Mobile ATM: A Generic and Flexible Network Infrastructure for 3G Mobile Services

Jun Li, Roy Yates, and Dipankar Raychaudhuri

Abstract: **This paper presents the concept of "mobile ATM," a proposal for third-generation (3G) mobile communication network infrastructure capable of supporting flexible evolution of radio technologies from today's cellular and data services towards future wireless multimedia services. Mobile ATM provides generic mobility management and QoS-based transport capabilities suitable for integration of multiple radio access technologies including cellular voice, wireless data, and future broadband wireless services. The architecture of a mobile ATM network is outlined in terms of the newly-defined "W-UNI" interface at the radio link and "M-UNI" and "M-NNI" interfaces within the fixed network. The concept of a "proxy-UNI" interface which supports unified access for WATM and non-ATM mobile terminals through corresponding interworking functions (IWF) is explained, leading to an understanding of how different radio access technologies are supported by the same ATM-based core network infrastructure. Generic mechanisms for handoff and location management within the core mobile network are discussed, and related protocol extensions over the "W-UNI" and "M-UNI/NNI" interfaces are proposed. The issue of "crossover switch (COS)" selection in mobile ATM is considered, and a unified handoff signaling syntax which supports flexibility in COS selection is described. Typical signaling sequences for call connection and handoff using the proposed protocols are outlined. Experimental results from a proof-of-concept mobile ATM network prototype are presented in conclusion.**

Index Terms: **Third generation mobile, mobile ATM, mobility management.**

I. INTRODUCTION

This paper presents the concept of "mobile ATM," a proposal for third-generation (3G) mobile network infrastructure. Mobile ATM was proposed in [1]–[4] as a potential core mobile network solution for IMT-2000 scenarios involving a gradual migration from today's cellular and data services towards future wireless multimedia. The basic idea of mobile ATM is to use a generic, mobility-enhanced broadband network infrastructure to provide QoS-based transport for wireless services, thus taking it possible for multiple radio technologies to coexist on the same common infrastructure. It is noted that a similar generic architecture for 3G networks can also be developed using IP/MPLS with LDP/CR-LDP signaling rather than ATM as the foundation. The idea is to enhance the switching/routing capabilities of the core mobile network to incorporate generic mobility management capabilities such as location management and handoff. This facilitates a decoupling of radio access and core network technologies, thus permitting gradual evolution of radio technology towards broadband services without constant changes to the infrastructure.

The deployment of the IMT-2000 network based on the third-generation (3G) mobile communication system is expected in next two to three years. Research and standardization activities have been conducted intensively in recent years [5]–[9]. A third-generation core mobile network will be required to support the following major requirements:

1. integrated, QoS-based transport capabilities suitable for the range of connection-oriented and connectionless voice, data and multimedia services associated with 3G scenarios;
2. generic mobility management capabilities including location management and handoff, with performance appropriate for voice, data and multimedia services;
3. flexible architecture capable of supporting a mix of radio access technologies without requiring separate overlay networks in each case;
4. scalability both in terms of address space and cost/performance necessary to support the fast-growth in mobile services;
5. architectural alignment with future broadband networks so as to provide seamless services to users with a mix of fixed and wireless service requirements.

The mainstream 3G network architecture follows proposals in 3GPP (third generation partnership project) [10], which is evoluted from GSM and GPRS networks, containing a circuit switching (CS) domain and a packet switching (PS) domain. A third domain, IP multimedia domain is proposed but nothing has been yet defined inside it. People believe this three domain architecture can evolve to an all IP network architecture for IMT-2000 to meet the above requirements. However, it is still unclear how an all IP solution can be achieved. Mobile ATM is an all ATM solution. Conceptually, it is similar to the all IP solution as mobility functions are integrated in the network layer. The mobile ATM can be used as the access network architecture for IMT-2000 and is a good reference model for the all IP solution.

The mobile ATM approach described in [3], [4], [11] substantially meets all the above requirements. Although ATM is designed for supporting QoS and connection-oriented services, it can easily be extended to provide connectionless services for IP packets with proper IP over ATM configurations. Given that ATM is being used extensively as a foundation for broadband and Internet access networks, we believe that this approach has

Manuscript received January 17, 2000.

J. Li is with NEC USA, C&C Research Laboratories, e-mail: lijun@ccrl.nj.nec.com.

R. Yates is with the WINLAB, Rutgers University, e-mail: ryates@ccrl.nj.nec.com.

D. Raychaudhuri is with NEC USA, C&C Research Laboratories, e-mail: ray@ccrl.nj.nec.com.

several practical advantages. It is remarked here that our presentation of the mobile ATM approach should be viewed as an architectural idea and study rather than a definitive proposal for IMT-2000 infrastructure which will obviously require much further standards and development work before being finalized.

The rest of this paper is organized as follows. Section II presents an outline of the mobile ATM network architecture when applied to multiple radio access technologies. This section identifies the new "W-UNI," "M-UNI," and "M-NNI" interfaces in a mobile ATM network, along with the interworking function (IWF) necessary to convert between specific radio protocols and the common core network protocols. In Section III, we provide a description of mobile ATM network functionality including location management and handoff. Alternatives methods for identifying a "cross-over switch (COS)" for handoff are discussed. Section IV explains the protocol syntax and typical signaling sequences required for connection establishment and handoff, and includes some typical examples of the protocol's operation. Finally, Section V provides a summary of experimental experience in building a proof-of-concept mobile ATM prototype that was developed at NEC USA's C&C Research Lab in Princeton, NJ during 1996–98. Sections VI, VII, and VIII contain concluding remarks, acknowledgements and references respectively.

II. MOBILE ATM ARCHITECTURE

The goal of a mobile ATM network is to provide transport for ubiquitous mobile services with different radio technologies and QoS requirements. Since mobile ATM is a compatible upgrade to a conventional ATM network, it can support both mobile and fixed hosts on the same infrastructure. In order to support mobile services, the ATM network protocols must be enhanced to support three basic mobility functional components: wireless access, location management, and handoff control.

A regular ATM network has ATM switches and ATM hosts. In a private ATM network, a user to network interface (UNI) between an ATM host and an ATM switch, and a private network to network interface (PNNI) between any two ATM switches have been defined. The standard ATM signaling specifications define the signaling messages over UNI and PNNI interfaces and the signaling procedures for connection establishment and release.

In addition to regular ATM switches, a mobile ATM network has radio access points (basestations) for mobile host accessing. The interface between a mobile host and a basestation is a UNI interface, if the radio technology is wireless ATM. We define it as a *W-UNI* interface, i.e. a wireless UNI inerface.

Within the fixed part of the mobile ATM network, the regular ATM interfaces are enhanced with the mobility support. We call them *M-UNI* and *M-PNNI* interfaces. They stand for *mobility-enhanced UNI and PNNI*, extended from regular ATM UNI and PNNI interfaces.

A mobile ATM network architecture based on the newly defined interfaces is shown in Fig. 1. It contains ATM switches and basestations. It can support various kinds of mobile hosts. In a basestation (BS), first, the interworking functions (IWFs) will map a specific non-ATM radio protocol (such as GSM, CDMA, WLAN, etc.) to the ATM network protocol, then, an ATM proxy signaling agent (PSA) will provide a UNI signaling function on behalf of a mobile terminal. It is noted here that this mobile ATM network architecture is fully distributed and has no centralized devices such as the "mobile switching center(MSC)" typically associated with second generation cellular systems. Mobility is handled by ATM switches and location servers (LS) in an integrated manner, providing important scalability and cost/performance advantages over earlier approaches.

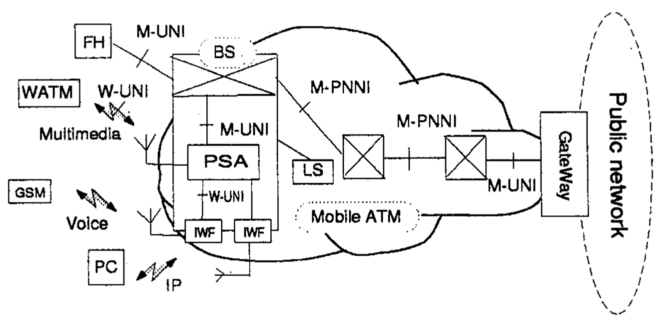


Fig. 1. Mobile ATM system architecture.

A mobile terminal can obtain mobility support from W-UNI, M-UNI and M-PNNI signaling if it roams within a mobile ATM cloud. If the mobile terminal requires a service function outside the mobile ATM cloud, such as a call to the public network, a service specific gateway interworking function must be provided (GateWay). For example, a 3G mobile terminal that makes a phone call to a regular phone in the PSTN network must go through a mobATM-PSTN gateway.

The installation of mobile ATM networks can start from a local area network, such as a university campus, hospital, airport or factory, providing both fixed services and various kinds of mobile services. For example, GSM/GPRS can use the mobile ATM network as an office LAN and the gateway IWF as a PBX, which localize the internal calls and multiplex the external calls. By leasing a small number of radio frequencies and/or operating at a lower power, the service throughput per unit cost can reach to a much higher rate. Such a mobile ATM network can also be used as office LAN with IPOA to provide Internet access.

III. MOBILE ATM FUNCTIONS

The features of 3G mobile networks require the access network to support wireless access and mobility management independent of the type of service. Since 3G systems integrate both voice and data services while evolving toward multimedia, it is desirable that the mobile ATM network, as the core access network, has sufficiently generic interfaces and functionalities to be independent from the type of services as well as the type of radio access technologies. This requirement should be satisfied by each functional unit of the mobile ATM network, including wireless access, location management and handoff control.

A. Wireless Access

The wireless access interface of a mobile ATM network must accept different radio access protocols, both ATM based and non-ATM based, in a uniform manner. We propose the follow-

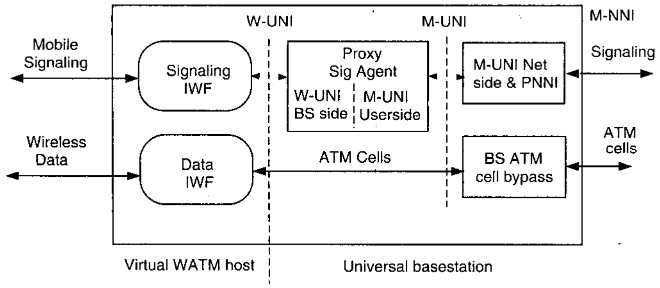


Fig. 2. Wireless access of mobile ATM.

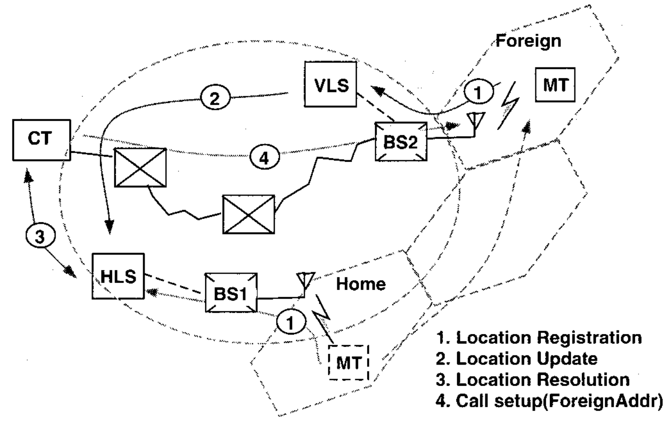


Fig. 3. Location management in wireless ATM network.

ing components at the W-UNI interface on basestations to meet these requirements.

A.1 Proxy Signaling Agent (PSA)

In a mobile ATM network, one basestation may serve multiple mobile terminals at the same time, which implies multiple W-UNI interfaces should be dynamically created at the mobile terminal's registration time. As shown in Fig. 2, we propose to have a *Proxy Signaling Agent* (PSA) in each basestation. It creates one proxy UNI object (ProxyUNI) for every mobile terminal at its registration time. In this way, regardless of whether the mobile terminal is ATM or not, an ATM signaling entity acts on behalf of it. The design of signaling protocol within the fixed network can then be unified. More importantly, the protocol operations over wireless interface and within the fixed network can be separated, which is desirable because the control requirements are different between radio links and fixed links. We will see that this approach leads to a clean design of ATM signaling protocol extensions for mobility management.

A.2 Interworking Function (IWF)

If a non-ATM radio access technology is used, an interworking function (IWF) must be supported at the W-UNI interface. This can be achieved by a universal basestation with variety of radio access plug-ins. As shown in Fig. 2, a *signaling interworking function* (SigIWF) is required to map mobile signaling from the mobile service specific radio format to the W-UNI format, and a *Data interworking Function* (DataIWF) is required to convert radio format data to ATM cells.

B. Location Management

In a network (ATM or otherwise), each end-user (terminal) is identified by a network address. This address serves two purposes: 1) as an endpoint identifier for the terminal, e.g. transport protocols use such endpoint identifiers to establish transport-level connections and 2) as a location identifier, i.e. the address often also implicitly identifies the network route to reach the endpoint.

Traditionally, a single identifier served both purposes since the location of a terminal relative to the network does not change.

With the introduction of mobile terminals, it is no longer possible for a single identifier to serve both purposes. The location of a mobile terminal will change often, and so will its location identifier. On the other hand, its endpoint identifier, i.e. its name, should remain unchanged so that other terminals can identify the mobile terminal regardless of its attachment point to the network.

B.1 Home and Foreign ATM Addresses

In a mobile ATM network, for each mobile terminal, there are two ATM addresses associated with it. One is the *Home* ATM address which serves as the mobile terminal's host identification. The other is the *Foreign* ATM address which serves as the mobile terminal's location identification.

B.2 Home and Visit Location Servers

An ATM address of a mobile terminal is assigned by a location server (LS). The location server that covers a mobile terminal's home area is called home location server (HLS) and assigns a home ATM address for the mobile terminal. The location server that covers a mobile terminal's current visit area is called visit location server (VLS) and assigns a foreign ATM address for the mobile terminal. In a mobile ATM network, in general, there is no restrictions on where location servers are located and how large area each location server should cover. The location management scheme can be either centralized, distributed or in between, i.e. hierarchical. The extreme examples are 1) completely centralized: only one LS per mobile ATM cloud and 2) fully distributed: one LS per basestation.

In general, location management function can be illustrated as in Fig. 3. An MT registers to a HLS or a VLS, depending whether it is at home. If it is away from its home, a location update will be done from VLS to HLS. When a CT call for the MT, it gets the MT's location resolved by consulting the associated HLS. The CT will call directly to MT's current location based on location resolution.

C. Handoff Control

In a connection-oriented network like ATM, the services attached to a mobile terminal are provided through connections with specific routes in the network. If the mobile terminal changes its access point, the network services will lose their

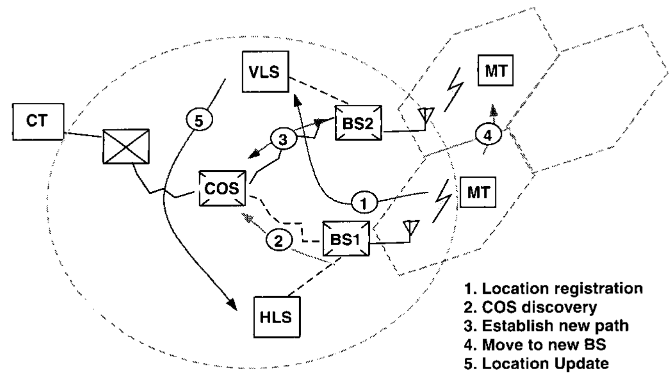


Fig. 4. Handoff control in mobile ATM network.

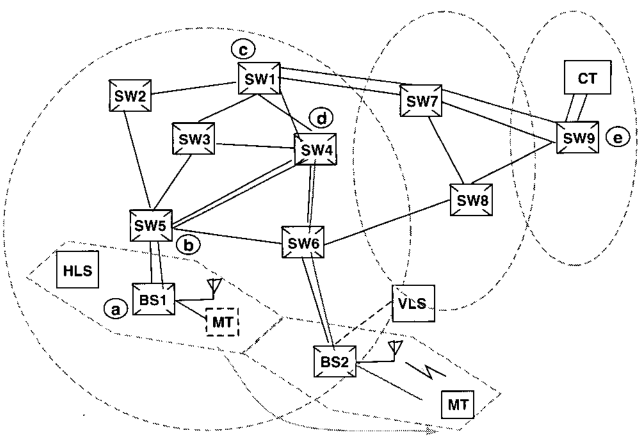


Fig. 5. Alternative path rerouting algorithms in mobile ATM networks.

contact to the mobile terminal unless those connections can be rerouted to the mobile terminal through the new access point. Handoff control is a process to seamlessly reroute the mobile terminals' connections. The essential concerns in designing handoff control are 1) processing delay and 2) the efficiency of network resource utilization after handoff.

C.1 Incremental Path Re-routing

Handoff is usually a more frequent event than call setup in a micro cellular system. Thus it is desirable for a handoff control process to use a partial rerouting instead of connection re-establishment, especially when the call is long distance since handoff is always local. For handoff control, an incremental (dynamic) path re-routing mechanism is commonly agreed in the standards and research community.

C.2 Alternative Path Re-routing Algorithms

In the incremental path re-routing approach, there is an anchor switch from which the connection path is re-routed. We define this switch as crossover switch (COS). The criteria of selecting a COS can lead to different performance for the path re-routing process, i.e., different type of COSs lead to different path re-routing algorithms. A path rerouting algorithm can be based on the following criteria [12].

1. Latency of completing handoff path rerouting,
2. link utilization of rerouted connections,
3. QoS guarantees during the handoff procedure,
4. QoS guarantees of rerouted connections, and
5. complexity and reliability of the handoff procedure.

Based on various performance concerns, there have been many handoff path rerouting proposals [3], [13]–[17]. All of proposals use different terms to define each type of path rerouting algorithm. We can summarize these approaches as follows:

Path Extension: The simplest way to reroute the connection path is to extend the path from the old basestation to the new basestation, as shown in Fig. 5 where the COS is labelled **a**.

First Divergence Rerouting: With a little modification from path extension, the first divergence rerouting algorithm selects the first switch from which the extended path to the new basestation diverges from the original connection path. As shown in Fig. 5, the COS is **b**.

This approach sacrifices the simplicity of path extension algorithm but improves resource utilization by removing loops in extended connections.

Fixed Anchor Rerouting: Similar to GSM, a fixed anchor may be chosen for path rerouting. The anchor switch is fixed during the lifetime of the call. For example, in a PNNI hierarchical network, the COS can be the first switch at border of a peer group that a connection traverses. As shown in Fig. 5, the COS is **c**. There is no need to select a COS for every handoff, however, it may lead to similar performance problem as the path rerouting in GSM or path extension algorithm.

Shortest NewPath Rerouting: It is possible to configure the shortest new path to the old path. As shown in Fig. 5, a COS may be at **d**, which gives the shortest new path. Since the major delay in path rerouting is contributed by the new path establishment, the shortest new path rerouting may give the lowest latency in the handoff control process.

Last Divergent Rerouting: This algorithm targets optimizing the rerouted connection path for each handoff. As shown in Fig. 5, the divergent point between the new path from the new basestation (BS2) toward CT and the original path is the selected as COS **e**. Although this algorithm gives the best performance in terms of resource utilization, the latency of the handoff control process could be large since the variation of the new path length can be large.

A mobile ATM network can employ one or more alternative path rerouting algorithms. For those services which require fast response, path extension or first divergence rerouting algorithm may be used. For those services with long holding time and infrequent handoff, optimal path rerouting may be considered.

IV. PROTOCOL DESIGNS

Many mobility management protocols for mobile ATM networks have been proposed [3], [12]–[14], [17]–[22] since the early 90s when wireless ATM was first proposed [23]. These protocols are each provided to satisfy the designer's specific performance criteria, such as complexity, delay, latency and/or network resource utilization. However, in IMT-2000 networks, the service performance criteria can be quite diversified. A 3G mobile terminal may have phone calls, web browsing and/or video conferences. The services may be delay sensitive or loss sensitive. Therefore, the concerns of performance and complexity for each application can be quite different. Our goal is to design a unified protocol for mobility management functions which are independent of service specific performance and complexity requirements.

A. W-UNI Interface

A mobile terminal needs to gain access to the network at W-UNI interface either for new registration, for roaming or for handoff.

A.1 Access Registration

In regular ATM networks, an ATM host uses ILMI (interim local management interface) query to get access to the network at a UNI interface. Similarly, in a mobile ATM network, a mobile terminal will also use ILMI queries to access to the network at a W-UNI interface. However, a mobile terminal's registration to the network is done through the proxy UNI object generated by the PSA.

In cases of new registration and roaming, the same registration request can be used as in regular ATM networks. However, in the case of handoff, the protocol of registration need to be extended to meet the special QoS concerns of handoff. The reason is as follows.

If a mobile terminal loses its connectivity before the handoff process starts, it must perform a *forward handoff* scheme, which directly registers to the new basestation. Otherwise, it is possible to realize seamless handoff, i.e., a mobile terminal can keep the connectivity to the old basestation while registering to a new basestation. The de-registration from the old basestation happens only after the registration to the new basestation is completed. This requirement leads to a *backward handoff* scheme, which performs registration to the new basestation via connectivity of the old basestation. For a successful handoff, the registration to the new basestation means not only radio access but also re-routing of the existing connections through the new basestation. For this reason, signaling for backward handoff at the W-UNI interface is more than just de-registration at the old basestation and registration at the new basestation.

A.2 Signaling Extensions

In Table 1, new messages, equivalent to new contents in ILMI queries, are defined to represent the transition states during a backward handoff process.

The main idea of introducing transition states is to keep the current connectivity while connection path re-routings are carried on for a handing off mobile terminal.

HO_REQUEST is essentially a registration request to the new basestation. However, since it is a handoff action, the registration will be done only if the handoff path re-routing procedure is completed. HO_RESPONSE will be sent back with the resource availability (if there is enough radio bandwidth) on the new basestation and HO_READY will be sent back with the resource availability within the network (if the path re-routings are successful).

Table 1. Signaling messages over W-UNI interface.

Message	Direction	Process state
HO_REQUEST	MT to ProxyUNI	handoff initiation
HO_RESPONSE	ProxyUNI to MT	handoff proceeding
HO_READY	ProxyUNI to MT	reroute completed
HO_COMFIRM	MT to ProxyUNI	start path replacement
HO_COMPLETE	ProxyUNI to MT	mobile can leave cell
HO_VALIDATE	MT to ProxyUNI	mobile enter cell
HO_VALID	ProxyUNI to MT	handoff complete

Once handoff through the new basestation is guaranteed with enough resource on radio and network links, the mobile terminal can determine the time to move to the new basestation. HO_CONFIRM says that the mobile terminal wants to move and HO_COMPLETE says that the mobile terminal can do so.

Finally, when the mobile terminal switches to the new frequency provided by the new basestation, it sends a HO_VALIDATE to confirm its presence at the new basestation. A HO_VALID is replied to complete the whole handoff control procedure.

B. M-UNI and M-PNNI Interfaces

With the introduction of interworking functions (IWFs) and proxy signaling agent (PSA) at a W-UNI interface, the protocol extension for M-UNI and M-PNNI interfaces can be designed without considering radio links.

The protocol extension needs to address two major problems, one is reachability and the other connectivity. The first problem is in regard to location management of mobile terminals with the main concern on how a caller can reach a callee transparent to its location changes. The second problem related to handoff control is concerning how efficiently a path re-routing can be done. Our protocol design aims to answer these questions. So we have proposals of using an integrated location resolution scheme and a unified handoff path re-routing scheme.

B.1 Integrated Location Resolution

As we mentioned in mobile ATM architecture, location servers for mobile terminals can be centralized, distributed or hierarchical. At M-UNI interface, every mobile terminal, represented by a proxy UNI object, is assigned with a *Home ATM Address* by its home location server (HLS) when it accesses via a basestation to its home area. It will need to obtain a *Foreign ATM address* by a visitor location server (VLS) when it accesses through a basestation outside its home area. To obtain an ATM address from a LS (HLS/VLS), the proxy UNI object needs to have a registration procedure with M-UNI and M-PNNI signaling messages. The location registration within the mobile ATM network follows the access registration at W-UNI interface. Whenever a mobile terminal changes its access basestation, a new registration is required to the current VLS. A new ATM address may or may not be assigned if the VLS is unchanged, but the M-PNNI routing information about the mobile terminal is updated for each registration regardless of roaming or handoff. Only if the mobile terminal's foreign ATM address is changed, an M-PNNI signaling message is sent from the VLS to the HLS to update the mapping between its home ATM address to its current foreign ATM address.

Table 2. Control functions for location management.

Function	Message	Information
Registration	UNI/NNI LOC_REG	ATM address & routing
Resolution	UNI/NNI SETUP/RELEASE	cause: MOBILE_AWAY and Foreign Address IE
Update	UNI/NNI LOC_UPDATE	ATM address mapping

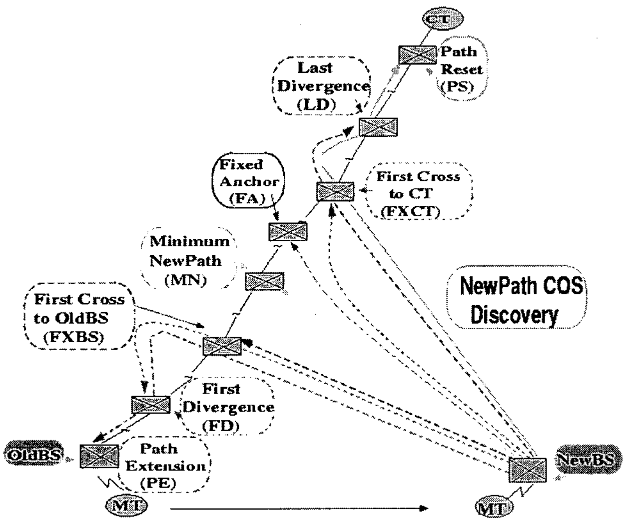


Fig. 6. Crossover switch (COS) discovery in handoff control.

The location information of a mobile terminal maintained by HLS and VLS is used for PNNI routing to determine where to send signaling messages in the call control process. When a mobile terminal is called by a caller, its home ATM address is first used. A SETUP message will be sent to the area covered by HLS. If the mobile terminal is away from its home area, the HLS must have updated the PNNI routing table where the mobile terminal is marked as unreachable in this subnet and its current Foreign ATM address is available. A RELEASE message will be sent back to the caller with an IE containing the Foreign ATM address of the mobile terminal. Then another SETUP message will be sent to the mobile terminal's current VLS area. The connection path can be established from the caller directly to the mobile terminal in its current foreign location. It is possible that there is a race condition that the mobile terminal has moved away again when the SETUP message comes to its VLS area. This problem can be solved again by a RELEASE message containing an IE either with the mobile terminal's current foreign ATM address or simply indicating mobile terminal is unreachable in this area. The PNNI signaling at the caller side will determine whether another attempt to the mobile terminal's HLS is necessary. The timer at the UNI interface can limit how many times of location resolution a call setup can attempt. The M-UNI and M-PNNI signaling extensions for location management are summarized in Table 2.

B.2 COS Discovery

As we have pointed out in last section, there are many ways to reroute a call path, each corresponding to a path rerouting algorithm. A path rerouting algorithm can be specified by the *crossover switch* (COS), that is the anchor switch of both the old path and the new path under a given criterion.

Table 3. Destination address and COS discovery function in SETUP message.

COS	Destination Address	AM_I_COS()
Path Extension (PE)	OldBS	dummy
First Divergence (FD)	OldBS	check routes of Connections /w GCID
First Cross to OldBS(FXBS)	OldBS	check GCID
Min. NewPath (MNP)	multiple switches on OldPath	check GCID compare distances
Fixed Anchor (FA)	a fixed ATM switch	dummy
First Cross to CT(FXCT)	CT	check GCID
Optimal (OP)	CT	check routes of Connections /w GCID
Path Reset (PRST)	CT	dummy

As shown in Fig. 6, a COS can be chosen at any node along the route from the old basestation (oldBS) to the corresponding terminal (CT). Since a COS is the cross point of both the old path and the new path, the COS can be discovered by signaling from either the oldBS or the newBS. In Fig. 6, COS discovery through new path is illustrated. A signaling message is sent from the newBS towards a destination according to given path rerouting algorithm. When the message traverses a mobile enhanced ATM switch, a function called AM_I_COS() is executed to check the COS condition. The COS condition can vary upon the selected path rerouting algorithm, which is based on the service or network performance criteria. For example, Path Extension(PE) algorithm is selected for simplicity and Last Divergence(LD) algorithm is selected for network resource utilization.

If COS discovery signaling is from the newBS, a global connection identifier (GCID) must be introduced for each handoff connection in order to determine the connection pass through any specific ATM switch. In current ATM UNI/NNI signaling specifications, connection ID has only local significance, meaning it is unique only for one switch. With GCID, a connection can be identified from a path rather than the original connection path. The GCID was introduced in a ATM Forum wireless ATM group meeting [24] for handoff control. A similar ID (NCCI, network correlated connection ID) was introduced in PNNI 2.0 [25], which is used for link failure recovery. Since handoff is considered as a process to recover the radio link failure, GCID performs a similar function as NCCI.

The signaling syntax of the new path COS discovery can be realized in a unified way. The differences in the COS condition (the choice of path rerouting algorithm) can be indicated in the COS discovery signaling message with 1) the destination address in calling party's address IE and 2) the type of COS in a handoff control IE (HCIE). Depending on the type of COS, the COS condition function AM_I_COS() can have different criteria to make decision. In Table 3, we list destination addresses and COS condition functions AM_I_COS() for the path re-routing algorithms with different COS types. Based on this, a unified signaling syntax for handoff control in the network can be designed.

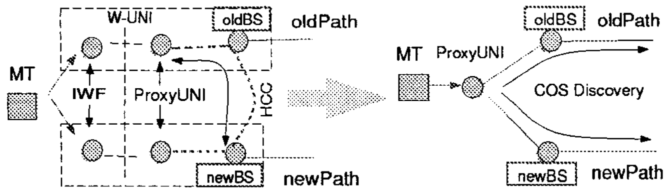


Fig. 7. ProxyUNI and virtual MT to the mobile ATM network.

Table 4. Signaling messages for path rerouting.

Message	Direction	Function
SETUP	ProxyUNI to COS	establish new subpath
CONNECT	COS to ProxyUNI	establish new subpath
(HO_CONFIRM)	ProxyUNI to COS	trigger path change at COS
RELEASE	COS to ProxyUNI	release old subpath

B.3 Unified Handoff Path Re-routing

Usually, a handoff request is initiated by a mobile terminal. When a mobile terminal detects fading of the link quality to the current basestation, it may request a handoff initiation to the oldBS or directly ask to register to a newBS. To initiate handoff from the old basestation, the handoff is a backward handoff, otherwise, from the new basestation, it is a forward handoff.

Whether a handoff is backward or forward , the mobile terminal may need to contact to both basestations during the handoff control process for either operation or performance reasons. For examples, to keep track of the GCID, a mobile terminal may need its proxy UNI object migrating to the new basestation even if forward handoff is performed. To reduce cell loss rate in transition period, the transmission through the old basestation must continue while the new subpath is being established through the new basestation. For this reason, we propose to have a handoff control channel (HCC) between the old basestation and the new basestation. As shown in Fig. 7, a mobile terminal can be virtually connected to both basestations at the same time through HCC.

This could simplify the handoff path re-routing procedure. Regardless of forward or backward handoff, the same description is given as follows.

Starting at the proxy UNI object, a handoff path rerouting procedure is initiated over the M-UNI and M-PNNI interfaces. The enhanced M-UNI/M-PNNI is shown in Table 4. The basic path re-routing function is performed by existing messages SETUP/CONNECT/RELEASE with new IE extension. And in addition, at mobility enhanced ATM switches, a PNNI signaling module called AM_I_COS() should be defined and executed whenever SETUP message traverses them. A new signaling message HO_CONFIRM is introduced as an option to delay the replacement of the old data path by the new sub-path. This option can reduce the data loss, that is, only if the mobile terminal is ready to move to the new basestation, the old sub-path is replaced by the new sub-path. The mobile terminal can keep the service on through the old path until all VCs' path re-routings are completed.

C. Overall Signaling Syntax

Based on the signaling extensions over W-UNI, M-UNI, and M-PNNI interfaces, we give a signaling sequence as shown in Fig. 8. New messages and modified messages are shown in time sequence. This signaling sequence is aligned with a four phase signaling syntax.

1. **Register Resource on NewBS**: The first phase is performed by the HO_REQUEST and HO_RESPONSE messages at W-UNI interface. After the Proxy UNI object receives a HO_REQUEST from W-UNI interface, it will start a registration process to the new basestation on behalf of the mobile terminal through ILMI signaling (create or update MIBs). The process needs not be a full registration process as when the mobile terminal boots up, but the resources availability must be checked before data VC's handoff. New ILMI objects are defined for all new resources, such as ATM address, signaling PVCs, radio frequency and bandwidth. After registration, HO_RESPONSE is sent back to mobile ATM terminal through W-UNI interface.
2. **Re-route Data VCs**: The second phase is performed by the SETUP and CONNECT messages within the fixed network. In this phase, SETUP/CONNECT with new information elements (IEs) are used for the new sub-path establishment and COS discovery for each data VC's path rerouting. A GCID IE is used for COS discovery which can identify a connection from a path different from the original path. An HCIE (handoff control IE) is used in the SETUP and CONNECT messages to indicate the message is for handoff control. The HCIE can carry the algorithm related information, such as COS type and rerouting options. COS discovery is performed in this phase while the newPath is being set up. As shown in Fig. 8, the rerouting for all connections for the MT can be done in parallel. And after all connections have been rerouted, a CONN_READY message is set to the MT to indicate all connections are ready.

   We may define the duration of the first and second phases as *Handoff Latency*, which is the time from a MT requesting a handoff to the MT being ready to handoff. The major contributor to handoff latency is the signaling process for setting up new paths. During the first and second phases, for backward handoff, the MT is still connected to the old basestation so the service can be continued. However, for forward handoff, the MT may lose its connection right from the beginning, so the handoff latency is critical to maintain the quality of service.
3. **Switch Re-routed Path on COS**: The third phase is introduced in order to reduce the data loss. The handoff connection path will not be replaced by the new path until a HO_CONFIRM message is received by the COS. With this optional phase, the data transmission can continue while the new path is established. Only after HO_CONFIRM is sent out, the MT can be ready to handoff to the new basestation. Fig. 8 shows a HO_READY message which indicates receiving of HO_CONFIRM by the old basestation.
4. **Active Re-routed Path at NewBS**: The fourth phase, performed by new messages—HO_VALIDATE / HO_VALID, is used for assuring the completion of the connection path before data transmission starts on the newBS. If data loss is not an issue at the cell level, this phase can also be omitted by letting the MT and the newBS transmit data immediately after the data VCs are opened on either side.

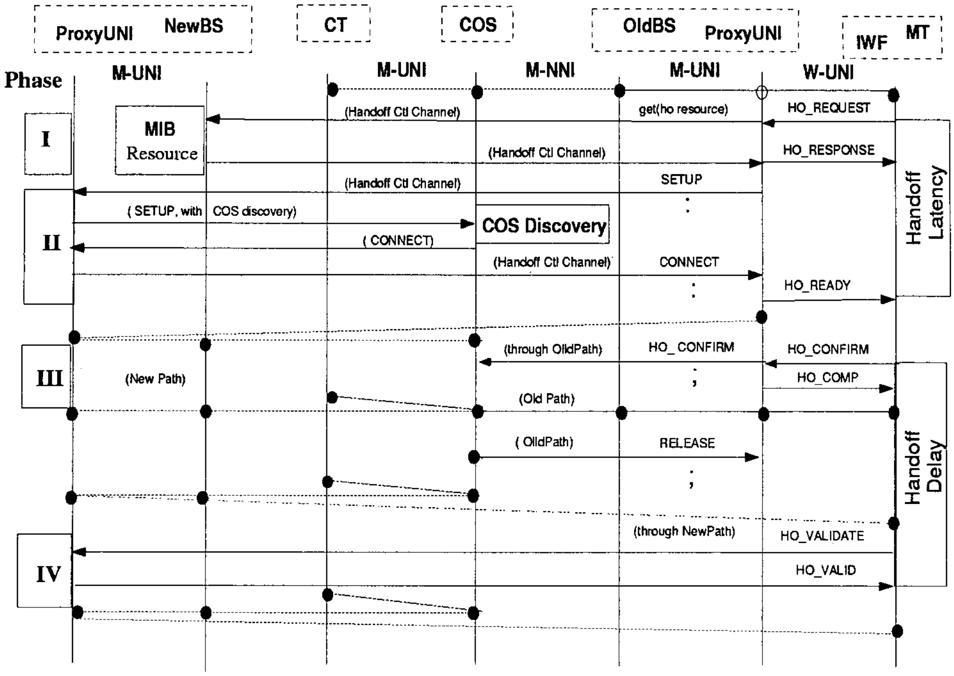


Fig. 8. Signaling sequence of mobility management.

In the third and fourth phases, there is a potential of data loss for backward handoff. We define this period as the *Handoff Delay*. It is a transition period during which the MT loses its connectivity with both the old basestation and the new basestation.

It is also shown in Fig. 8, after handoff is completed, a mobile terminal is at a new location and an LOC_UPDATE message should be sent to its VLS. If the mobile terminal's ATM address is changed, the new ATM address is informed to its HLS. If the mobile terminal retains the same ATM address, the PNNI routing tables on switches in the current VLS area must be updated to reflect the mobile terminal's current access point. When a caller tries to reach this mobile terminal, a SETUP message will first go to its HLS, obtaining the current foreign ATM address. The call is then re-routed to mobile terminal's current location. There is a critical period during which the caller may still get the mobile terminal's old ATM address. In this situation, the call may fail again and the caller needs to try the mobile terminal's HLS again. Unless a VLS confirms that the update on the HLS is done, it does not assign the old ATM address to other mobile terminals. There is a trade off between updating the ATM address without changing routing tables and updating routing tables without changing ATM address. It is a topic for future work.

V. PROTOTYPE IMPLEMENTATION AND EXPERIMENTS

An implementation of handoff control protocols in mobile ATM networks will be described in the context of the system integration of wireless access, interworking functions and ATM signaling extensions. We have built a prototype mobile ATM system with IP waveLAN radio access, wireless ATM access and GSM access. As an example, we will present the basic protocol stacks and software components and describe configurations for the system with IP waveLAN access. Some experimental results, such as handoff delays and data loss during handoff are given.

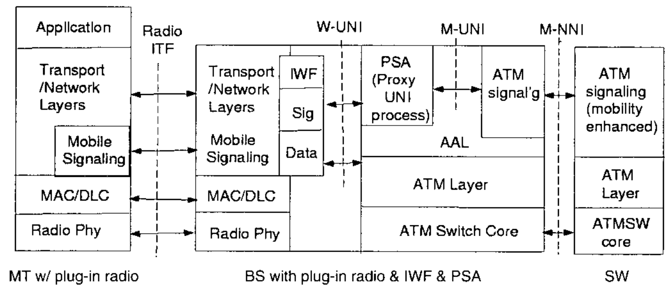


Fig. 9. Protocol stack of mobile ATM networks.

A. Protocol Stacks

Mobile ATM protocol stacks on mobile terminals, basestations and ATM switches are shown in Fig. 9. In general, a mobile terminal (MT) runs a service specific (*Mobile Signaling*) protocol over a radio link. On the basestation, the mobile signaling is converted to ATM W-UNI signaling by an interworking function (IWF). The ATM W-UNI signaling is handled by the proxy signaling agent (PSA), which further generates mobility enhanced ATM signaling over the M-UNI interface to the network side signaling process on the basestation. The IWF converts data between the radio specific data format and the ATM cell format.

B. Prototype Implementation

As shown in Fig. 10, our prototype system consists of a network with three ATM switches (2.4 Gbps NEC Model 5S), running GSMP (general switch management protocol) [26]. Each

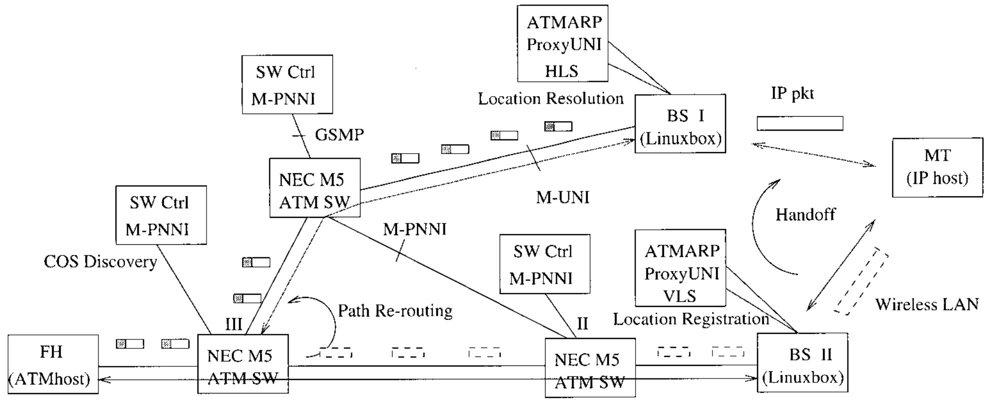


Fig. 10. Mobile ATM network prototype.

switch is controlled through an external 200 Mhz Pentium PC, running Linux as the operating system. GSMP is a simple client-server protocol that allows the switch controller (client) to add/delete/modify VCs on the switch (server). No signalling software runs on the switches themselves. The M-NNI signalling software runs on the external switch controllers; when the signalling for connection setup/deletion is completed, the switch controller issues a GSMP message to add/delete entries in the VC tables within the switch.

Each basestation's radio port provides a WaveLAN radio interface, using an off-the-shelf WaveLAN card with IEEE 802.11 standard radio technology. An NEC Versa laptop with a PCMCIA WaveLAN card acts as the mobile terminal, that runs only the IP protocol stack. Mobility between basestations is emulated by changing the default gateway of the mobile terminal to be one or the other basestation. The prototype network thus support two radio coverage areas, under basestation 1 and 2; both operate in the same frequency, and overlap in space.

In our specific prototype system, a Base Station is a Linux machine with both waveLAN and ATM NIC cards back to back. The PSA function is performed by ATM UNI signaling modules and the IWF function is performed by IP over ATM process. There is no switching function on the basestation as shown in Fig. 9. At the IWF interface, for waveLAN IP service, the first IP packet to a destination will trigger ATM signaling to establish an IP-over-ATM VC. Subsequent IP packets are segmented to ATM cells and transmitted to that VC.

As shown in Fig. 10, the mobile terminal (MT) by default has BS I as its home. It is given a home ATM address by a HLS sitting on BS I. When the mobile terminal roams from BS I to BS II, it registers to the VLS and obtains a Foreign ATM address. When a fixed host (FH) tries to call the MT, it uses ATMARP to get a translation from the MT's IP address to its home ATM address and then triggers an IP over ATM VC towards BS I. At BS I, the SETUP for ATM VC fails and the MT's foreign ATM address is carried back by a RELEASE message. The SETUP message will use the Foreign ATM address to SETUP the connection again to BS II where the call can get through. When a call is in session, handoff can be performed by M-UNI/M-PNNI signaling. In this prototype, we artificially create a handoff trigger by signaling from the MT to basestation to change the default gateway for packet transmission. The signaling triggers the IP-over-ATM connection SETUP from the new basestation. When the connection SETUP crosses the old connection path at the ATM switch III, that switch is discovered as the COS, and the connection path is re-routed. The sub-path from the switch III to BS I is released. The data flow then goes from FH to MT through BS II. An important issue here is the mapping of an IP session to an ATM connection with certain GCID. We proposed to use a mapping from IP session information to BLLI for setting up an IP-over-ATM VC. Different IP sessions will have different BLLIs which map to different GCIDs of ATM VCs. With this solution, forward handoff should also acquire the proxy UNI object migrating to the new basestation which can provide GCID for corresponding IP session.

The prototype system demonstrates the feasibility of interworking between IP and mobile ATM. Although the MT is an IP terminal, mobile ATM has been used to provided mobility management including location management and dynamic handoff. Note that mobile IP is not required within the mobile ATM cloud, but can be used for global mobility [27].

C. *Experiments in the Prototype System*

We have done preliminary experiments of measuring the signaling delays for new call establishment and handoff call path re-routing.

In the hardware and software environment as in Fig. 10, we have measured the signaling delays, summarized in Table 5. They are approximate values based on about a hundred sample runs and may be largely dependent upon switch and BS processing speed in this configuration. The values can give us a rough idea the time scale of the signaling delays. We divide the delay into different parts.

- The signaling delays on an entity are roughly the processing delay for SETUP and CONNECT messages (T_s, T_c).
- GSMP delay (T_g) is the delay of adding a connection path on the switch through GSMP protocol running on the switch controller. If the ATM signaling waits for the ACK of each GSMP request before it goes to the next hop, we call it synchronous GSMP; otherwise, we call it asynchronous GSMP operation. In this case, the GSMP delay is negligible.
- The REROUTE delay (T_r) is the processing time on the COS for re-routing the OldPath to the NewPath. It is roughly same

Table 5. Signaling delay per control function.

Device	SETUP T_s	CONNECT T_c	REROUTE T_r	GSMP T_g	ACCEPT T_a	TOTAL T
BS	0.5 ms	0.5 ms	–	–	–	1 ms
SW	1.3 ms	1 ms	–	4 ms	–	6.3 ms
COS	–	–	2.4 ms	4 ms	–	6.4 ms
CT	0.5 ms	0.5 ms	–	–	7 ms	8 ms

Table 6. Signaling delays for call setup and handoff path re-routing.

Delays T_d	Signaling w/ Synch GSMP	Signaling w/ Asynch GSMP
Per Switch	6.3 ms	2.3 ms
New Call establishment	21.6 ms	13.6 ms
Handoff Path Rerouting	13.7 ms	5.7 ms

as the sum of the processing time for SETUP and CONNECT.

- The ACCEPT delay (T_a) is the processing time for a call acceptance procedure in the server on the CT. Since the procedure goes to the kernel and then to the user space on the CT, the delay is expected to be relatively large.

In Table 6, the delay per switch can be used to provide an estimate of the maximum number of new calls or handoff calls per second that can be processed. In our experiment, the switch is controlled by an outside controller through GSMP protocol, no more than 400 calls (new + handoff) per second can be handled by a switch even when signaling uses asynchronous GSMP. Also, each switch contributes a delay for call setup or call handoff which should be as little as possible since handoff requires fast response. This justifies the necessity of using incremental path re-routing for handoff control protocols. Using our unified handoff control protocol, the processing delays (latencies) for various re-routing algorithms can be compared within the same framework. However, the differences will be significant only in a large scale network. We plan to conduct simulations and experiments for larger network topologies in the future.

VI. CONCLUSIONS

In this paper, we have presented an ATM-based architecture for 3G core mobile networks which provides important advantages of QoS support, generic mobility management support including dynamic handoff, scalability, etc. A generic core network of this type can support coexistence of current and future radio access technologies such as cellular, wireless data, and broadband wireless access. Our work on protocol design and proof-of-concept prototyping demonstrates that the proposed concept is feasible and can support different wireless services with necessary QoS and handoff performance. This paper focuses on the mobility support mechanisms for QoS-based connection-oriented services. More generic mechanisms for a mobile ATM network also supporting mobility for best-effort and Diffserv IP packet services are left for our future work. The ATM Forum's WATM specification due in early 2000 should provide a standardized basis for the architecture outlined here [11]. We believe that mobile ATM is thus a useful building block for future 3G mobile systems, even though it is not the only possible solution.

ACKNOWLEDGEMENT

The authors gratefully acknowledge valuable discussions with Dr. Arup Acharya and Dr. Bala Rajagopalan, both of whom have made independent contributions to the topic of mobile ATM.

REFERENCES

[1] D. Raychaudhuri and Y. Furuya, "ATM-based wireless personal communication system with migration to broadband services," *Americas Telecom'96*, May 1995.

[2] D. Raychaudhuri, "Wireless ATM networks: Architecture, system design and prototyping," *IEEE Personal Commun.*, vol. 12, no. 8, Dec. 1996.

[3] A. Acharya, J. Li, and D. Raychaudhuri, "Mobility management in wireless ATM networks," *IEEE Commun. Mag.*, vol. 35, no. 11, pp. 100–109, Nov. 1997.

[4] J. Li, A. Acharya, and D. Raychaudhuri, "Protocol design and implementation for mobility management in mobile ATM network," in *Proc. International Conf. Commun. Technol.*, Beijing, China, Oct. 1998.

[5] K. Buchanan *et al.*, "IMT-2000: Service provider's perspective," *IEEE Personal Commun.*, pp. 8–13, Aug. 1997.

[6] T. Ojanperä and R. Prasad, "An overview of third-generation wireless personal communications: A european perspective," *IEEE Personal Commun.*, pp. 20–29, Dec. 1998.

[7] A. Sasaki and S. Inada, "The current situation of IMT-2000 standardization activities in Japan," *IEEE Commun. Mag.*, pp 145–153, Sept. 1998.

[8] R. Pandya *et al.*, "IMT-2000 standards: Network aspects," *IEEE Personal Commun.*, pp. 20–29, Aug. 1997.

[9] T. Ojanperä and R. Prasad, *Wideband CDMA for Third Generation Mobile Communcations*, Artech House Publishers, 1998.

[10] 3GPP: Technical Specification Group Services and Systems Aspects, "Network architecture," 3G TS 23.002 V3.1.0, Dec. 1999.

[11] ATM Forum, "Draft wireless ATM capability set 1 specifications," ATM Forum/BTD-WATM-01.13, Feb. 2000.

[12] C. K. Toh, "Crossover switch discovery for wireless ATM LANs," *ACM/Baltzer Mobile Networks and Nomadic Appl.*, vol. 1, no. 2, Dec. 1996.

[13] R. Yuan *et al.*, "A signaling and control architecture for mobility support in wireless ATM networks," *ACM/Baltzer Mobile Networks and Appl.*, vol. 1, no. 3, Dec. 1996.

[14] B. Akyol and D. Cox, "Rerouting for handoff in a wireless ATM network," *IEEE Personal Commun.*, vol. 3, no. 5, Oct. 1996.

[15] O. T. W. Yu and V. C. M. Leung, "Signaling network architecture and transaction protocols to support realtime connection rerouting in ATM/B-ISDNs," in *Proc. INFOCOM'96*, San Francisco, CA, Mar. 1996.

[16] Y. Choi, M. Song, and C. Kim, "Connection rerouting in general-topology connection-oriented mobile networks," in *Proc. IEEE VTC*, Atlanta, CA, Apr. 1996.

[17] M. Veeraraghavan, M. Karol, and K. Eng, "Mobility and connection management in a wireless ATM LAN," *IEEE J. Select. Areas Commun.*, vol. 15, no. 1, pp. 50–68, Jan. 1997.

[18] A. Acampora and M. Naghshineh, "An architecture and methodology for mobile-executed handoff in cellular ATM networks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1365–1375, Dec. 1994.

[19] H. Mitts *et al.*, "Lossless handover for wireless ATM," *ACM/Baltzer Mobile Networks and Appl.*, vol. 1, no. 3, Dec. 1996.

[20] S. Seshan, H. Balakrishnan, and R. H. Katz, "Handoffs in cellular wireless networks: The daedalus implementation and experience," *IEEE Wireless Personal Commun.*, vol. 4, no. 2, pp. 141–162, Mar. 1997.

[21] M. Singhal, G. Dommety, and M. Veeraraghavan, "A route optimization algorithm and its application to mobile location management in ATM networks," *IEEE J. Select. Areas Commun.*, vol. 16, no. 6, pp. 890–908, Aug. 1998.

[22] I. Chalamtac, T. Liu, and P. Bahl, "Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks," *IEEE J. Select. Areas Commun.*, vol. 16, no. 6, pp. 922–936, Aug. 1998.

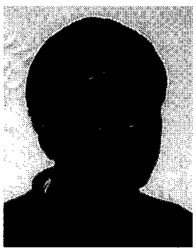
[23] D. Raychaudhuri and N. D. Wilson, "ATM-based transport architecture for multiservices wireless personal communication networks," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1401–1414, Dec. 1994.

[24] ATM Forum, "Baseline text for wireless ATM specifications," ATM Forum/BTD-WATM-01.1, Feb. 1998.

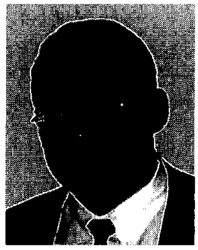
[25] ATM Forum, "Private network-network interface (PNNI) specification version 2.0," ATM Forum/BTD-PNNI 02.00, Sept. 1997.

[26] P. Newman *et al.*, "Ipsilon's general switch management protocol specification," version 1.1, *RFC 1987*, Aug. 1996.

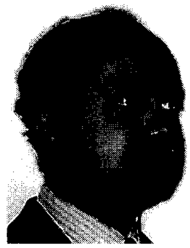
[27] A. Acharya and B. Rajagopalan, "Integrated location management for mobile ATM: Internetworking with mobile telephony and mobile IP," *ATM Forum Contribution/97-0851*, Sept. 1997.



Jun Li received his B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1983 and 1985, and he received his Ph.D. degree from Rutgers University, in 1999, all in electrical engineering. He was with Beijing Information Technology Institute as a research engineer from 1986 to 1990, conducting research on speech synthesis and recognition. During October 1987 to December 1988, he visited NEC's C&C Central Labs, Kawasaki, Japan, as a visiting scholar. Since 1990, he worked as a software engineer for several networking software companies in Japan and US, developing networking software products. He has started work with NEC C&C research laboratories in Princeton since the summer of 1995 while he was persuading his Ph.D. Degree in WINLAB, Rutgers University. Now he is a research staff member in IP networking group of C&C research laboratories of NEC, working on network protocol designs for next generation mobile communication systems.



Roy Yates received the B.S.E. degree in 1983 from Princeton University, and the S.M. and Ph.D. degrees in 1986 and 1990 from M.I.T., all in Electrical Engineering. Since 1990, he has been with the Wireless Information Networks Laboratory (WINLAB) and the ECE department at Rutgers University where he is currently an associate professor. His research interests include power control, interference suppression, and media access protocols for wireless communications systems.



Dipankar Raychaudhuri received the B. Tech. (Honors) degree from the Indian Institute of Technology, Kharagpur, India in 1976, and the M.S. and Ph.D. degrees in Electrical Engineering from SUNY, Stony Brook in 1978 and 1979 respectively. He was with the David Sarnoff Research Center (formerly RCA Laboratories), Princeton, NJ, as Member of Technical Staff (1979–87), Senior Member of Technical Staff (1988–89) and Head, Broadband Communications Research (1990–92). At Sarnoff, he worked on a range of R&D topics including: very small aperture terminal (VSAT) satellite networks, direct broadcast satellite (DBS), packet video, digital HDTV and wireless data networks. During the period 1990–92, he led the multi-company project team responsible for system design and specification of the "Advanced Digital HDTV" prototype tested by the U. S. FCC in 1992.

Since January 1993, he has been with NEC USA, C&C Research Laboratories, Princeton, NJ, where he is currently Assistant General Manager & Department Head (Systems Architecture), with focus on multimedia networking technologies including IP and ATM switching/protocols, broadband wireless and distributed multimedia software. In 1995, his research group demonstrated one of the first proof-of-concept wireless ATM networks ("WATMnet") capable of delivering multimedia services to portable computing devices. He is an active participant in related standardization activities, and has been Vice-Chair of the ATM Forum's Wireless ATM Working Group since its inception in 1996.

Dr. Raychaudhuri is a Fellow of the IEEE, and has authored over 100 technical papers and 13 U.S. patents. He is currently a Technical Editor of the IEEE Multimedia Magazine, and has previously served as Editor, IEEE Trans. on Networking (1993–98), Editor, IEEE Trans. on Communications (1992–93), Technical Editor, IEEE Communications Magazine (1980–94), and IEEE Communications Society Distinguished Lecturer (1992–96).