS Domains

Though tunnelling has been adopted as the standard mechanism for redirecting packets in Mobile IP, it has certain constraints if used in conjunction with the QoS frameworks currently being developed. In applying RSVP to Mobile IP, it is assumed that RESV messages follow the inverse path of PATH messages. However, this is not the case for the base specification of Mobile IP as it uses triangle routing. Moreover, when RSVP packets enters a tunnel, they are encapsulated with a tunnel header that does not carry the Router-Alert option. Consequently, these packets are only regarded as best effort traffic by all RSVP routers along the path.

Even if this option were copied over during encapsulation, the tunnel header would still need to compensate for the offset of the port number fields, which form part of the flow identifier. Without further modifications, RSVP routers are not capable of differentiating packets belonging to different flows, and hence their per-flow state resource reservations become useless cannot be applied.

To allow intermediate RSVP routers to handle tunnelled traffic with minimal modifications, it has been proposed to encapsulate such data packets with properly defined IP and UDP headers, and to use the UDP port numbers to distinguish packets from different flows [24]. 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 traffic.

RSVP is proposed to be tightly connected to the IntServ architecture, but it is also considered as the signalling protocol to be used with DiffServ [14]. In this case, the same shortcomings of QoS and tunnelling will apply to

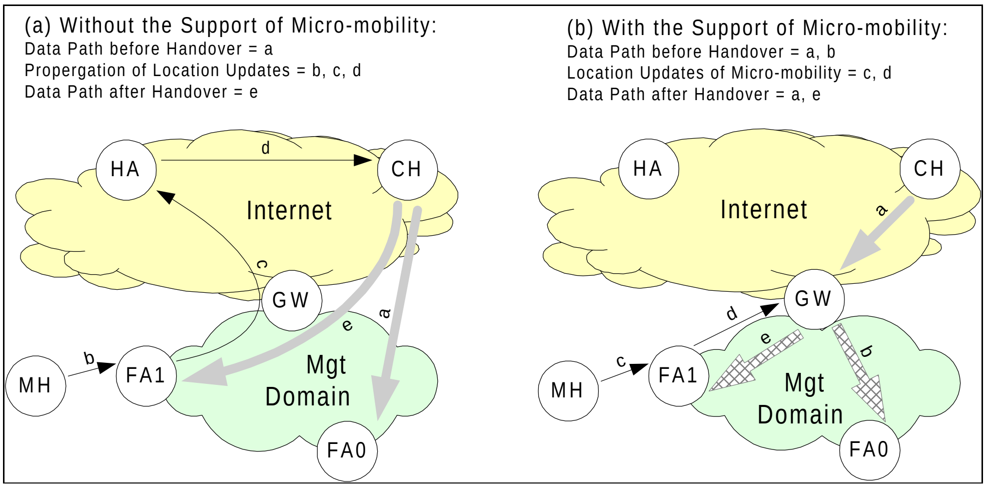


Fig. 2 – Improvements in Location Update and Re-routing with Micro-mobility

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. Like 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.

Regional Mobility Management

Mobile IP does not mandate fast handover and location updates. Each time a mobile host moves from one subnet to another, it needs to register its new location with the home agent. Thus if the visited network is some distance away from the home network, the signalling delay for these reregistrations can become large and consequently many packets could be mis-routed. In addition, Mobile IP does not require a mobile host to inform its previous foreign agent when it moves. Therefore, the former foreign agents is not able to re-route packets to the current foreign agent. The extension of routing optimisation in Mobile IP allows the mis-routed packets to be tunnelled from the old to the new foreign agents. However, this requires the mobile host to inform the previous foreign agent of its current location, and the mobility agents to deal with all the associated security issues. Even if all these problems were properly resolved, it would still take require a significant long some time in QoS capable networks to allocate and re-establish the routing path between the corresponding host and the mobile host (see Fig. 2 (a)).

Lately there have been many attempts in the Internet community to resolve these problems. These proposals develop methodologies for micro-mobility, which try to efficiently handle packet routings and location updates within a single management domain. For compatibility to the rest of the Internet, the functionality of Mobile IP is preserved to deal with roaming across management domains (macro-mobility). To co-ordinate micro and macro mobility, they use a gateway entity to convert control messages and to establish routing paths for the local management domain (see Fig. 2 (b)). These proposals can be divided into three categories.

- **Regional Tunnelling** — A logical solution to location registration latency resulting from distance is to perform registrations locally in the visited domain, which can be achieved through a hierarchy of foreign agents with tunnelling in between as described in [25]. This approach reduces the number of signalling messages to the home network and shortens the signalling latency when moving from one foreign agent to another. However, in order to facilitate this, it is necessary to introduce new registration messages for local mobility. Furthermore, a special gateway entity (gateway foreign agent) is required to handle and transform these regional registrations, and to dynamically manage regional tunnels for the mobile host in both the forward and reverse directions. Therefore, changes to Mobile IP entities (e.g. mobile host and mobility agents) are necessary. These changes ensure that the HA will always see this gateway foreign agent as the current location of a mobile host, regardless of which FA is serving the mobile host in the visited domain. This regional tunnelling approach can also be found in other proposals such as Regional Aware Foreign Agent [26] and Transparent Hierarchical Mobility Agents [27]. Instead of modifying those well-defined mobility agents of Mobile IP, these proposals introduce several new entities in their frameworks for handling the regional tunnel management.

- **Regional Host Routing** — As an alternative to the regular IP routing, the concept of dynamic host routing can be deployed in a limited geographical area, i.e. between different subnets within the same management domain. Because of the limited scope and single ownership of the management domain, it minimises problems associated with scalability and compatibility. With this scheme, since the IP address of a mobile host has no location significance inside the domain, neither tunnelling nor address conversion is necessary during packet delivery. For example in HAWAII [28], path refresh messages are used to establish and update some host specific forwarding entries in those routers between a gateway entity called the Domain Root Router and the base station.

The role of a base station in HAWAII is twofold. Firstly, it emulates a FA for replying to Mobile IP registration messages, thus making HAWAII entities transparent to mobile hosts that use the Mobile IP protocol. Secondly, it converts Mobile IP registration updates into HAWAII refresh messages, which in turn either revives refreshes or creates forwarding entries in the routers depending on whether a subnet handover has taken place.

A similar approach is used in Cellular IP [29]. Cellular IP also requires special routers that can set up, refresh or modify their host specific forwarding entries by examining control packets. Besides control packets, routers also utilise user data packets to refresh forwarding entries. To cater for large-scale deployment, special paging packets and paging caches are integrated such that the gateway route can efficiently locate any idle mobile hosts.

- **Regional Overlay Routing** — To overcome the inadequacies of IP routing for mobile users, regional overlay routing applies an overlay model where IP packets are either segmented or encapsulated into another packet format for local delivery. Since the data forwarding mechanism is no longer IP-based, address conversions need to be done at the gateway entity and the base stations, and all mobility support issues have to be resolved by the overlay network. One example of this approach is IP over Mobile ATM [30], where IP packets are segmented into ATM cells and delivered by virtual connections between the gateway and the base station across a mobility-enabled ATM network. When the mobile host roams from one base station to another, its on-going virtual connections need to be re-routed to its latest location.

  Another example of overlay routing can be considered as a MPLS for Mobile IP [31], where IP packets are encapsulated with a label that directs the forwarding path to a base station. In fact, this is very similar to the regional tunnelling scheme mentioned earlier, except that the regional IP tunnel is now replaced by a Label-Switched-Path across the MPLS domain.

Arrival Detection

In Mobile IP, the foreign agents advertise their availability through router advertisements that are periodically transmitted. A mobile host can detect that it has moved from one subnet to another in two ways. In the first method, the mobile host uses the lifetime field of the foreign agent advertisement to refresh its association with that foreign agent. If the lifetime expires without receiving another advertisement, the mobile host will attempt to register with a new agent. In the second method, the mobile host compares the subnet prefixes of agent advertisements. If the prefixes differ from its current care-of-address, the mobile node may assume that it has moved.

Since agent advertisements are either broadcast or multicast, to minimise the radio bandwidth usage they cannot be transmitted too often (it was once per second in the original specification). However, this can make the registration process unacceptably long for both reliable data transfers using TCP and real-time communications during inter-subnet handovers [32], [33]. Although this is more an implementation rather than a design issue, it will impact the provision of real-time services in a wireless Internet environment.

Discussion of Mobility Management Issues

Future IP mobility frameworks should consider the QoS constraints of a connection more closely while handling the usual requests of handover and rerouting. Several optimisations can be done to improve the overall mobility performance. First, the handover latency should be improved significantly if a link layer hint is given to the IP layer for arrival detection. Second, it is beneficial to forward packets directly between the corresponding and mobile hosts, so that the resultant QoS path is optimised. However, it is unclear when all corresponding hosts would become mobility aware in the near future. Third, we can assign a gateway entity near the mobile host to handle micro-mobility, in which techniques such as regional tunnelling, regional host routing and regional overlay routing have been proposed. Nevertheless, it is still debatable as to which approach is the most appropriate. Moreover, if tunnelling is used to redirect packets in the future wireless Internet environment, its integration into the IntServ and DiffServ framework demands more attention and new solutions.

INCORPORATION OF USER MOBILITY INTO INTERNET QOS

Through the deployment of IntServ, DiffServ, MPLS or their hybrids, it is possible to provide deterministic QoS guarantees to the users in a wireline network. On the other hand, it is generally difficult to promise a specified level of QoS to a mobile user since there may not be enough resources in the part of the network that the mobile user is moving into. However, in order to provide real-time services, it is necessary to deliver some level of QoS guarantees to the mobile users, i.e. the mobile service models need to honour, at least statistically, the QoS agreements between a mobile host and the serving network. The approaches taken to this problem are based on advanced reservation. Advanced reservation needs to deal with two aspects, namely how to configure resources in advance, and where to pre-allocate resources for mobile users.

How to Configure Resources in Advance When Mobile

It is necessary to configure resources in advance for mobile users because resources may not be available in the location where they are moving into. Furthermore, the latency associated with new resource allocation may be prohibitively large for continuous operation of real-time services.

Recently this issue has been addressed by several proposals using a combination of Mobile IP and IntServ models. The pre-configuration strategies of these proposals can be divided into three main schemes (see Fig. 3)

- **Pre-Configured Anchor Rerouting** — In the MRSVP protocol [34], a mobile host specifies and dynamically maintains a set of locations, from which it wishes to make advance reservations to its HA (i.e. the anchor point). This set of location is known as the MSPEC. Special routing entities, called proxy agents, are provided at the locations specified in the MSPEC to make reservations on behalf of the mobile host. To allow for better link utilisation, reservations made by these proxy agents are classified as either active or passive, depending on whether the reserved resources are used for the data flow or can be temporarily borrowed by other lower priority services. Of all proxy agents associated with a MSPEC, only the one currently serving the mobile host is allowed to make active reservations. The others will remain as passive reservations until the mobile host moves into their wireless region. Similar mechanisms can also be found in other proposals such as RSVP-A [35] and Mobile Extensions to RSVP [36].

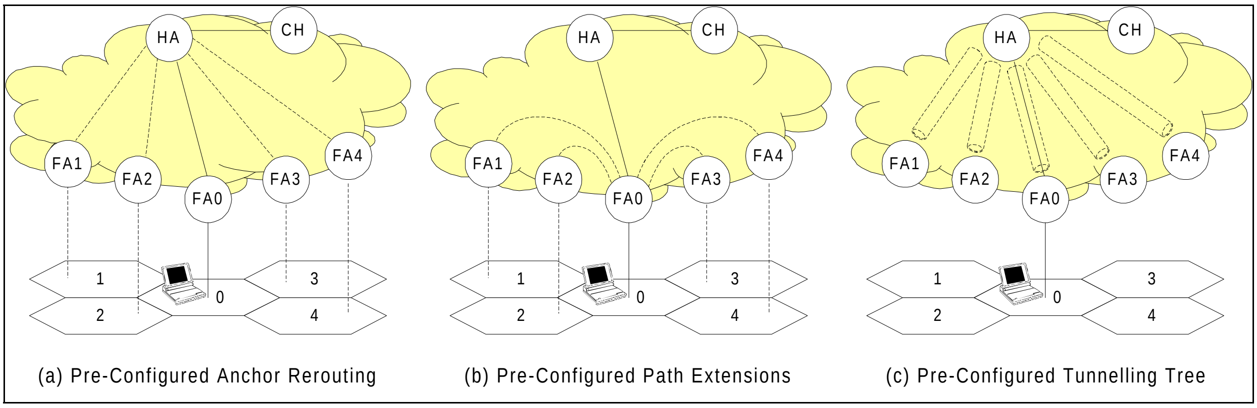
- **Pre-Configured Path Extensions** — Advanced Reservation Signalling [37] also uses a similar concept to passive reservation. However, instead of making multiple reservation paths connecting the HA with other foreign agents, the advanced reservation signalling simply extends the existing RSVP data path from the current position of a mobile host to all its adjacent base stations. Depending on whether the adjacent base station shares the same FA as the current base station or not, passive reservations can be initiated from either the current base station or the current FA to this adjacent location. As soon as the mobile host moves to one of its neighbours, one of these pre-configured path extensions becomes activated and all other extensions can be deleted.

- **Pre-Configured Tunnelling Tree** — The proposal in [38] does not rely on the notion of passive reservations. Instead, it requires RSVP-capable tunnels [24] established between the HA and other foreign agents. Unlike ordinary IP encapsulations, these RSVP tunnels are pre-provisioned with certain levels of resources while accommodating multiple end-to-end RSVP sessions. When the resources consumed by mobile hosts visiting the FA exceed the reserved amount, this FA can request an incremental block of resources to be added to the RSVP tunnel.

Where to Pre-Allocate Resources for Mobile Users

It is non-trivial to determine where to pre-allocate resources because of the difficulties of user movement prediction. Despite the challenges, many resource pre-allocation algorithms have been proposed in the literature as an attempt to safeguard the QoS agreements of mobile services. These algorithms can be classified into three main categories.

- **Neighbourhood-Based Allocation** — This scheme pre-allocates network resources between an anchor node and a set of base stations surrounding the mobile host. The number of base stations involved in the pre-allocation process depends on how far ahead in time the network is willing to support a



(a) Pre-Configured Anchor Rerouting (b) Pre-Configured Path Extensions (c) Pre-Configured Tunnelling Tree

Fig. 3 – Various Schemes to Configure Resources in Advance when Mobile

mobile service. For example, Virtual Connection Tree [39] configures resources in advance between a root switch and each base station in the management domain upon the admission of a call. This implies that the network may support this mobile host as long as it stays within the domain, but the network may also have low utilisation of resources. Advanced Reservation Signalling [37], in contrast, reserves resources only between the current location and all adjacent cells. Thus the network guarantees continuity of services after the next handover, but its further commitments are subject to successful reservations at the new neighbouring cells.

- **History-Based Allocation** — Through modelling and simulation, many proposals have shown that the user mobility history can be helpful in predicting the future movements of a mobile host. Depending on the service commitment to mobile users, these proposals pre-allocate resources at various levels in advance along the predicted path. For instance, to obtain mobility independent service guarantees, the MRSVP and other similar protocols [34], [35], [36] attempt to make resource reservations at each location a mobile host may visit during the lifetime of a session. Shadow Cluster concept [40], on the other hand, estimates future locations of a mobile host in the short term rather than long term. Based on the probabilities of a visit in the past and the current trajectory of a mobile host, network resources are reserved near its present location and along its direction of travel.

  A less ambiguous scheme can be found in the Profile Based Next-Cell Prediction [41], where resources are reserved only at the most likely visiting cell, and further QoS commitments depends on the reservation process after the next handover. It is noticeable that the further ahead a scheme tries to predict the movement, the more likely a network can support the lifetime of a session. However, this is achieved at the expense of overall network utilisation because of poor prediction accuracy.

- **Coarse-Grained Allocation** — This scheme does not reserve resources on a per-user or per-cell basis, but works on a logical model called the Virtual Bottleneck Cell [42] that treats a cluster of base stations as an aggregate virtual system. It is believed that by controlling the parameters and functions of a virtual bottleneck cell, the QoS agreements at each base station inside the cluster can be satisfied, even in environments with heterogeneous demands among base stations. However, it is not obvious how to decide the boundary of a virtual bottleneck cell, so that it is large enough for users to stay for a sufficient duration but small enough to accurately reflect the characteristics of underlying base stations. Moreover, because of its design philosophy, it is difficult to integrate this aggregated admission control with a flow-specific mobility protocol such as MRSVP [34] or Advanced Reservation Signalling [37].

Discussion of QOS Management Issues

From the above discussion, future mobile service models should make a compromise between the continuity of QoS support and the risk of over-reservations in the mobile network. The coarse-grained allocation appears to be a scalable approach to this problem, but the feasibility of aggregated functions and the scope of a virtual bottleneck cell require further investigation. The neighbourhood-based allocation is the simplest scheme to be implemented. Nevertheless, resources are likely to be over-subscribed because mobile users are seldom walking randomly in real life. By applying user mobility patterns, the history-based allocation scheme reserves resources in selective surrounding cells, and thereby attempts to minimise the probability of over-reservation in the mobile network. This view has been supported by simulation results from various sources studies [8], [40], [43], but its usefulness in real life cannot be fully verified unless the actual user mobility under wireless networks is better understood [44], [45], [46].

Implementations of advance reservation have mostly been done in IntServ rather than DiffServ environments. Although the DiffServ model is at a relatively early stage of design and definition, a simple implementation of static Service Level Agreement (SLA) will not be able to meet the dynamics of resource requirements caused by user mobility. Even with the assistance of Bandwidth Brokers [16], it is still unclear how resources should be pre-allocated in this architecture. Nevertheless, since there is no requirement for explicit signalling on a per-flow level inside a DiffServ domain, pre-configurations at the base stations or edge routers would become comparatively easier using a Bandwidth Broker [47].

CHARGING AND ACCOUNTING FOR MOBILE MULTIMEDIA SERVICES

Besides the technical perspective of integrating QoS support into IP mobility management, wireless Internet access also creates certain challenges to the current techniques of charging and accounting, which 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.

Authentication and Reconciliation of Roaming Service

As mobile users are roaming 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 consumed. The current approach of establishing this trust relationship is based on initial strong identity verification, credit history checks, and online authentication at the start of a session [48].

Furthermore, because of the liability of 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 defined time interval to the home ISP for reconciliation [49]. 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”[48] has been proposed which acts as a settlement agent, providing a common point of contact for charging and settlement services for various administrative domains.

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 basis because users are likely to ask for the best service quality at all times. Therefore, many alternative or complementary pricing schemes [50], [51] 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 them. While these schemes have their attractive properties, their successful implementation relies on the availability of efficient traffic measurement and network monitoring mechanisms. It is still an issue for further research to find out a solution that is easy to implement by ISPs and simple enough for the average user to understand.

CONCLUSION

The IETF Mobile IP protocol is a simple and scalable solution for IP mobility, but currently it has some drawbacks in the areas of performance and interoperability while serving users with high mobility and QoS expectations. This paper highlighted the problems associated with the integration of IP mobility, QoS management, and charging issues into wireless Internet environments for providing real-time services. It has outlined various potential solutions, and identified several open issues for future research.

ACKNOWLEDGEMENTS

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

REFERENCES

1. P. Bhagwat, C. Perkins, and S. Tripathi, “Network Layer Mobility: An Architecture and Survey,” *IEEE Pers. Commun. Mag.*, Jun. 1996, pp. 54-64.
2. C. Perkins, "IP Mobility Support for IPv4, revised" *IETF Internet Draft*, Oct. 1999.
3. S. Faccin, L. Hsu, R. Koodli, K. Le, and R. Purnadi, “GPRS and IS-136 Integration for Flexible Network and Services Evolution,” *IEEE Pers. Commun. Mag.*, Jun. 1999.
4. A. Furuskär, S. Mazur, F. Müller, and H. Olofsson, “EDGE: Enhanced Data Rates for GSM and TDMA/136 Evolution,” *IEEE Pers. Commun. Mag.*, Jun. 1999.
5. T. Ojanperä and R. Prasad, “An Overview of Third-Generation Wireless Personal Communications: A European Perspective,” *IEEE Pers. Commun. Mag.*, Dec. 1998.
6. R. Braden, D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: an Overview,” *IETF RFC 1633*, Jun. 1994.
7. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differential Services,” *IETF RFC 2475*, Dec. 1998.
8. C. Oliveira, J. Kim, and T. Suda, “An Adaptive Bandwidth Reservation Scheme for High-Speed Multimedia Wireless Networks,” *IEEE JSAC*, 16(6), Aug. 1998, pp. 858-874.
9. C. Perkins, “Mobile networking in the Internet,” *ACM/Baltzer MONET*, 3(4), 1998, pp. 319-334.
10. X. Xiao and L. Ni, “Internet QoS: A Big Picture,” *IEEE Network*, Mar. 1999.
11. M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, “SIP: session initiation protocol,” *IETF RFC 2543*, Mar. 1999.
12. S. Deering and R. Hinden, “Internet Protocol, Version 6 (IPv6) Specification,” *IETF RFC 2460*, Dec. 1998.

13. G. Montenegro, "Reverse Tunneling for Mobile IP," *IETF RFC 2344*, May 1998.
14. Y. Bernet, R. Yavatkar, P. Ford, F. Baker, L. Zhang, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, "A Framework for Integrated Services Operation over DiffServ Networks," *IETF Internet Draft*, Sep. 1999.
15. D. Awduche, L. Berger, D-H Gan, T. Li, G. Swallow, and V. Srinivasan, "Extensions to RSVP for LSP Tunnels," *IETF Internet Draft*, Sep. 1999.
16. B. Teitelbaum, S. Hares, L. Dunn, R. Neilson, V. Narayan, and F. Reichmeyer, "Internet2 QBone: Building a Testbed for Differentiated Services," *IEEE Network*, Sep. 1999.
17. T. Li, "MPLS and the Evolving Internet Architecture," *IEEE Comm. Mag.*, Dec. 1999.
18. C. Perkins and D. Johnson, "Route Optimization in Mobile IP," *IETF Internet Draft*, Feb. 1999.
19. D. Johnson and C. Perkins, "Mobility Support in IPv6," *IETF Internet Draft*, Oct. 1999.
20. I. Akyildiz, J. McNair, J. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Current and Future Communications Networks," *IEEE Network*, Jul. 1998.
21. R. Jain, T. Raleigh, C. Graff and M. Bereschinsky, and M. Patel, "Mobile Internet Access and QoS Guarantees Using Mobile IP and RSVP with Location Registers," in *Proc. IEEE ICC'98*, Jun. 1999.
22. E. Wedlund and H. Schulzrinne, "Mobility Support using SIP," in *Proc. ACM WoWMo'99*, Aug. 1999.
23. B. Landfeldt, T. Larsson, Y. Ismailov, and A. Seneviratne, "SLM, A Framework for Session Layer Mobility Management," in *Proc. IEEE ICCCN*, Oct. 1999 ..
24. A. Terzis, J. Krawczyk, J. Wroclawski, and L. Zhang, "RSVP Operation over IP Tunnels," *IETF RFC 2746*, Jan. 2000.
25. E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regional Tunnel Management," *IETF Internet Draft*, Aug. 1999.
26. S.F. Foo and K.C. Chua, "Regional Aware Foreign Agent Scheme for Mobile-IP," in *Proc. MoMuC'99*, Nov. 1999.
27. P. McCann, T. Hiller, J. Wang, A. Casati, C. Perkins, and P. Calhoun, "Transparent Hierarchical Mobility Agents (THEMA)," *IETF Internet Draft*, Mar. 1999.
28. R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "IP Micro-mobility Support using HAWAII," *IETF Internet Draft*, Jun. 1999.
29. A. Valkó, "Cellular IP: A New Approach to Internet Host Mobility," *ACM Comp. Commun. Review*, Jan. 1999.
30. A. Acharya, J. Li, F. Ansari, and D. Raychaudhuri, "Mobility Support for IP over Wireless ATM," *IEEE Comm. Mag.*, Apr. 1998.
31. R. Zhong, C. Tham, C. Foo, C. Ko, "Integration of Mobile IP and MPLS," *IETF Internet Draft*, Jun. 2000.
32. A. Fladenmuller and R. De Silva, "The Effect of Mobile IP handoffs on the Performance of TCP," *ACM/Baltzer MONET*, 4(2), 1999, pp. 131-135.
33. N. Fikouras, K. El Malki, S. Cvetkovic, and C. Smythe, "Performance of TCP and UDP during Mobile IP Handoffs in Single-Agent Subnetworks," in *Proc. IEEE WCNC'99*, Sep. 1999.
34. A. Talukdar, B. Badrinath, and A. Acharya, "MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts," to appear in *ACM/WINET*, May 1999.
35. A. Pajares, N. Neriér, L. Wolf and R. Steinmetz, "An Approach to Support Mobile QoS in an Integrated Services Packet Network," in *Proc. IQWiM Workshop*, Apr. 1999.
36. D. Awduche and E. Agu, "Mobile Extensions to RSVP," in *Proc. IEEE ICCN'97*, 1997.
37. I. Mahadevan and K. Sivalingam, "Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ," to appear in *ACM/WINET*, Jul. 1999.
38. A. Terzis, M. Srivastava, and L. Zhang, "A Simple QoS Signaling Protocol for Mobile Hosts in the Integrated Services Internet," in *Proc. IEEE INFOCOM'99*, Mar. 1999.
39. A. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks," *IEEE JSAC*, 12(8), Oct. 1994, pp. 1365-1374.
40. D. Levine, I. Akyildiz, and M. Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept," *IEEE/ACM Trans. on Networking*, 5(1), Feb. 1997, pp. 1-12.
41. V. Bharghavan and J. Mysore, "Profile Based Next-Cell Prediction in Indoor Wireless LANs," in *Proc. IEEE SICON'97*, Apr. 1997.
42. R. Jain, B. Sadeghi, and E. Knightly, "Towards Coarse-Gained Mobile QoS," in *Proc. ACM WoWMoM'99*, Aug. 1999.
43. P. Ramanathan, K. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless Networks," *IEEE JSAC*, 17(7), Jul. 1999, pp. 1270-1283.
44. J. Chan, S. Zhou and A. Seneviratne, "A QoS Adaptive Mobility Prediction Scheme for Wireless Networks," in *Proc. IEEE GLOBECOM'98*, Nov. 1998, pp. 1414-1419.

45. J. Chan and A. Seneviratne, "A Practical User Mobility Prediction Algorithm for Supporting Adaptive QoS in Wireless Networks," in *Proc. IEEE ICON'99*, Sep. 1999, pp. 104-111.
46. J. Chan, B. Landfeldt, A. Seneviratne and P. Sookavatana, "Integrating Mobility Prediction and Resource Pre-allocation into a Home-Proxy Based Wireless Internet Framework," to appear in *Proc. IEEE ICON 2000*, Sep. 2000.
47. J. Chan, B. Landfeldt, R. Liu, and A. Seneviratne, "A Home-Proxy Based Wireless Internet Framework in Supporting Mobility and Roaming of Real-Time Services," submitted for review, Jul. 2000.
48. S. Glass, T. Hiller, S. Jacobs and C. Perkins, "Mobile IP Authentication, Authorization, and Accounting Requirements," *IETF Internet Draft*, Mar. 2000.
49. J. Arkko, P. Calhoun, P. Patel, and G. Zorn, "DIAMETER Accounting Extension," *IETF Internet Draft*, Mar. 2000.
50. M. Borella, V. Upadhyay, and I. Sidhu, "Pricing Framework for a Differential Services Internet," *European Tran. on Telecomm.*, 10(3), May 1999, pp. 275 – 288.
51. G. Fankhauser, B. Stiller, and B. Plattner, "Arrow: A Flexible Architecture for an Accounting and Charging Infrastructure in the Next Generation Internet," *Baltzer Netnomics*, 1(2), Feb. 1999, pp. 201 – 223.

THE AUTHORS

***Jonathan Chan** received his Bachelor of Electrical Engineering with honours at the University of New South Wales in 1990. Since then he has worked at Macau Telecom, Macau Electricity, Andersen Consulting (Sydney), and Dow Jones Telerate (Australia). He is currently completing his Ph.D. at the same University. His main areas of research are in QoS management and mobile computing. He may be contacted through Email - jchan@ee.unsw.edu.au*

***Aruna Seneviratne** is Professor in the School of Electrical Engineering and Telecommunications, University of New South Wales. He completed his Ph.D. at the University of Bath, UK, in 1983. Since graduating he has held academic appointments at the University of Bradford (UK), Curtin University, the Australian Defence Force Academy (UNSW) and the University of Technology, Sydney. Outside academia he has worked at the Standard Telecommunication Laboratories (UK), Muirhead (UK), and Telecom Australia. He may be contacted through Email - aruna@ee.unsw.edu.au*

***Björn Landfeldt** was born in Stockholm, Sweden. He started studying electrical engineering at the Royal Institute of Technology, Stockholm, and completed his Ph.D. in Electrical Engineering at the University of New South Wales in 2000. Currently he is with Ericsson Research, Networks and Systems, Sweden and his areas of interest are mobile networking and QoS management. He may be contacted through Email - Bjorn.Landfeldt@era.ericsson.se*