PROTOCOL DESIGN AND PERFORMANCE ANALYSIS FOR HANDOFF CONTROL IN MOBILE ATM NETWORKS

BY JUN LI

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Professor Roy Yates

and approved by

New Brunswick, New Jersey

October, 1999

ABSTRACT OF THE DISSERTATION

Protocol Design and Performance Analysis for Handoff Control in Mobile ATM networks

by Jun Li

Dissertation Director: Professor Roy Yates

Future access networks are expected to provide universal access for a variety of network services, especially for mobile services. The concept of *Mobile ATM* has been proposed to meet this requirement. A mobile ATM network consists of a core network with regular ATM switches plus wireless access points of different kinds. The core ATM network is enhanced with mobility support functions for location management and handoff control, while supporting heterogeneous mobile services such as GSM, GPRS, UMTS and wireless ATM. This thesis studies the handoff control process in mobile ATM networks from four aspects – system architecture, protocol design, performance analysis and system implementation.

In system architecture design, we propose a W-UNI (wireless user-to-network) interface at radio link, in addition to the mobility enhanced M-UNI (mobile user-to-network) and M-NNI (mobile network-to-network) interfaces within the wired ATM network. All signaling messages of nonATM mobile terminals are converted to the W-UNI interface through an interworking function. Chapter 2 introduces the mobile ATM architecture by describing how each mobility support functions work over W-UNI, M-UNI and M-NNI interfaces.

In a mobile ATM network, a particular alternate path rerouting algorithm will be

preferable for some mobiles services but not others. Thus, to support a variety of mobile services it is essential that the protocol design for handoff control allow multiple alternative path rerouting algorithms. Hence we propose a *unified handoff control protocol* which has a common ATM signaling syntax for alternative path rerouting algorithms. In Chapter 3, the detailed handoff control procedure is described along with signaling sequence and state diagram illustrations. This work has been done in support of the mobile ATM baseline in the wireless ATM group of ATM Forum. Additional design work in Chapter 4 focuses on handoff path rerouting in PNNI hierarchical networks, lossless control in the handoff transition period, and handoff control for multicast connections.

User mobility and handoff path rerouting may cause changes in the network traffic distribution that affect the overall network utilization. In Chapter 5, we propose a *Flow Model* to analyze the network traffic distribution in a mobile environment. The model defines network traffic as stochastic processes (flows) for a given set of OD (origin-destination) pairs over the network. For every OD pair, the flow has a routing function which describes the fraction of the flow carried by each network link. User mobility is characterized by a transition probability between flows. Handoff path rerouting algorithms are defined by transformations of routing functions among different flows. Using this model, we observe a smoothing effect when the mobility pattern is balanced and a clustering effect when the mobility pattern is unbalanced. Based on the flow model with routing functions, we present an analytical methodology to evaluate handoff path rerouting algorithms. Three typical path rerouting algorithms are compared through both analysis and simulations.

Based on the above proposed protocols, proof-of-concept research prototypes have been implemented. In these systems, the mobility of both IP terminals and ATM terminals can be supported through a common mobile ATM core network. The prototype system has been used to validate the proposed protocols, gain experimental insight into system performance, and support example mobile applications. In Chapter 6, we present the software architecture for the prototype systems and give some experimental results on the handoff control delay within the mobile ATM network.

Acknowledgements

I would like to thank my thesis advisor, Professor Roy Yates. He showed great patience in improving the quality of my work. His insightful advise and continuous support guided me through my Ph.D. thesis work.

I would like to thank my mentor Dr. Dipankar Raychaudhuri. It is a great opportunity for me to work with him during the most important period of my career development. His encouragement and supervision made it possible for me to get into a challenging new field and complete this Ph.D. thesis. I believe my whole career will benefit on my achievement under his management.

My thanks are also expressed to my thesis committee members, Professor David Goodman, Professor Narayan Mandyam and Professor Ivan Marsic. Their precious time and feedback are greatly appreciated. I am proud to be the last Ph.D. in WINLAB during Professor Goodman's ten year serving as Director.

I like to thank the support from NEC C&C Labs. It is my privilege to work with outstanding researchers in the labs while completing my Ph.D. thesis. I benefited from discussions and practical help offered during both research and project development. I especially thank Arup Acharya, Bala Rajagopalan, Thomas Kheunel, Partha Narasimhan, Daniel Rieninger and Hang Liu for their invaluable help.

This thesis is dedicated to my family. I thank my wife Yan for her love and sacrifice in taking on most of the family work while pursuing her own career. I thank my daughters Shirley and Emily for their love and patience. I also give thanks to my parents for their encouragement and faith, and to my parents-in-law for their continuous caring and understanding.

I am glad this six year long ordeal is over. It was a great challenge for me to complete my Ph.D. with a full time job and a family with two little kids. Without the support

from all of the above people, I cannot imagine how I could have gone through this long journey. Again, thanks to all of the people who supported my study at Rutgers University.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ATM Asynchronize transfer mode

GPRS General packet radio service

GSM Group special mobile (European cellular system)

HLR Home location register

IE Information element

ILMI Interim local management interface

MIB Management information base

MS Mobile station (in GSM system)

NCCI Network correlated connection identification

NIC Network interface card

OAM Operation and Management (ATM cell)

PNNI Private network to network interface

PSTN Public switched telephone network

UMTS Universal mobile telecommunications system

UNI User to network interface

VLR Visitor location register

ATMARP ATM address resolution protocol

AP Access point

BS Basestation

COS Crossover switch

CT Corresponding terminal

FD First divergent (path rerouting algorithm)

FXCT First cross to CT (path rerouting algorithm)

FXBS First cross to (old) BS (path rerouting algorithm)

FRR Flat rerouting

GCID Global connection identification

HCIE Handoff control information element

HCC Handoff control channel

HLS Home location server

HRR Hierarchical rerouting

IWF Interworking function

LD Last divergent (path rerouting algorithm)

M-NNI Mobility-enhanced NNI

M-UNI Mobility-enhanced UNI

MT Mobile terminal

OD Origin destination (pair)

PE Path extension (path rerouting algorithm)

PMP Point to multipoint (ATM connection)

PSA Proxy signaling agent

RT Root (node for PMP connection)

UBS Universal basestation

VLS Visitor location server

WATM Wireless ATM

W-UNI Wireless link UNI

Chapter 1

Introduction

1.1 Technology background

As the Internet grows rapidly, it is expected that the conventional public services, such as telephony, will be integrated with data networking services. That is, a broadband, integrated service network is going to replace the currently separated telephone and data networks. Although the capacity of the backbone for both networks is increasing dramatically every year, the merging of two networks really depends on technology development in the access networks. Wireless access technologies have long been developed as cellular telephony systems. With the success of the second generation mobile system – digital telephony systems (GSM, IS-95), the need of data services through wireless access has become more and more desirable. Based on the existing and emerging technologies, such as GPRS, wireless LAN, UMTS and wireless ATM, wireless access for integrated services can be provided at various quality of service (QoS) levels in the near future.

A wireless access network, as depicted in Figure 1.1, has radio access networks and serving networks. A radio access network (RAN) provides physical media access for mobile terminals. It has radio specific networking functions, including radio resource management, media access control and channel control etc. The serving network provides all non-radio specific functionalities, including security, pricing, inter-RAN mobility management etc. For example, in GSM, each RAN is rooted by a mobile switching center (MSC) and a serving network connects one or multiple MSCs to the backbone network. A wireless access network may be designed independent of the backbone network. Regardless of whether the backbone is the PSTN or the Internet, the wireless access network can access the backbone through a gateway with certain interworking

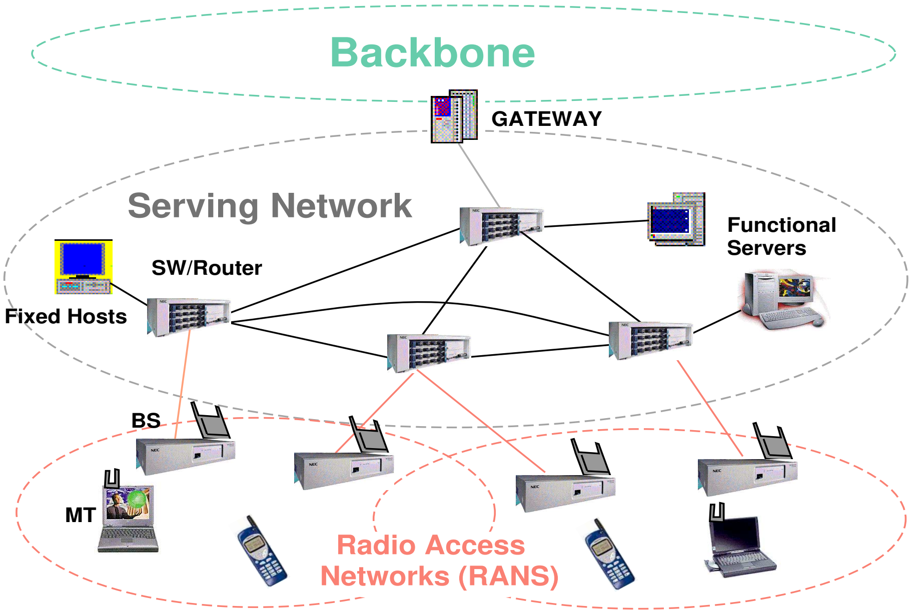

Figure 1.1: Mobile access network

functions. In order to avoid confusion, we will call a general wireless access network with radio access and serving network as a *Mobile Access Network* rather than wireless access network because wireless and radio are used often interchangeably.

Future mobile access network should be able to provide broadband services, unified IP and/or ATM networking capabilities for both fixed and mobile users and scalable mobility functions of handoff and location management for increasing percentage of mobile users. A *Mobile ATM Concept* is proposed [6, 46] as a solution to meet these requirements. It provides broadband services with quality of service (QoS) for both fixed and mobile users through a common core ATM network with enhanced signaling protocols for mobility management functions which can be used for various kinds of mobile services with different radio access technologies, such as GSM, GPRS, wireless LAN, UMTS and wireless ATM. This thesis aims to design mobility support protocols for such a *mobile ATM* network, with particular emphasis on the design of handoff control protocols. Historically, mobility support protocols are designed for specific services. Typical examples are given as follows.

1.1.1 Service model I: cellular telephony

Public mobile telephone service in the form of the analog FDMA/ AMPS has served as a major mobile communication system for many years. Only in recent years, have the second generation digital mobile service, such as GSM and IS-136 with TDMA technology, and IS-95 with CDMA technology, been deployed. Mobile telephony systems all use cellular networks in order to increase system capacity through frequency/code reuse over space. A cellular network introduces mobility management problems because the mobile terminal can change its network access point randomly and dynamically with very high speed. Finding a mobile terminal's current access point is called the location resolution problem. To maintain a mobile terminal's connectivity to the network leads to the handoff path rerouting problem. In general, the needs of location management and handoff control become key features in a mobile network system.

Previous research on cellular system mainly focuses on the throughput of the radio access network [18, 19, 12, 37, 38, 39]. These works deal with the problems of prioritized handoff call admission and dynamic channel allocation to lower the handoff dropping probability. In the fixed network, the mobility support functions have been considered only at the protocol level. Since the service model is uniquely for telephony, the mobility support uses a static mechanism, which is described as follows based on [36].

In traditional cellular systems, the mobility support functionalities are handled by overlaying a mobility control network consist of specialized control entities. A typical example is the GSM system. As shown in Figure 1.2, there is a mobile switching center (MSC) for each radio access network (RAN). All mobility functions are handled by MSC. A RAN can have multiple basestation controllers (BSC) in which link layer protocols (MAC/DLC) are handled. Each BSC can control multiple basestation transceivers (BTS) and each BTS covers one cell in a cellular system. In this thesis, we will use the term – mobile station (MS) for GSM mobile terminals.

In this model, the mobile network is completely separated from the PSTN by the MSC. A mobile station must have a phone number that belongs to the partition (area code) that a MSC owns. Through an MSC, mobile stations can interwork with the

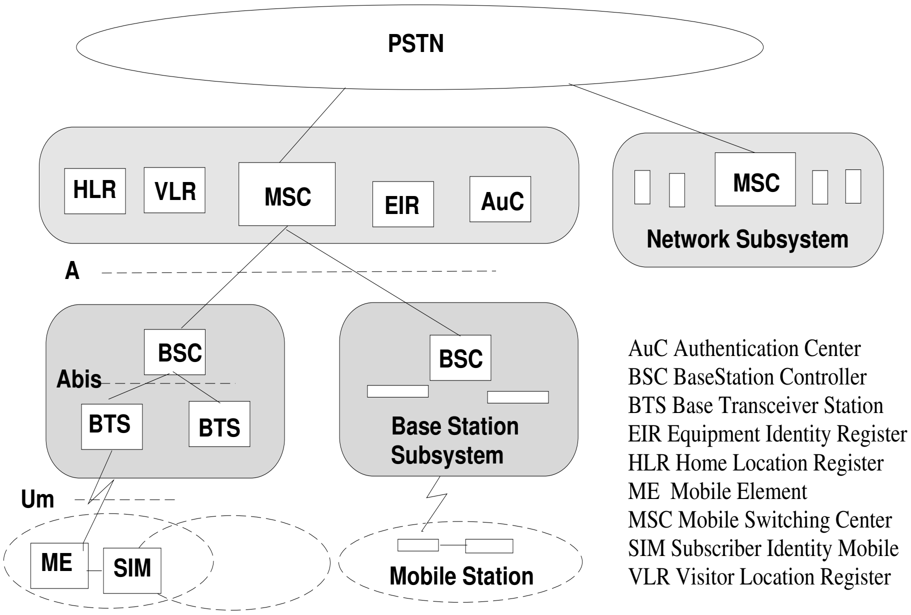

Figure 1.2: Architecture of GSM system

regular phones in the PSTN network. In the current installation of GSM systems, an MSC can cover a very big area, as big as a small or mid size country in Europe. note that an MSC controls all mobility functions in the network even when MS has only an intra-BSC handoff.

The location management in a GSM system is shown in Figure 1.3. Every MS has a home location register (HLR) which resides on the MSC that owns the area code for the phone number of the MSC. The HLR keeps track of a mobile station's location through a location update protocol. Whenever a mobile station moves to a new location, it tries to contact the local MSC in the area. The location register on the local MSC becomes a visitor location register (VLR) for the MS. The local MSC first contacts the MS's HLR to authenticate the visit of the MS, then the HLR updates the MS's current address. Whenever a caller tries to connect to a MS, the call setup signaling first goes to the MS's HLR, which forwards the call through MS's current MSC to reach the MS.

The handoff control in a GSM system is shown in Figure 1.4 [36]. When a mobile station moves from one access point to another while a call is in session, a handoff control process is required to reroute the call connection path. In GSM, the path rerouting is always done at the MSC. Since an MSC covers a big area, in most situations, an MS will

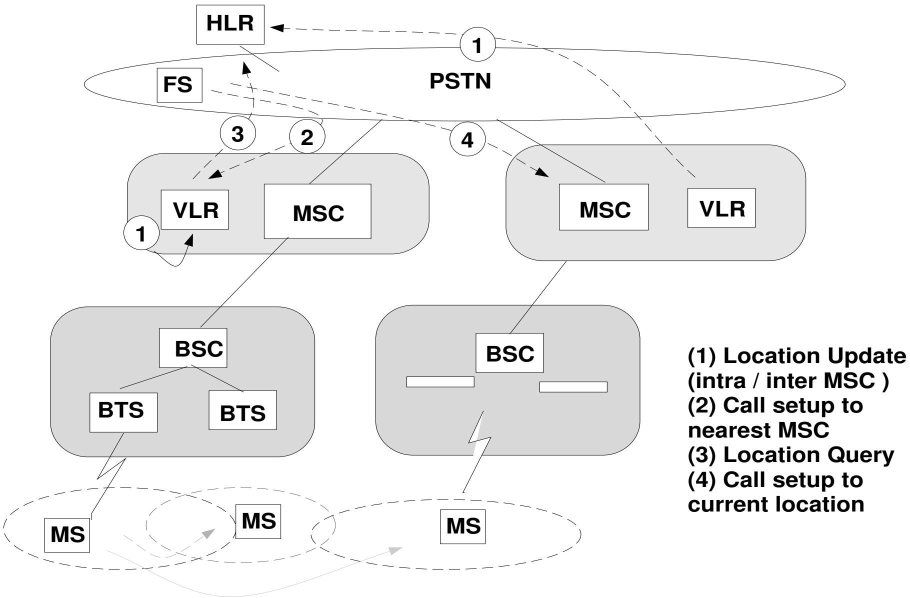

Figure 1.3: Location management in GSM system

have a intra-MSC handoff, i.e. a handoff between two BTSs under the same or different BSCs but under the same MSC, as shown in Figure 1.4 (1) on the left. In this case, the MSC will re-establish a new path to the MS through the new BTS. The original path is released. After the handoff is done, the location information is updated at the VLR on the MSC. It is not necessary to update the HLR since the MS is still at the same location represented by VLR. However, it is also possible for a MS to move from the area under one MSC to another during a call. Then a a handoff between two MSCs is required, as shown in Figure 1.4 (2). In this case, a new subpath from the old MSC to the new MSC is extended, and from the new MSC a path to the MS is established. The the part of original path under the old MSC is released. No matter how many times a MS having an inter-MSC handoff, the first MSC always acts as an anchor point for the extended path. After the MS's handoff is done, the VLR on the new MSC needs to report the MS's current location to the home MSC so that the address pointer is updated on the HLR, as shown in Figure 1.4 (3).

The GSM mobility support mechanism could be considered a candidate solution for mobile ATM networks. However, since the network architecture of a mobile ATM network is distributed, no single ATM switch can serve as a central MSC through which

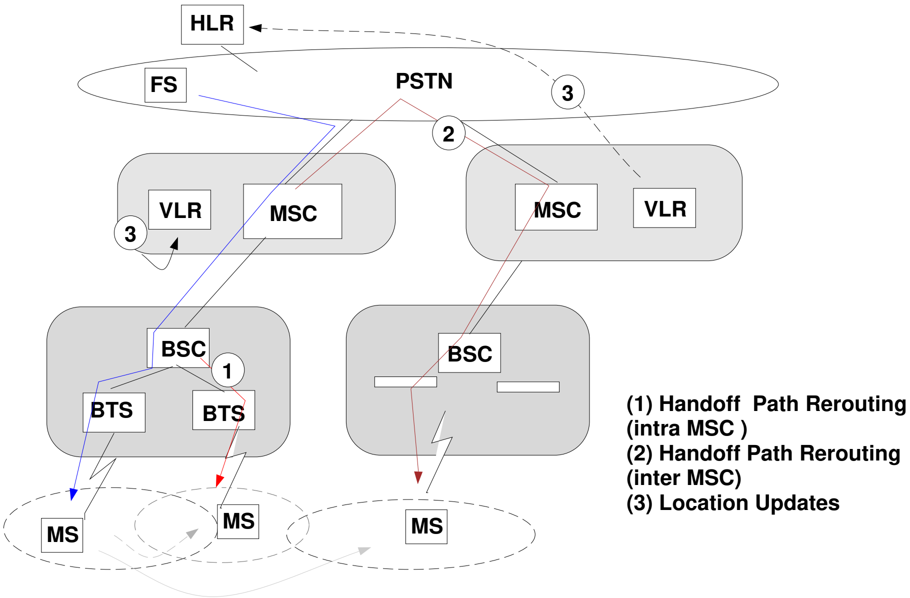

Figure 1.4: Handoff control in GSM system

all connection paths go. Only through the radio access point does every connection to an MS pass. Therefore, the centralized functions performed by the MSC can only be performed by radio access points in a distributed manner. Since mobile ATM networks may serve multiple types of mobile services at the same time, if MSC functions are not located on the access points, address partitioning is required within the network. That is, the network has to know whether the called party is a MS or not by looking only at the MS's address (phone number). When multiple services exist in a mobile ATM network, giving each service an address partition is not an easy work to do, especially dynamic service volume allocation is required.

1.1.2 Service model II: Mobile IP

Mobile IP represents the typical solution for mobility support in data networks. Mobile IP is an extension to the IP protocol that allows an IP node continuously receive datagrams no matter which subnet it attaches to. To make this happen, as shown in Figure 1.5, at least one IP node should act as a *location agent* for each subnet. Location agents will advertise themselves through an Internet Control Message Protocol (ICMP) to IP terminals. Every mobile IP terminal will choose a location agent as its home

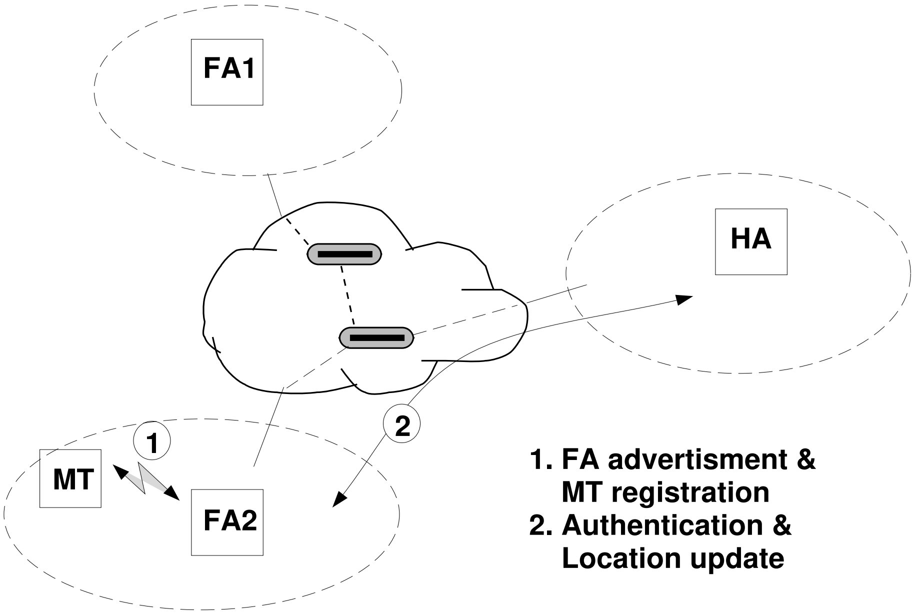

Figure 1.5: Mobile IP architecture

location agent (HA), from where a permanent IP address from the subnet is assigned to the terminal. If the IP terminal is within its home subnet, everyone can reach the terminal by using its home IP address. If the IP terminal roams out of its home subnet, it has to choose another location agent in its current subnet as its foreign location agent (FA). The IP terminal has to be authenticated through its HA and then be able to register on an FA. The FA will give a care-of-address (foreign address) to the mobile IP terminal. A care-of-address could be either the IP address of the FA or a dedicated temporary IP address for the mobile terminal. The HA will keep track of the mobile terminal movement by updating the care-of-address as the mobile terminal moves.

All packets sent to the mobile terminal will go to its HA first and then be encapsulated with its current care-of-address. This is called IP-tunneling. The packets will be redirected to the terminal's current location through the tunnel from the HA to the FA, as shown in Figure 1.6. The packets are de-encapsulated at the FA and forwarded to the mobile IP terminal.

In this service model, the locations of location agents are quite flexible. There can be one location agent per subnet, or one per access point. As long as the HA knows where to forward the packets to the mobile terminal's current FA and the FA knows

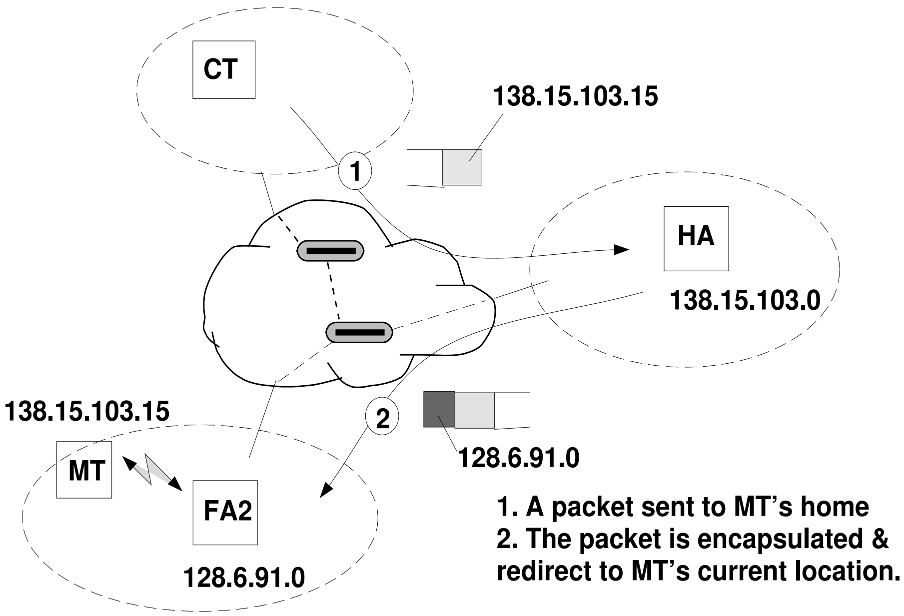

Figure 1.6: Packet encapsulation in mobile IP

where to forward the packets to the mobile terminal. The mobile IP mechanism will work no matter how far a mobile terminal is from its current location agent. However, due to performance concerns, it is always good to have the location agent closed to the MT.

Although Mobile IP has been standardized through IETF since October 1996, it has not been widely installed for either public or enterprise Internet environments. In the meantime, some extensions are proposed in addition to the basic functionalities [35].

First, the performance issue has been considered. When a mobile user is far away from its home agent, route optimization mechanisms have been proposed to minimize the packets forwarded by the home agent in-order to increase the network utilization, and more importantly, the extra packet delivery delay. The route optimization can be done by informing the sender with the mobile's current location, as shown in Figure 1.7 (1). As soon as the sender knows the mobile's current location, it can directly send packets through a IP tunnel from sender to FA without going through the mobile's HA. This route optimization mechanism may introduce packet loss or mis-ordering in the transition period. However, it is worth to do so if the service continues for a sufficient time after mobile moves into the new location.

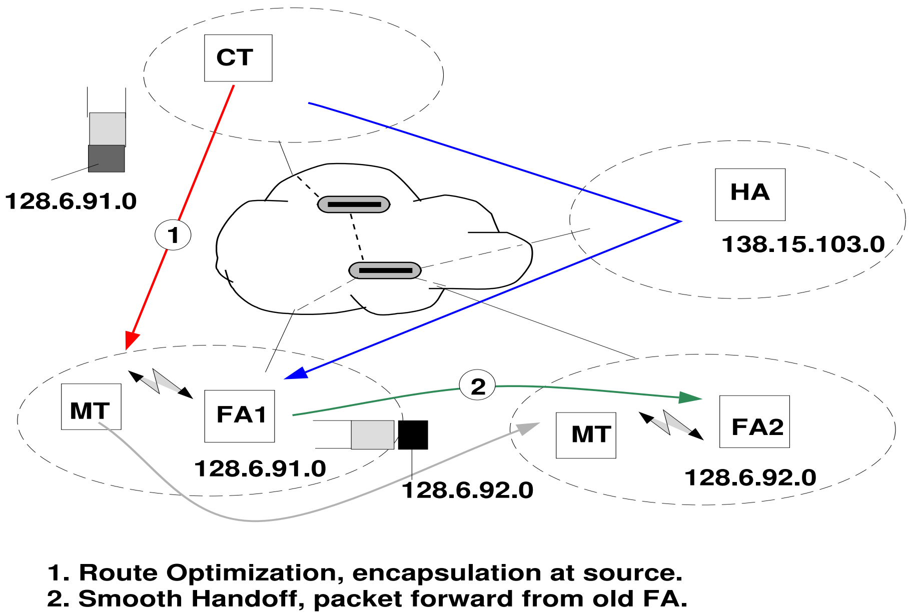

Figure 1.7: Protocol extensions in mobile IP

Second, the handoff scenarios have been considered. Since at the IP layer, there are no service sessions defined, handoff, a location change while a service is still on a session, is not distinguishable from a location change while a computer is shut down. However, if packet loss and delay are critical concerns for the applications, an extension to the original mobile IP protocol is necessary to reduce the transition period of handoff control. Mobile IP extensions in [35] have considered packet forwarding from one foreign agent to another foreign agent while a mobile terminal hands off from one area to another. This is known as *smooth handoff* and is depicted in Figure 1.7 (2). If there are some packets reaching the foreign agent FA1 after the MT moves to FA2, the packets are encapsulated with the new care-of-address given by FA2. Thus it is required that FA2 reports the MT's new care-of-address not only to the HA but also to FA1.

With route optimization and handoff extension, a mobile terminal needs to do an end-to-end location update to the senders of alive sessions after every handoff. In addition, it needs to inform the previous location agents. Thus it is not very efficient if frequent handoff is required in a micro cellular system.

The key advantage of mobile IP is the integration of the mobility support functionality into the conventional Internet. Unlike the cellular telephone model, here it is not necessary to partition the mobile network from the fixed network. The same core network can serve both fixed and mobile IP terminals at the same time. The mobile IP service model is suitable for the fixed network service provider to upgrade its service to mobile service.

The integration feature of Mobile IP is very promising for mobile ATM networks, however, the triangle problem and/or end-to-end location update requirement makes the protocol not scalable. So we would like to have a protocol combine the integration and hierarchy features of Mobile IP and GSM.

1.1.3 Service model III: wireless ATM

Wireless ATM is used for the integrated services of both real time service like telephony, and best effort service like IP. Since an ATM network has a switched virtual circuit (SVC) for each call, regardless of the type of service, the mobility management of the wireless ATM will at least to have a path rerouting mechanism for handoff SVCs, which is similar to the approach used by the GSM system.

In wireless ATM radio access, the research mainly has focused on media access control (MAC) and reliable channel coding (DLC) in order to flexibly schedule multimedia services from multiple users with high efficiency [1, 2, 7, 4, 5]. In the fixed network, studies have emphasized protocol designs for location management and handoff controls [8, 9, 10, 11, 13, 14, 15, 16, 17]. Here we will review the basic system architecture in the following reference system.

A. Wireless Access

Wireless ATM terminals can get access through the access points (basestations) at the edge of the core access network. The basestations are topologically distributed without necessarily being attached to any centralized ATM switch.

There are three fundamental differences between wireless ATM access and wired ATM access. First, the wireless link is a shared media for multiple users. A media access control (MAC) protocol is required to schedule the wireless ATM cell transmission for different users. Second, the wireless link is an unreliable media with higher bit error rate. A data link control (DLC) protocol is required to recover the data to improve the

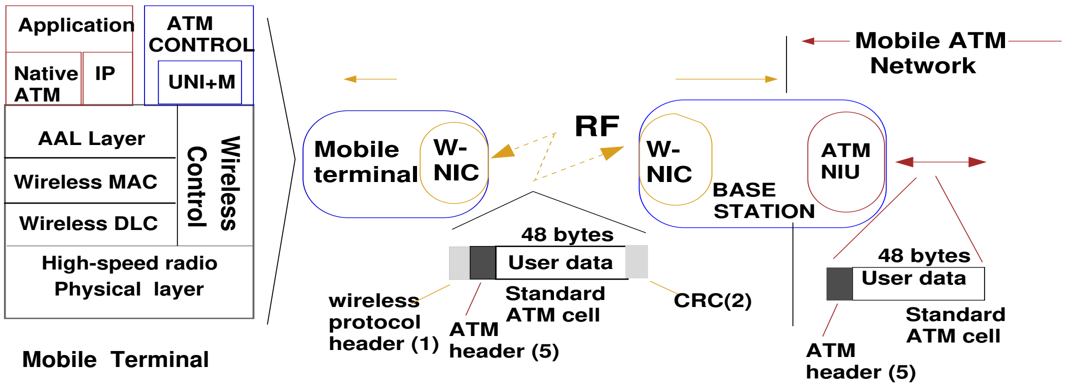

Figure 1.8: Wireless ATM access

quality of channel. Third, the access point of a mobile can be dynamically changed. A wireless control protocol is required for a mobile's registration and handoff control. Figure 1.8 shows the protocol stacks of NEC's WATMnet, which has a TDMA/TDD MAC layer and FEC+ARQ DLC layer. A six byte wireless ATM cell header plus two byte CRC replaces the five byte ATM cell header for greater error protection.

The wireless ATM access provides end-to-end QoS guaranteed ATM service capability over the unreliable media. It is obvious that, with the dynamic media access control and resource allocation by MAC layer protocol, the bandwidth efficiency at the link layer could be much higher than an IEEE 802.11 system which operates in contention mode.

B. Location Management

Location management functions enable the network to track the mobile terminals' current locations to facilitate terminals receiving calls. Figure 1.9 depicts location management functions in wireless ATM networks.

1. **Location registration:** Every mobile terminal (MT) is assigned with a *Home ATM Address* by registering to a Home Location Server (HLS). Whenever the MT moves out of the basestations covered by the HLS, it first registers on a Visitor Location Server (VLS) through its currently accessed basestation and gets a *Foreign ATM Address*.

2. **Location update:** After the MT is registered at the VLS, the foreign ATM address given by VLS is reported to the MT's HLS through a location update

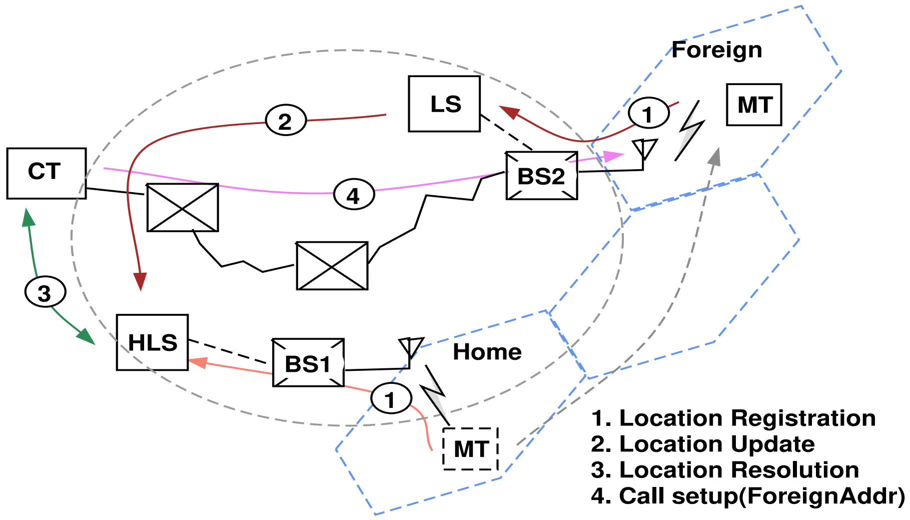

Figure 1.9: Location management in wireless ATM network

protocol between location servers. After a handoff, an MT needs to update its location at its HLS.

3. **Location resolution:** A caller (CT) always needs to resolve an MT's current location before it can send a SETUP message to establish a connection to the MT. In general, the resolution mechanism is performed externally before the ATM signaling call control starts. It is also possible to implement an integrated approach which resolves the MT's current location through ATM signaling protocol for call control. We will discuss this in detail in the next chapter.

4. **Call setup:** The CT sends a SETUP message to the MT using the foreign ATM address for connection establishment.

In a wireless ATM network, the location servers can be managed in different ways. For example, they can be totally distributed, i.e. one location server per access point. Whenever MT changes its access point, it will register to a new VLS and update its location at the HLS. Another way is to have hierarchical location servers. A regional VLS may take care of multiple access points within a subnet, only the regional VLS should be updated if a MT changes its location within the subnet.

C. Handoff Control

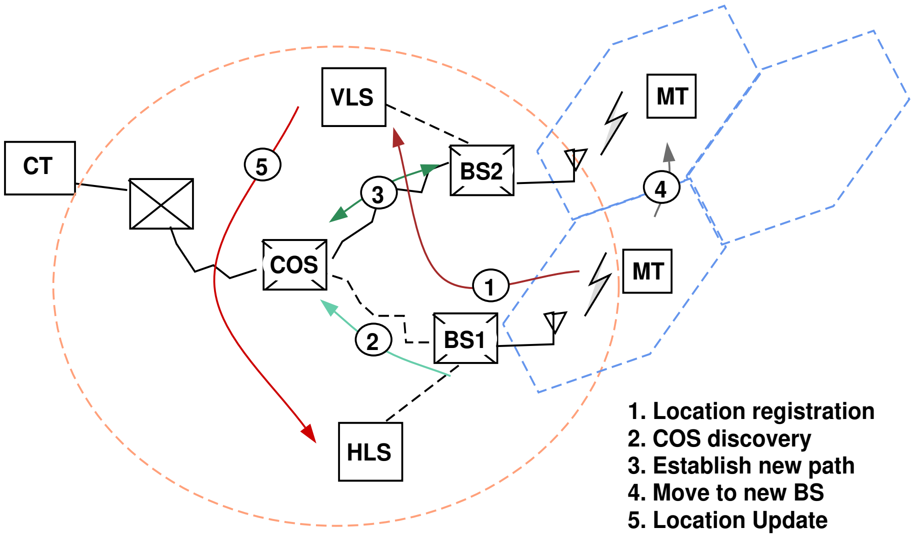

Figure 1.10: Handoff control in wireless ATM network

The handoff control function maintains a mobile's network connectivity with QoS guarantees. If a mobile terminal moves while a call is in session, a handoff control process is required to reroute the connection path. However, unlike GSM, since there is no centralized ATM switch through which all connection paths must go, the anchor of connection rerouting cannot be fixed in wireless ATM networks. As shown in Figure 1.10, the anchor, we call it a *crossover switch* (COS),can be anywhere in the network, as long as it is on the original path. Different choices of COSs have different performance and complexity criteria in handoff control, which leads to alternative path rerouting algorithms. A handoff control process can be described as follows.

1. The mobile terminal (MT) registers to a location server which covers the new cell. An authentication procedure is performed between the VLS and the HLS to make sure the MT is a valid user.

2. Based on a given path rerouting algorithm, a COS discovery procedure is performed to locate the anchor switch where the connection path should be rerouted.

3. Between COS and the new basestation (newBS), a new subpath is established, replacing the old subpath between COS and the old basestation (oldBS).

4. The MT moves into the new cell and continuously receives service from the newBS.

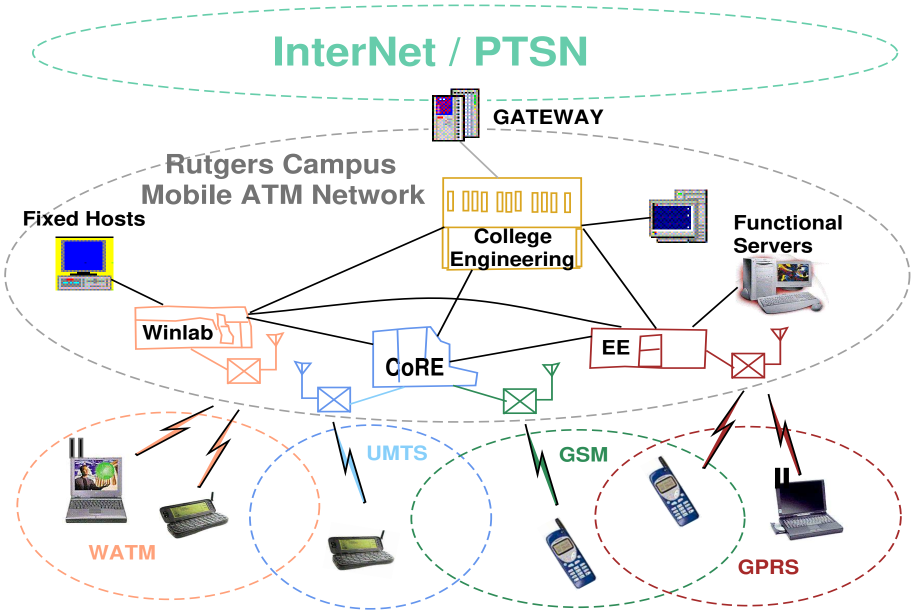

Figure 1.11: A Rutgers campus mobile ATM network

5. The MT's current location is updated through a location update protocol between the HLS and the VLS of the MT.

The design of the mobility support protocol in wireless ATM networks borrows from both GSM and mobile IP models. The handoff control is similar to GSM. However, the path is rerouted at a dynamically changed anchor (COS). The location management is similar to mobile IP since the location servers are distributed. However, the MT's new location is resolved in a signaling phase rather than during every packet's transmission.

1.1.4 Universal solution: mobile ATM

The mobility support mechanism in wireless ATM can be generalized into a mobile ATM network which can provide an infrastructure with mobility functions for ubiquitous mobile services.

We call this ATM based core network with mobility enhancement a *mobile ATM* network. An ATM based core network can be used as a campus wide network for hosts through various wireless access technologies. The same campus wide core network can also be shared by the fixed hosts. We can imagine, as shown in Figure 1.11, that mobile ATM network serving as a core network for the Rutgers campus can provide networking services of GSM telephony, GPRS, wireless LAN, UMTS or wireless ATM. With one core network investment, the University can utilize its resource for the fixed service as well as various kinds of mobile services.

Each type of mobile services can obtain the network access through a plug-in NIC with interworking functions (IWFs) which convert signaling/data of each mobile service to ATM signaling/data. The mobility functions are transparently supported within the ATM core network in a uniform way.

1.2 Thesis outline

In this thesis, we will discuss mobile ATM networks from different aspects, including system architecture, protocol design, performance analysis and system implementation. In particular, we focus on one of the key mobility management functions – handoff control in mobile ATM networks.

1.2.1 Mobile ATM system architecture

In Chapter 2, we will outline the system architecture of mobile ATM by defining a new interface called W-UNI (wireless user to network interface). The W-UNI interface works with the extended ATM UNI/PNNI interface and leads to an efficient protocol for handoff control in mobile ATM networks. The relationship among mobility related components, including radio access, interworking, location management and handoff control are addressed.

1.2.2 Protocol design

Previous work on handoff protocol design mainly has been considered for wireless ATM radio access networks. In most design proposals, a single path rerouting algorithm is preferred and a specific signaling syntax and sequence is given for the proposed protocol. In Chapter 3, we will propose a "unified handoff control protocol" for mobile ATM networks that supports a variety of radio access methods. Because the *mobile ATM* concept is to provide mobility functions for diversified mobile services, the performance and complexity requirements of handoff path rerouting may be quite diverse as well. A single path rerouting algorithm may not meet the criteria of all different services. For example, if a service is short-lived but requires high reliability, a fast, low complexity algorithm may be favorable. On the other hand, if a service is long-lived and delay sensitive, an optimal resource utilized path rerouting algorithm may be favorable. Features of the access network may also affect the choice of path rerouting algorithms, such as, network topology, cell size etc. In general, our design goal is to let a mobile ATM network be flexible to the variations of different types of services and access technologies by providing a generic handoff control protocol that adapts to the service and network properties.

The proposed unified handoff control protocol has a generic signaling syntax and sequence, supporting alternative handoff path rerouting algorithms. Such a protocol can provide a common software architecture for system implementation, a common measurement base for performance evaluation, and a common signaling syntax for standardization for different path rerouting algorithms.

Besides the basic path rerouting function, further functions are considered in Chapter 4 for protocol design. We discuss (1) an extension of the protocol in PNNI hierarchical structures, (2) a lossless control mechanism to avoid cell loss and misordering during handoff transition period, and (3) a protocol for point-to-multi-point connection handoff control.

1.2.3 Performance analysis

A common performance issue for a network service provider is the resource utilization, or throughput of the network. In the past, the performance analysis for a wireless system is only conducted at the radio access network since the radio link is considered as the bottleneck of the whole network. For example, in a cellular system, it is assumed that a handoff call is dropped only because of lack of radio resource. In order to improve the throughput of a cellular system, prioritized handoff call admission control or dynamic channel allocation can be implemented [39]. Although the argument that radio link is the bottleneck in cellular systems is true in general, the effect of user mobility on the traffic distribution of the fixed network and the influence of the mobility management approach on the resource utilization over the fixed network are also interesting significant issues.

Traditionally, the performance analysis of a fixed network is conducted through a static flow model characterized by long term traffic averages [28]. The objective is to route traffic optimally in order to maximize the network throughput with minimum cost. There has been little work done for traffic distributions of a fixed network in a mobile service environment. This thesis deals with the problem for mobile ATM networks.

In a mobile ATM network, the traffic distribution will be affected by the variation of multimedia service and the re-allocation (rerouting) of mobile terminals. Multimedia brings diversity in bandwidths of network flows while mobility results in these flows being rerouted dynamically. For these reasons, in Chapter 5, we propose a *Flow Model* that represents network traffic as stochastic flows transferable from one route to another. The user mobility is represented by the transfer rates between traffic flows. Using this model, the traffic distributions versus both mobility patterns and alternative path re-routing algorithms are studied. Numerical examples are given in a simple network topology. The results can be used for re-designing network topology to meet the extra capacity demands in the mobile ATM network. In order to be able to evaluate the network performance in various network topologies and with specifiable mobility patterns, we have built a PNNI network simulator. A simulation has been conducted which gives statistics of path lengths of traffic flows in three situations with each of them using one of three typical path re-routing algorithms.

1.2.4 Implementation and experimental study

In chapter 6, we present a software architecture of the mobile ATM system for the implementation purpose. The protocol stacks and software components of a mobile ATM system are described. We have presented our prototype system and the pre-commercial trial system separately. The prototype system supports IP-over-ATM with off-the-shelf WaveLAN radio access. The trial system is a complete end-to-end wireless

ATM system. With these systems, we have conducted some experiments, including delays of call setup and handoff and data loss rate during handoff transition, etc.

Chapter 2

Mobile ATM Architecture

A mobile ATM network is designed for ubiquitous mobile services. It is based on a regular ATM network and shared by both mobile hosts and fixed hosts. Although handoff control is the main focus of this thesis, in order to give an outline of mobile ATM architecture, the general functions of all three mobility support components, wireless access, location management and handoff control, are defined and alternative approaches for each function are discussed.

2.1 System architecture

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, [54, 55], define signaling messages over UNI and PNNI interfaces for call control, and the signaling procedures for connection establishment and release.

A mobile ATM network has radio access points (basestations) in addition to a regular ATM network. Mobile hosts access the ATM network through radio access points. When radio access uses wireless ATM technology, for the interface between a wireless ATM terminal and a wireless ATM basestation, we defined a *W-UNI interface*, which stands for *wireless user to network interface*. In addition, the mobility control functions, such as location update and handoff path rerouting, are introduced at regular UNI and PNNI interfaces in the network. So we define *M-UNI* and *M-PNNI* interfaces, which stand for *mobility enhanced UNI* and *mobility enhanced PNNI*. The signaling procedures for location management and handoff control will be specified through our

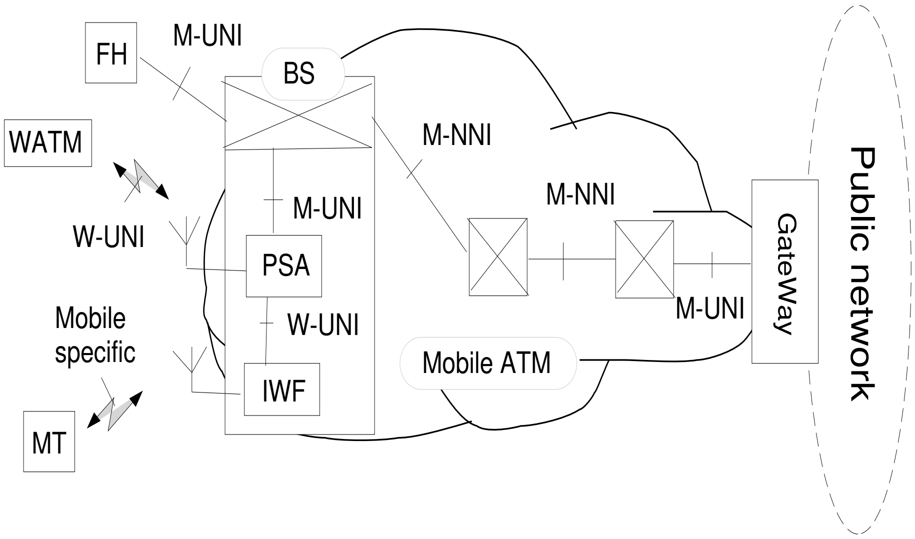

Figure 2.1: Mobile ATM system architecture

protocol design.

The system architecture can be shown in Figure 2.1. In general, if the radio access uses non-wireless ATM technology, an interworking function (IWF) must be provided to map non-ATM signaling/data to ATM signaling/data. This is performed at a proxy W-UNI interface, i.e. the non-ATM signaling/data are converted to ATM signaling/data at W-UNI interface, from where, any type of mobile hosts can be viewed by a mobile ATM network virtually as a wireless ATM host. For example, a GSM phone is considered as an wireless ATM host by a mobile ATM network.

If a mobile host obtains services from a server beyond the mobile ATM cloud, a gateway interworking function (GW-IWF) may be used to convert ATM signaling/data back to the mobile specific signaling/data. For example, a mobile ATM network can provide access for IP mobile terminals as follows. The IP packets will be converted to ATM cells by the IWF at radio access points and use IP-over-ATM in the mobile ATM cloud. At a gateway, it is again converted back to IP packets by the GW-IWF for the Internet. As long as the IP mobile terminal is within the mobile ATM cloud, the mobility functions are supported by mobile ATM.

A mobile ATM system can serve as an ATM based LAN for a university campus, hospital, airport or factory, providing fixed services as well as diversified mobile services are offered. It can provide private services as well as the integration of public services.

For example, GSM/GPRS can use the mobile ATM network as an office PBX to localize the internal traffic and multiplex the external traffic. By leasing small number of radio frequencies and/or operating at lower power, the service throughput per unit cost can be realized at a much higher rate.

This general purpose mobile ATM system requires interworking at wireless access and mobility control in the network. To design a mobile ATM system, each functional component must be well understood.

2.2 Wireless access

2.2.1 Definitions

A wireless access point (basestation) provides an ATM W-UNI interface to ATM mobile terminals. In addition to normal call control functions, such as connection establishment and release, mobility related control functions, such as location registration, location update and handoff control, are supported through this interface.

Over W-UNI interface, the signaling protocol should be re-designed to fit the characteristics of radio links. Since regular ATM UNI signaling is quite heavy weighted, a direct extension may not be appropriate over wireless channels. A simple, lightweight signaling protocol should be designed for W-UNI interface, at least, the mobility control part should be simple since it could be a more frequent operation than regular call control operations.

Our proposal is to use a *Proxy Signaling Agent* (PSA) [54] on each basestation. Whenever a mobile terminal comes to the basestation, a PSA will generate a ProxyUNI object for the mobile terminal. For mobility management functions, the signaling protocol between the mobile terminal and PSA, i.e. at the W-UNI interface, can be designed on per mobile terminal base instead of per connection base, through which signaling delay can be greatly reduced. In addition, a uniform control procedure can be defined inside the fixed part of the mobile ATM network, which is a basic requirement for mobile ATM system architecture.

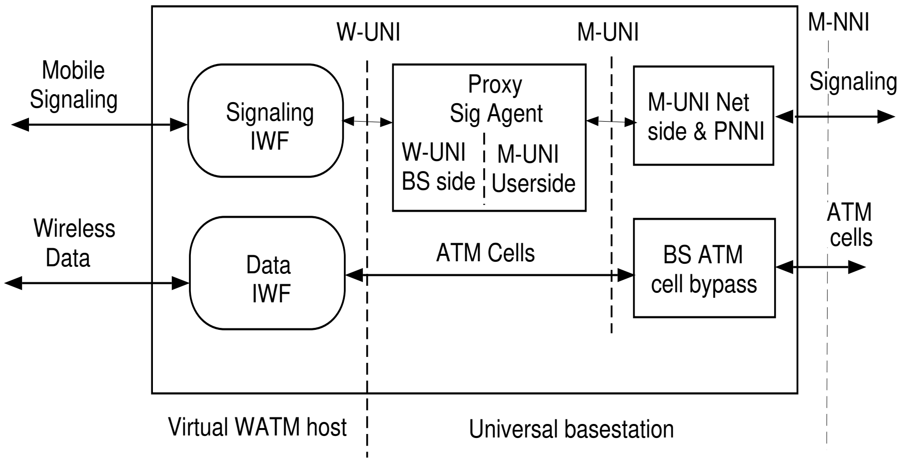

Figure 2.2: Wireless access of mobile ATM

2.2.2 Interworking: Universal basestation and virtual WATM host

If a non-ATM radio access technology is used, at W-UNI interface, a proper interworking function (IWF) must be supported. This can be achieved by more than one approach. A single basestation can be designed per radio access technology or a universal basestation can be designed for variety of radio access plug-in. As shown in Figure 2.2, 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.

We call a basestation with a W-UNI interface and equipped with multiple wireless access technologies a *universal basestation* (UBS). Our view of a UBS is a special ATM switch with plug-in slots for various network interface cards (NICs). Each plug-in NIC card embeds a SigIWF and a DataIWF, which output ATM W-UNI signaling and ATM cells, respectively. Through a UBS, a mobile ATM network can be used with all kinds of mobile terminals, whether it be a GSM phone, an IP host or a wireless ATM terminal. We call any such terminal a *virtual WATM host.* As long as the mobile terminal is within this mobile ATM cloud, the mobility functions, i.e. location management and handoff control, can be supported through the mobile ATM network. Always assuming a proper IWF between non-ATM and ATM exists, we will not especially consider non-ATM radio aspects , since all types of radio access can be treated as virtual wireless ATM host through the IWF.

2.3 Location management

2.3.1 Definitions

In a network (ATM or otherwise), each end-user (terminal) is identified by a network address. This address serves two purposes:

- as an endpoint identifier for the terminal, e.g. transport protocols use such endpoint identifiers to establish transport-level connections
- 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 didn't change.

With the advent of mobile terminals, it is no longer possible for a single identifier to support both functions. The location of a mobile terminal will change often, and so will its location identifier (since it determines reachability to the mobile terminal). 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.

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 for the mobile terminal's host identification. The other is the *Foreign* ATM address which serves as the mobile terminal's location identification. The function of location management is described as follows.

Capability: A location management process enables a mobile ATM network to provide the current location of a mobile terminal so that a caller can setup a connection to a mobile terminal transparently to user applications regardless of the location of the mobile terminal.

Participating Entities: As shown in Figure 1.9, the network entities involving a location management process are

1. Caller terminal or corresponding terminal (CT)
2. Mobile terminal (MT)
3. Mobile's home location server (HLS) and
4. Mobile's visitor location server (VLS)

In general, a mobile terminal is a virtual WATM host, but for convenience, we still simply call it an MT, providing the location servers (HLS or VLS) can have non-ATM to ATM address mapping for any non-ATM mobile terminals.

Primary functions: A location management process requires following primary functions to maintain the accessibility of a mobile terminal in the network:

1. home address registration at HLS
2. location resolution through HLS
3. visit location registration at VLS and
4. location update from the VLS to HLS.

2.3.2 Location management schemes

Like all location management schemes, in mobile ATM networks, it is required that some location servers (LS) manage the mobile terminals' Foreign ATM address through appropriate registration and update procedures. It is possible that these location servers can be either centralized, distributed or in between, i.e. hierarchically organized. Among the location servers in the network, each mobile terminal may select one as its home (HLS) which makes a mapping from mobile terminal's user id (Home ATM address) to its location id (Foreign ATM address).

In general, the location management functions can be designed as a separate service from the call control process. We call this an external location management scheme. A caller will resolve the callee's current foreign ATM address before call control process (SETUP) starts, or in another words, through a separate request to its HLS which is

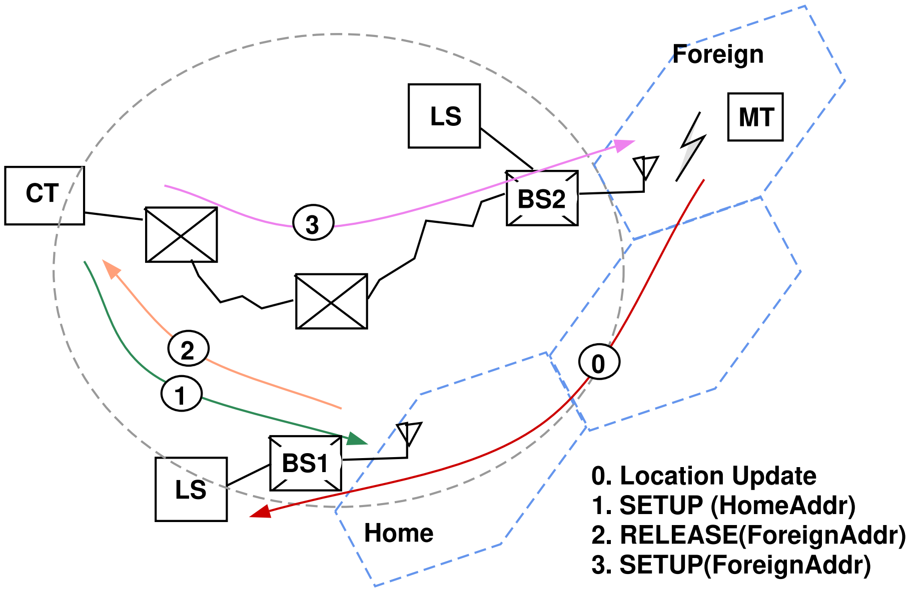

Figure 2.3: Location management in mobile ATM networks

equivalent to a DNS in the Internet. This scheme requires address partitioning between mobile and non-mobile terminals, because non-mobile terminals do not need a specific request for location management.

We have proposed an integrated scheme [50] that uses call control signaling for location resolution as well. The scheme is shown in Figure 2.3. The main idea is to integrate location resolution function into call control signaling process. A caller is sending a call setup with the callee's home address regardless whether the callee is mobile or not. If the callee is non-mobile or it is at home, the call (SETUP) will succeed. Otherwise, a reject signaling (RELEASE) containing the mobile terminal's current foreign ATM address is returned to the caller and the call setup can be automatically rerouted to the callee's current location. This approach doesn't need address partitioning within the network for mobile and non-mobile terminals as long as every call to a mobile terminal can pass through its HLS, from where mobile and non-mobile are distinguishable.

Although the location servers are distributed at each basestation in this integrated approach, it is also possible to have a centralized location server that communicates with the distributed location clients on basestations and performs other centralized functions, such as billing etc.

The additional advantages of the integrated scheme include:

- Unlike mobile IP or cellular telephony, the end-to-end connection path setup by the integrated scheme can be optimized.
- The size of the network routing tables is unaffected by the number of mobile terminals within the system. Each radio-port advertises reachability to a single network prefix (similar to a switch with only wireline interfaces). The size of network routing tables is proportional to the number of switches in the system.

The control functions for location management are performed through extensions in ILMI signaling and UNI/PNNI signaling. In addition, a protocol between location servers for location update. They are listed as shown in Table 2.1.

- The location registration function is performed through ILMI signaling, similar to a boot up process for a fixed host. The extension consists of new MIB objects required for a mobile terminal to store new information, such as radio frequency, home address and foreign address etc.
- In the integrated approach, the location resolution function is implemented through UNI/NNI signaling with the extension in RELEASE message which can carry a foreign address back to caller. In addition, the signaling module should be able to handle (1) a SETUP message when mobile terminal is not home and (2) a RELEASE message with a cause of MOBILE_AWAY and a foreign address IE.
- The location update function is external to existing ATM signaling. However, the updates change the contents of MIB objects which are used by ATM signaling.

2.4 Handoff Control

2.4.1 Definitions

In a connection-oriented network like ATM, the services attached to a mobile terminal are connections established with specific routes in the network. If the mobile terminal

Function	Message	Requirement
Registration	MIBs get/set	create MT object
Resolution	UNI/NNI SETUP/RELEASE	Cause: MOBILE_AWAY and Foreign Address IE
Update	Specified by LS	update MIB's MT object

Table 2.1: Control functions for location management

changes its access point to the network, the network services will lost their contacts 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 through the new access point.

Since the connection route is established by the ATM signaling protocol, the rerouting in mobile ATM networks must be done through extensions of the ATM signaling protocols. In general, regardless of the services supported by a mobile ATM network, the service region is likely to be relatively small. A LAN environment could be a typical area that a mobile ATM cloud may serve. Within this service area, multiple access points with a number of micro cells form a cellular access network. At a higher hierarchical level, multiple mobile ATM clouds may interwork through PNNI protocols regardless if they are directly connected or tunneled through other networks, such as the Internet or the PSTN.

In a microcell system, handoff is likely to be a more frequent event than call setup. Therefore a specific handoff control process which does connection path rerouting instead of connection re-establishment is desired. It is possible for a handoff control process to perform faster and to be cheaper than a regular connection establishment process because handoff is also a geographically local activity. Two connection routes from two neighboring access points to a common corresponding terminal will statistically share most of their paths in common. That an incremental (dynamic) path re-routing process should be applied for handoff is well accepted and it is agreed by major ATM Forum participants in the baseline text [51]. The following definitions describe the functional requirements of a handoff control based on dynamic path re-routing process.

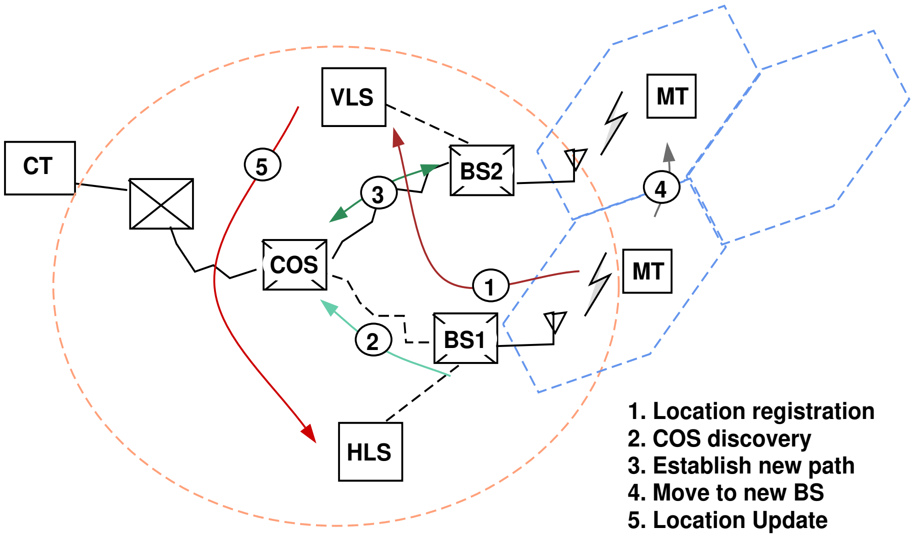

Figure 2.4: Handoff control in mobile ATM network

Capability: A handoff control process enables a mobile ATM network to seamlessly re-route the path of a connection between a mobile terminal and a correspondent terminal when the mobile terminal moves from one wireless access point to another wireless access point.

Participating Entities: The network entities involved in a handoff control process are illustrated in Figure 2.4. They are

1. Mobile Terminal (MT).
2. Corresponding Terminal (CT): the host at the other end of the connection. In most of cases, the CT is not aware of the handoff control procedure.
3. Original Basestation (oldBS): it covers the cell where the MT is currently in.
4. New Basestation (newBS): it covers the cell where the MT is about to enter.
5. CrossOver Switch (COS): the divergence point of the current connection path and a new connection path through the newBS.

Primary functions: A handoff control process requires following primary functions to maintain the connectivity of a mobile user in the network:

1. handoff initiation,

2. selection of a path re-routing algorithm,
3. a COS discovery process to find the COS,
4. establishment of a new path (NewPath) between COS and newBS,
5. replacement of the old connection path (OldPath, between COS and oldBS) with the NewPath and
6. release of the OldPath.

2.4.2 Alternative path rerouting algorithms

There are various protocol designs of dynamic path rerouting for handoff control. Described in a performance analysis paper, a path rerouting algorithm can be based on the following criteria [10].

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 [11, 13, 14, 15, 16, 45]. All of proposals use different terms to define each type of path rerouting algorithm. We can summarize these approaches as follows:

A. Path Extension

The simplest way to reroute the connection path is to extend the path from the oldBS to the newBS, as shown in Figure 2.5 where the COS is labelled (**a**). The advantages of this algorithm are:

- The switches in the network except the basestations, which are special switch nodes, do not need to be mobility-capable.

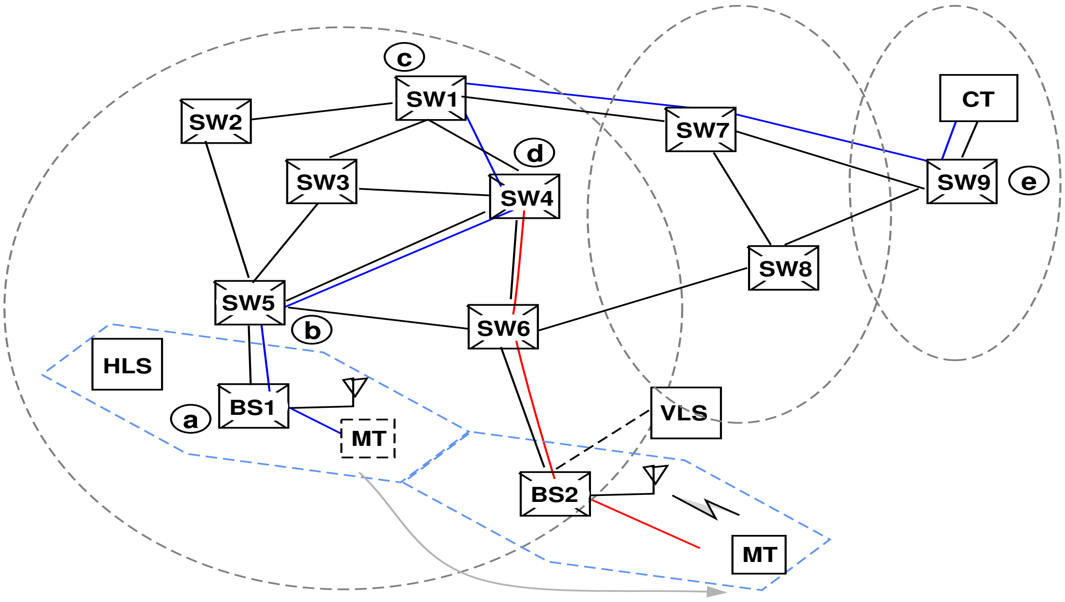

Figure 2.5: Alternative path rerouting algorithms in mobile ATM networks

- All connections of a mobile have a common anchor, the oldBS, regardless of the location of corresponding terminals for these connections.
- No synchronization is required for cells at the oldBS and the newBS for loss-free control.

The significant disadvantage of the path extension algorithm is that the length of the connection path keeps growing which consumes additional network resources and increases the transmission delay.

B. 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 newBS diverges from the original connection path. As shown in Figure 2.5, the COS is **b**.

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

C. 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 Figure 2.5, 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.

D. Shortest NewPath Rerouting

It is possible to configure the shortest new path to the old path. As shown in Figure 2.5, 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.

F. Last Divergent Rerouting

This algorithm targets optimizing the rerouted connection path for each handoff. As shown in Figure 2.5, the divergent point between the new path from the newBS 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.

Chapter 3

Design of Unified Handoff Control Protocol

3.1 Motivation and methodology

Since wireless ATM has been proposed [1] in early 90s, many researchers have studied the handoff control protocols [9, 10, 11, 16, 17, 45, 13, 12] in ATM networks. The protocols are each designed to satisfy the designer's specific performance criteria, such as complexity, delay, latency and/or network resource utilization. However, these criteria may not fit all type of mobile services. In addition, the traffic condition may be dynamic, which requires dynamic criteria. For example, the mobility pattern may vary during the day. A balanced mobility pattern may prefer the last divergence algorithm more than an imbalance mobility pattern. We will explain this in the performance analysis of Chapter 4. In general, a handoff control protocol should be flexible enough to allow variations in network construction and operation. Since the mobility functions have been introduced into the ATM Forum standardization, we have made efforts on the acceptance of a unifying protocol design [49, 50, 51, 25, 53].

With the introduction of interworking functions (IWFs) and proxy signaling agent (PSA) for a W-UNI interface, it is possible for us to design a handoff path rerouting process independent of the radio access technology. However, our design goal is to make the handoff path rerouting process also independent of the alternative path rerouting algorithms. Since a mobile ATM network is a general purpose core access network, it is possible for one mobile service to prefer a given path rerouting algorithm. The choice of path rerouting algorithm may also depend on the network topology or other performance concerns. Therefore, it is important to design a handoff control protocol for mobile ATM networks that can support various kinds of path rerouting algorithms.

Each path rerouting algorithm has a specific type of crossover switch (COS) which

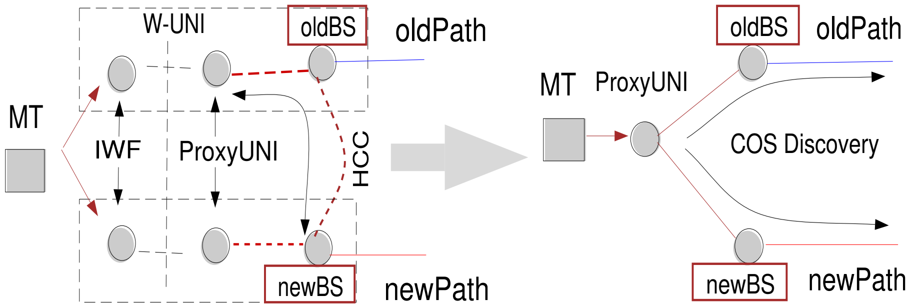

Figure 3.1: ProxyUNI and virtual MT to the mobile ATM network

satisfies a given criterion. Since a COS is a common switch on both the old path and the new path, the COS can be found by a signaling procedure along with either the old path or the new path. We call these *oldPath COS Discovery* and *newPath COS Discovery*, respectively.

A mobile terminal can access the network through either the oldBS or the newBS. In order to simplify the protocol design, we consider a virtual MT that can logically connect to both basestations at the same time, since there is a ProxyUNI object at each basestation and a handoff control channel (HCC) from the oldBS to the newBS. As shown in Figure 3.1, for example, upon receiving a handoff request at W-UNI interface, a ProxyUNI object at the oldBS can start the handoff path rerouting procedure on behalf of the MT. From the ProxyUNI object, the virtual MT has two interfaces, one to the oldBS and another to the newBS. The ProxyUNI can use either oldPath COS discovery or newPath COS discovery, regardless of which BS that the MT is currently accessing.

Since the COS discovery is basically a process based on routing decisions, two COS discovery signaling approaches may result in fairly different signaling syntax due to the signaling sources are different. For the oldPath COS Discovery, the source is the oldBS. The COS_DISCOVERY message is sent toward a destination, such as the newBS, along with the original connection path. After the COS is located, another signaling message is sent from the COS to the newBS for the newPath establishment. For the newPath COS Discovery, the source is the newBS. The COS_DISCOVERY message is sent toward a destination, such as the oldBS or the CT. When the message reaches the COS, based on given criteria, the newPath has also been established with the same signaling message. We observe that most types of COSs can be discovered with the newPath COS discovery approach. For this reason, we propose to use newPath COS discovery and based on that a single signaling syntax for the handoff control protocol is defined. Therefore, the handoff control protocol based on this method is unified for various path rerouting algorithms.

The unified handoff control protocol based on the newPath COS discovery will provide additional advantages of a common software architecture for system implementation, and a common measurement base for performance evaluation.

3.2 Control procedure

Usually, a handoff request is initiated by the MT. When the MT detects the 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 oldBS is called *backward handoff* while from the newBS, it is called *forward handoff* [51]. In the case of an MT totally losing its connectivity to the oldBS, the only choice for the MT is to do forward handoff. As illustrated in Figure 3.1, since a ProxyUNI object is generated for an MT, an MT can be seen logically connecting to both the oldBS and the newBS, via the handoff control channel. For backward handoff, the ProxyUNI for the MT is generated at the oldBS. For forward handoff, the ProxyUNI is generated at the newBS. Eventually, the ProxyUNI object at the oldBS will migrate to the newBS as the mobile terminal moves into the area of the newBS. From the perspective of ProxyUNI object, the control procedure is the same regardless whether backward handoff or forward handoff is used. In addition, by using newPath COS discovery, the starting point of the newPath is always fixed at the newBS. So no matter what kind of path rerouting algorithm is being used, the control procedure can be the same.

The ProxyUNI will see a virtual WATM host through the W-UNI interface regardless of the W-UNI signaling is actually from a real WATM MT or an IWF for different types of MT. For convenience, we view the MT as a WATM host. The control procedure, as shown in the Figure 3.2, has following steps:

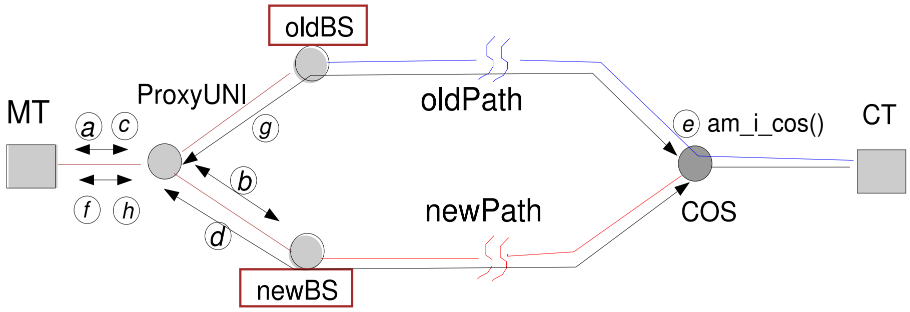

Figure 3.2: Handoff control process

- The ProxyUNI receives a handoff initial request from W-UNI interface.
- The ProxyUNI will check the resource availability on the NewBS.
- A response to the handoff initial request is sent back to W-UNI interface to the MT with all resource information (ATM address, signaling PVC number, bandwidth at the radio link etc). If the resource registration successes, the procedure continues, otherwise, the procedure stops and handoff fails.
- ProxyUNI starts the handoff path rerouting procedure for each data VC starts setting up new subpath toward the COS.
- COS discovery is performed along with the new path setup for each data VC. An local function am_i_cos() is used on each switch for identifying if the switch itself satisfies the conditions to be the selected type of COS.
- After all path reroutings are done, the ProxyUNI sends a handoff complete message, with all data VC's information, such as VCI, to the MT.
- The replacement of the oldPath with the newPath is confirmed by signaling through the oldPath.
- The ProxyUNI object on the NewBS receives an activate request for the MT accessing at the newBS.

This control procedure applies to both backward handoff and forward handoff. And we will see that a single signaling syntax can be designed based on the control procedure for various path rerouting algorithms.

Message	Direction	Function
HO_REQUEST	MT to ProxyUNI	handoff initiation
HO_RESPONSE	ProxyUNI to MT	handoff proceeding
CONN_READY	ProxyUNI to MT	reroute completed
HO_COMFIRM	MT to ProxyUNI	start path replacement
HO_READY	ProxyUNI to MT	mobile can leave cell
CONN_ACTIVATE	MT to ProxyUNI	mobile enter cell
CONN_ACTIVE	ProxyUNI to MT	handoff complete

Table 3.1: Signaling messages over W-UNI interface

3.3 Definition of signaling syntax

3.3.1 W-UNI signaling

The W-UNI interface is a newly defined interface which does not yet exist in ATM specifications. Over the W-UNI interface, several new messages are introduced for handoff control, as listed in Table 3.1.

The signaling messages are divided into four phases.

1. **Initiation:** The mobile terminal, through an IWF, initiates handoff by sending a HO_REQUEST message to ProxyUNI object. The ProxyUNI will respond after checking the radio resource availability with a HO_RESPONSE message. To avoid unnecessary retransmission of HO_REQUEST, the HO_RESPONSE is sent back before data VC's path rerouting starts.

2. **Path rerouting:** After HO_RESPONSE is sent out, the ProxyUNI will start rerouting data VCs one by one. After all VCs are rerouted, a CONN_READY is sent back to the MT. The details of path rerouting involve network UNI/NNI signaling extensions and will be given in next two subsections.

3. **Confirmation:** The mobile terminal, through an IWF, decides to start handover by sending a HO_CONFIRM message. Until now, the original connectivity of MT may still be available for data transmission. So the connection path for the MT will only be switched to the new path when HO_CONFIRM is received. In this phase, the network can reduce the data loss or eliminate data loss by integrating a

zero-cell-loss scheme, which will be discussed in the next Chapter. A HO_READY is replied by ProxyUNI after the connection paths are switched to the new paths. This phase is optional when data loss is not an issue. In that situation, each connection path is switched to the new path at the same time when the new path is established. In particular, forward handoff, handoff initiation from the newBS, will not need this phase anyway because the existing radio link to the oldBS is not available and data loss starts right at the beginning.

4. **Handover:** When the mobile terminal receives HO_READY message, it can change the radio frequency and directly start data transmission to the newBS. However, a CONN_ACTIVATE message is used to avoid unnecessary data loss. Only if the newBS receives this message, the connections related to the MT is enabled; otherwise, the data should be buffered or otherwise dropped. On the other hand, when CONN_ACTIVE is received by the MT, MT can start data transmission to the newBS. Again, this phase can be eliminated when data loss is not an issue or the application is delay sensitive, requiring MT to start data transmission as soon as possible. For example, for cellular telephony, it is desired that transmission start right after the frequency is changed without relying on any further signaling exchange.

3.3.2 Path rerouting signaling

Following the steps of the handoff control procedure in Figure 3.2, path rerouting signaling starts from the ProxyUNI object at the oldBS, for backward handoff, and from the ProxyUNI object at the newBS, for forward handoff. The path rerouting process begins with a SETUP message through the newBS toward the COS. A COS discovery procedure will be executed on every switch that the new path traverses. When the COS is discovered, a CONNECT message is sent back to complete the new subpath. Similar to W-UNI messages, a HO_CONFIRM is optional to delay the time that subpath is switched to the new one. This option can allow the old subpath to be alive until all connection path reroutings are done in order to reduce the data loss.

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

Table 3.2: Signaling messages for path rerouting

The path rerouting signaling is per connection based while the W-UNI signaling is per MT based. Whenever the ProxyUNI receives the HO_REQUEST from the MT and sees the radio resource on the newBS available for the MT, it will starts the path rerouting procedure for each connection that belongs to the MT. Whenever path rerouting for all connections is done, the ProxyUNI will send a CONN_READY to MT.

From Table 3.2, we can see the path rerouting signaling syntax is simple and applies to all algorithms. The difference for path rerouting algorithms is the COS discovery function, which is a local function run on switches along the new path.

3.3.3 COS discovery function

As shown in Figure 3.3, a COS can be chosen at any node along the route from the oldBS to the corresponding terminal (CT). Depending on the algorithm, the destination of the new path is first specified. For example, for path extension, the oldBS is the destination, while for optimal path rerouting, the CT is the destination. With the destination address set in the SETUP message, the new path is routed toward the destination based on conventional PNNI routing algorithm. However, on each ATM switch, a local function Am_I_COS() is executed to check the COS criteria according to the given path rerouting algorithm. For example, for optimal path rerouting, the function will check if the old path traverses the switch and if the new path has the same outgoing port as the old path at the switch, i.e. if the new path merges with the old path at the switch.

To check if a switch is on the old path, a global connection identifier (GCID) must be introduced for each connection. In current ATM UNI/NNI signaling specifications,

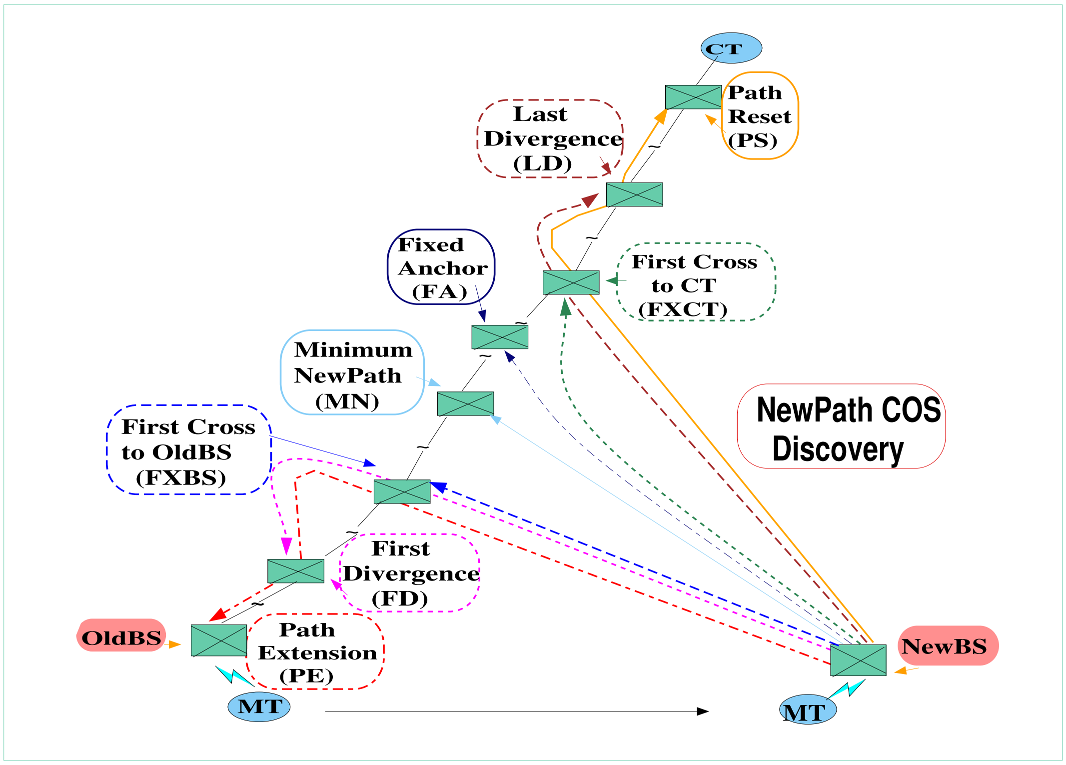

Figure 3.3: Crossover switch (COS) discovery in handoff control

connection ID has only local significance, meaning it is unique only for the switch. With GCID, a connection can be identified from a path rather than the original connection path. The GCID was introduced in a recent ATM Forum meeting [53] for handoff control. A similar ID (NCCI, network correlated connection ID) was introduced in PNNI 2.0 [55], 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.

In Table 1, we list destination addresses and COS detection functions Am_I_COS() for the re-routing process using all different COS types. The destination address in SETUP message and Am_I_COS() function to be used on switches are determined according to a handoff control information element (HCIE) in the VC_HANDOFF_REQUEST 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

Table 3.3: Destination address and COS discovery function in SETUP message

3.4 Handoff signaling sequence

Based on the control procedure in Figure 3.2 and the signaling syntax defined in extensions of ATM UNI/NNI signaling, we give a signaling sequence as shown in Figure 3.4. New messages and modified messages are shown in time sequence. This signaling sequence is aligned with the four phases signaling syntax.

1. **Resource registration on NewBS**

   The first phase is performed by the HO_REQUEST and HO_RESPONSE messages. After the ProxyUNI on the oldBS (or the newBS) receives a HO_REQUEST, it will start a registration process to the newBS on behalf of the MT. The process needs not be a full registration process as the MT 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. In case of the backward handoff, the ProxyUNI on the OldBS will forward the ILMI messages to the NewBS using handoff control channel (HCC). After registration, HO_RESPONSE is sent back to mobile ATM terminal through

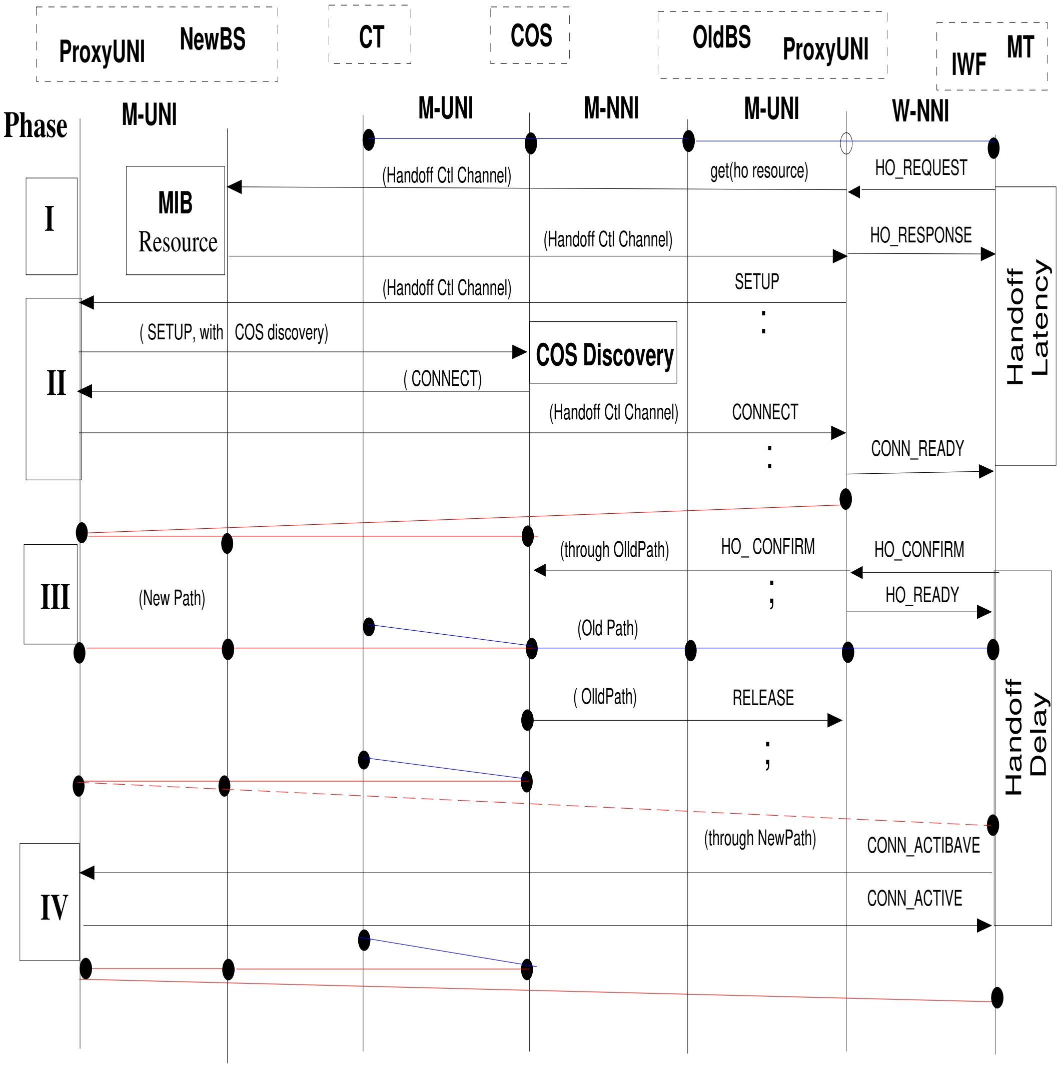

Figure 3.4: Signaling sequence of handoff control messages

W-UNI interface.

2. **Data VCs' path re-routing**

   The second phase is performed by the SETUP and CONNECT messages. In this phase, SETUP/CONNECT with new information elements (IEs) are used for the NewPath establishment and COS discovery for each data VC's path rerouting. A GCID IE is used for NewPath 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 Figure 3.4, 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 contribution of handoff latency is the signaling process for setting up new paths. If the cell size is small, little handoff latency can be tolerated. 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 rerouted 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 NewPath until a HO_CONFIRM message is received by the COS. With this optional phase, the data transmission can be continued while the NewPath is established. Only after HO_CONFIRM is sent out, the MT can be ready to handoff to the new basestation. Figure 3.4 shows a HO_READY message which indicates receiving of HO_CONFIRM by the oldBS.

4. **Active rerouted path on newBS** The fourth phase, performed by new messages – CONN_ACTIVATE / CONN_ACTIVE, 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 MT and newBS transmit data immediately after the data VCs are opened on either side.

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

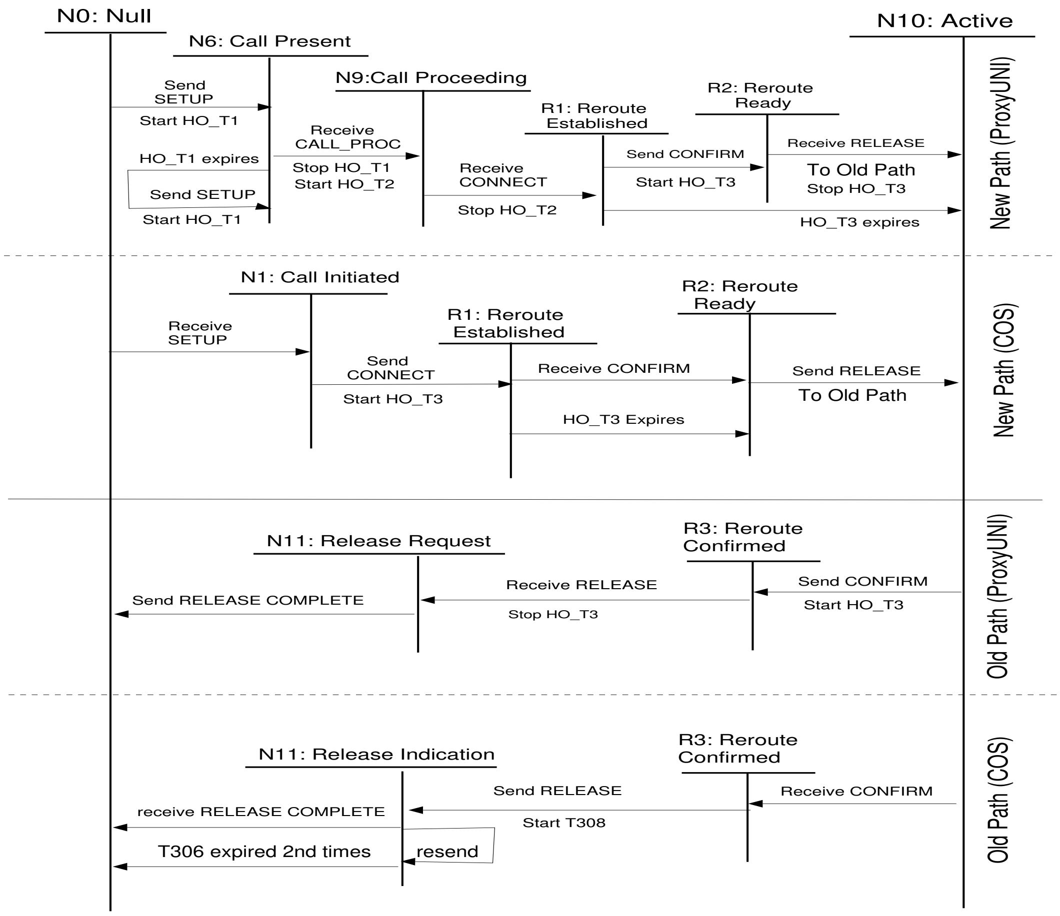

Figure 3.5: Signaling states of handoff control process

3.5 Handoff signaling state

During the handoff process, there will be transition state of the handoff connection. We introduce new states in the state machines at the ProxyUNI and COS for implementation purpose. Figure 3.5 illustrates the new states and their transitions in the signaling state machine. On both ProxyUNI and COS, for each connection to be handed-off, two states are associated for the oldPath and the newPath, respectively. The state of the newPath goes from $N0$ – *null* to $N10$ – *active* and the state of the oldPath goes from $N10$ to $N0$. In the figure, all $N*$ are standardized states and all $R*$ are new states we introduced, representing the transition of path replacement process. For the newPath, we introduce $R1$–*reroute established* and $R2$– *reroute ready*. For the oldPath, we introduce $R3$–*reroute confirmed*.

On the COS, the state of handoff connection goes to $R1$ for the newPath, after

CONNECT message is sent out to the newPath. If HO_CONFIRM is used and received, the state goes to $(R2, R3)$ for the newPath and the oldPath, respectively. After RELEASE is sent out, the state goes to $(N10, R3)$ for the newPath and the oldPath. The state of the oldPath goes to $N0$ when RELEASE_COMP is received.

At the ProxyUNI, when CONNECT is received, the newPath enters $R1$ state. Then when ProxyUNI send a HO_ CONFIRM to COS, the state of connection goes to $(R2, R3)$ for the newPath and the oldPath, respectively. When RELEASE is received from COS, the state of connection enters $(N10, N11)$. When RELEASE_COMP is sent out by ProxyUNI, the state of connection enters $(N10, N0)$.

If a signaling message gets lost for some reason, it must be either retransmitted or aborted from the action. This is controlled by the timers associated to each state. New timers are introduced for handoff path rerouting state, represented by HO_T*. Due to the restriction of handoff latency, the timeout value of the timers HO_T1 and HO_T2, which are for SETUP and CONNECT messages, could be less than the value for ordinary connection SETUP and CONNECT messages. Timer HO_T3 is designed explicitly for handoff path rerouting to assure no deadlock in path replacement process due to any message loss. The values of timers for handoff signaling may vary from system to system depending on performance specifications.

Chapter 4

Protocol Design Extensions

In Chapter 3, we proposed a unified handoff control protocol for mobile ATM networks which is independent of radio access, handoff initiation and path rerouting algorithms. The protocol can be used in a campus wide ATM based core network for generic mobile service purposes. To make the protocol more generic, we have to consider additional features. In this chapter, we discuss (1) handoff path rerouting in PNNI hierarchical networks, (2) zero-cell-loss control in handoff path rerouting and (3) handoff path rerouting for point-to-multipoint connections.

4.1 Path rerouting in PNNI hierarchical networks

The first topic is path rerouting in PNNI hierarchical networks. Since handoff path rerouting involves routing protocols, if it operates in a PNNI hierarchical network, the PNNI routing protocol will have influence on the handoff path rerouting procedure. In the unified handoff control protocol, COS discovery is integrated with the SETUP message for the newPath establishment. There is a requirement that on every switch that SETUP message traverses, an Am_I_COS() function should be executed to see if the COS condition is satisfied. Since the newPath COS discovery needs to check GCID to find the handoff connection, if there are thousands of connection on each switch, the cost of COS discovery could be high. However, by using PNNI hierarchical routing information, we can avoid checking COS condition on every switch on the newPath. We will propose a *hierarchical rerouting* (HRR) extension to extend the *flat rerouting* (FRR) mechanism used in the unified handoff control protocol [22].

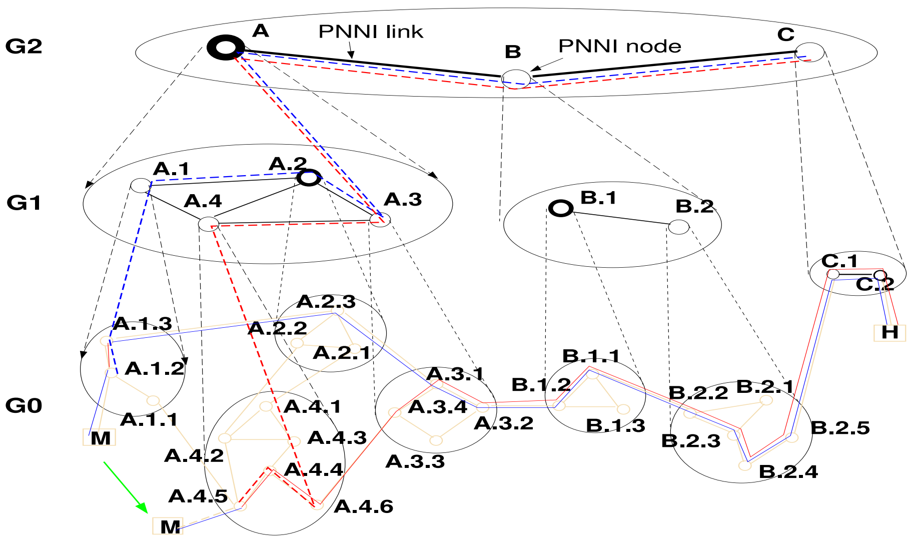

Figure 4.1: A sample PNNI network

4.1.1 PNNI hierarchy and source routing

An example of a PNNI network is shown in Figure 4.1. ATM switches in PNNI networks are clustered into peer groups hierarchically so that the network is scalable with capacity and traffic information aggregated into peer groups level-by-level. A PNNI routing protocol uses source routing mechanism which is performed hierarchically based on the network information associated with each peer group.

Each ATM switch maintains the information for its own peer group and its parent peer groups in order to make routing decisions when it initiates a source routing procedure.

For example, in Figure 4.1, switch $A.4.5$ has routing information of its peer group $A.4$ and its parent peer groups A and R (the top group). If a host at $A.4.5$ wants to have a connection to a host at $C.2$, a *PNNI source routing* procedure is performed on $A.4.5$ as follows. A *Designated Transit List* (DTL) $[A, B, C]$ is generated based on the topology and QoS information of group R, giving a path from A to the destination $C.2$ in the top peer group R. This DTL is pushed into a DTL stack. Then a DTL $[A.4, A.3]$ is generated in peer group A, giving a path from $A.4$ to B. Again the DTL is pushed into the DTL stack. And then a DTL $[A.4.5, A.4.4, A.4.6]$ is generated within peer group $A.4$, giving a path from $A.4.5$ to $A.3$. The DTL stack $[A.4.5, A.4.4, A.4.6], [A.4, A.3], [A, B, C]$

at the source routing switch $A.4.5$ forms a PNNI hierarchical route guiding to $C.2$, as shown with dashed line in Figure 4.1. After source routing is done, the DTL stack is carried by the SETUP message and updated at every *Entry Border Node*, – the first node of a new peer group on the way toward destination, until the destination switch is reached. For example, when SETUP message gets on $A.3.1$, the DTL for peer group $A.4$ is popped out and the DTL for peer group $A.3$ – $[A.3.1, A.3.2, A.3.3]$ is pushed on, shown in dashed line in group $A.3$. It continues until the the last peer group C is reached and PNNI routing ends at $C.2$ with a DTL stack $[C.1, C.2], [A, B, C]$. An actual connection path is established following the guide of the DTL stacks generated by the PNNI routing protocol, as shown in solid line in Figure 4.1.

For a conventional connection, a DTL stack is required for PNNI routing only when SETUP message traverses each source routing switch (first switch and each entry border switch). Although the DTL stack provides the routing information of a path, a regular ATM connection may not need to store it since it does not change its path during its active session. However, for a mobility capable connection, the routing information for the original connection path is useful for path rerouting algorithm. For this reason, the DTL stacks used for PNNI routing are assumed to be available during and after the connection path is setup, i.e. the DTL stacks for the current connection path is stored at each entry border node in every peer group from the directions from both end hosts since both ends can be mobile terminals. For the connection is Figure 4.1, for example, the DTL stacks are stored at $A.4.5$, $A.4.6$, $A.3.1$, $A.3.3$, $B.1.2$, $B.1.1$, $B.2.2$, $B.2.4$, $C.1$ and $C.2$.

In the unified handoff control protocol presented in Chapter 3, COS discovery does not utilize PNNI routing information. The flat rerouting (FRR) scheme is used which checks COS condition on every switch the SETUP message traverses. With proper utilization of the PNNI routing information, the unified handoff control protocol can be more efficiently applied to a mobile ATM network with PNNI hierarchical structure.

4.1.2 Hierarchical rerouting scheme

In the unified handoff control protocol we have proposed, the COS discovery signaling is integrated with the process of newPath establishment starting from the ProxyUNI. For all path rerouting algorithms, a SETUP message is sent to a destination through the newBS. At the newBS, based on the destination ATM address, a DTL stack is generated for source routing. We call this DTL stack the newDTL. On the other hand, we assume that a DTL stack at the oldBS is available for the original connection path (oldDTL). Using the oldDTL and the newDTL, a hierarchical rerouting scheme is proposed as follows:

- Comparing the oldDTL and the newDTL at the newBS, a *Candidate COS* is selected as the common node in both DTL stacks. The Candidate COS can be a complex node representing a peer group.
- The SETUP message is sent toward the candidate COS.
- The Am_I_COS() function at the *Exit Border Node* of the candidate COS, checks the criteria for the given type of COS.
- If the COS is found, goto the next step. Otherwise, update the candidate COS by comparing the updated newDTL with the oldDTL at the entry border node to the next peer group, then go back to the second step.
- The CONNECT message is sent back from the real COS to the newBS and then the rest of the rerouting procedure in the unified handoff control protocol is followed.

Let us look at an example. Assuming a connection from a host at the source switch $A.1.2$ to a host at the destination switch $C.2$ is being established, and a DTL stack (oldDTL) is generated, as $[A.1.2, A.1.3]$, $[A.1, A.2, A.3]$, $[A, B, C]$, shown in the solid line of Figure 4.2. Suppose the switch $A.4.2$ is a neighboring basestation of $A.1.2$, when a handover happens, a new DTL stack (newDTL) is generated to establish a new path for path rerouting. At the newBS, the newDTL is determined by the destination of the newPath as shown in Table 1.

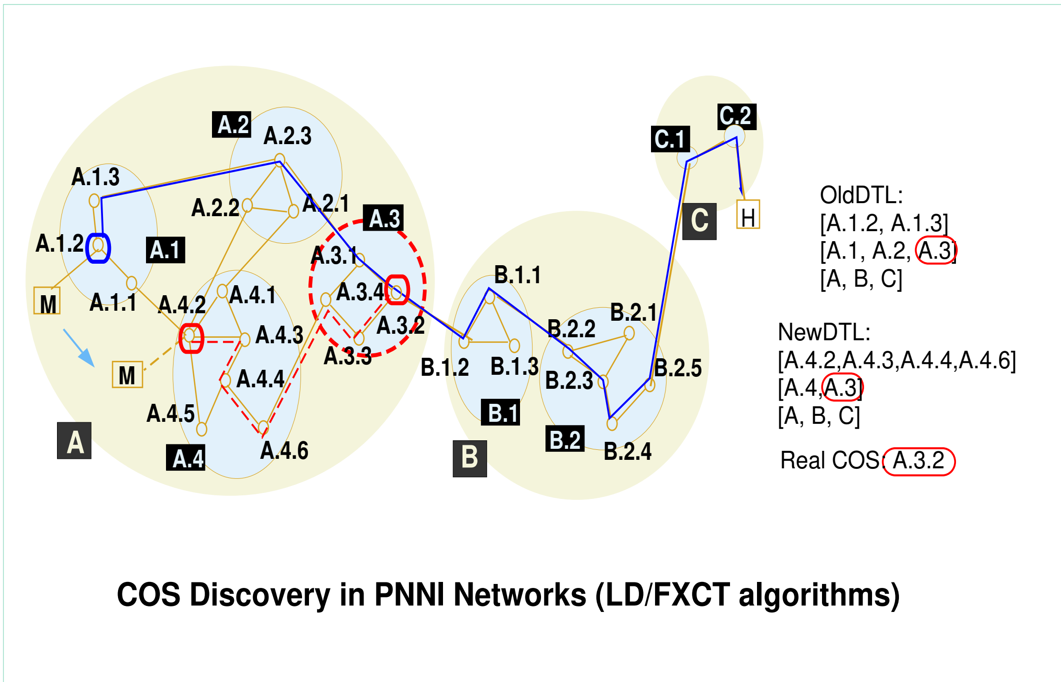

Figure 4.2: COS discovery in PNNI hierarchical networks (1)

In *First-Cross-to-CT* (FXCT) algorithm, for example, the destination for the new-Path is CT, then a DTL stack $[A.4.2, A.4.3, A.4.4, A.4.6]$, $[A.4, A.3]$, $[A, B, C]$ is generated, as shown in the dashed line of Figure 4.2. A DTL comparison function compares the newDTL with the oldDTL, and thus we can identify the first common node $A.3$, which is selected as the candidate COS. A real COS can only be at or after this node. Before the SETUP message gets into the peer group $A.3$, there is no need to invoke Am_I_COS() function. When the SETUP message enters the peer group $A.3$ at the entry border node $A.3.1$, a DTL $[A.3.1, A.3.3, A.3.2]$ is generated. At the exit border node $A.3.2$ toward the CT, the Am_I_COS() is executed to check the GCID. If GCID matches, the COS is discovered. Otherwise, the next candidate COS is selected by comparing the oldDTL and the newDTL. In this example, the next candidate COS is the node B. This process can be repeated until the destination switch is reached.

The *Last Divergence* (LD) algorithm has the same hierarchical rerouting scheme as the FXCT algorithm since their destinations are the same as the CT. However, the Am_I_COS() functions are different.

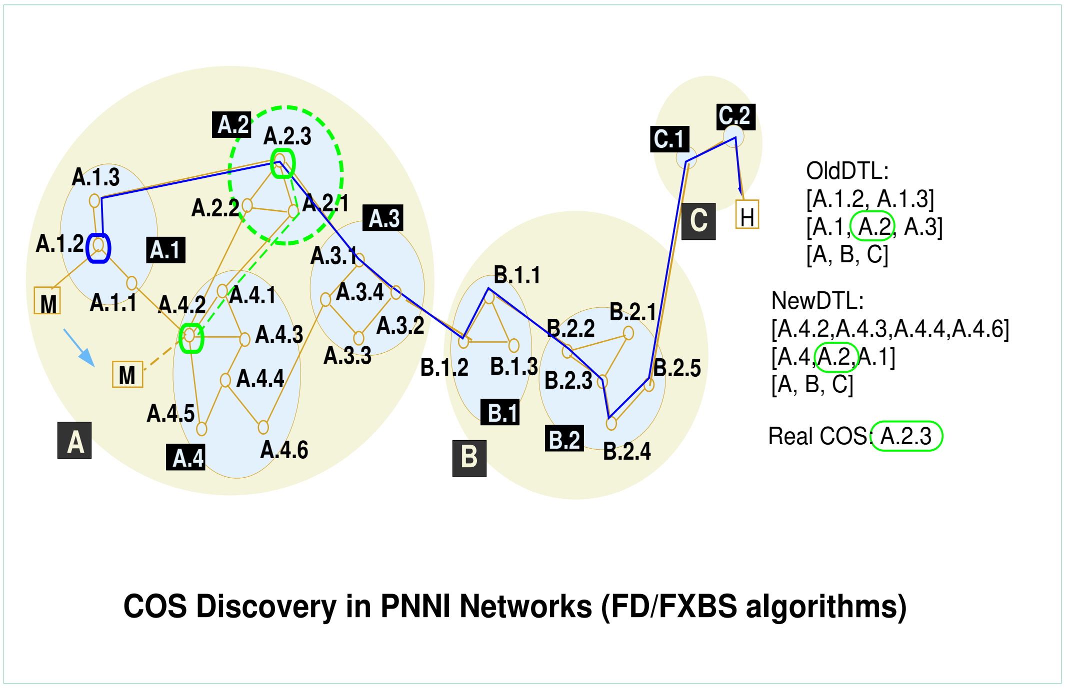

Figure 4.3: COS discovery in PNNI hierarchical networks (2)

If the destination is the oldBS, as in *First Divergence* or *First-Cross-to-BS* (FXBS) algorithms, the DTL comparison in the hierarchical rerouting scheme is very similar. By comparing the newDTL from left to right and top-down with the oldDTL, the first common node is selected as the candidate COS . In the example, the newDTL is $[A.4.2, A.4.1]$, $[A.4, A.2, A.1]$. The candidate COS is $A.2$ not $A.1$ although the later is also in common in both the oldDTL and the newDTL. The real COS that ultimately will be discovered is $A.2.3$, which is an exit border node of $A.2$ toward the oldBS.

This hierarchical rerouting scheme can considerably reduce time spent for COS discovery. The number of COS condition checks will be in the order of number of source routing for the new path setup. In a PNNI network, a source routing process is executed for every new peer group that SETUP message enters. Suppose the number of source routings for the new path setup is bounded by N_{RT}, the number of COS condition checks, N_{COS}, for COS discovery will be no more than N_{RT}. In the flat network path rerouting scheme, the N_{COS} is on the order of the network size, so it is unbounded and non-scalable.

The cost of limiting N_{COS} is that the discovered COS is not optimal with respect to the route obtained from a FRR scheme. We can see that unless the real COS is at the same peer group as the oldBS, it is not guaranteed that the COS is the first switch through which the new path crosses to the old path with the COS condition satisfied. This is obviously true since the COS condition is only being checked at the exit border node of the candidate COS while the real COS can be anywhere within the candidate COS. For example, in Figure 4.3, in the flat network, the real COS is $A.2.1$ while using the hierarchical rerouting scheme, the COS is discovered as $A.2.3$. Consequently, the connection path after handoff may have loops (duplicated part) which degrade the efficiency of the network service. Even if the suboptimal COS does not cause loops in the connection path, it may lead to suboptimal path or simply increase the processing cost for newPath setup and oldPath release.

The problem of suboptimal COS of the hierarchical rerouting, comparing with a flat rerouting scheme, is caused due to COS being checked only at the exit border node of the candidate COS. There is a way to improve the HRR scheme. Suppose the oldBS can keep all routing details of the original connection path. For example, for connection M to H, $[A.1.2, A.1.3]$,$[A.2.3]$,$[A.3.1, A.3.2]$, $[A.1, A.2, A.3]$, $[B.1.2, B.1.1]$, $[B.2.2, B.2.3, B.2.4, B.2.5]$, $[C.1, C.2]$, $[A, B, C]$, then it is possible for HRR scheme to find the COS that is same as what found by FRR. However, one may argue that it is not scalable to keep all source routing information for a connection path. This is because the PNNI routing itself is not scalable. If the PNNI network can be configured with scalable routing, the routing information – DTLs themselves can also be scalable. A PNNI network with scalable routing configuration can provide a path through a peer group with PVC without a network having PVCs at each level which provide path for a connection by passing a peer group without detailed source routing into that peer group. For example, the connection between M and H has routing information of $[A.1.2, A.1.3]$, $[A.1, A.2, A.3]$, $[A, B, C]$, $[C.1, C.2]$. There must be PVCs from the entry border node to the exit border node for peer groups, $A.2, A.3$ and B. In this case, the optimal COS is at exit border node of the candidate COS, because inside the candidate COS, there is no routing done for original connection, only PVC is used. Therefore, the HRR scheme will be equivalent to the FRR scheme when network is configured for scalable PNNI routing.

4.2 Zero-cell-loss handoff control

The handoff control process performs a replacement of an oldPath (between MT through oldBS to COS) with a newPath (between MT through newBS to COS). In order to guarantee no cell loss during the replacement process, a zero-cell-loss handoff control is considered. We propose to use inband signaling, i.e. using OAM (operation and management) cells to control cell buffering and transmission so that the cell streams of the handoff VCs can be rescheduled in a way that has no loss, duplication or misordering.

The problem of zero-cell-loss handoff has been studied previously. A typical example is the work done by H. Mitts et al. [17]. In their paper, a buffering mechanism at wireless access points is provided. This buffering mechanism does require the cell level interactions between the old basestation and the new basestation. Our approach is based on the local functions at cell level on each ATM device (switch or NIC card) [24]. The functions are used for re-scheduling cell transmission on the switch or the NIC card and can be embedded into the switch and the NIC hardware.

4.2.1 Cell level control functions on ATM devices

An ATM connection is by default a duplex connection with forward and backward directions. However, in a path re-routing process, the procedure for replacing the oldPath with the newPath is symmetric. As described below, three fundamental cell level control functions are required for zero cell loss control, from the upstream node to the downstream node along with the ATM cell transmission flow at each direction.

- **FUNCTION-I: Mark/redirect cells at upstream node**

  At the upstream node, the replacement of the old path with a new path can be performed in following way: (1) An End-of-Stream (EOS) OAM cell followed a Start-of-Stream (SOS) OAM cell are inserted into the VC to be re-routed. (2) The EOS cell will trigger the control program to replace the old path by the new

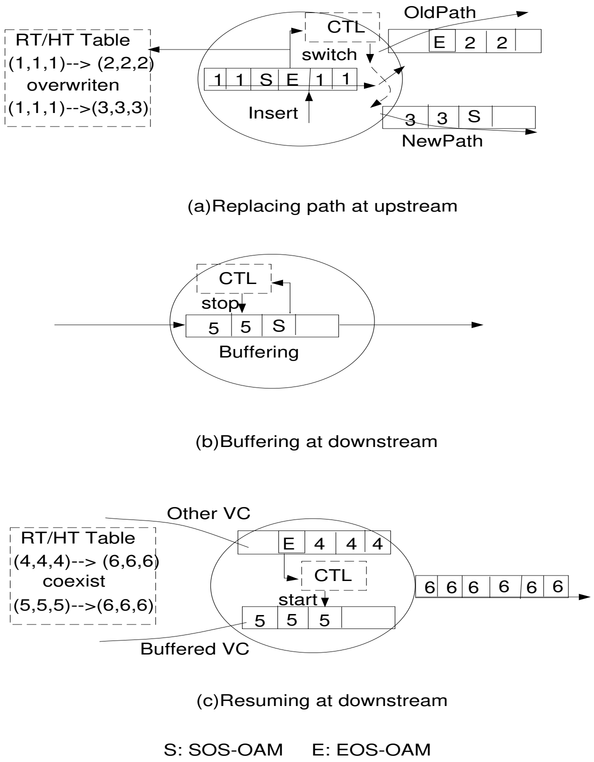

Figure 4.4: Cell level scheduling functions for generic path re-routings

path. In Figure 4.4(a), the notation (x, y, z) means a virtual circuit with VPI x, VCI y at port number z. A switch path $(1,1,1) \longrightarrow (2,2,2)$ is replaced with $(1,1,1) \longrightarrow (3,3,3)$, which will update the header translation (HT) table on the switch. (3) All the cells after EOS including SOS cell will be sent through the new path $(1,1,1) \longrightarrow (3,3,3)$.

- **FUNCTION-II: Buffer cells at downstream node**

  At the downstream node of the paths, as shown in figure 4.4(b), when an SOS is received, the transmission at the new path, i.e. the VC $(5,5,5)$, is stopped and cells are buffered into a given per-VC queue.

- **FUNCTION-III: Release cells at downstream node**

  At the downstream node of the path, the transmission in the new path, the VC $(5,5,5)$, can be enabled upon receiving EOS from the old path, as shown in Figure 4.4(c).

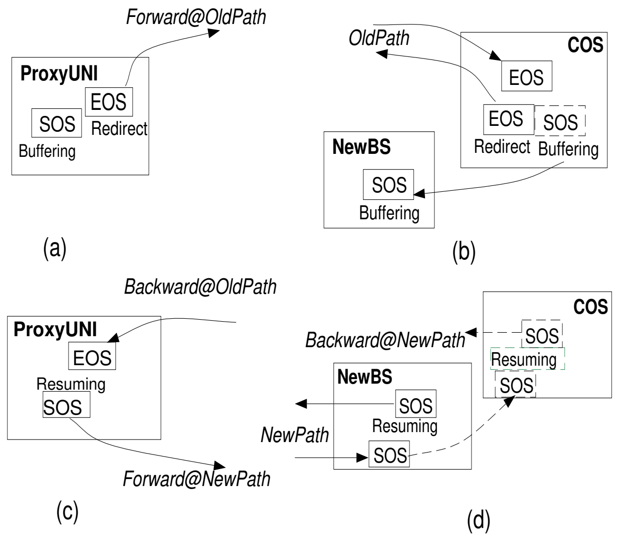

Figure 4.5: Cell level scheduling during path replacement

In handoff control, one end of the path to be rerouted is the mobile terminal (MT) and the other end is the Crossover Switch (COS). For the forward direction, from MT to COS, the upstream node is the MT and the down stream node is the COS. For the backward direction, from the COS to the MT, the upstream node is the COS and the down stream node is the MT. Since the new path between the COS and the MT is not completely established before the MT moves into new basestation's area, for the new path, the downstream node for buffering cells will become the last available node at the downstream. For example, the new BS is the downstream node for backward direction of the new path.

4.2.2 Cell level rescheduling during path replacement

In our unified handoff control protocol, at the end of phase I of the control process, a new path is setup between the COS and the newBS but not yet to the MT. Therefore, the forward cell stream (from the MT to the COS) must be buffered on the MT and the backward cell stream (from the COS to the MT) has to be buffered either at the newBS. The cell level scheduling procedure is described in the following four steps.

- **Inserting OAMs and buffering cell stream at the MT**

Illustrated in figure 4.5(a), at MT, the forward data stream is re-directed to the new path by inserting an EOS-OAM and a SOS-OAM. The EOS-OAM will trigger the VC switch to the new path (FUNCTION-I) and the SOS-OAM will disable the transmission at the MT and buffer the cells after SOS-OAM on a specific queue in the ATM NIC card (FUNCTION-II). We see that the second function is performed at the upstream node for the forward connection because the newPath is not ready in this direction.

- **Inserting OAMs at the COS and buffering data stream at the newBS**

  After the COS receives the EOS cell from the old path in the forward direction, an EOS-OAM followed by a SOS-OAM are inserted into the backward connection. Similarly, the EOS-OAM will trigger the backward VC switch to the new path (FUNCTION-I) and SOS-OAM will disable the transmission at the newBS (FUNCTION-II). The process is shown in Figure 4.5(b).

- **Enabling data transmission on MT**

  As shown in Figure 4.5(c), at the MT the cell transmission for the forward direction of the new path can be enabled by the EOS-OAM received from the backward direction of the old path (FUNCTION-III). At the cell level, if the cell stream is transmitted or not depends on whether the wireless link is available or not at the moment.

- **Enabling data transmission on the newBS**

  The cell stream of the backward connection is buffered at the newBS and the transmission is resumed when the SOS-OAM in the forward direction of the new path is received (FUNCTION-III). This is shown in Figure 4.5(d).

The in-band signaling using OAM cells guarantees the handoff connections have no cell loss, duplication or misordering, assuming the buffer size is reasonably large. Integrated with the out-band ATM signaling, a handoff control protocol with zero-cell-loss can be implemented.

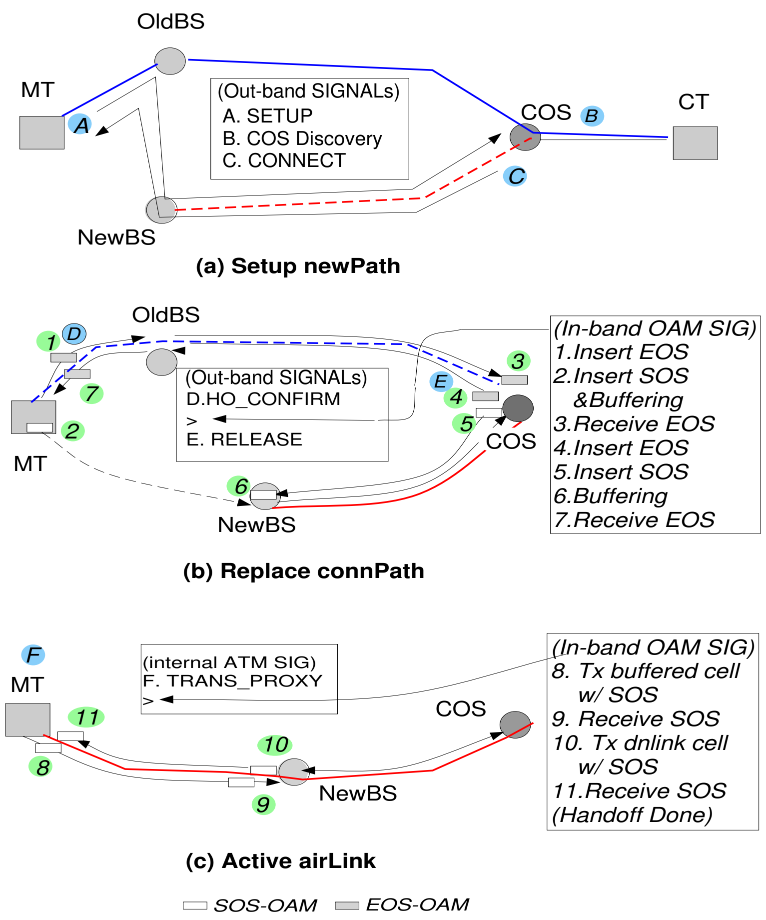

Figure 4.6: Control process for zero-cell-loss handoff

4.2.3 Protocol integration for zero-cell-loss

In the unified handoff control protocol, there are four phases in a handoff control process. We will describe how the in-band signaling works with each of them.

The messages and OAM cells are listed in Table 4.1 and the corresponding signaling sequence is shown in Figure 4.7.

In the first phase, from HO_REQUEST to HO_RESPONSE, no data VC has been processed yet, so lossless control does not start.

After the new path of all data VCs has been established through handoff SETUP /CONNECT signaling, the proxyUNI will send a CONN_READY message to the MT and let the MT decide when to handover to the newBS. A HO_CONFIRM message will trigger ATM in-band signaling starting from the ProxyUNI. For each VC, an EOS-OAM will be available at the MT (virtual WATM host). In the case of wireless ATM,

Message	Type	Source and Destination
HO_REQUEST	W-UNI SIG	MT to ProxyUNI
get(Resource)	ILMI REQ	ProxyUNI to MIB newBS
HO_RESPONSE	W-UNI SIG	ProxyUNI to IWF
SETUP	M-UNI SIG	ProxyUNI to COS via newBS
CONNECT	M-UNI SIG	COS to ProxyUNI via newBS
CONN_READY	W-UNI SIG	ProxyUNI to MT
EOS	ATM in-band SIG	MT to COS via oldPath
SOS	ATM in-band SIG	at MT via newPath
EOS	ATM in-band SIG	COS to ProxyUNI via oldPath
SOS	ATM in-band SIG	COS to newBS via newPath
RELEASE	M-UNI SIG	COS to MT via oldPath
HO_READY	W-UNI SIG	MT to ProxyUNI
TRANS_PROXY	internal SIG	oldBS to newBS)
CONN_ACTIVATE	W-UNI SIG	MT to ProxyUNI
CONN_ACTIVE	W-UNI	ProxyUNI to MT

Table 4.1: Signaling messages for Zero-cell-loss Handoff

the EOS-OAM is inserted at the MT after CONN_READY is received at the MT. In case of a nonATM radio, the EOS-OAM is generated by the IWF based on a message from MT, indicating it is leaving the cell.

After the newPath is established at the ProxyUNI, the connection branches to two output VCs, one is the old path and the other is the new path. The EOS-OAM will switch the data path from the oldPath to the newPath, meaning all succeeding cells will go to the newPath.

When the COS receives the EOS-OAM, it will insert an EOS-OAM and a SOS-OAM at the downlink direction. Similarly, the EOS-OAM will switch the connection path to the newPath and the SOS-OAM along with succeeding cells go to newPath. When the SOS-OAM reaches the newBS, it enables buffering of the succeeding cells.

When the MT receives the EOS-OAM from COS for all data VCs, it issues a HO_READY message to the oldBS, then moves to the newBS.

After CONN_ACTIVATE and CONN_ACTIVE are received by the newBS and the MT, the radio access and ProxyUNI are confirmed to be available. Both the newBS and the MT can release their buffers which are held by SOS-OAM cells, for all data VCs.

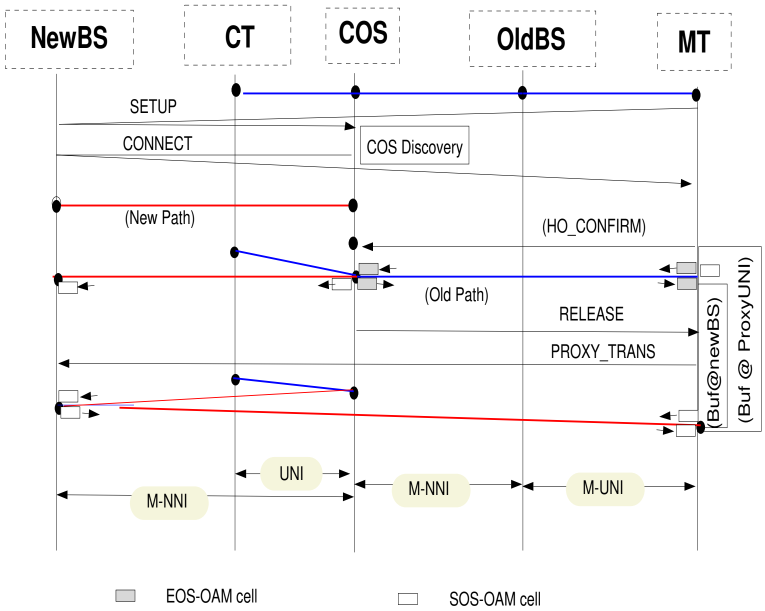

Figure 4.7: Signaling sequence of zero-cell-loss handoff

We have seen that the in-band signaling can replace the HO_CONFIRM and HO_READY messages; see Figure 3.1. However, the messages can be used for protecting cell loss. If HO_CONFIRM is used, an alternative sequence of in-band signaling can be applied. In this case, in-band signaling at the COS can start when the HO_CONFIRM is received. The EOS/SOS are first inserted into the backward direction of the paths. When the EOS-OAM is received by the MT, the EOS/SOS are inserted at the MT into the forward direction of the paths. The cell stream followed the SOS is buffered at the MT until the new radio link to the newBS is ready. In this sequence, the buffer size on the MT can be reduced while in the previous scheme, shown as Figure 4.7, the MT has to buffer upstream for a round-trip delay period for in-band signaling between the MT and the COS. Using this alternative sequence, we expect to observe less delay jitter for real-time services because of reduced handoff delay.

4.3 Path rerouting for point-to-multipoint connections

In the ATM Forum, specifications for point-to-multipoint (PMP) connection call control procedure have been standardized. When a mobile terminal on a PMP connection needs handoff, the mobile terminal has to be added at the newBS and dropped at the oldBS.

The handoff control protocol can be as simple as doing nothing by just dropping the mobile terminal first and then adding it at the newBS. However, it will only work for one type of PMP connection. Our objective is to propose a mechanism which works for all three types of PMP connections specified in ATM Forum and support zero-cell-loss control [23].

4.3.1 Point to multipoint connections

A PMP connection has one source and multiple destinations for a data communication service. The source is referred to as the *root* and each destination is referred to as a *leaf* of the connection. The paths of a PMP connection forms a tree rooted by the root node. According to ATMForum specifications [54, 55], there are three types of PMP connections as follows.

- **Root Initiated** the PMP connection is created by the root and only the root can initiate the signaling procedure by sending an ADD_PARTY message toward a new leaf.
- **Root LIJ** (root prompted leaf initiated join) A leaf can request to join the PMP connection by sending a LEAVE_SETUP_REQUEST message to the root. Upon receiving the message, the root can start the signaling procedure to add this leaf.
- **Network LIJ** (leaf prompted leaf initiated join) A leaf can request to join the PMP connection by sending a LEAVE_SETUP_REQUEST message *toward* the root. The network can identify a *proxy root* for the leaf from which an ADD_PARTY message is sent toward the leaf, so the nodes at upstream of the proxy root may not notice the leaf's join.

A PMP connection in a PNNI ATM network is shown in Figure 4.8. In a PMP connection, the branch overlapping between any two PNNI nodes must be avoided. It means that, at any hierarchical level, the PMP connection must form a tree. In other words, no loops or double links can be formed at any level. This requires that only the first leaf in a peer group can be added with a branch crossing the border of the peer

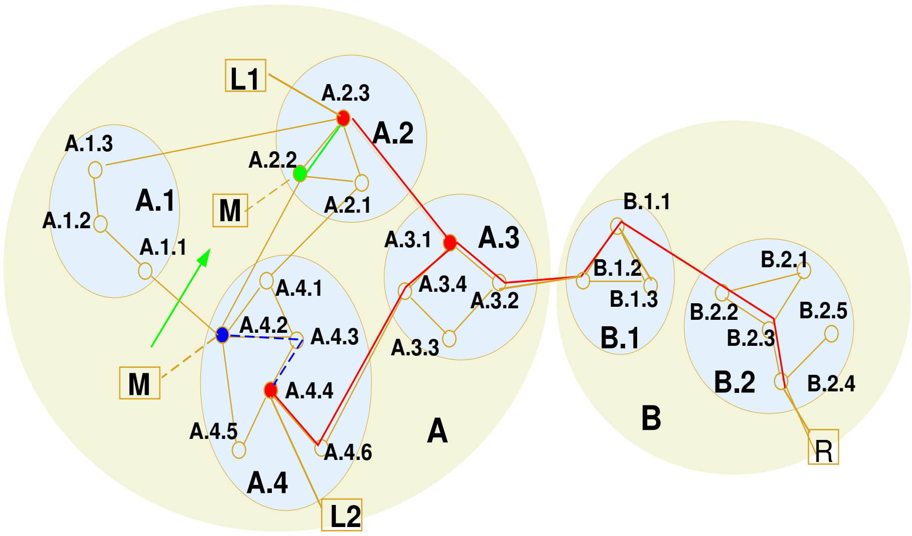

Figure 4.8: A PMP connection in a PNNI network

group to the PMP tree. The rest of leaves in the peer group must be added to the branches within the peer group. In Figure 4.8, for example, leaf $L2$ at $A.4.4$ is added to the PMP tree through $A.4.4, A.4.6, A.3.4, A.3.1$, which is the first leaf in the group that belongs to the PMP connection. If a new leaf M at $A.4.2$ is trying to join the tree, it is not allowed to be added through $A.2.2$ although that is the closest point on the connection to the leaf.

In order to avoid branch overlapping in a PMP connection, a *Connection View* is established for each node on the PMP tree [54, 55]. The connection view of a node is a tree graph rooted at the node. The PNNI routing algorithm can use a node's connection view to determine the DTLs along with the existing path on the tree toward the new leaf to be added. Without the connection view, a parallel path to the existing path may be created for adding a new leaf. Actually, it is not necessary for all the nodes on the connection to have a connection view. Only those nodes responsible for PNNI source routing must have it. This lead to another important entity in PMP connections – the *Entry Border Node*.

For a PMP connection, the entry border node of a peer group is, from the upstream of the connection with respect to the root, the first node in the peer group which is on the PMP tree. For example, $B.1.1$ is the entry border node for peer group $B.1$ and $A.3.2$ is the entry border node for peer group $A.3$ and A. Since root and entry border nodes are responsible for PNNI source routing, those are the nodes that need to maintain a connection view.

The protocol design for handoff control of PMP connections is intended to perform incremental path re-routing for mobile terminals, as in point to point connection handoff; however, the non-overlapping property has to be maintained through by using the available topology information provided by the node's connection views.

4.3.2 Handoff control functions

We will only focus on a leaf's handoff. We believe that in most situations, the root must be selected as a fixed host for a PMP connection. If in case the root is also a mobile terminal, a fixed host on the tree must be selected as the root's mirror side during the PMP connection session. Otherwise, the non-overlapping guarantee of a PMP connection could be very costly, if not impossible. We will leave the root's handoff as future work. We define the capability and functions of a leaf's handoff for PMP connections as follows.

Capability:

A mobile terminal, as a leaf of a PMP connection, can seamlessly become a new leaf of the PMP connection through a handoff control process when the mobile terminal moves from one wireless access point to another wireless access point.

Participating Entities:

The network entities involved a handoff control process for a PMP connection are illustrated in Figure 4.9. They are:

1. mobile terminal (MT)
2. root or proxy root (RT) of the MT
3. old basestation (oldBS), covering the current cell of MT
4. new basestation (newBS), covering the cell where MT is about to enter
5. crossover switch (COS): a node on the PMP tree from where the MT is added to the newBS

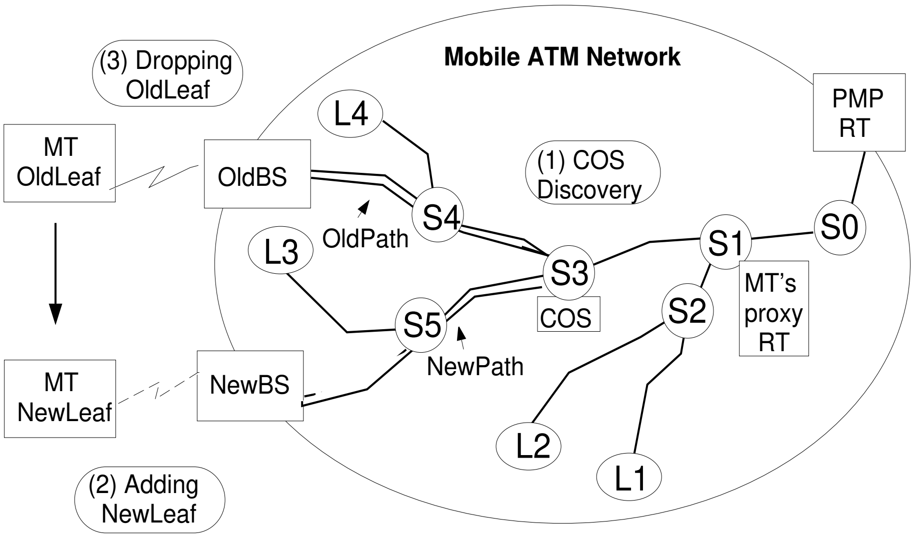

Figure 4.9: Handoff control functions and participating Entities

We refer to the MT at the oldBS as an old leaf (oldLeaf) and at the newBS as a new leaf (newLeaf). The connection path between the oldBS and the COS is referred to as an old path (oldPath) and that between the newBS and the COS is referred to as a new path (newPath). In Figure 4.9, $S1$ is marked as the proxy root for the MT. In practice, it is possible that a COS is an upstream, downstream or sibling node to the MT's proxy root.

Primary functions:

A handoff control process requires the following primary functions to maintain the connectivity of a mobile user to a PMP connection in the network: (1) a COS discovery process, (2) adding the MT as the newLeaf from the COS, (3) dropping the MT as the oldLeaf from the COS.

When adding the newLeaf, it is possible that a part of the newPath is already on the PMP tree (from $S3$ to $S5$). When dropping the oldLeaf, the oldPath is released to the point that no more leaf exists downstream (between the oldBS and $S4$ if no more mobile at the oldBS).

COS discovery:

The key function is the COS discovery which is different from either the proxy root discovery for Network LIJ or the COS discovery for a point to point connection. New procedure must be designed.

In the handoff process for a point to point connection, a COS can be selected anywhere on the original path from the oldBS to the last switch attached to the fixed host [20]. However, the COS selection for a PMP connection is very restrictive due to the non-overlapping feature of the PMP connection. A node has to be an entry border node whose peer group covers the newBS, in order for the node to be qualified as a COS.

In order to differentiate a handoff process from a two-step add/drop process which can be supported by existing protocols in the Network LIJ connection procedure, we define a COS as an entry border node whose peer group covers both the oldBS and the newBS. We will see with this requirement, the handoff control process is consistent for all three PMP connections and the lossless control is straight forward.

4.3.3 Handoff control process

Three types of PMP connections differ in (1) who initiates the joining request (root or leaf) and/or (2) who initiates the add procedure (root or proxy root). Given the tree graph of a PMP connection, the connection view of a node may be different in different types of connections. For example, a leaf added in a Network LIJ connection through a proxy root may not be in the view of some upstream nodes, while in other two types, every downstream leaf is in the view of an upstream node.

However, we have found that without leaf information, the connection view of a node must be the same for all three types of PMP connections. Since this view is used for PNNI source routing without causing branch overlapping, we define it as the *PNNI Route View* of a node. For all three type of PMP connections, a PNNI route view must be at least maintained at both the root and every entry border node of the connection.

We will design a common handoff control protocol for all three types of PMP connections, using the common available information – the node's PNNI Route View.

A. COS discovery

A crossover switch (COS) is a node on the PMP tree which initiates the adding procedure for the newLeaf. It is intuitive that a COS must satisfy the condition of a proxy root for the newLeaf, which is an entry border node of a peer group covering the newBS. However, the procedure to discover a COS may be different from finding a proxy root in a Network LIJ connection.

Just as in point to point connection handoff, we have two ways of signaling for COS discovery, i.e. signaling either from the newBS or from the oldBS. In both cases, the signaling message for COS discovery must pass to an entry border node of the peer group which covers the newBS.

In the case that the signaling message is initiated from the oldBS, the message with a Leaf Locator IE (information element) containing the newLeaf's ATM address is sent upstream following the PMP tree until it reaches an entry border node which covers the newBS. For example, in Figure 4.8, the message goes from the oldBS ($A.4.2$) to $A.3.2$, the entry border node of A, which covers the newBS ($A.2.2$). The discovered COS is an entry border node whose peer group covers both the oldBS and the newBS. In a Network LIJ connection, the COS is not equivalent to the proxy root of a leaf being added at the newBS. In the example, the proxy root of the MT is the $A.2.3$ while the COS is the $A.3.2$.

In the second case, the signaling message for COS discovery is initiated from the newBS. The procedure is similar to a proxy root discovery in Network LIJ. However, because the oldBS is also within the picture, their processes are slightly different.

In a Network LIJ connection, a LEAF SETUP REQUEST, which is responsible for finding a proxy root, can hit the PMP tree at the either a downstream, upstream or sibling node to the proxy root. After it hits the PMP tree, it is transfered toward upstream or downstream to locate the proxy root – the entry border node for the lowest peer group which covers the new leaf. For example, in Figure 4.8, when M at $A.4.2$ initiates a LEAF SETUP REQUEST towards R at $B.2.4$, the message may hit $A.2.3$ (a sibling node) on the connection first. The message need to traverse upstream to $A.3.2$ and then downstream to $A.4.6$ which is found to be the proxy root for the new leaf M at $A.4.2$.

Similarly, the signaling message for COS discovery initiated from the newBS may hit the PMP at the downstream, upstream or sibling node of the proxy root for the newLeaf. The message may need to traverse upstream and/or downstream to locate the COS – an entry border node of the peer group which covers both the oldBS and the newBS. In the example, the mobile terminal at the newBS ($A.2.2$) hits the PMP tree at the $A.2.3$. Although $A.2.3$ is a proxy root of the newLeaf, the COS is the node $A.3.2$, which covers both $A.4.2$ and $A.2.2$.

B. Re-routing procedure

In a PMP connection, although all the leaves share same data stream, each leaf is uniquely identified in the PMP tree with an *Endpoint Reference* for connection trace purposes. When a leaf is added on the tree, an endpoint reference is stored in the nodes from the root (proxy root) to the node to which the leaf is attached. The endpoint reference is not necessarily an end-to-end identifier. An upstream root (proxy root) may assign a number for a leaf as the endpoint reference but that number may be in use for another leaf on a downstream link (in Network LIJ). In this case, the endpoint reference can be mapped to an unused number at a node. In general, the endpoint reference can be a hop-by-hop identifier. Using the endpoint reference, the root can quickly trace a leaf along the tree without invoking the PNNI routing procedure, which can be used for any control signaling, such as dropping a leaf.

The endpoint references for a leaf forms a path from the root or proxy root to the leaf. We call this as a *Leaf's Path* on the PMP tree. For the PMP connection handoff, if the oldPath of the MT (see Figure 4.9) is available at the COS, the dropping procedure of the oldLeaf can be initiated from the COS. However, in a Network LIJ connection, the MT's oldPath may not be available because its endpoint reference cannot be identified locally on the COS. This is because (1) the endpoint reference is not end-to-end between the MT and its proxy root, and (2) the COS may be found upstream of the MT's proxy root, which is beyond the scope of the MT's oldPath. Although this problem can be solved by introducing new IDs and/or extending the endpoint reference from the proxy root to the COS, the simplest way to avoid it is to initiate dropping procedure from the MT itself. We will see how this can be done in conjunction with lossless control.

The control process of a PMP connection handoff using the oldBS COS discovery is illustrated in Figure 4.10. It includes the following steps.

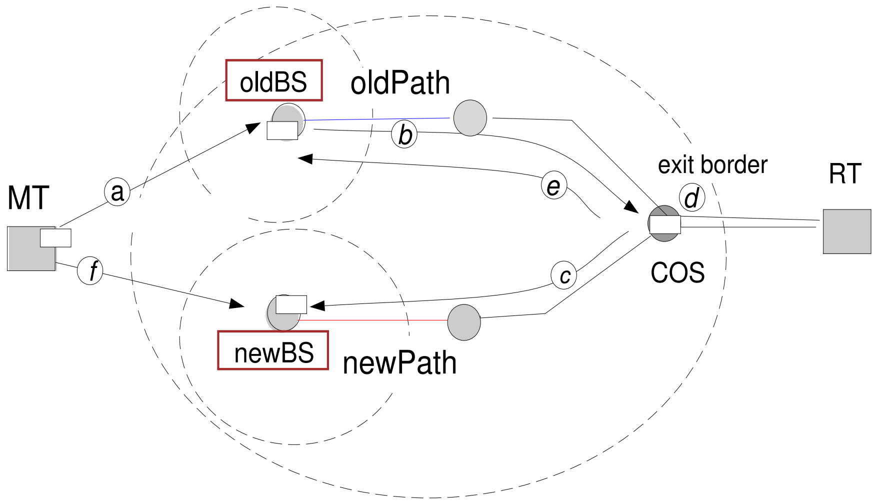

Figure 4.10: Handoff control process in PMP connections

- Handoff initiation: a handoff request is issued from the MT at the oldBS towards the RT.
- COS discovery: the request traverses an entry border node whose peer group covers the newBS.
- Adding the newLeaf: setup a path to the newBS, with necessary endpoint reference mapping.
- Lossless control: using in-band signaling to synchronize stream.
- Dropping the oldLeaf: release the path to the oldBS, delete the endpoint reference.
- Handoff completion: the MT leaves the oldBS's cell and enter the newBS's cell.

An important issue is lossless handoff. Just as in point to point connections, this can be realized through in-band signaling with OAM cells. Since a PMP connection has only downstream data transmission to the MT, the lossless control can be performed as follows. A marker OAM cell can be inserted at the COS. When the MT receives the marker OAM cell from the oldBS, it can request dropping the oldLeaf and moves to the newBS. At the newBS, when the marker OAM cell is received, cell buffering starts until the newBS receives another OAM cell showing that the MT is ready.

C. Signaling syntax and sequence

Based on the control process in Figure 4.10, we define a signaling mechanism by extending ATM signaling at the UNI and NNI interfaces. This includes new and modified ILMI (Interim Local Management Interface) and Q2931 signaling messages. We call the extension sets ILMI+ and Q2931+, respectively, as shown in Figure 4.11. New messages and modified messages are shown in time sequence. The signaling sequence has the following aspects.

- **Resource registration on the new BS**

  Through ILMI+ signaling, a mobile ATM terminal registers on the newBS to obtain resources, such as ATM address, signaling PVCs and radio channel. The mobile may not need a full registration to the newBS as it boots up, but the resources availability must be checked before data connections' handoff. The ILMI+ messages MT_HO_REQUEST /RESPONSE are used for resource acquiring from newBS, MT_LEAVE_CELL releases the resources on oldBS and MT_ENTER_CELL activates the using of the resources on newBS.

- **COS discovery**

  A new message PMP_HO_REQUEST is defined for initiating the handoff control process and discovering the COS. In the message, a Leaf Locator ID information element for the newBS is included. The COS is selected when the message traverses an entry border node whose peer group covers the newBS.

- **Adding the newLeaf**

  After the COS is discovered, an ADD PARTY / SETUP message is sent toward the newBS. A *Handoff Control* information element (HCIE) is included in the message to indicate the message is for handoff control. A regular SETUP message ends at the user side UNI interface. But the SETUP with HCIE can end at the network side UNI interface on the newBS. It is only responsible to establish the path from the COS to the newBS. Although SETUP does not reach the MT, which is not yet at the newBS, the MT is virtually added through the newBS.

- **Cell synchronization**

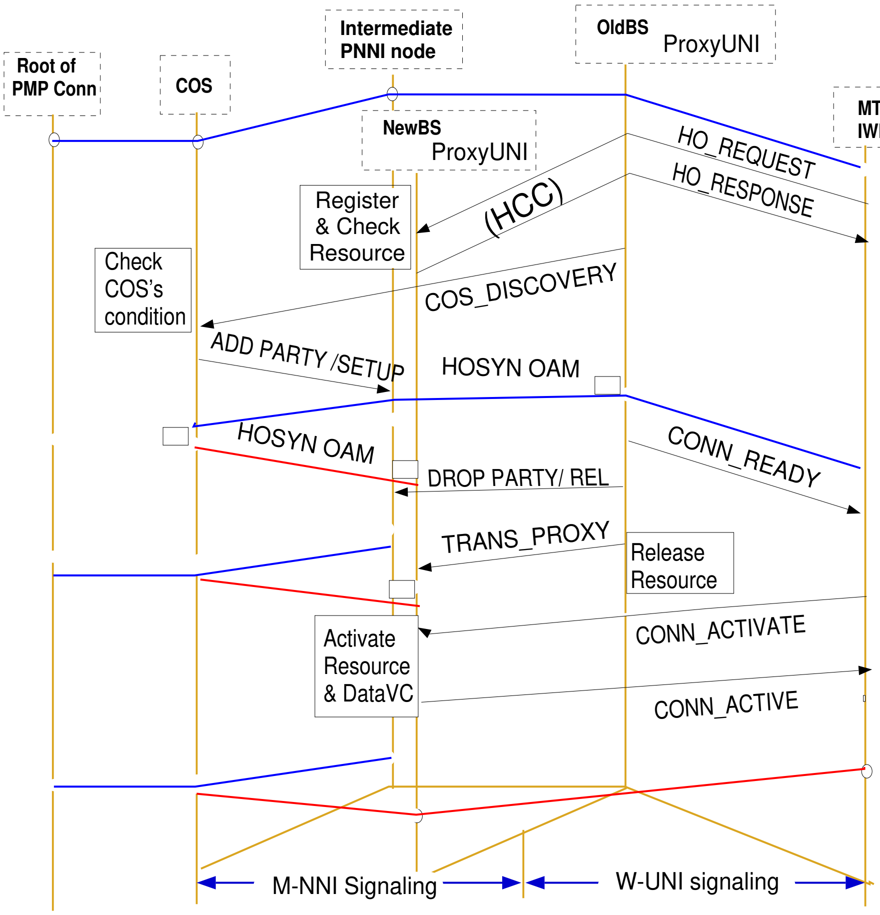

Figure 4.11: Signaling sequence of handoff control in PMP connections

In order to avoid cell loss or duplication, we introduce an OAM cell called the *Handoff Cell Synchronization* (HOSYN) OAM. After the new path for the MT is added from the COS to the newBS, i.e., when COS received a CONNECT / ADD PARTY ACK, the HOSYN OAM cell is inserted to the input VC at the COS. Both the oldBS and the newBS can receive this marker cell. When the newBS receives the HOSYN, it knows a specific mobile terminal is in handoff process. It will buffer the cell stream after the HOSYN until a control message is received from the MT.

- **Dropping the oldLeaf**

When the MT at the oldBS receives the HOSYN, it can send a DROP PARTY request to the root. This message has also a HCIE indicating it is a handoff control message. The message goes up to the COS or MT's proxy root, whichever is reached first. Another option is that the COS send the DROP PARTY message to the oldLeaf after HOSYN is sent out. This option requests that the endpoint reference for the MT is available at COS, possibly extended by PMP_HO_REQUEST from downstream.

- **Completing the new path**

  Note that the newLeaf is added prior to the MT moves to the newBS. If lossless control is required, the data cell is buffered at the newBS before the new path is complete to the MT. A new message HO_ACTIVATE is used to release the cell stream buffering on the newBS. The message is sent by the MT to the newBS. One HO_ACTIVATE can be used for all VCs of the MT. Upon receiving the message, the newBS release all the buffering for the MT, including both point to point and point to multipoint connections.

Chapter 5

Performance Analysis on Handoff Path Rerouting

In mobile ATM networks, the traffic distribution is affected by user mobility and handoff path rerouting. Extra traffic load may be added to network links according to mobility patterns and path rerouting algorithms. This requires higher network link capacity and possible topology re-engineering in order to support the same QoS for mobile services. To capture the dynamic variations of traffic due to user mobility and path rerouting, we propose to use a *Flow Model* for mobile ATM networks. The model represents the mobile-generated traffic as a set of stochastic flows over a set of OD (Origin-Destination) pairs. The user mobility is defined by transfer probabilities of the flows and the handoff path rerouting algorithm is modeled by a transformation between the routing functions for traffic flows. The analysis shows that user mobility may cause temporal variations as well as smoothing effects on the network traffic. Using the flow network model, typical handoff path rerouting algorithms are evaluated through both analytical and experimental approaches. The evaluation methodology can be used for either redesigning the network topology for a given path rerouting algorithm or selecting a path rerouting algorithm for a given network topology under a specific mobile service scenario.

5.1 Introduction

Traditionally, in a circuit network such as telephony system, a queueing model is used to describe traffic distributions. In such a queuing model, the traffic at burst level is constant bit rate (CBR) and the traffic at call level is assumed as a Poisson arrival process. Although an ATM network uses virtual circuits for each call, its traffic distribution is largely different from telephony system at both call level and burst level. The integrated services offered by an ATM network produce calls with various distributions of inter-arrival time and call lifetime, which result in diversified traffic statistics at the call level. At the burst level, the bit transmission rates of calls may include a variety of types, such as CBR, VBR, ABR or UBR varying over a very large range.

In mobile ATM, two new features, multimedia and mobility, raise basic difficulties in traffic modeling and network performance analysis. Multimedia connections imply a large number of traffic classes with heterogeneous service rates and terminal mobility causes traffic flows to migrate from one network link to another.

In a traditional traffic modeling hierarchy [34], the most detailed model is the queuing model which gives probability distribution over all states. The next level is a diffusion model which characterizes the traffic flows by mean and variance. The third level is fluid approximation which takes time-varying averages and analyzes the non-stationary behavior of the traffic. The flow model with long term averages of traffic flows, which is used for network design, is perhaps the most abstract. Based on the above hierarchy, multimedia traffic is often described by hybrid models which characterize traffic flows at different time scales with varying degrees of detail [33]. These include fluid flow models that uses a queuing model at call level and a fluid approximation at burst level [26]. Another example is Markov-Modulated Poisson Process(MMPP) [27] which uses a histogram approximation at burst level.

Traffic modeling can be used for either off-line network design or online traffic control. Two problems may be decoupled and use different traffic models. For network design, a static flow model is often used [28] to construct a multi-commodity problem based on network topology and offered traffic. For the online traffic control in a broadband network, traffic models with bursty level distributions are required to estimate the QoS over network links [29, 30, 32, 33]

In mobile ATM networks, we wish to use traffic models for both network design and traffic control. First, we want to know if mobility support results in higher network capacity requirements or in general, different network topology design. This is a network design problem. Second, we want to control QoS guarantees for each connection in the network, especially to relieve the QoS degradation caused by handoff. On the one hand, a simple static flow model won't reflect the dynamic route assignments and correlations of traffic flows caused by user mobility. On the other hand, as we pointed out, a multi-class queuing model is complicated and needs strong assumptions when mobility is incorporated. Therefore, we propose a network *Flow Model* which has moderately detailed statistics. In the model, the overall bandwidth of the aggregated traffic on each OD pair is represented by a continuous random process (stochastic flow) and the user mobility is captured by the migration of the traffic flows on different OD pairs.

By providing integrated data services, the time scale for call level and burst level in an ATM network becomes vague because a user can request, at a short time, a burst of call setups with very small bandwidth requirements, or in a long time, a call of only one burst with very large bandwidth requirements.

In this case, if a traditional queuing model is used, a number of traffic classes should be specified and each class may have a specific mobility pattern associated with a type of mobile service. Using a queuing model, the performance analysis will become extremely complicated. Fortunately, our target for traffic modeling is not to know the number of calls in each traffic class as traditional call admission control (CAC) needs. Our target is to estimate overall bandwidth requirement of the integrated service under a mobile environment, in order to re-engineer the network topology or link capacity. Therefore, we consider a *Flow Model* representing the distribution of integrated traffic over the mobile network. A *flow* is a continuous function of continuous time, representing the bandwidth consumption of offered traffic. Without distinguishing the call level and the burst level, we claim that the statistical variation of a flow may be caused by both call arrivals, departures, handoffs and/or rate variation of each call at the same time scale. That is, in any time interval (Δt), the variations caused at call level and at the burst level are at the same order of Δt.

We do not address the optimization problems in network topology or routing design. Instead, we evaluate the extra cost that a core access network must pay in order to support mobile services. In addition, we identify the relative costs of different path rerouting algorithms. The purpose of performance evaluation for handoff path rerouting algorithms is twofold. First, the results can be used for re-engineering the network topology and link capacities of the ATM core network according to the user mobility pattern and path rerouting algorithm. Second, for a given mobile ATM core access network, the results can be used for selecting a path rerouting algorithm or altering user mobility patterns for a specific mobile service.

5.2 Flow model for mobile ATM networks

5.2.1 Definition of mobile ATM network

A mobile ATM network consists of a core ATM network and wireless access points. The core ATM network has ATM switches as *nodes* and fiber optic links between switches as *edges*. The network shown in Figure 5.1 is a mobile ATM network. Some of the nodes (switches $A.4.5, A.4.6, A.3.4, A.3.3$) are radio access points (basestations). Each radio access point may serve several cells with multiple antennas. If there is a traffic flow from a mobile user M at the access point $A.3.3$ to a fixed host H, the flow may take a route as shown in the figure. When the mobile user moves from cell $a4$ to cell $a0$, it is necessary for the network to reroute the flow to a route which goes through the new access point $A.4.6$. This requires a handoff control process and path rerouting algorithm. Figure 5.1 depicts three new routes. Each has a different anchor node (crossover switch or COS) to reroute the connection path, corresponding to a particular path rerouting algorithm. First is the *Last Divergence* (LD) algorithm with the COS at $A.2.3$. It aims to improve the efficiency of the network resource utilization by optimizing the path length after each handoff. Second is the *Path Extension* (PE) algorithm with the COS at $A.3.3$. It aims for low complexity by just extending the original path from the old access point to the new access point. Third is the *First Divergence* (FD) algorithm with the COS at $A.3.4$. It is a tradeoff between complexity and performance by avoiding the duplicate paths generated by PE algorithm.

5.2.2 Network topology and mobility pattern

A mobile ATM network has a *topology* which can be defined as a graph $G = [X, C]$, where $X = \{1, 2, ..., N\}$ are N *nodes* in the network and $C = \{c_{ij}, i, j \in X\}$ is a *link*

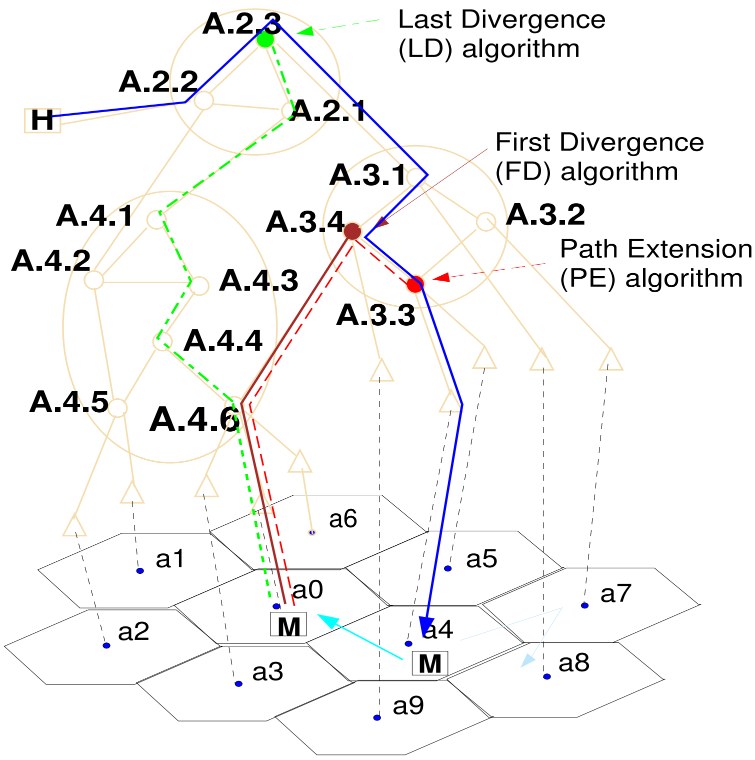

Figure 5.1: A mobile ATM network and path rerouting

capacity matrix. If $c_{ij} > 0$, we say there is a link, or a C-*edge* (Capacity edge) from node i to node j.

In addition, the network topology of a mobile ATM network has a *mobility pattern*, $G_w(t) = [Y, D(t)]$, where, $Y \subset X$ is a set of wireless access points and $D(t) = \{d_{ij}(t)\}$ is the *User Transition Matrix* of the network. At time t, a mobile user departing from access point i will enter node j with probability $d_{ij}(t)$. If $d_{ij}(t) > 0$, we say there is an M-*edge* (Mobility edge) from node i to j, i.e. node j is a neighbor of node i at time t. The user transition matrix $D(t)$ will depend on the users' call lifetimes and access point dwell times.

5.2.3 Traffic flows and handoff

For a given mobile ATM network with topology G, the network traffic can be specified as bandwidth requirements over the OD pairs. Each requirement can be represented by a random process (flow) in the unit of bandwidth per second. A flow consists of infinite numbers of infinitesimal calls over the OD pair. Each call can be viewed as a contributor of a *mini flow*, which may take a specific route in the network. The bandwidth of a traffic flow in a mobile ATM network is randomly changed because of call arrivals and departures, call handoffs and call bursts. In general, a multimedia traffic flow may be quite complicated statistically. However, in this paper, we assume that the bandwidth requirements of traffic flows are memoryless, meaning the holding time is exponentially distributed. In this case, the traffic flow $\psi_k(t)$ on OD pair k is a Markov process which satisfies the following stochastic differential equations, for $k = 1, ..., K$:

$$\Delta\psi_k(t) = -\mu_k(t)\psi_k(t)\Delta t + \sum_{l}^{M} p_{lk}(t)\mu_l(t)\psi_l(t)\Delta t + \Delta\phi_k(t) \tag{5.1}$$

In equation (5.1), $\phi_k(t)$ is the *traffic demand* representing new bandwidth requirement (in bits/sec) over the time period Δt, $\mu_k(t)$ is the service rate of traffic flow $\psi_k(t)$ on OD pair k, and $p_{lk}(t)$ is the *handoff transfer probability*, representing the probability that a call in the traffic flow finished its service in an OD pair o_l and moves to a neighboring OD pair o_k. The probability that a departure call from o_l terminates is $p_{l0}(t)$. Clearly, $\sum_k p_{lk}(t) = 1$. With $\Psi(t) = [\psi_1(t), \ldots, \psi_K(t)]^T$ and $\Phi(t) = [\phi_1(t), \ldots, \phi_M(t)]^T$, equation (5.1) becomes in matrix form,

$$\dot{\Psi}(t) = A(t)\Psi(t) + \dot{\Phi}(t) \tag{5.2}$$

where $A(t)$ is *flow transition matrix* with $A_{ij} = -\mu_i(t)$ for $i = j$ and $A_{ij} = \mu_i(t)p_{ij}(t)$ for $i \neq j$. The flow transition matrix $A(t)$ can be derived from the user transition matrix $D(t)$ given the call lifetime and user dwell times at the wireless access points, see Appendix A.

5.2.4 Routing functions and path rerouting algorithms

In a mobile ATM network, each traffic flow $\psi_k(t)$ has a *routing function* defined by a vector $V_k(t) = [v_{k,1}(t), \ldots, v_{k,M}(t)]^T$, where $v_{k,u}(t)$ is the probability that the link u is being used by any call in the flow $\psi_k(t)$. For example, in Figure 5.2, suppose ψ_1 is the flow for OD pair AD. One part of the flow takes the route ABD and rest of the flow takes the route $ACBD$. Corresponding to the order of the network links $[AB, AC, CB, CD, BD]$, one possible routing function is $V_1 = [0.8, 0.2, 0.2, 0, 1]^T$.

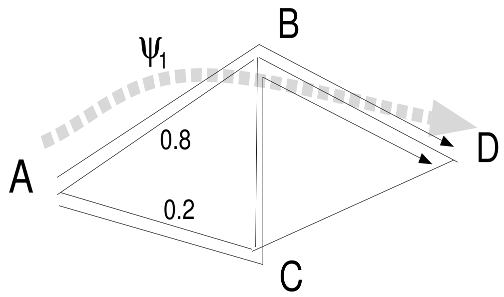

Figure 5.2: Illustration of flow assignments by routing function

The routing function for each traffic flow may be static, as a result of network synthesis, or dynamic, as a result of dynamic routing process. In either case, the routing function may be varied by the path rerouting process for handoff control.

In particular, we assume that the traffic demand $\phi_k(t)$ is initially routed according to the routing function $Z_k(t) = [z_{k,1}(t), \ldots, z_{k,M}(t)]^T$ where $z_{k,u}(t)$ is the fraction of the traffic demand $\phi_k(t)$ on C-edge u. In addition, we define $w_{lk,u}(t)$ as the fraction of the handoff traffic on C-edge u based on a path rerouting algorithm. Suppose flow $\psi_k(t)$ on a network link u is $v_{k,u}(t)\psi_k(t)$. For all k and u, it must satisfy the traffic balance equation

$$\Delta(v_{k,u}(t)\psi_k(t)) = -\mu_k(t)v_{k,u}(t)\psi_k(t)\Delta t + \sum_{l}^{M} w_{lk,u}(t)p_{lk}(t)\mu_l(t)\psi_l(t)\Delta t + \Delta(z_{k,u}(t)\phi_k(t)) \tag{5.3}$$

To characterize $w_{lk,u}(t)$, we define

$$W_{lk}(t) = [w_{lk,1}(t), \ldots, w_{lk,M}(t)]^T = H_{lk}(t)V_l(t) \tag{5.4}$$

where the $M \times M$ matrix $H_{lk}(t)$ is called a *path rerouting transform matrix* , which re-assigns the handoff flow, from $\psi_l(t)$ to $\psi_k(t)$, to a new route according to given path rerouting algorithm.

Let us give an example in Figure 5.3. The network has four nodes and two OD pairs with flows ψ_1 and ψ_2. The initial routing functions are $Z_1 = [10001]$ and $Z_2 = [00101]$, respectively, for links $[AB, AC, CB, CA, BD]$. A handoff path rerouting algorithm defined by $W_{12} = H_{12}V_1 = V_2$ and $W_{21} = H_{21}V_2 = V_1$ assigns handoff flows to the same routes as offered flow. In this case, the path rerouting transform matrix can be defined

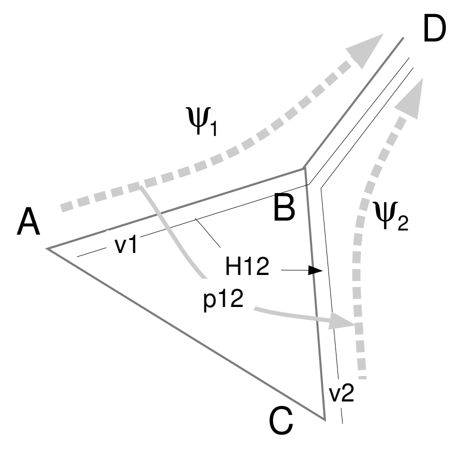

Figure 5.3: Illustration of path rerouting algorithms

as

$$H_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad H_{21} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In H_{12}, $[H_{12}]_{11} = 0$ means all handoff traffic from AB goes away while $[H_{12}]_{31} = 1$ means all handoff traffic from AB goes to CB.

In general, the rerouting of the handoff flows can be either static and based only on fixed network topology or dynamic and based on both network topology and traffic loads. The $H_{lk}(t)$ may be deterministic functions of time or random processes. If $H_{lk}(t)$ is deterministic, for given path rerouting algorithm (given rules of rerouting), it is not difficult, though complicated for a large network, to obtain $\{H_{lk}(t)\}$. Otherwise, it is possible to estimate $\{H_{lk}(t)\}$ through statistical measurement at the time scale with any significant change of user mobility pattern. With the path rerouting transform matrix $\{H_{lk}(t)\}$, we can solve for the routing functions $\{V_k(t)\}$ by following equations.

$$\Delta(\psi_k(t)V_k(t)) = -\mu_k(t)\psi_k(t)V_k(t)\Delta t + \sum_{l}^{K} p_{lk}(t)\mu_l(t)\psi_l(t)W_{lk}(t)\delta t + \Delta(\phi_k(t)Z_l(t)) \tag{5.5}$$

We call the vector $\psi_k(t)V_k(t)$ a generic *network flow* of an OD pair o_k which carries its routing information. Solving the balance equations (5.2) and (5.5), we can evaluate the traffic loads over network links.

To summarize, a mobile ATM network is modeled by a flow network model with a fixed network topology G, an access network topology (mobility pattern) $G_w(t) = (G, D(t))$, network traffic flows $\{\psi_k(t)\}$, input traffic flow $\{\Phi(t)\}$, traffic service rate $\{\mu_k(t)\}$, routing functions $\{V_k(t)\}$, and path rerouting transformations $\{H_{lk}(t)\}$. The traffic distribution over the network depends on all factors; however, our analysis will focus on the effects of the user mobility ($D(t)$) and the handoff path rerouting ($\{H_{lk}(t)\}$).

5.3 Analysis on traffic distribution

We have given the flow balance equations at two levels. One is at OD pair level and the other is at the link level. In this section, the user mobility effects on traffic flows on both levels are analyzed.

5.3.1 Traffic flows and user mobility

A. Mean values and temporal variation:

Suppose the traffic demands $\{\phi_k(t)\}$ are differentiable functions and let $d\Phi(t)/dt = \Omega(t)$, where $\Omega(t) = [\omega_1(t), \ldots, \omega_K(t)]^T$ represents the rates of traffic demand variation. If a steady-state can be reached, the mean values of traffic flows can be solved as

$$m_\Psi(t) = -A^{-1}(t) m_\Omega(t) \tag{5.6}$$

We still denote the mean value as the function of time because of hour by hour variations of mobility pattern $A(t)$ can yield different steady-state operating points.

The solution above gives the properties of traffic flows as follows.

- The total traffic load, represented by $\sum_k m_{\psi_k}(t)$, is conserved regardless of user mobility.
- For a balanced OD pair o_l satisfying $\sum_k p_{lk}(t)\mu_l(t)\psi_l(t) = \sum_l p_{kl}(t)\mu_k(t)\psi_k(t)$, the traffic load is independent of user mobility. A balanced OD pair has the arrival rate of handoff flow equal to the departure rate of handoff flow. Equivalently,

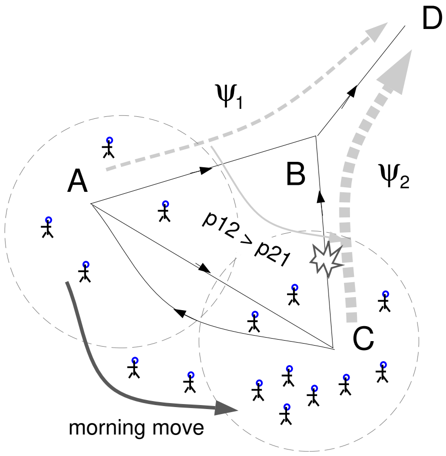

Figure 5.4: Example of temporal variations

a balanced OD pair has the flows from new calls balanced with the flows from terminated calls, i.e. $\omega_k(t) = (1 - \sum_l p_{kl}(t))\mu_k(t)\psi_k(t)$.

- The traffic load on an unbalanced OD pair depends on user mobility. As user mobility increases, the traffic load may increase, if more flows handoff in than flows handoff out, or decrease, if more flows handoff out than flows handoff in.

These properties are quite intuitive and so the proofs are not given. The third property tells us that at a given time period, the traffic requirement over an OD pair may be larger or smaller because user mobility causes temporal variations on network traffic loads. Although the total traffic load is conserved at any given time, the network capacity requirements could be larger than those for non-mobile-generated traffic, since user mobility may cause traffic requirements to vary from time to time but the network topology and link capacity may not permit the corresponding changes.

Previous studies show that the dynamic routing schemes can help the network adapt to traffic load variations. Similarly, we will see the temporal variations can be relieved by dynamic path rerouting algorithms for handoff control.

Consider a simple example from Figure 5.4. Suppose A is in a residential area while C is in a business area. D is attached with a newspaper server. Mobile users are reading newspapers while they are commuting from home to work or from work to home. In the morning ($t = t_m$), more users move from A to C than from C to A.

In the afternoon ($t = t_a$), more users move from C to A than from A to C. They are reflected by $p_{12}(t_m) > p_{21}(t_m)$ and $p_{12}(t_a) < p_{21}(t_a)$, respectively. Suppose, the new requests from A and C are symmetric and constant, i.e. $m_{\omega_1}(t) = m_{\omega_2}(t) = m_0$. The mean value of flows on OD pairs are

$$\begin{aligned} m_1(t) &= \frac{(1+p_{21}(t))(1-p_{12}(t))}{(1-p_{12}(t)p_{21}(t))}Tm_0 \\ m_2(t) &= \frac{(1-p_{21}(t))(1+p_{12}(t))}{1-p_{12}(t)p_{21}(t)}Tm_0 \end{aligned} \tag{5.7}$$

where T is the average service time for traffic flow. When user mobility is symmetric, i.e. $p_{12}(t) = p_{21}(t)$, the traffic flows on both OD pairs are balanced, then the traffic loads are independent of user mobility, $m_1(t) = m_2(t) = Tm_0$. Otherwise, the mean values of flows $\psi_1(t)$ and $\psi_2(t)$ depend on the user mobility, as reflected by the handoff probability p_{ij}. For example, if in the morning, $p_{21}(t_m) = 0.1, p_{12}(t_m) = 0.9$, then $m_2(t_m) = 1.8Tm_0$ and $m_1(t_m) = 0.2Tm_0$. Suppose traffic flow $\psi_1(t)$ always takes the route ABD, it is necessary to increase the capacity of AB to meet the traffic with 80% higher mean value. If in the afternoon, the user mobility is in the opposite direction, it requires the increment of the capacity of CB. As a result, to support user mobility, both AB and CB need more capacity!

B. Covariance and the smoothing effect:

In a small network with few subscribers, the temporal variations caused by user mobility may be significant. However, if the number of subscribers is large, even in a short time period the traffic flow over an OD pair may be balanced by the movement of mobile users. However, even if the traffic flows are balanced, the variations in the traffic flows can still be observed through the second moments – covariances.

If the rate of traffic demand $\Omega(t)$ is a white noise, or equivalently, the traffic demand $\Phi(t)$ is a Brownian motion process, the covariance of the traffic flows satisfies

$$A(t)R_\Psi(t) + R_\Psi(t)A^T(t) + \sigma_\Omega(t)\sigma_\Omega(t)^T = 0 \tag{5.8}$$

The proof see Appendix B. The covariances of the traffic flows always depend on the user mobility regardless of the balance of traffic flows. However, when traffic flows

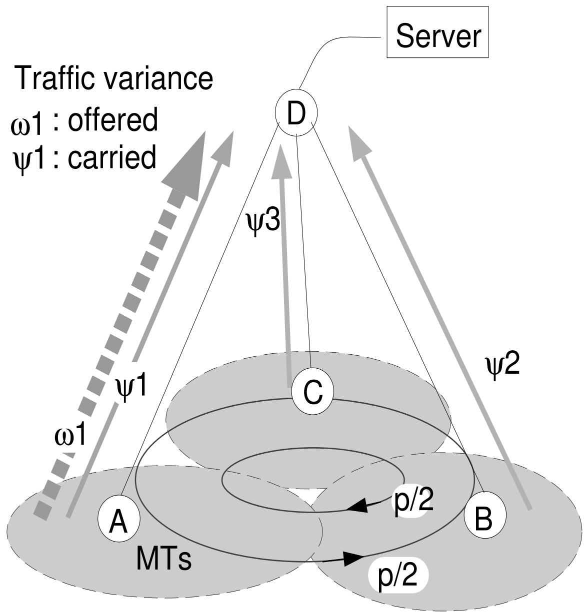

Figure 5.5: Example for the smoothing effect

are balanced, a *smoothing effect* can be observed, in that the variances of the traffic flows decrease as the user mobility increases.

Without loss of generality, we show this smoothing effect through an example in Figure 5.5. The rate of traffic demands are iid processes with mean m_ω and variance σ_ω, and user mobility is symmetric among three OD pairs with all $p_{kl} = p/2$. The traffic flows ψ_1, ψ_2, ψ_3 will be same on all three OD pairs because of the symmetry and will have mean and variance

$$m_\psi = \frac{1}{\mu(1-p)} m_\omega = T m_\omega \tag{5.9}$$

$$\sigma_\psi^2 = R_{kk} = -\frac{\sigma_\omega^2}{2\mu(-1+\frac{p^2}{2-p})} = T\sigma_\omega^2 \frac{2-p}{2(2+p)} \tag{5.10}$$

The correlation coefficient between traffic flows is

$$\rho = R_{kl}/R_{kk} = p/(2-p) \tag{5.11}$$

As the user mobility increases with increasing p, the variances of traffic flows ψ_k decrease compared with the traffic demands $T\omega_k$, i.e. the traffic flows are being smoothed.

In Figure 5.6, the variance vs.user mobility is shown, where user mobility $L = 1/(1-p)$ is the average number of cells a mobile user traverses during an average call

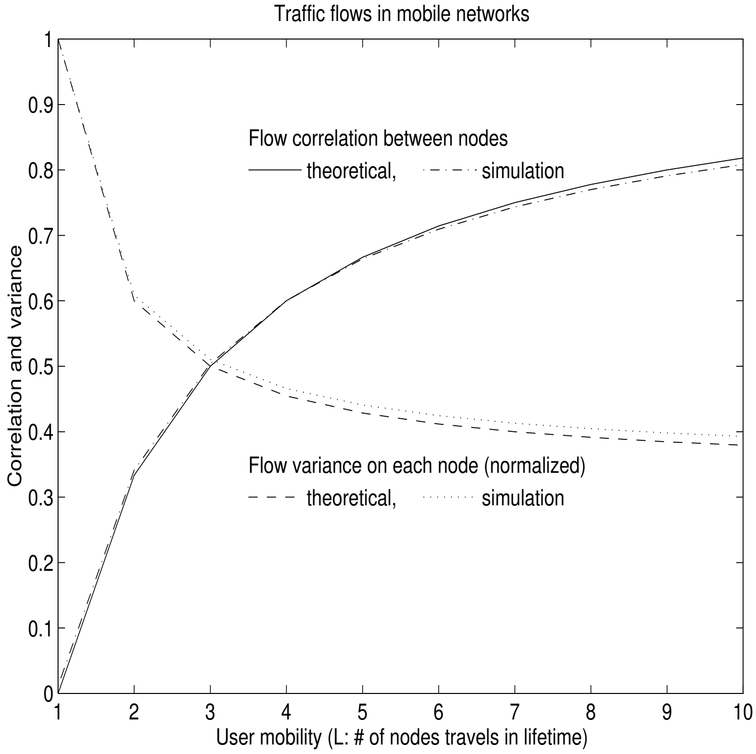

Figure 5.6: Variance and correlation of traffic flows

lifetime. In addition to the variance, the correlation between traffic flows is shown. As the user mobility increases, the traffic flow correlation increases also. The simulation results (dashed lines) come from the model's simulation based on Equation (5.1), which verifies the correctness of Equation (5.8).

Since the flow variances decrease as the user mobility increases, the traffic overflow (data loss rate) will decrease on those links with limited capacity. In Figure 5.7, we show the loss rate vs. user mobility by assuming traffic flows obey a Gaussian distribution. In general, the loss rate is determined by the tail distributions (or effective bandwidth) of the traffic flows. It is expected that the effective bandwidth of the traffic flows will decrease if the variances of the flows decrease. Thus the smoothing effect of user mobility should reduce the loss rate. The simulation result (dashed line) measure the average traffic overflow over a link with limited capacity.

From the analysis of the means and covariances of the traffic flows, we observed that user mobility causes temporal variations in the network traffic loads and has the smoothing effect on network traffic distribution. The results tell us that if in any period

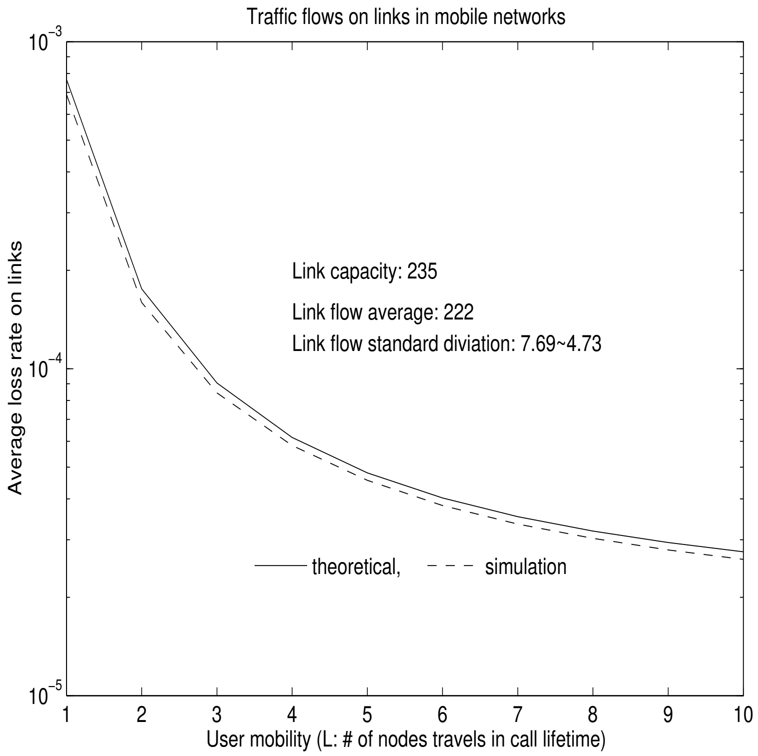

Figure 5.7: Data loss rate reduced by smoothing effect

(short or long), the users' movement among cells is balanced, the traffic distribution over OD pairs is smoothed due to user mobility. In other words, the network service quality may be improved due to the user mobility. However, if the users' movements are not balanced among cells in any period, the traffic load on some OD pairs may increase while the others may decrease. As the result, the network service quality may be degraded due to the user mobility.

The smoothing effect seems contradictory to the analysis result of using traditional queuing model, in which user mobility always degrades system performance. The reason is that the higher is the user mobility, the higher the handoff frequency and the higher of call drop probability. In traditional queuing model, which is used for real-time service like telephony, the call drop rate is an important performance criterion. However, in our model, no call level QoS is specified, that is why we can see improvement of system performance in terms of data loss rate. It is reasonable to assume a service scenario that it so-called session-oriented service. In this case, a call is a session which has certain

QoS guarantees when it is active, but a session can go into a suspended mode if it fails to be admitted either initially or during handoff. A suspended session can become active when it is admitted later when resources are available. For real-time services, this is an almost unacceptable scenario, but for mobile computing, this can be tolerated by many applications. If the majority of applications are these kind of session-oriented services, the smoothing effect can help to improve network performance.

5.3.2 Handoff flows and path rerouting algorithms

When a mobile user has a handover from one wireless access point to another wireless access point, all traffic flows associated with the user must migrate to routes linked to the new access point.

Equation (5.5) can be solved when routing functions $\{V_l(t)\}$ are independent of traffic flows $\{\psi_k(t)\}$, meaning the path rerouting transform matrices $\{H_{lk}(t)\}$ are deterministic and given by the path rerouting algorithm. In other words, alternative rerouting which dynamically adapts traffic distributions is not used. In this case, we can apply statistical averaging to Equation (5.5), which yields, for $k = 1, ..., K$

$$\mu_k(t) m_{\psi_k}(t) V_k(t) = \sum_{l}^{K} p_{lk}(t) \mu_l(t) m_{\psi_l}(t) W_{lk}(t) + m_{\omega_k}(t) Z_k(t) \tag{5.12}$$

Substituting solution of Equation (5.6) into Equation 5.12, we can have the solution of $V_l(t)$.

With the solution of $\{V_l(t)\}$, we can obtain the traffic flow on each network link u,

$$y_u(t) = \sum_{k}^{K} v_{k,u}(t) \psi_k(t) \qquad u = 1, \ldots, M \tag{5.13}$$

Both means and covariances of $\{y_u(t)\}$ can be obtained based on the solutions for $\{\psi_k(t)\}$.

Let us look at the example in Figure 5.3 again, when we have different path rerouting algorithms. A call in flow ψ_1 at A can rerouted to the flow ψ_2 at either B or C. If the anchor point is chosen as B, the link CB will take the handoff traffic flow from ψ_1. As we pointed out, the handoff traffic may not be balanced, for example, in the morning, there are more handoffs into C than handoffs out of C. The link capacity of CB must be increased to maintain the same QoS. However, it is possible to reroute the handoff traffic at other anchor nodes. For example, reroute the handoff traffic at A, i.e. the handoff flow from ψ_1 can take route $CABD$ instead of CBD. In this case, the link capacity of CA (or AC) needs to be increased to cover the handoff traffic flow while the link capacity of CB can remain the same. If the costs of links CA and AC are lower than the costs of links CB and AB, it is worthwhile to build CA (and/or AC) instead of increasing the capacity of AB and CB.

This motivates the use of alternative path rerouting algorithm for handoff traffic control. A path rerouting algorithm for handoff control specifies where and how a connection path is rerouted. We have given three typical path rerouting algorithms in section II, as illustrated by Figure 5.1. Theoretically, path rerouting algorithm can be defined by the transform matrix $H_{lk}(t)$ with following features for each algorithm.

- *Last Divergence (LD) Algorithm:* Using the LD algorithm, statistically, all traffic flows for OD pair o_k will take the same set of routes $V_k(t)$ regardless if they are from new calls or handoff calls, since handoff calls also take the optimal route. This requires the path rerouting transform matrix to satisfy $H_{lk}(t)V_l(t) = V_k(t)$.

- *Path Extension (PE):* In the PE algorithm, the path rerouting transform matrix satisfies $W_{lk} = H_{lk}(t)V_l(t) = (I+\Delta_{lk}(t))V_l(t)$, where $\Delta_{lk}(t)$ represents the handoff flows transfered to the links on a extended path. For a mobile ATM network, if the extended paths are long, the performance of PE algorithm will be low. The $\Delta_{lk}(t)$ is always positive which makes the connection path longer and longer after each handoff.

- *First Divergence (FD):* the FD algorithm uses the same extended route as in PE for every handoff, however, the duplicated part is removed. Suppose the path rerouting algorithm satisfies $W_{lk}(t) = H_{lk}(t)V_l(t) = (I + \Delta_{lk}(t))V_l(t)$, where $\Delta_{lk}(t)$ represents the handoff flows transfer away from the links on the old path, and to the links on the new path.

We will outline an analysis procedure to evaluate the performance of these three typical path rerouting algorithms through a simple example. Although the example

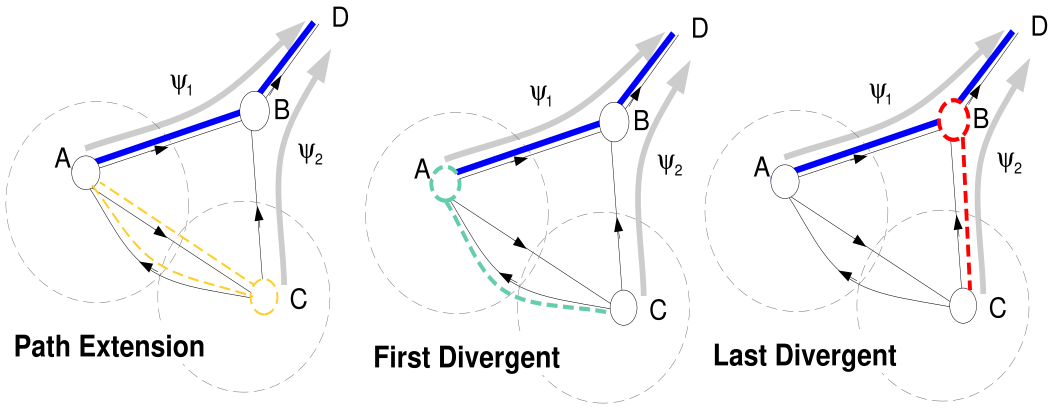

Figure 5.8: Performance analysis of path rerouting algorithms

is simple, as shown in Figure 5.8, the methodology can be used in general situations. In Figure 5.8, for the LD algorithm, the COS is always at B, so the handoff traffic flow from ψ_1 takes only the optimal route CBD after handoff. For the PE algorithm, the handoff traffic flow from ψ_1 will take an extended path CA, resulting in the route becoming $CABD$. In case a call is a handoff from A to C and then back from C to A, the handoff flow of the call will take an overall extended path ACA which forms a loop. The route for the call becomes $ACABD$. For the FD algorithm, the COS is at either A or C and the handoff traffic flow from ψ_1 takes the extended path CA which makes the route become $CABD$. The FD algorithm takes the same extended path as the PE algorithm, however, it does not form a flow loop.

To identify the traffic load induced by path rerouting, we must

1. find the initial routing functions,

2. find the path rerouting matrix,

3. solve the balance equation on network links to obtain the routing functions and

4. evaluate the incremental traffic load.

First, we determine the initial routing functions, i.e. the routing functions without user mobility. In this example with five network links AB, AC, CB, CA, BD, they are

$$Z_1 = [1,0,0,0,1]^T \quad Z_2 = [0,0,1,0,1]^T \tag{5.14}$$

Second, we find the path rerouting transform matrices H_{12} and H_{21}. These will be different for three path rerouting algorithms.

1. *LD algorithm:* for the LD algorithm, by definition, the handoff flow in ψ_2 will take the route for flow ψ_1, i.e. $H_{21}V_2 = V_1$. Similarly, we have $H_{12}V_1 = V_2$. The path rerouting transform matrix can be represented as

$$H_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\Delta_{lk} = H_{lk} - I$, represents the change in routing function due to handoff path rerouting. For example, $\delta_{lk,ij} = 1$ means the handoff flow from flow ψ_l on link i will be added to flow ψ_k onto link j.

The routing function $V_l(t)$ can be obtained by solving both balance equations for OD pairs and network links, Equation (5.2) and (5.5). In this case, we have $V_l = Z_l$, meaning the routing function does not change due to handoff.

In the example, the mean and variance of traffic flows on network links are

$$\begin{aligned} m_Y &= [m_1, 0, m_2, 0, m_1 + m_2] \\ \sigma_Y^2 &= [\sigma_{\psi_1}^2, 0, \sigma_{\psi_1}^2, 0, \sigma_{\omega_1}^2 + \sigma_{\omega_2}^2] \end{aligned}$$

Combined with the statistics we previously obtained for traffic flows, we can analyze the traffic distributions on network links. For example, if the mobility pattern is balanced, the only thing affected by the user mobility is the variances of the flows on AB and CB and they decrease as user mobility increases. If the mobility pattern is unbalanced, it is possible that $m_1 > m_2$, which requires AB to have higher capacity when user mobility exists.

2. *PE algorithm:* For the PE algorithm, the COS is always at the OldBS and every handoff extends the previous connection path by one more hop. This implies

$$H_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Solving (5.5) based on the solution of Equation (5.6), we find that the routing functions for the PE algorithm are

$$[V_1 V_2] = \begin{pmatrix} \frac{1}{1+p_{21}} & \frac{p_{12}}{1+p_{12}} & : AB \\ \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & \frac{p_{12}(1+p_{21})}{(1+p_{12})(1-p_{12}p_{21})} & : AC \\ \frac{p_{21}}{1+p_{21}} & \frac{1}{1+p_{12}} & : CB \\ \frac{p_{21}(1+p_{12})}{(1+p_{21})(1-p_{12}p_{21})} & \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & : CA \\ 1 & 1 & : BD \end{pmatrix}$$

The routing functions depend on user mobility and they show how much of a given flow uses a specific network link. We can get the means and variances of link traffic flows $\{y_u(t)\}$ based on the solution of Equation (5.2), for traffic flows on OD pairs. In our example, the mean traffic flow on AB is $m_{AB} = v_{1,1}m_1 + v_{2,1}m_2$. By substituting m_1 and m_2 from Equation (5.8), we have $m_{AB} = Tm_0$, where T is the average call lifetime. It is independent of user mobility. While the mean traffic flow on AC is

$$m_{AC} = v_{1,2}m_1 + v_{2,2}m_2 = Tm_0 \frac{p_{12}(1+p_{21})}{(1-p_{12}p_{21})} \tag{5.15}$$

which will increase greatly as user mobility increases. The cost of the PE algorithm is the traffic flows on links AC and CA. To reduce the traffic flows on extended paths, we may use the FD algorithm which removes possible duplication.

3. *FD algorithm:* For the FD algorithm, the COS is at the OldBS (or the NewBS) and a handoff may increase or decrease the path length by one. Thus,

$$H_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Given H_{12} and H_{21}, the solution of (5.5) and (5.6) yields

$$[V_1 V_2] = \begin{pmatrix} 1/(1+p_{21}) & p_{12}/(1+p_{12}) & :AB \\ p_{21}/(1+p_{21}) & 0 & :AC \\ p_{21}/(1+p_{21}) & 1/(1+p_{12}) & :CB \\ 0 & p_{12}/(1+p_{12}) & :CA \\ 1 & 1 & :BD \end{pmatrix}$$

The routing functions depend on user mobility. However, unlike the PE algorithm, the traffic load on links AC and CA is not high since there is no duplicate usage of the links. The mean of traffic flow, $m_{AC} = Tm_0(v_{1,2}m_1 + v_{2,2}m_2)Tm_0p_{21}(1 - P_{12})/(1 - p_{12}p_{21})$, won't grow as fast as the traffic load for the PE algorithm. On the other hand, the total traffic loads on links AB and CB are same as that of the PE algorithm, which are independent of user mobility. So the FD algorithm gets the advantage of the PE algorithm with a much lower cost.

5.3.3 Performance evaluation

The solution of the routing functions $V(t)$ can help us re-design the network topology or, if the network topology is fixed, evaluate which path rerouting algorithm should be chosen. As we have done in the example, we can analyze the mean and variance of the traffic flows on links. However, we may want to have a general idea how good or bad is a path rerouting algorithm in a given network. One criterion is the sum of traffic flows on all links, or *network traffic volume*, represented

$$|Y(t)| = \sum_{u}^{M} y_u(t) \tag{5.16}$$

A good path rerouting algorithm should have little increment in network traffic volume due to handoff control. In the example of Figure 5.8, when handoff traffic flows are balanced, that is $p_{12} = p_{21} = p$, the network traffic volumes are $2(\psi_1 + \psi_2)$, $(2 + p/(1 - p))(\psi_1 + \psi_2)$ and $(2 + p/(1 + p))(\psi_1 + \psi_2)$, for LD, PE and FD algorithms, respectively.

More generally, without the knowledge of traffic requirements $\psi_k(t)$, we may use link costs to evaluate a network topology. In particular, given g_u, the weight representing the cost of link u, we define the *route lengths* of flow l over OD pair o_l as

$$|V_l(t)|_g = \sum_{u}^{M} g_u v_{l,u}(t) \tag{5.17}$$

In Figure 5.8, the cost of link AB may be greater than the cost of link AC. In our example, assuming $g_u = 1$ for all u, the route lengths of both OD pairs are 2, $2+p/(1-p)$ and $2 + p/(1 + p)$, for LD, PE and FD algorithms, respectively. Figure 5.9 shows the route length versus the user mobility, as measured by $L = 1/(1-p)$ which is the average number of cells a mobile users visits during a call lifetime. It is obvious that the PE path rerouting has the route length linearly increased as user mobility increases. Using FD algorithm, only a slightly increment in route length due to user mobility. While using the LD algorithm, the route length is supposed to be a constant relative to the user mobility.

When link costs g_u are not uniform, for example $g_{AB} = g_{CB} = 5, g_{AC} = g_{CA} = 1, g_{BD} = 1$, the lengths of traffic flows for LD, PE and FD algorithms will be 6, $6+p/1-p$ and $6+p/1+p$. The relative increasing rate of the lengths for the PE and FD algorithms will be smaller compared with the original length of traffic flows.

In the case of unbalanced traffic, the performance cannot be evaluated only by the route length of OD pairs. The traffic flows ψ_1 and ψ_2 may reach their peak values at different times. If the new path which takes handoff traffic flow is long, the extra traffic load caused by user mobility will be large. For example, if the peak value of m_1 is $1.8Tm_0$, the extra cost of using LD algorithm is $0.8g_{AB} = 4$ while only $0.8g_{AC} = 0.8$ for

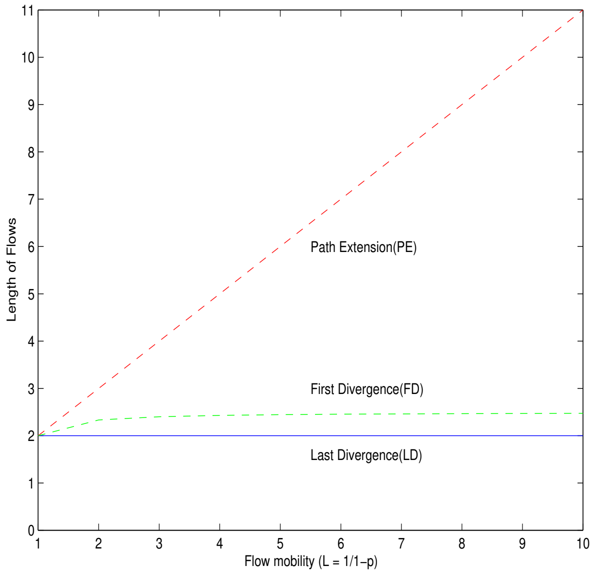

Figure 5.9: The route lengths

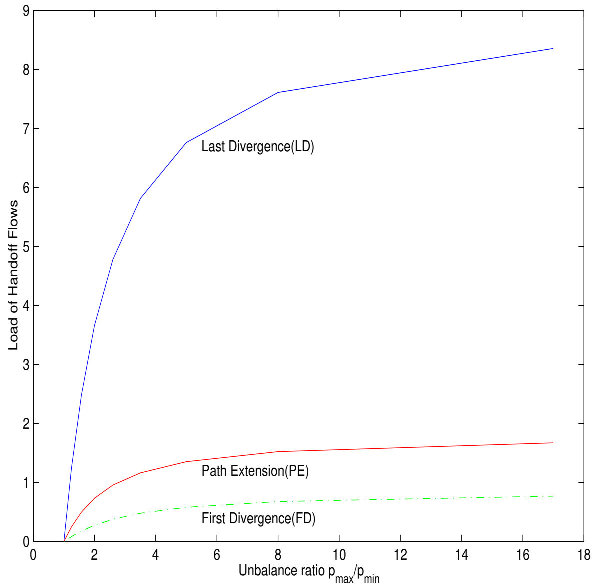

Figure 5.10: The new path lengths

Analytical analysis on example in Figure 5.8

FD algorithm. Thus in the case of unbalanced traffic, we can evaluate the performance of path rerouting algorithms using the route length of handoff flow

$$|U_l(t)^+|_g = |(\sum_{k=0}^{K} \Delta_{lk}(t)V_l(t))^+|_g \tag{5.18}$$

where $+$ means only the positive elements are considered. In our example, suppose at time t_m, $p_{12} = p_{max}$ and $p_{21} = p_{min}$, at time t_a, $p_{12} = p_{min}$ and $p_{21} = p_{max}$. And $g_{AB} = g_{CB} = 5, g_{AC} = g_{CA} = 1, g_{BD} = 1$. We also call the route length of handoff flow as handoff new path length, which is shown in Figure 5.10 as a function of the imbalance ratio p_{max}/p_{min}. The route length of handoff flow can also reflect the delay of handoff path rerouting. That is the longer the new path, the greater the delay.

For a given path rerouting algorithm, it is desirable to have both the route length and new path length minimized in order to reduce the extra network resource demand for supporting handoff control. Through simple examples, we have obtained analytical results on the performance of typical path rerouting algorithms. They are reflected by the route lengths of OD pairs and the length of handoff new paths. We have observed that the PE algorithm is not a good choice if the user mobility is high. And the LD algorithm may not be a good choice if traffic is unbalanced. For both concerns, it

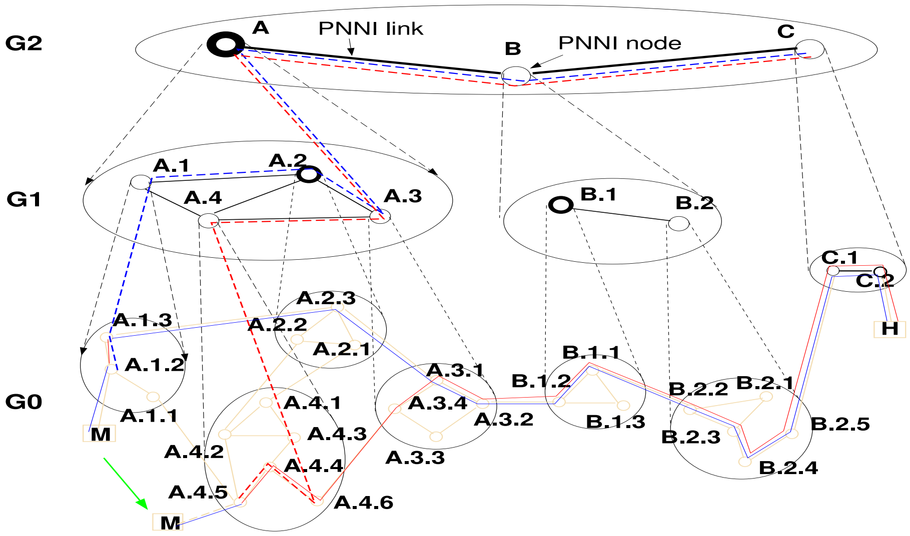

Figure 5.11: A sample PNNI network

seems the FD algorithm is a good choice since it has the smallest handoff new path route length and only a slightly greater overall route lengths in OD pairs. We now verify the analytical observations through simulations of a larger network configuration.

5.4 Experiments through Simulations

In the previous chapter, we have seen that the route lengths of OD pairs and the route lengths of handoff new paths can measure the performance of path rerouting algorithms. The route lengths of OD pairs reflect overall traffic load in the network and the route lengths of handoff new paths reflect the extra traffic load when handoff traffic flows are unbalanced.

The method we are going to use is to have several calls active in the network. For each call, one party is mobile terminal in the mobile ATM subnetwork and the other party is the fixed host sitting outside of the mobile ATM subnetwork. Then we will let the mobile terminals handoff to another cell according to certain path rerouting algorithm. The length of the whole connection path for each call is measured after each handoff, in order to estimate the route lengths of OD pairs. And length of the new path of each call is measured, in order to estimate the route lengths of handoff new paths.

The network used in the simulation is as shown in Figure 5.11. It is an example in

ATM Forum PNNI specifications [55] for PNNI hierarchical structure. In this network, there are 27 switches and three hierarchical levels. We use peer group C as a fixed backbone network and peer group A and B as the mobile ATM access network. Every switch in group A and B is considered a basestation. Mobile terminals can be attached through any switch in A, B and move (handoff) from one to another.

5.4.1 Simulation results with fixed mobility pattern

In the first simulation, we assume the user mobility pattern matches the network topology, i.e. any two nodes within A, B are neighboring access points (having an $M - edge$) if there is a link ($C - edge$) between them. We call this setup as fixed mobility pattern.

The simulation is conducted as follows. Three mobile terminals starting calls at $A.1.2$, $A.3.4$ and $B.2.3$ to a remote server at $C.2$ in the fixed network. Each call starts with an exponential distributed lifetime represented by t_f with $E[t_f] = L$ seconds and performs a handoff for every second. Hence the average number of handoff for a call lifetime is L, which is also representing the call or flow mobility as $L = 1/1 - p$. At the access point where a call is terminated, a new call is started. The procedure continues until each mobile terminal performs 10,000 handoffs. A mobile user will randomly handoff to one of the neighboring cells with equal probability.

In the simulation, we use three path rerouting algorithms, Path Extension (PE), First-Cross-to-CT (FXCT, as an approximation of LD algorithm, which uses first crossover switch toward the CT as the COS) and First-Cross-to-OldBS (FXBS, as an approximation of FD algorithm, which uses first crossover switch toward the OldBS as the COS).

Figure 5.4.1 shows the route lengths for the three path rerouting algorithms. It is obvious that PE has linear increasing in route length. Higher user mobility implies more handoffs per call lifetime. It is not surprising that FXBS has only a little extra path length over the FXCT algorithm and is almost constant as mobility increases. This is consistent with our analytical example in Figure 5.9. Using the FXBS algorithm, the network needs about 10% more capacity than using FXCT algorithm, however, the capacity increment does not change much as the user mobility increases.

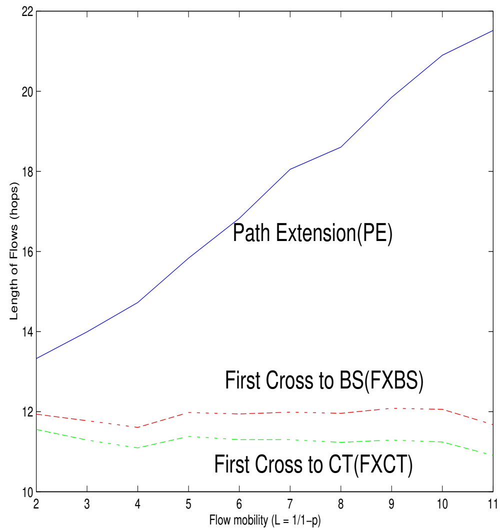

Figure 5.12: Average OD pair length

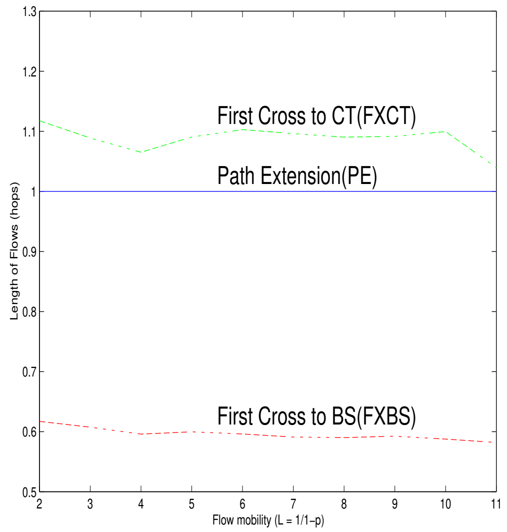

Figure 5.13: Average new path length

Figure 5.13 shows the route lengths of handoff new paths for the three algorithms. They do not change much with the user mobility. However, they will cause the handoff traffic flow to increase at different rates when imbalance ratio increases. The FXCT algorithm has a new path length about 0.5 hops more than that of the FXBS algorithm. When handoff flows are not balanced, the capacity increment on the new path for the FXCT algorithm can be much larger than for the FXBS algorithm. Since the new path takes the handoff flows, if handoff traffic flows double the original traffic load on the new paths, the FXCT algorithm will need one more hop $(2 * 0.5)$ on average than the FXBS algorithm, which is also about 10% of total average path length. So depending upon the mobility pattern, one algorithm may consume less of the network resources.

5.4.2 Simulation results with random mobility pattern

In the second simulation, the neighborhood of the access network is randomly generated among the nodes in peer groups A, B. For each node x, a node y is a neighbor of x with a probability of $16/N$, $4/N$ and $1/N$, if y and x belong to same peer group at $G0$, $G1$ and $G2$ level, respectively, where N is the total number of nodes. Based on this selection, we have each node having about 4 neighbors. If a node x is closer to y in terms of PNNI hierarchy, the probability that x is a neighbor of y is higher. The

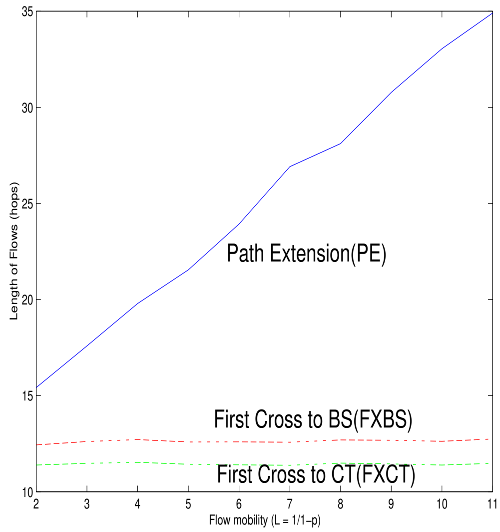

Figure 5.14: Average whole path length

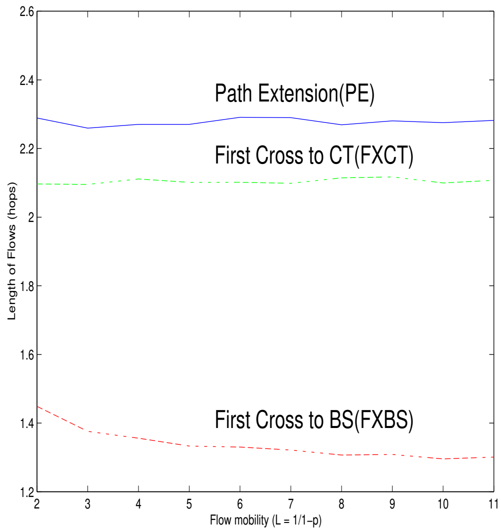

Figure 5.15: Average new path length

probability a handoff is to which neighbor of a node is again same as in last example. That is 1/number of neighbors. We call this neighborhood setup as random mobility pattern.

From Figure 5.14, we can see the route lengths for FXCT and FXBS are about the same as the first example. The FXBS needs about 10% more hops than FXCT algorithm. However, in Figure 5.15, we observe the difference between the route length for handoff new paths for FXCT and FXBS is bigger about 0.7 hops. Therefore when handoff flows are imbalance, the handoff flows may result in greater capacity requirements for the FXCT algorithm than the FXBS algorithm.

Again in this experiment, we can see the PE algorithm is costly. Since the neighborhood is random, not only the whole path length is linearly incrementing with user mobility, but also the average new path length is the highest within three algorithms. For FXCT algorithm, the whole path length is the shortest but the new path is larger. The overall performance of the FXBS algorithm is good since it keeps both the whole path length and the new path length relatively short, which implies that the extra network resources required for supporting handoff is small and the handoff latency is small among three algorithms.

Chapter 6

Implementation and Experiments

In this chapter, the implementation of handoff control protocols in mobile ATM networks is described in the context of the system integration of wireless access, interworking functions and ATM signaling extensions. We have built a prototype system with waveLAN radio access and a pre-product trial system with wireless ATM access. We will present the basic protocol stacks and software components and describe configurations of those systems. Some experimental results, such as handoff delays and data loss during handoff are given.

6.1 Software architecture

6.1.1 Protocol stacks

Mobile ATM protocol stacks on mobile terminals, basestations and ATM switches are shown in Figure 6.1. 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 radio specific data format and ATM cell format.

We can see the complexity of the implementation is in the basestation which is responsible for the interfaces between wireless and wired networks for both signaling messages and data.

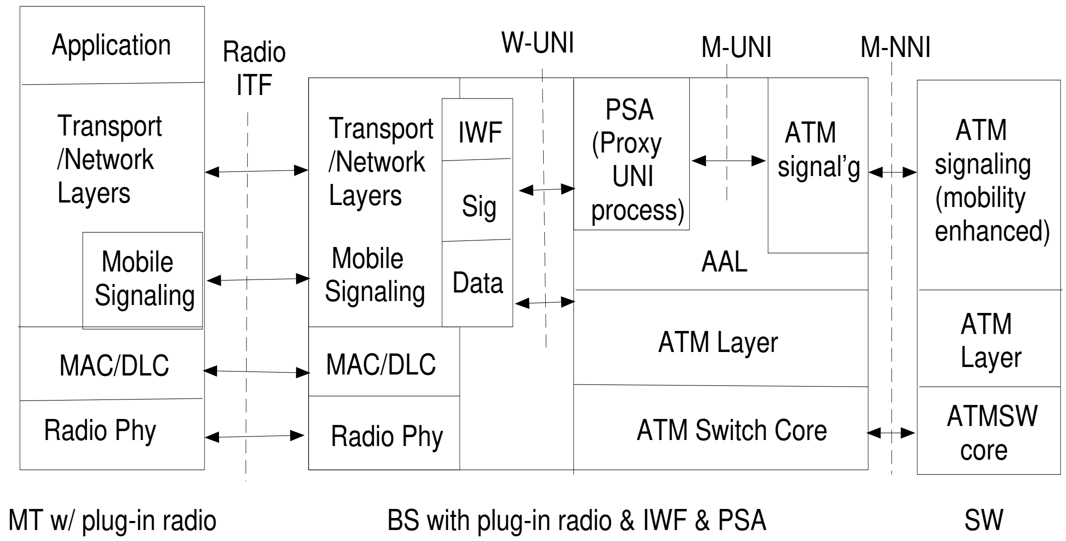

Figure 6.1: Protocol stack of mobile ATM networks

6.1.2 Software components

Based on the protocol stack given in Figure 6.1, we have defined the following software components for mobile terminals, basestations, including virtual WATM host and universal basestation components, and ATM switches, as shown in Figure 6.2.

On a mobile terminal, proprietary equipment and software are installed for a specific radio access. This must include a mobile signaling module and a wireless control module. The wireless control will communicate with mobile signaling module when the MT powers on or detects the need for a handoff.

On a basestation, similar proprietary radio equipment and software are installed. However, the MAC layer protocol on a basestation will schedule data transmissions from/to multiple mobile terminals at the same time. As we have discussed as shown in Figure 6.1, IWF, PSA and M-UNI/M-NNI signaling and PNNI routing modules are required. The handoff control module will especially handle the handoff related messages which need to be sent through the handoff control channel between the oldBS and the newBS.

On an ATM switch, the mobility enhanced M-NNI signaling module is installed.

The software operation sequence can be roughly described as follows.

1. The MT powers on. A wireless control protocol starts. Through wireless control,

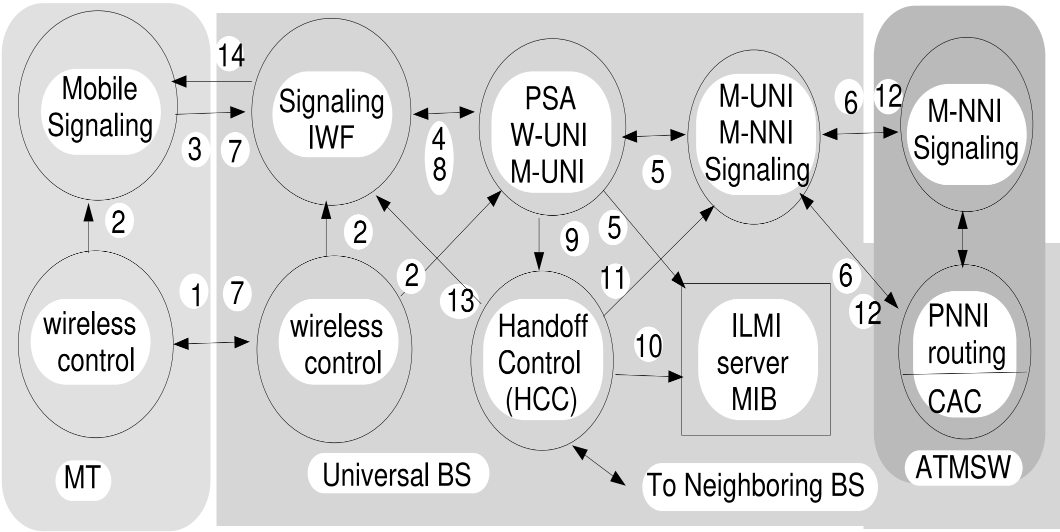

Figure 6.2: Software components of mobile ATM networks

the MT got registered.

2. The signaling module on the MT starts. The signaling IWF and PSA start on the BS start.
3. Call establishment starts at mobile signaling.
4. Call establishment is passed by the IWF.
5. Call establishment starts at the M-UNI interface.
6. Call establishment starts to the M-NNI interface.
7. Link fading is detected by wireless control and a handoff request is issued by mobile signaling on the MT.
8. The handoff request is passed by the IWF.
9. The handoff request is handled by handoff control. The MT is registered to the neighboring basestation – the newBS.
10. At the newBS, the resource availability is checked at the MIB server.
11. The handoff path rerouting starts at the newBS.

12. The handoff path rerouting starts to the network through the ATM M-UNI/M-NNI signaling and PNNI routing.

13. We omit the corresponding procedure for the handoff path rerouting. After it is done for all the VCs. The MT is ready to move to the newBS.

6.2 Prototype implementation

In order to prove the concept of mobile ATM, we have built a prototype system which uses waveLAN radio access. The data format at radio link is IP packet and the data is converted to ATM cells through IP over ATM.

6.2.1 System configuration

Our prototype implementation consists of a network with three ATM switches (2.4 Gbps NEC Model 5S), running GSMP (general switch management protocol) [56]. Each switch is controlled through an external 200Mhz 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 using 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 the prototype system shown in Figure 6.3, hardware and software configurations are listed in Table 6.1.

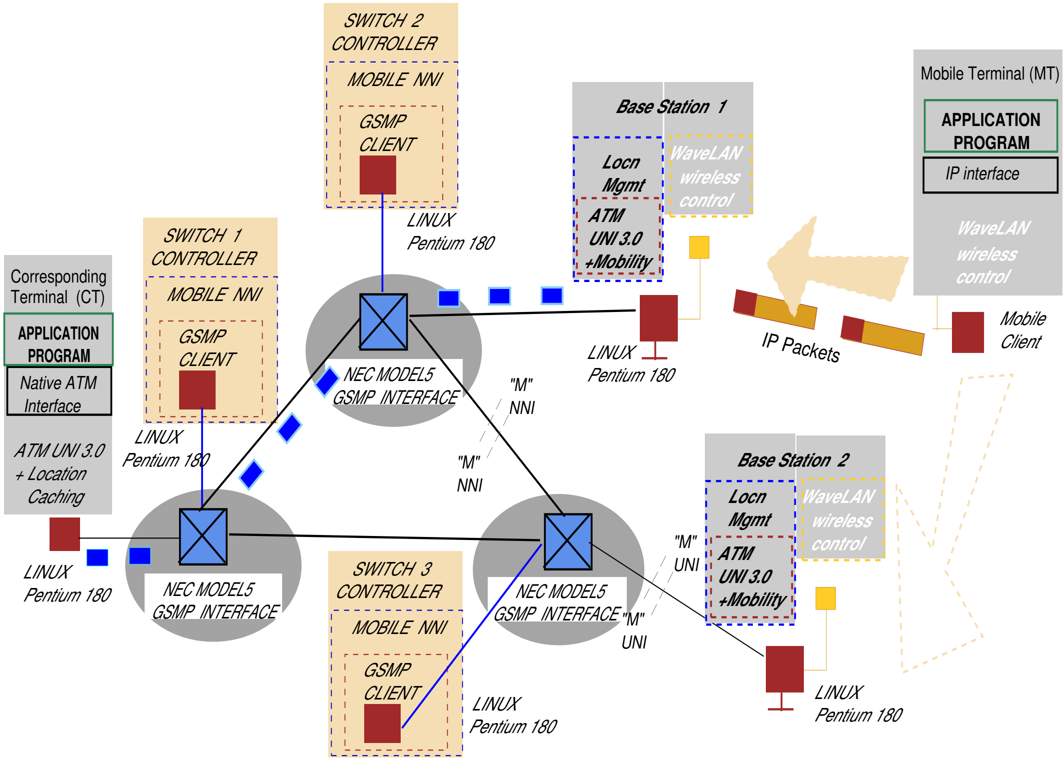

Figure 6.3: Mobile ATM network prototype

In essence, the prototype shows that mobility support in IP may be obtained by mapping IP to mobile ATM, which is a generic network, instead of using IP specific mobility support (mobile IP).

6.2.2 Mobility support functionalities

The IWF on the basestation converts IP packets to ATM cells. Since WaveLAN uses IP packets over IEEE 802.11, there is no call control signaling per se, the IWF on basestation is an IP over ATM function which works as follows. For each IP destination, the IWF will establish an ATM connection when first packet is received. The connection will be removed if no packet for the IP destination is transmitted for a timeout period. The IWF can obtain an ATM address for each IP address through an ATMARP (ATM address resolution protocol) table. It can be stored in the MIB server, too. The detailed control procedure is described as follows.

The mobile terminal is associated with two addresses that never change: a home IP address and a home ATM address. The binding between the two addresses is static

ATM switches	NEC Model 5, 2.4Gbps
Switch Controller	HP Desktop with 180MHz Pentium Pro
Basestations	HP Desktop with 180MHz Pentium Pro
Mobile Terminal	NEC Versa 6200MX with Pentium 133MHz
Fixed Host (CT)	HP Desktop with 180MHz Pentium Pro
ATM interface	Zeitnet 155Mbps ATM NIC card
Radio Access	NCR WaveLAN 2.0
OS on PCs	Linux 2.0.29 with ATM stack 0.26
ATM Signaling	UNI/NNI 3.1 with mobility extensions
Switch Control	GSMP server embedded in the switch
PNNI Routing	not used (static routing is applied)
Handoff Re-routing	First-Cross-To-CT algorithm

Table 6.1: Hardware and software configuration of the prototyping system

and location-independent: the IP address can be resolved to the corresponding ATM address using an address resolution protocol such as ATMARP[24]. In the prototype, basestation 1 is the "home switch" for the mobile terminal, and basestation 1 advertises reachability to the terminal's home ATM address. As the mobile terminal moves within the mobile ATM network, it acquires a second ATM (foreign) address, that reflects its current basestation. When the fixed terminal initiates a TCP connection to the mobile terminal's IP address, the IP address is first resolved to the home ATM address. Then, using the integrated location resolution scheme outlined in Chapter 2 , an ATM connection is set up to the mobile terminal's current basestation; thus, the location management is provided by mobile ATM and is completely transparent to IP (i.e. to the TCP connection).

In this prototype, handoff is initiated after the mobile terminal moves to the new basestation, that is the forward handoff. It supports first cross to the CT (FXCT) path rerouting algorithm. Consider an active connection to the mobile terminal from the fixed terminal, while it is located under the basestation 1. The ATM data path consists of the CT – switch 1 – switch2 – basestation 1. The MT then changes its default (IP) gateway on its wireless LAN interface, to point to basestation 2. When the first data packet (on the active connection) is received by basestation 2, it sends a setup with the ATM address of the CT. The SETUP message also includes a global connection-id

(GCID) which is based on the home ATM address assigned to the MT. This message is routed along the path basestation 2 – switch 3 – switch 1. At switch 1, the GCID within the SETUP matches that of an existing connection path; consequently, (a) switch 1 sends a Q.2931 connect() reply on the new sub-path, (b) sends a GSMP message to the switching fabric to change datapath, and (c) sends a release() down the existing sub-path towards basestation 1. As with location management, the re-routing of the ATM connection path is transparent to IP, and is handled within mobile ATM.

The prototype system shows the interworking between IP and mobile ATM. Although MT is an IP terminal, we use mobile ATM instead of mobile IP for mobility management.

6.3 Trial system implementation

The trial system is a pre-commercial wireless ATM network product. It has wireless ATM radio access points, embedded MAC and DLC layer protocols on PCMCIA size WATM NIC card and mobility enhanced ATM software for location management and handoff control.

6.3.1 System configuration

A demo configuration of the trial system is illustrated in Figure 6.4. It has two ATM switches, three basestations and five mobile terminals. ATM switches are controlled by outside controllers through the GSMP [56] protocol. On the switch controllers, the enhanced mobile ATM M-UNI signaling and M-PNNI signaling routing modules are running. Basestations are also configured as a switch node, which runs ATM M-UNI signaling and M-PNNI signaling/routing modules. Each basestation or mobile terminal has a PCMCIA size wireless ATM NIC card. The mobile terminal runs ATM M-UNI signaling and operates at one of three frequencies of three basestations.

The system operates in the 5.1GHz unlicensed band. A single carrier can have up-to 25 Mbps capacity shared by multiple mobile terminals through the scheduler of MAC layer protocol.

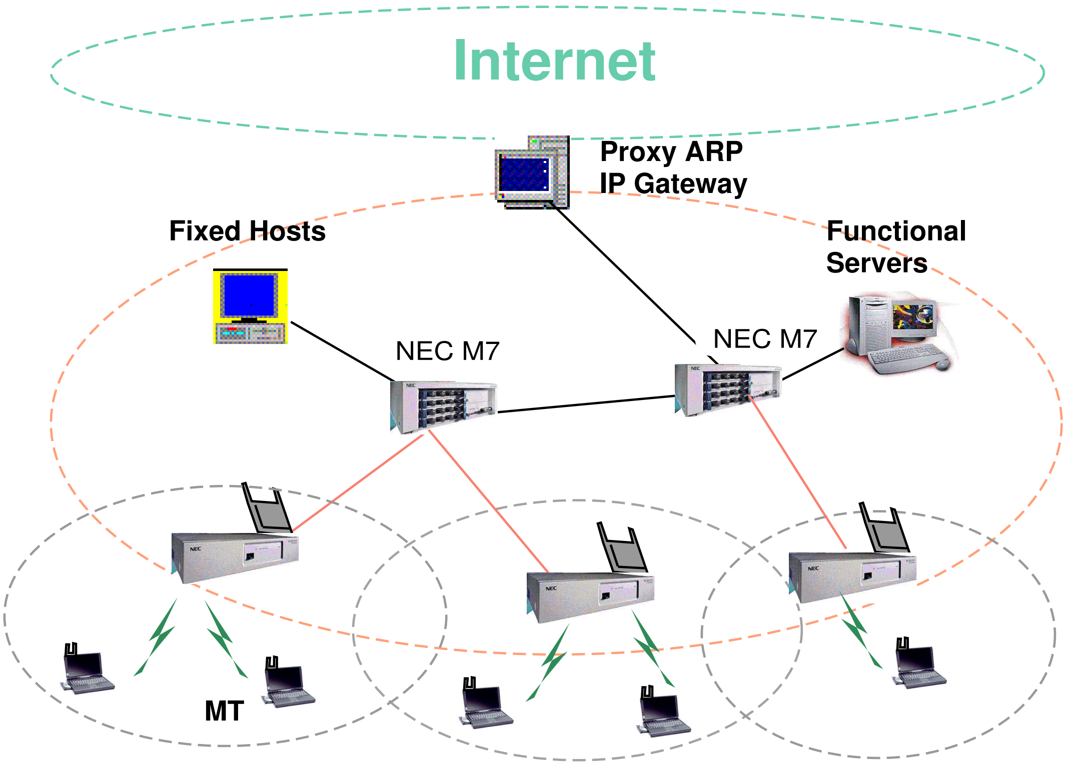

Figure 6.4: Wireless ATM network trial system

Table 6.2 shows the hardware and software configuration.

6.3.2 Mobility support functionalities

The trial system supports end-to-end ATM signaling so there is no interworking function for the ATM signaling. However, for the data transmission, data interworking function is needed to convert WATM cells to regular ATM cells on basestations.

A basestation does not have an ATM switch core, instead, it acts as a bridge between a WATM NIC card from the radio link and a regular ATM NIC card to the fixed network. Logically, it switches a VC from one NIC to another VC in the other NIC. However, the switch is done at the AAL layer. The WATM cells are assembled in WATM NIC and then passed to AAL layer. The assembled data is again segmented to ATM cells at the ATM NIC to the network side. The interworking function at AAL layer is performed by a module in the kernel which interface the output/input of WATM and ATM NIC cards.

On both mobile terminals and basestations, there is a wireless control module in

ATM switches	NEC Model 7, 2.4Gbps
Switch Controller	NEC Powermate with 400MHz Pentium II
Basestations	Dedicated machine runs pSoS
Mobile Terminal	NEC Versa
Fixed Host (CT)	NEC Powermate with 400MHz Pentium II
ATM interface	Zeitnet 155Mbps ATM NIC card
Radio Access	WATMnet 2.0 PCMCIA NIC
OS on SW controller	Linux 2.0.29 with ATM stack 0.26
ATM Signaling	UNI/NNI with mobility enhancement
Switch Control	GSMP server embedded in the switch
PNNI Routing	not used
Handoff Re-routing	First-Cross-To-CT algorithm

Table 6.2: Hardware and software configuration of the trial system

the user space of the operating system. It has an API to WATM NIC driver and sends/receives wireless control messages for mobile registration and handoff detection.

The integrated location management approach is used. A mobile terminal is always associated with a home ATM address at one basestation. Whenever it moves, it uses M-ILMI protocol to pre-register on the new basestation before handoff. After handoff, the current location of the mobile is updated on the MIBs of MT's home ILMI server (HLS) by a location update VC from the MT to the HLS.

The unified handoff protocol is implemented based on newBS COS discovery signaling syntax. First cross to CT (FXCT), an approximation of last divergence (LD) path rerouting, is used as the primary handoff path rerouting algorithm.

6.4 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 Figure 6.3, we have measured the signaling delays, illustrated in Table 6.3. They are approximate values based on about a hundred sample runs and may be largely dependent upon the processing capacity 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.

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

Table 6.3: Signaling delay per control function

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

Table 6.4: Signaling delays for call setup and handoff path re-routing

- 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 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 6.3, 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. On the other hand, 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 will conduct the simulations and experiments for larger network topologies in the future.

Appendix A

Derivation of the flow transition matrix from the user transition matrix

Usually, the mobility is measured through the users' movement among access points. And the mobility pattern is specified only by user transition matrix D. The flow transition matrix A is not given directly. However, we can derive A from D as follows. In Figure A.1, traffic flows are directional. From nodes 1 and 2 to nodes 3 and 4. A part of traffic flow ψ_1 can be handed off to traffic flow ψ_2, ψ_3 or ψ_4 due to (1) end user at access point 1 moves to 2, (2) end user at access point 4 moves to 3 or (3) the end users at both 1 and 4 move to 2 and 3 at the same time.

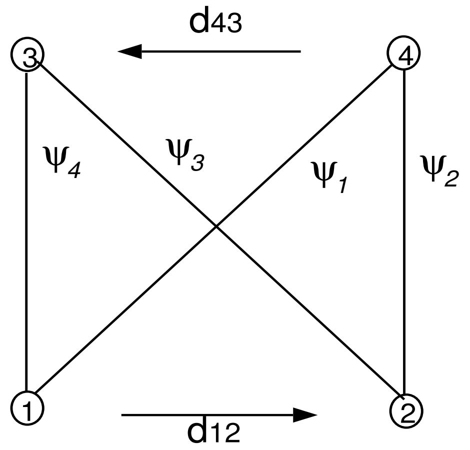

Figure A.1: Transition of traffic flows

Suppose an active user in node i departs to node j with probability d_{ij}. The time period that a user stays at an access point i is a random variable, T_i with the exponential distribution $F_{T_i}(t) = 1 - e^{-\gamma_i t}$. Note that γ_i is the *user departure rate* at access point i. Suppose a typical call's departure rate is Γ. A call departing from the access point 1 is either because of the termination or handoff, so a call's departure rate from the access point 1 is

$$\nu_1 = \gamma_1 + \Gamma \tag{A.1}$$

Therefore, the traffic flow ψ_1 has departure rate $\mu_1 = \nu_1$. Within the departure traffic flow from the access point 1, a fraction Γ/ν_1 is terminated and a fraction γ_1/ν_1 is added to the flows at other access points, we have $p_{k0} = \Gamma/\nu_i$. And the flow transition probability to a flow $2, 3, 4$ are

$$\begin{aligned} p_{12} &= \frac{\gamma_1}{\nu_1} d_{12} = \frac{\gamma_1}{\gamma_1 + \Gamma} d_{12} \\ p_{13} &= \frac{\gamma_1}{\nu_1} d_{12} d_{43} = \frac{\gamma_1}{\gamma_1 + \Gamma} d_{12} d_{43} \\ p_{14} &= \frac{\gamma_1}{\nu_1} d_{43} = \frac{\gamma_1}{\gamma_1 + \Gamma} d_{43} \end{aligned} \tag{A.2}$$

respectively, assuming the user movement from access point 1 to 2 is independent from the user movement from access point 4 to 3. Given mobility pattern of an access network $G_w = [Y, D]$, and the average call lifetime $T(1/\Gamma)$, the flow transition matrix can be obtained.

Appendix B

Solutions of the balance equations

For traffic flows $\Psi(t)$ over network OD pairs, balance equation is satisfied.

$$\dot{\Psi}(t) = A(t)\Psi(t) + \dot{W}(t) \tag{B.1}$$

When $W(t)$ is a K-dimensional Brownian motion, i.e. the arrival traffic flow is a K-dimensional Gaussian random processes $\Omega(t) = \dot{W}(t)$. By applying statistical expectation on both side of Equation (B.1) and exchanging derivative and expectation, we have the mean vector of traffic flows is

$$\dot{m}_{\Psi}(t) = A(t)m_{\Psi}(t) + m_{\Omega}(t) \tag{B.2}$$

The covariance of $\Psi(t)$ is defined as

$$R_{\Psi}(t) = E[(\Psi(t) - m_{\Psi}(t))(\Psi(t) - m_{\Psi}(t))^T] \tag{B.3}$$

By applying derivative to both sides and exchange with the expectation operator, we obtain derivative of the covariance matrix satisfying

$$\dot{R}_{\Psi}(t) = E[(\dot{\Psi}(t) - \dot{m}_{\Psi}(t))(\Psi(t) - m_{\Psi}(t))^T + (\Psi(t) - m_{\Psi}(t))(\dot{\Psi}(t) - \dot{m}_{\Psi}(t))^T] \tag{B.4}$$

The first term on the right side is

$$\begin{aligned} E[(A(t)\Psi(t) + \Omega(t) - A(t)m_{\Psi}(t) - m_{\Omega}(t))(\Psi(t) - m_{\Psi}(t))^T] &= \quad \text{(B.5)} \\ E[A(t)(\Psi(t) - m\Psi(t))(\Psi(t) - m\Psi(t))^T + (\Omega(t) - m_{\Omega}(t))(\Psi(t) - m\Psi(t))^T] &= \\ A(t)R_{\Psi}(t) + E[\int_0^s (\Omega(s) - m_{\Omega}(s))(\Omega(t) - m_{\Omega}(t))^T ds] & \end{aligned}$$

Since $W(t)$ is Brownian motion, the derivative process $\Omega(t)$ is a white noise. Hence the power spectrum is an impulse. The integration becomes

$$\int_0^s \frac{1}{2}\sigma_\Omega(s)\sigma_\Omega^T(t)\delta(t-s)ds = \frac{1}{2}\sigma_\Omega(t)\sigma_\Omega^T(t) \tag{B.6}$$

The second term is symmetric to the first one, so we have

$$\dot{R}_\Psi(t) = A(t)R_\Psi(t) + R_\Psi(t)A^T(t) + \sigma_\Omega(t)\sigma_\Omega^T(t) \tag{B.7}$$

Appendix C

Routing function after applying path rerouting transform

Suppose the traffic flow is stationary when t is sufficiently large, $\dot{m}_\Psi(t) \equiv 0$ and A is constant. The solution of mean traffic flow can be obtained from Equation (5.5),

$$m_\Psi = A^{-1} m_\Omega \tag{C.1}$$

By constraining this solution on Equation (5.12), for $k = 1, ..., K$,

$$\mu_k m_{\psi_k} V_k = \sum_{l}^{K} p_{lk} \mu_l m_{\psi_l} H_{lk} V_l + m_{\omega_k} Z_k \tag{C.2}$$

From equation (C.2), we can find the solution of the set of V_k.

For the example of Figure 5.8, we have

$$\begin{aligned} \mu_1 m_{\psi_1} &= p\mu_2 m_{\psi_2} + m_{\omega_1} \\ \mu_2 m_{\psi_2} &= p\mu_1 m_{\psi_1} + m_{\omega_2} \end{aligned} \tag{C.3}$$

For flow balance on links, we have

$$\begin{aligned} \mu_1 m_{\psi_1} V_1 &= p\mu_2 m_{\psi_2} H_{21} V_2 + m_{\omega_1} Z_1 \\ \mu_2 m_{\psi_2} V_2 &= p\mu_1 m_{\psi_1} H_{12} V_1 + m_{\omega_2} Z_2 \end{aligned} \tag{C.4}$$

Given

$$Z_1 = [1, 0, 0, 0, 1]^T \quad Z_2 = [0, 0, 1, 0, 1]^T \tag{C.5}$$

and H_{12}, H_{21} for a specific path rerouting algorithm, applying Equation (C.4) to Equation (C.5), we can solve for V. For example, using the path extension algorithm on the example of the Figure 5.8.

$$H_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad H_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Since ψ_1 and ψ_2 are symmetric in this example, $m_\psi = m_\omega/\mu(1-p)$. Equation (C.5) becomes,

$$\begin{aligned} V_1 &= \frac{p}{1-p} H_{21} V_2 + Z_1 \\ V_2 &= \frac{p}{1-p} H_{12} V_1 + Z_2 \end{aligned} \tag{C.6}$$

therefore, $[V_1, V_2]$ can be solved as a function of flow mobility p as,

$$[V_1 V_2] = \begin{pmatrix} \frac{1}{1+p_{21}} & \frac{p_{12}}{1+p_{12}} & : AB \\ \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & \frac{p_{12}(1+p_{21})}{(1+p_{12})(1-p_{12}p_{21})} & : AC \\ \frac{p_{21}}{1+p_{21}} & \frac{1}{1+p_{12}} & : CB \\ \frac{p_{21}(1+p_{12})}{(1+p_{21})(1-p_{12}p_{21})} & \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & : CA \\ 1 & 1 & : BD \end{pmatrix}$$

References

[1] D. Raychaudhuri and N. D. Wilson. ATM-based transport architecture for multiservices wireless personal communication networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1401–1414, December 1994.

[2] D. Raychaudhuri. Wireless ATM networks: Architecture, system design and prototyping. *IEEE Personal Communications*, 12(8), December 1996.

[3] D. Raychaudhuri. Wireless ATM networks: Architecture, system design and prototyping. *IEEE Personal Communications*, August 1996.

[4] D. Skyrianoglou N. Passas, L. Merakos. MAC protocol and traffic scheduling for wireless ATM networks. *ACM/Baltzer Mobile Networks and Applications*, pages 275–292, September 1998.

[5] J. Mikkonen and L. Nenonen. Wireless ATM MAC performance evaluation, a case study: HIPERLAN type 1 vs. modified MDR. *ACM/Baltzer Mobile Networks and Applications*, pages 247 – 259, September 1998.

[6] A.Iwata D.Raychaudhuri, R.Yuan and H.Suzuki. Rationale and framework for wireless ATM specification. ATM Forum/95-1646/PLEN, 1995.

[7] M. Srivastava P. Agrawal, P.P. Mishra. Network architecture for mobile and wireless ATM. In *Proceedings of the 16th International Conference on Distributed Computing System*, Hong Kong, May 1996.

[8] B. Samadi S. Nanda, S. Dravida. Handoff management and performance for wireless access using metropolitan area networks. In *Proceedings of IEEE Vehicle Technology Conference*, Secaucus, NJ, May 1993.

[9] A. Acampora and M. Naghshineh. An architecture and methodology for mobile-executed handoff in cellular ATM networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1365–1375, December 1994.

[10] C. K. Toh. Crossover switch discovery for wireless ATM LANs. *ACM/Baltzer Mobile Networks and Nomadic Applications*, 1(2), December 1996.

[11] R. Yuan, S.K.Biswas, L.J.French, J.Li, and D.Raychaudhuri. A signaling and control architecture for mobility support in wireless ATM networks. *ACM/Baltzer Mobile Networks and Applications*, 1(3), December 1996.

[12] S. Seshan, H. Balakrishnan, and R.H. Katz. Handoffs in cellular wireless networks: the daedalus implementation and experience. *IEEE Wireless Personal Communications*, 4(2):141–162, March 1997.

[13] B.Akyol and D. Cox. Rerouting for handoff in a wireless ATM network. *IEEE Personal Communications*, 3(5), October 1996.

[14] 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 *Proceeding of INFOCOM '96*, San Francisco, CA, March 1996.

[15] Y.Choi M.Song and C.Kim. Connection rerouting in general-topology connection-oriented mobile networks. In *Proceeding of IEEE VTC*, Atlanta, CA, April 1996.

[16] M. Veeraraghavan, M.Karol, and K.Eng. Mobility and connection management in a wireless ATM LAN. *IEEE Journal on Selected Areas in Communications*, 15(1):50–68, January 1997.

[17] H. Mitts, H. Hansen, J. Immonen, and S. Veikkolainen. Lossless handover for wireless ATM. *ACM/Baltzer Mobile Networks and Applications*, 1(3), December 1996.

[18] A. Yener and C. Rose. Genetic algorithms applied to cellular call admission: local policies. *IEEE Transactions on Vehicular Technology*, 46(1):72–79, Feburary 1997.

[19] T.P. Chu. and S.S. Rappaport. Overlapping coverage with reuse partitioning in cellular communication systems. *IEEE Transactions on Vehicular Technology*, 46(1):41–46, Feburary 1997.

[20] J.Li, A. Acharya, and D. Raychaudhuri. A signaling mechanism for handoff control in mobile ATM networks. In *Proceedings of the 12th International Conference of Information Networking*, Tokyo, Japan, January 1998.

[21] J.Li, R. Yates, and D. Raychaudhuri. Performance analysis on path rerouting algorithms for handoff control in mobile ATM networks. In *Proceedings of INFOCOM '99*, New York City, NY, March 1999.

[22] J.Li, R. Yates, and D. Raychaudhuri. Handoff control in PNNI hierarchy of mobile ATM networks. In *Proceedings of the 32th Hawaii International Conference on System Sciences*, Maui, Hawaii, January 1999.

[23] J.Li and A. Acharya. Handoff control in point to multi-point connections in mobile ATM networks. In *Proceedings of IEEE GLOBECOM '98*, Sydney, Australia, November 1998.

[24] J.Li, A. Acharya, and D. Raychaudhuri. Zero-cell-loss handoff in wireless ATM networks. In *Proceedings of 5th International Workshop on Mobile Multimedia Communications*, Berlin, Germany, October 1998.

[25] J.Li, R.Yates, and D. Raychaudhuri. Unified handoff control protocol for dynamic path rerouting in mobile ATM networks. In *Proceedings of the Ninth IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Boston, MA, September 1998.

[26] B. Maglaris et al. Performance models of statistical multiplexing in packet video communications. *IEEE Trans. on Communications*, 36:834–844, 1988.

[27] P.A. Skelly et al. A histogram-based model for video traffic behavior in an ATM multiplexer. *IEEE/ACM Trans. on Networking*, 1:446–459, 1993.

[28] M.Minoux. Network synthesis and optimum network design problems: Models, solution methods and applications. *Networks*, 19:313–360, 1989.

[29] J.Y.Hui. Resource allocation for broadband networks. *IEEE JSAC*, 6:1598–1608, December 1988.

[30] R. Guerin et al. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE JSAC*, 9:968–981, September 1991.

[31] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag Press, 1988.

[32] Keith.W.Ross. *Multiservice Loss Models for Broadband Telecommunication Networks*. Springer, 1995.

[33] M. Schwartz. *Broadband Integrated Networks*. Prentice Hall PTR, 1996.

[34] J. Filipiak. *Modeling and Control of Dynamic Flows in Communication Networks*. Springer-Verlag, 1988.

[35] C. E. Perkins. *Mobile IP, Design Principles and Pratices*. Addison Wesley, 1997.

[36] A. Mehrotra. *GSM System Engineering*. Artech House Publishers, 1996.

[37] D. McMillan. Traffic modeling and analysis for cellular mobile networks. In *Proceedings of the 13th International Telecommunication Conference*, Copehagen, 1991.

[38] D.L.Pallant and P.G.Taylor. Modeling handovers in cellular mobile networks with dynamic channel allocation. *Operations Research*, 43(1), January-February 1995.

[39] D.Hong and S.S.Rappaport. Priority oriented channel access for cellular systems serving vehicular and portable radio telephones. *IEE Proceedings, Part I*, 136(5):339–346, October 1989.

[40] D. Reininger, D.Raychaudhuri, and J. Hui. Dynamic bandwidth allocation for VBR video over ATM networks. *IEEE Journal on Selected Areas in Communications*, 14(6):1076–1086, August 1996.

[41] M. Singhal G. Dommety, M. Veeraraghavan. A route optimization algorithm and its application to mobile location management in ATM networks. *IEEE Journal on Selected Areas in Communications*, 16(6):890–908, August 1998.

[42] J. Naylon, D. Filmurray, J.Porter, and A. Hopper. Low-latency handover in a wireless ATM LAN. *IEEE Journal on Selected Areas in Communications*, 16(6):909–921, August 1998.

[43] I. Chalamtac T. Liu, P. Bahl. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal on Selected Areas in Communications*, 16(6):922–936, August 1998.

[44] M. Naghshineh and M. Willebeek-LeMair. End-to-end QoS provisioning in multimedia wireless/mobile networks using an adaptive framework. *IEEE Communication Magazine*, 35(11):72–81, November 1997.

[45] A. Acharya, J.Li, and D. Raychaudhuri. Mobility management in wireless ATM networks. *IEEE Communication Magazine*, 35(11):100–109, November 1997.

[46] A. Acharya, J.Li, A.Bakre, and D. Raychaudhuri. Design and prototyping of location management and handoff protocols for wireless ATM networks. In *Proceedings of ICUPC 1997*, San Diego, Sept 1997.

[47] J.Li and R. Yuan. Handoff control in wireless ATM networks: An experimental study. In *Proceedings of ICUPC 1996*, pages 387–391, Boston, Sept 1996.

[48] T.Phipps, J.Loraine, N.Shipp, L.Neal, and M.Leach. Wireless ATM physical layer requirements. ATM Forum/96-1004, August 1996.

[49] J. Li, A. Acharya, and D. Raychaudhuri. Signaling syntax extensions for handoff control in mobile ATM. ATM Forum/96-1625, December 1996.

[50] A.Acharya, J. Li, and D. Raychaudhuri. Primitives for location management and handoff control in mobile ATM networks. ATM Forum/96-1121, August 1996.

[51] B. Rajagopalan H.Mitts, K.Rauhala and G.Bautz. Proposed handover signaling architecture for release 1.0 WATM baseline. ATM Forum/97-0845, September 1997.

[52] A. Acharya and B. Rajagopalan. Integrated location management for mobile ATM: internetworking with mobile telephony and mobile ip. ATM Forum/97-0851, September 1997.

[53] The ATM Forum. Baseline text for wireless ATM specifications. ATM Forum/BTD-WATM-01.16, Feburary 1998.

[54] The ATM Forum. ATM User-Network Interface (UNI) Signalling Specification, Version 4.0. ATM Forum/af-sig-0061, July 1996.

[55] The ATM Forum. Private Network-Network Interface Specification (PNNI) Version 2.0. ATM Forum/BTD-PNNI 02.00, September 1997.

[56] P. Newman et. al. Ipsilon's general switch management protocol specification, version 1.1. RFC 1987, Augest 1996.

[57] IETF. IP mobility support. RFC 2002, October 1996.

Vita

Jun Li

1985 M.E. in Electrical Engineering, Xidian University, Xi'an China.

1983 B.E. in Electrical Engineering, Xidian University, Xi'an China.

1995-now Research Engineer, C&C Research Labs, NEC USA Inc. Princeton, New Jersey.

1993-1995 Research Assistant, CAIP Center, Rutgers University. Piscataway, New Jersey.

1992-1993 Senior System Engineer, Data Control Limited, San Francisco, Califonia.

1990-1992 Software Engineer, Data Control Limited, Tokyo, Japan

1989-1990 Research Associate, Beijing Information Technology Institute, Beijing, China

1987-1988 Visiting Scholar, C&C Research Labs, NEC Corp. Kawasaki-si, Japan.

1986-1987 Research Associate, Beijing Information Technology Institute, Beijing, China

Publications

- Jun Li, Roy Yates and Dipankar Raychaudhuri, "Performance Analysis on path Rerouting Algorithms for Handoff Control in Mobile ATM Networks", In the Proceedings of IEEE INFOCOM '99, March 21-25, 1999, New York, USA.

- Jun Li, Roy Yates and Dipankar Raychaudhuri, "Handoff Control in PNNI Hierarchy of Mobile ATM Networks", In the Proceedings of the 32th Hawaii International Conference on System Sciences, January 5-8, 1999, Maui, Hawaii, USA.

- Jun Li and Arup Acharya, "Handoff Control for Point to Multi-point Connections in Mobile ATM Networks", Proceedings of IEEE GLOBECOM '98, pp2586-2591, November 8-12, Sydney, AUSTRALIA.

- Jun Li, Arup Acharya and Dipankar Raychaudhuri, "Protocol Design and Implementation for Mobility Management in Mobile ATM Networks", Proceedings of International Conference on Communication Technology, S31-08, October 22-24, 1998, Beijing, CHINA.

- Jun Li, Arup Acharya and Dipankar Raychaudhuri, "Zero-cell-loss Handoff in Wireless ATM Networks", Proceedings of 5th International Workshop on Mobile Multimedia Communications, pp349-352, October 12-14, 1998, Berlin, GERMANY.

- Jun Li, Roy Yates and Dipankar Raychaudhuri, "Unified Handoff Control Protocol for Dynamic Path rerouting in Mobile ATM Networks", Proceedings of Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, B5-1, September 8-11, 1998, Boston, USA.

- Arup Acharya, Jun Li and Dipankar Raychaudhuri, "Mobility Support for IP over Wireless ATM", IEEE Communication Magazine, vol.36, No.4, pp84-88, April, 1998.

- Jun Li, Arup Acharya and Dipankar Raychaudhuri, "Signaling Mechanisms for Handoff Control in Mobile ATM Networks", Proceedings of 12th International conference on Information networking, pp443-446, January 21-23, 1998, Tokyo, JAPAN.

- Arup Acharya, Jun Li and Dipankar Raychaudhuri, "Mobile ATM: Architecture, Protocols and Implementation", Proceedings of IEEE 2nd workshop on broadband switching systems, pp115-119, December, 1997, TAIWAN.

- A. Acharya, S. Biswas, L. French, J. Li and Dipankar Raychaudhuri, "Handoff and Location Management in Mobile ATM Networks", Mobile Communications, pp75-82, Plenum Press, 1997.

- Arup Acharya, Jun Li, Bala Rajagopalan and Dipankar Raychaudhuri, "Mobility Management in Wireless ATM Networks", IEEE Communications Magazine, vol.35, No.11, pp100-109, November, 1997.

- Arup Acharya, Jun Li, Arja Bakre and Dipankar Raychaudhuri, "Design and Prototyping of Location Management and Handoff Protocols for Wireless ATM Networks", Conference records of IEEE 6th International Conference on Universal Personal Communications, pp213-217, October 12-16, 1997, San Diego, USA.

- Jun Li and Ruixi Yuan, " Handoff Control in Wireless ATM Networks: An Experimental Study", Conference records of IEEE 5th International Conference on Universal Personal Communications, pp387-391, September 29 - October 2, 1996, Boston, USA.

- R.Yuan, S.K.Biswas, L.J.French, Jun Li and D.Raychaudhuri "A Signaling and Control Architecture for Mobility Support in Wireless ATM Networks" ACM/Baltzer Journal of Wireless Networks, Vol.1, No.3, pp287-298, December 1996.

- E. Ezhan, J. Hui, Jun Li and J. Zhang "Client-Server Synchronization and Buffering for Variable Rate Multimedia Retrievals" Journal of selected areas in communications, Vol.14, No.1, pp226-237, January 1996.

- J. Zhang, J. Hui and J. Li "Quality of Service Control in GRAMS for ATM Local Area Network", Journal of selected areas in communications, Vol.13, No.4, pp700-709, May 1995.

- A. Acharya, J. Li, D. Raychaudhuri "Primitives for Location Management and Handoff in Mobile ATM Networks", ATM Forum/96-1121, Aug 96.
- J. Li, A. Acharya, D. Raychaudhuri "Signaling syntax extensions for handoff control in mobile ATM",ATM Forum/96-1625, Dec 96.
- A. Acharya, J. Li, D. Raychaudhuri "Signaling syntax extensions for location management in mobile ATM",ATM Forum/96-1624, Dec 96.
- B.Rajagopalan, J. Li and A. Acharya, Signaling for Connection Rerouting for Handoff Control in Wireless ATM", ATM Forum Contribution/97-0338, May 1997.

Patents pending

- Jun Li and Arup Acharya " Handoff Control for Point to MultiPoint Connections in Mobile ATM Networks ", US Patent Application No. 09/145,225. September 1, 1998.
- Dipankar Raychaudhuri, Jun Li and Arup Acharya " Method for Zero Cell Loss Handoff in Mobile ATM Networks ", US Patent Application No. 08/948,612. October 10, 1997.
- Arup Acharya, Jun Li, Dipankar Raychaudhuri and Ruixi Yuan "Handoff Control Techniques for Wireless ATM ", US Patent Application No. 08/733,738. December 24, 1996.