THE UNIVERSITY OF CALGARY

A Handoff Rerouting Scheme for Wireless ATM Networks

by

Lei Fu

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

OCTOBER, 2000

0-612-64996-2

Abstract

With high demands of the market, the capabilities of wireless networks are progressing at a fast speed. Wireless ATM is one of the proposed solution technologies to meet the high bandwidth and QoS requirements. Handoff is one of several challenging issues to be resolved in operational wireless ATM networks. During handoff, a mobile's network connections need to be rerouted from one base station to another base station. Since handoff cannot happen in zero time, ATM cells need to be stored in a buffer during the connection. The handoff procedure can cause delay and loss of the ATM cells because of buffering. This thesis proposes a new handoff scheme called Soft Handoff Like Rerouting (SHLR). Soft handoff is utilized in radio links during network rerouting. This scheme aims to minimize latency due to connection rerouting and at the same time improves radio link quality during the handoff procedure. Results from both simulation and analytical models indicate that the SHLR scheme guarantees a lower cell loss ratio and a lower mean queue delay.

Acknowledgments

I would like to express my sincere appreciation to Dr. A. Fapojuwo for his guidance as my supervisor and for the invaluable knowledge he has shared with me. This thesis couldn't have been done without his understanding and helpful suggestions. I am very grateful to Dr. A. Fapojuwo for his continuous support on my path towards the completion of my thesis.

I greatly appreciate the financial assistance of the Natural Sciences and Engineering Research Council and of TR*Labs*.

I would like to thank Grant McGibney and Anthony Lo for their assistance on Lab support. I must also express my gratitude to professors at the University of Calgary, Mr. George Squires at TR*Labs*, Calgary, for their helpful assistance and good will.

Finally, I would like to thank my husband and our parents for their patience, understanding and support.

Dedication

To my family

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

a	offered traffic load
a_h	handoff traffic load
A	area covered by one base station
AAL	ATM adaptation layer
ATM	asynchronous transfer mode
AWA	ATM wireless access
$b^*(\theta)$	Laplace-Stieltjes transform of the service time distribution function
B(t)	service time distribution function
BAHAMA	broadband adaptive homing ATM architecture
BCR	boundary crossing rate
BCR_{BS}	boundary crossing rate to any other BS coverage area when mobile is on the phone
BCR_{on}	boundary crossing rate when mobile is on the phone
BS	base station
BSC	base station controller
c	server transmission speed
CBR	constant bit rate
CDMA	code division multiple access
dB	decibels
F(x)	distribution function
$f^*(x)$	Laplace-Stieltjes transform of the distribution function
FDMA	frequency division multiple access
FPLMTS	future public land mobile telecommunication systems
g(z)	probability generating function
h	call holding time
H	handoff rate
HHLR	hard handoff like rerouting

i	imaginary number
j	imaginary number
k	imaginary number
l	the mean number of waiting calls
L	the boundary length of the coverage area
LAN	local area network
lon	ON-OFF source ON period
loff	ON-OFF source OFF period
LS	local switch
M	number of MTs in on geographical cell
Mbps	megabits per second
MBR	mean bit rate
MES	mobility enhancement switch
M_{ho}	maximum number of mobiles that are allowed to perform handoff at on network node simultaneously
MSC	mobile switching center
MT	mobile terminal
NCNR	nearest common node rerouting
PASTA	Poisson Arrivals See Time Average
p_j	probability that j calls arrive during the service time
$p_j(t)$	probability density function
P_B	probability of blocking
P_h	probability that the traffic is due to handoff
P_{init}	handoff request initiation rate
P_j	probability that j calls exist at an arbitrary instant in steady state
P_{on}	probability that MT is moving crossing the boundary when mobile is on the phone
PBR	peak bit rate
PCS	personal communication service
PHS	personal handy phone system

PN	pivot node
RAL	radio access layer
RLLC	radio logical link control layer
RMAC	radio multiple access layer
RPHY	radio physical layer
s	number of virtual channels assigned to the transmission line
SHLR	soft handoff like rerouting
SNR	signal to noise ratio
TDMA	time division multiple access
v	MT moving speed
V_j	jth VBR source data speed
VBR	variable bit rate
W	mean waiting time
WAN	wide area network
y_j	jth VBR source data activity
Π_j	probability just prior to call arrival epoch
Π_j^*	probability that j calls exist just after a call departure in the steady state.
λ	call arrival rate
β	burstiness factor
Σ	summing symbol
$\int$	integrate symbol
!	factorial symbol

Chapter 1
Background – Wireless ATM

1.1 Introduction to Wireless ATM

The field of modern telecommunication is rapidly being redefined by events surrounding two evolving concepts: broadband networks and personal communication services. ATM networks are characterised by packet-based transport, bandwidth-upon-demand, and multimedia traffic integration. All types of telecommunication traffic (voice, data, image, video) are carried by ATM networks in a common fixed-length packet format, and the only distinction between low rate and high rate connections is the frequency with which such fixed-length packets are generated [31]. Network resources are statistically shared among users and are actively consumed only when packets are actually generated. Personal communication services are based upon the notions of tetherless access and networks that support connections between people or between people and places. Implicit here is the ability of the network to locate and communicate with a called person wherever that called person may be (as contrasted with today's networks which interconnect specific fixed ports regardless of whether or not the called person is in the vicinity of the called fixed port), and the ability of the network to handoff connections among network ports in response to user mobility. Also, central to the notion of personal communication service (PCS) are specific network services customised to the unique needs of a given user (e.g., filtering and forwarding of electronic mail).

A model for a ubiquitous telecommunication network that merges these two concepts is proposed in the literature. This model is based upon 1) a broadband wired infrastructure supporting Asynchronous Transfer mode (ATM) packet transport; 2) a system of radio base stations, each connected on one side to the wired infrastructure and each supporting an on-demand packet access shared radio channel serving all users within the vicinity or "geographical cell" surrounding the base station; and 3) portable communication units which transform user generated signals into ATM cells which are transferred to/from the base station over the shared radio channels. This W-ATM model generalises upon and exploits the concept of the ATM virtual connection in such a way as to accommodate a very high rate of connection handoff among base stations. This, in turn, allows the use of geographically small radio cells, a high degree of spatial frequency reuse, and the establishment of a high capacity environment as required to support bandwidth-intensive tetherless multimedia services.

1.2 Definition of the problem

The problem addressed in this thesis is focused on handoff issues in wireless ATM networks.

The design of W-ATM network raises a number of challenges, among which the mobility issues are addressed here. For Personal communication Services (PCS) mobility functions are basic. These include mobile registration, call delivery and handoff. Handoff is the process of switching the communication link from the old coverage area to the new coverage area when a wireless user moves during active communication.

In today's wired ATM environment, the user-network interface is a fixed port that remains stationary throughout the connection life time. At the connection set up time, a call processor establishes a path or a network route based on the connection traffic characteristics such that no network node or transmission link is overburdened and the quality of service guaranteed to the existing calls is maintained.

However, in a personal communication network, mobility or changing channel quality causes the users' access point to the wired network to change constantly, and a mobile user's call must be rerouted each time that the connection is handed over to a new base station. In a mobile network, as the mobile users move, segments of connections need to be torn down and reestablished. Meanwhile, maintaining cell sequence and connection QoS while performing handoffs are important requirements in a wireless ATM network.

During handoff procedure, ATM cells might need to be buffered in an ATM switch until a new route is available. The buffering requires network resource and may cause ATM cell loss and delay. To design good handoff schemes the above mentioned issues need to be considered.

The next section presents several solutions proposed by previous work in the literature. These proposals represent the main ideas in handoff protocol development.

1.3 Literature review

To reroute a mobile's connectivity to the ATM network, a rerouting switch should possess the ability to set up and release partial connections to the base station, keep the connection from the backbone network to the switch unchanged, and update the virtual path/connection routing table accordingly during rerouting. If the movement is still under the same switch, rerouting is done at the common switch. If the movement is across switches, there are several possible points in the ATM network where the connection can be rerouted. A number of schemes have been proposed and here are brief descriptions of five of them that represent current solutions to the handoff issues in W-ATM.

BAHAMA rerouting [34] scheme extended the old connection at the old BS to the new BS. This scheme does not require any rerouting, and thus modifications, at the ATM switches. However, the resulting path is not optimal; besides, more modifications are required at BSs. ATM cells may need to be buffered in the ATM switches along the connection path.

Virtual-Connection-Tree-Based Rerouting [18] utilizes a virtual connection tree. A virtual connection tree is formed by a root node of the backbone ATM networks. When the mobile moves from one node to another, no explicit connection reestablishment is required. The ATM cells will be rerouted to the new location simply by changing the VCI in the ATM cells by the root node of the tree. The mobile is utilizing only one leaf node at a time, whereas the rest of the multipoint connection is not utilized.

In Yuan-Biswas rerouting scheme [17] the connection between handoff switch (anchor switch) and the mobile is partially torn down and reestablished. ATM cells buffered at the old BS will be forwarded to the new BS through ATM nodes before any ATM cell arrived at new BS can be sent.

NCNR rerouting scheme exploits the anchor switch concept [2]. The connection between nearest common node (anchor switch) and the mobile is partially torn down and reestablished. The time-sensitive traffic and throughput dependent traffic have different rerouting algorithms. Buffering is performed for throughput dependent traffic only.

In Ajmone-Marsan, et al. 's rerouting scheme [8] (referred to as HHLR scheme in the later chapters of this thesis) the connection between pivot node (anchor switch) and the mobile is partially torn down and reestablished. Handoff buffers are installed in the new BS to preserve ATM cell sequence and minimize cell loss.

1.4 Contribution

The main contribution of this thesis is the proposal of a novel handoff procedure called Soft Handoff Like Rerouting (SHLR). SHLR is designed to perform connection reestablishment due to handoff event in a fast and QoS-guaranteed manner. SHLR enhances the current ATM signalling protocols by adding support to dynamically maintain an active user's connection. In the current version of ATM signalling protocols there is no provision for rerouting a connection once it is established. By using SHLR, multiple connections between a mobile terminal and the network end user can be rerouted seamlessly while the handoff procedure is being performed. A better radio link quality to meet QoS requirement is a result of adopting SHLR rerouting scheme in the W-ATM network. SHLR also minimises network resource (buffer) requirement. Detailed explanations of SHLR are presented in Chapter 3.

1.5 Thesis outline

The rest of the thesis is organised as follows. In chapter 2, the basic concepts of ATM network are provided, including wireless networks development, wireless channel characteristics, and handoff algorithms classification. This leads to chapter 3 where detailed SHLR scheme is presented and buffer performance is analysed. Also in chapter 3 a previously proposed handoff scheme by Ajmone-Marsan et al. [8] that is related to the SHLR scheme is discussed. Moreover, chapter 3 also gives the comparison of SHLR scheme with several handoff proposals in the literature. Chapter 4 gives the design of simulation modelling. In chapter 5, the simulation results of SHLR and HHLR schemes are first presented. The analytical results are then compared with simulation results for both SHLR and HHLR schemes. Chapter 6 concludes the thesis by outlining the main contributions and the suggestions on topics for future investigation.

Chapter 2

Wireless ATM and Handoffs in W-ATM

The development of wireline and wireless networks are two-sided but has one goal: to adapt to multimedia applications. The final trend is to have one integrated wireline and wireless network. This integrated network can be seen as fixed networks with wireless access or fixed networks with wireless extension. This chapter begins by introducing some ATM networks concepts, wireless channel characteristics and wireless access systems development. The current issues in W-ATM networks are then discussed. Emphasis is placed on handoff issues and an overview of the existing handoff proposals leads to the conclusion to this chapter.

2.1 ATM Network Concepts

2.1.1 ATM Network Applications

A basic driving force behind ATM is the recognition that narrowband circuit-switched connections, as deployed worldwide in support of voice, are inadequate to meet the needs of multimedia traffic [14]. Such multimedia traffic is characterized by continuous bit rate sources (e.g., low-speed digital voice, high-speed digital video) and bursty sources (e.g., data files and digital images wherein the need for bandwidth may vary over some fairly wide range while information is being sent, and falls to zero when no files or images are being sent).

Typical applications of multimedia services that can be supported by ATM are listed in Table 2.1 [1]. They include voice, data, still-picture, and motion-video. To support these applications, it is necessary to cover the wide range of service bit rates, from several tens of kilobits per second to 10 MB/s at the highest, and to support various service types like CBR (constant bit rate), VBR (variable bit rate), ABR (available bit rate), and UBR (unspecified bit rate). In addition, the required delay and error rate depends on the applications. As is well known, ATM is a transport technology characterised by its flexibility to accommodate any type of service as well as high-speed transport capability. In fact, ATM technology has been developed not for specific applications but for various types of applications. ATM is the only transport technology available that can support mixed multimedia information with a single user-network interface (UNI).

Table 2.1 Typical applications of multimedia services

Application	Type of Services	Delay	Required Error rate	Bit Rate
Voice/audio	CBR	Bounded	Medium	8-128 kb/s
Digital data	ABR/UBR	Unbounded	Low-medium	0.1-1.0 Mb/s
Video telephony	CBR	Bounded	Low	3874 kb/s
Motion-video (MPEG1/MPEG2)	CBR/VBR	Bounded	Low	1.5-6 Mb/s
File transfer	ABR/UBR	Unbounded	Low-medium	1.0-10.0Mb/s

2.1.2 ATM Network Elements

To handle such a wide variety of traffic types, all signals are first converted to a common format (a sequence of ATM cells), each of which is treated by the network as being a distinct data packet with its own routing header [14]. In ATM networks, the data is divided into small, fixed length units called cells. Note that the ATM cell is different from the geographical cell describing a base station coverage area, which will be mentioned later in this chapter. The ATM cell is 53 bytes long with 48 bytes payload. Each cell contains a 5-byte header that comprises identification, control priority and routing information. See Figure 2.1.

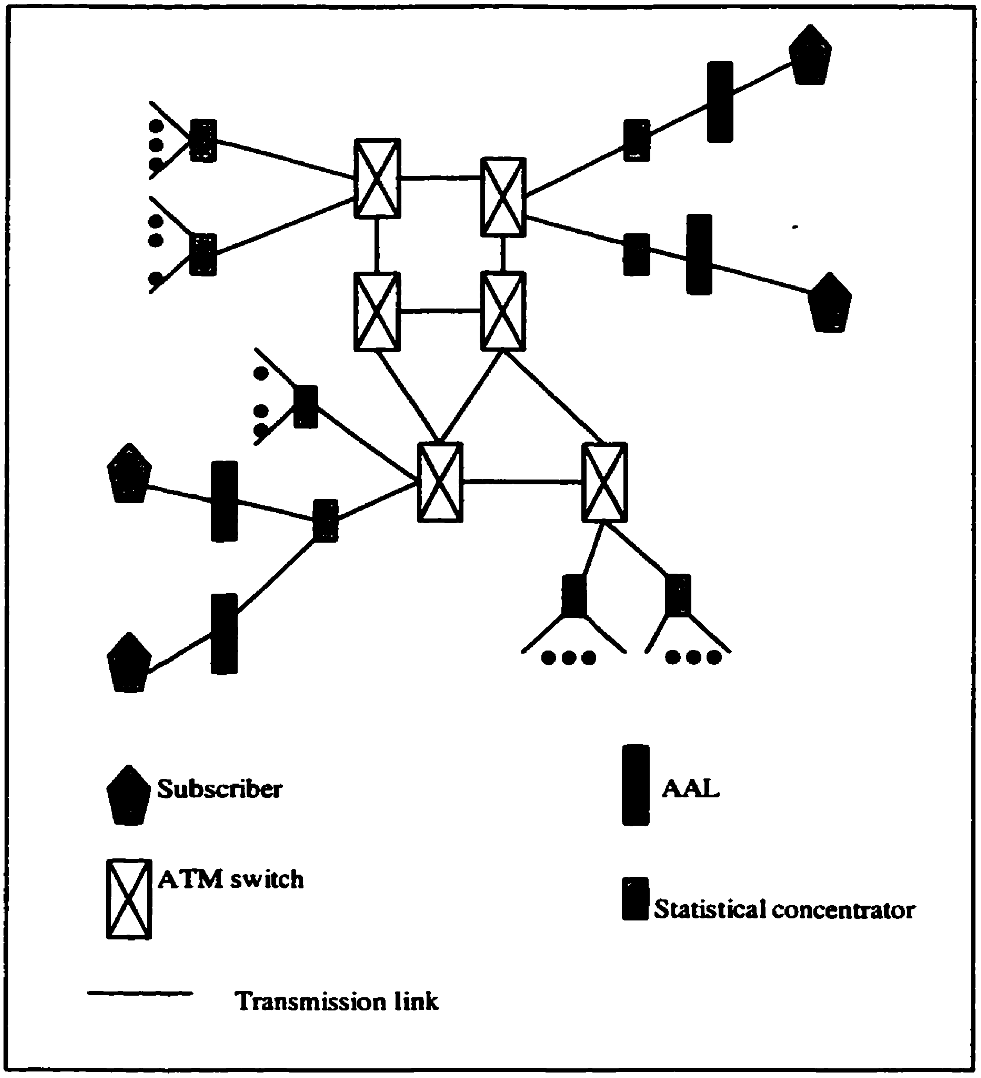

Figure 2.1 Elements in an ATM network

ATM cells arriving at the ATM switches are temporarily stored in smoothing buffers. As a function of time, the number of cells stored in any smoothing buffer will rise and fall in accordance with end-user cell generation patterns. The traffic intensity must be controlled such that QoS guarantees are maintained (e.g., low cell loss probability, minimal cell delay variation).

ATM is a connection-oriented service, meaning that prior to receiving service, a given user must request a "connection" to the intended receiver. Based on a set of traffic descriptors presented to the admission controller at connection set up, the admission controller will attempt to find a route through the network such that if the burden represented by the traffic descriptors is presented to each statistical concentrator and switch encountered along that route, QoS guarantees will be met. If such a route can be found, a virtual connection number is assigned for that call, and the routing tables of the intervening switches are provided with instructions for the routing of each ATM cell bearing that virtual connection number within its cell header. The user is now free to communicate over that newly established virtual connection. Once inside the network, all cells associated with a given virtual connection flow over the route assigned to that connection and are delivered in sequence. If no such route can be found, the user's connection request would be blocked (the equivalent of a busy signal, advising the user to try again later). The AAL (ATM Adaptation Layer) is responsible for converting a user's messages into a sequence of ATM cells (at the transmit end) and for reassembling ATM cells into complete message (at the receiving end). Here, a message may be an individual packet (data or image) or a continuous bit stream (voice or video).

In a real circuit-switched network (as opposed to a virtual circuit network), each connection enjoys exclusive access to the resources reserved for that connection. Thus, once a user's connection has been admitted, it is not possible for that user's traffic patterns to interfere with the service quality enjoyed by any other connection. In a virtual-connection-oriented network, however, resources are not assigned on an exclusive basis, but rather are statistically shared among multiple connections.

To avoid the need to assess the QoS impact of a newly requested connection, possibly on all currently admitted connections, an ATM network relies on virtual paths (VPs) to segregate the collection of virtual connections into independently manageable groups. In essence, real network resources are assigned for the exclusive use of a given VP, and a group of virtual connections are assigned to statistically share each VP. The virtual connections sharing a common VP are called virtual channels (VCs). Since real resources are assigned to each VP, it is not possible for the traffic associated with virtual channels assigned to one VP to influence the QoS enjoyed by virtual channels assigned to other VPs. When processing new connection requests, the admission controller attempts to place a newly requested connection onto a pre-existing VP by assessing the QoS impact of that new connection on the other virtual channels using that VP. In the process, it may become necessary to assign greater bandwidth to the VPs or even to create a new VP between the same endpoints but flowing over a different set of real links. Thus, each VP appears like a "pipe" between endpoints.

ATM switches support two kinds of interfaces: user-network interface (UNI) and network-node interface (NNI). UNI connects ATM end systems (hosts, routers etc.) to an ATM switch, while an NNI may be imprecisely defined as an interface connecting two ATM switches together.

At the UNI, the connection is identified by two values in the cell header: the virtual path identifier (VPI) and the virtual channel identifier (VCI). Both these VPI and VCI combine together to form a virtual circuit identifier.

The relationship between VC, VP and physical transmission path is shown in Figure 2.2 [9]:

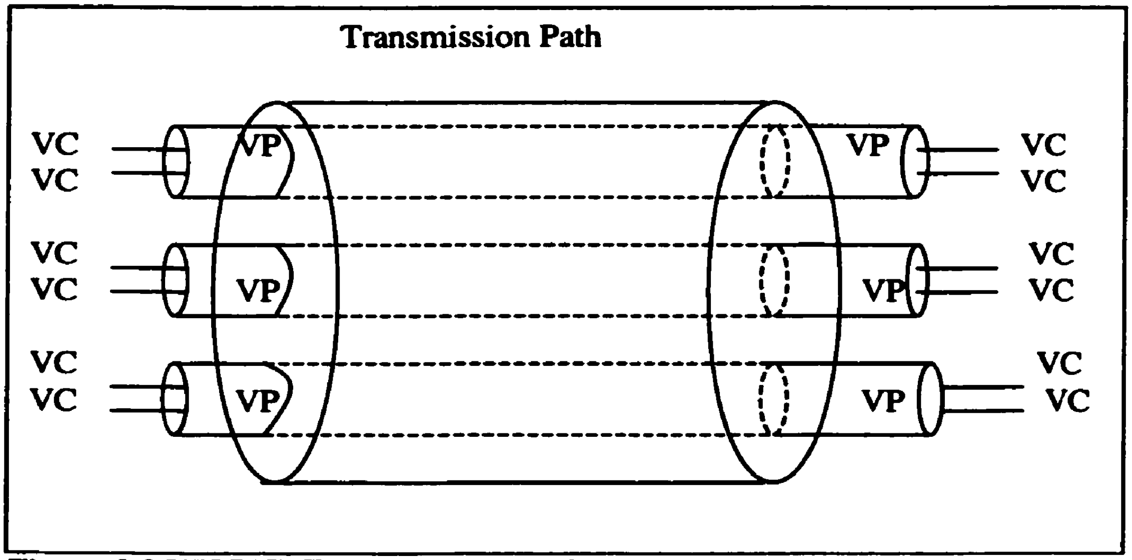

Figure 2.2 VPI/VCI/Transmission path concepts

Finally, it is not necessary for a given virtual connection to be identified by the same connection number on each link of its route. Rather, it is possible for a virtual connection to be identified by a sequence of numbers, one associated with each link, and for the switches to translate connection numbers from input to output. Thus, it is possible for the admission controller, at call setup time, to instruct a switch to route ATM cells arriving at input A and bearing connection number VPI1 VCI6 to output B (see Figure 2.3), and to translate the connection number VPI12 VCI15. The routing instruction for that virtual connection provided to the next switch along the route will refer to connection VPI12 VCI15 rather than to connection VPI1 VCI6. Although the ability to translate connection numbers is provided for reasons having nothing to do with wireless ATM, this property may be exploited to help enable wireless ATM and will be mentioned later in this chapter. Figure 2.3 illustrates VPI VCI usage [9].

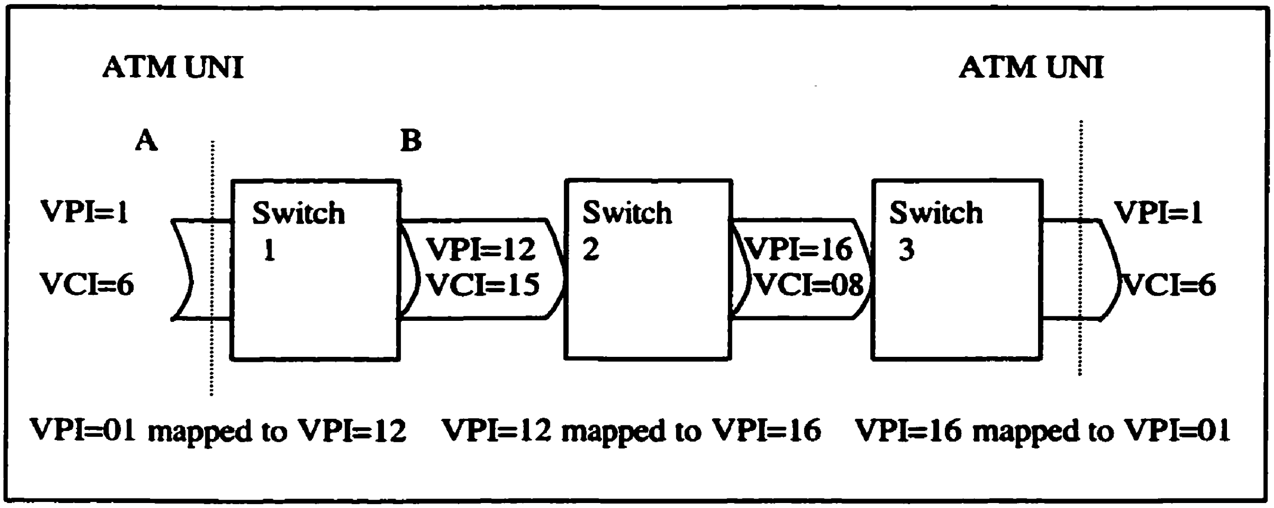

Figure 2.3 Illustration of VPI/VCI usage on link and end-to-end basis

2.2 Development of Wireless Access Systems

Since its inception in the early 1980s, wireless access systems have experienced 3 development phases.

First generation mobile system is characterised as macrocellular, analog, narrowband, and circuit-switched. A typical example of a first generation cellular telephone system is the North American Advanced Mobile Phone Service (AMPS) [29]. The system control for each Cellular network resides in the MSC, which maintains all mobile related information and controls each mobile handoff. The MSC also performs all of the network management functions, such as call handling and processing, billing and fraud detection within the network. MSC is interconnected with the PSTN via landline trunks and a tandem switch. First generation wireless systems provide analog speech and inefficient, low-rate, data transmission between the base station and the mobile user.

In second-generation wireless access systems, digital modulation is used. Advantages of digital modulation include signal robustness, higher capacity, greater compatibility etc. The primary multiple access strategy remains frequency division multiple access (FDMA), time division multiple access (TDMA), e.g., IS-54, Global System for Mobile -- GSM and code division multiple access (CDMA), e.g., IS-95. Second generation wireless systems have introduced new network architectures that have reduced the computational burden of the MSC. GSM has introduced the concept of a base station controller (BSC) which is inserted between several base stations and the MSC.

All second-generation systems use digital voice coding and digital modulation. The systems employ dedicated control channels within the air interface for simultaneously exchanging voice and control information between the subscriber, the base station, and the MSC while a call is in progress. In contrast to first generation systems, which were designed primarily for voice, second generation wireless networks have been specifically designed to provide paging, and other data services such as facsimile and high-data rate network access. The network controlling structure is more distributed in second generation wireless systems and the size of radio cells begins to diminish, thereby enabling higher capacity to be delivered per unit service area.

In third generation wireless access systems designed ultimately to carry multimedia traffic -- whether voice, video, images, files, or data, or combinations of these -- are under intensive study by the Commission of European Communities as well as the International Telecommunications Union (ITU-R) [15]. Third generation wireless systems will evolve from mature second-generation systems. Third generation systems will use the Broadband Integrated Services Digital Network (B-ISDN) to provide access to information networks, such as the Internet and other public and private databases. Packet radio communications will likely be used to distribute network control while providing a reliable information transfer. The RACE 2066 MONET project of the Commission of European Communities has dealt with network architecture and protocol issues arising in the design of the third-generation Universal Mobile Telecommunications System (UMTS), which is closely aligned to the ITU-R proposed standard IMT2000 for future public mobile telecommunications [15]. Thesé multimedia applications also require wireless systems to support a wide range of transmission speeds, flexible connections including asymmetric transmission speed at the forward and return links, and quality of service (QoS) control.

Wireless access to fixed ATM networks is one of the solutions to the integrated multimedia application services. In the following sections wireless ATM issues: W-ATM architecture, W-ATM radio link characteristics, W-ATM mobility issues are discussed.

2.3 Architectures for Wireless Access to ATM Systems

Wireless ATM is essentially the wireless extension of ATM networks.

There are three possible wireless to ATM internetworking architectures to support wireless mobility as indicated in [18]. They differ in the location of the ATM user-network interface (UNI). The air interface in the first two architectures can be any of the existing/emerging PCS standards: Personal Access Communication System (PACS), Global System for Mobile Communications (GSM), code-division multiple access (CDMA), and so on. In the third architecture, ATM cells are transmitted through air interface [8].

Architecture 1 – Mobility support is not required at the ATM switches. Mobility is still supported by the conventional mobile switching centres (MSCs). The MSCs are modified to accommodate an ATM interface to connect to an ATM switch. The signalling protocols – for example, GSM-MAP (Mobile Application Part) and IS41-MAP – remain unchanged. This architecture can be implemented based on the existing ATM standards. No particular standard extensions or modification are required. This architecture is preferable for upgrading the backbone of existing cellular/PCS networks into high-speed ATM links.

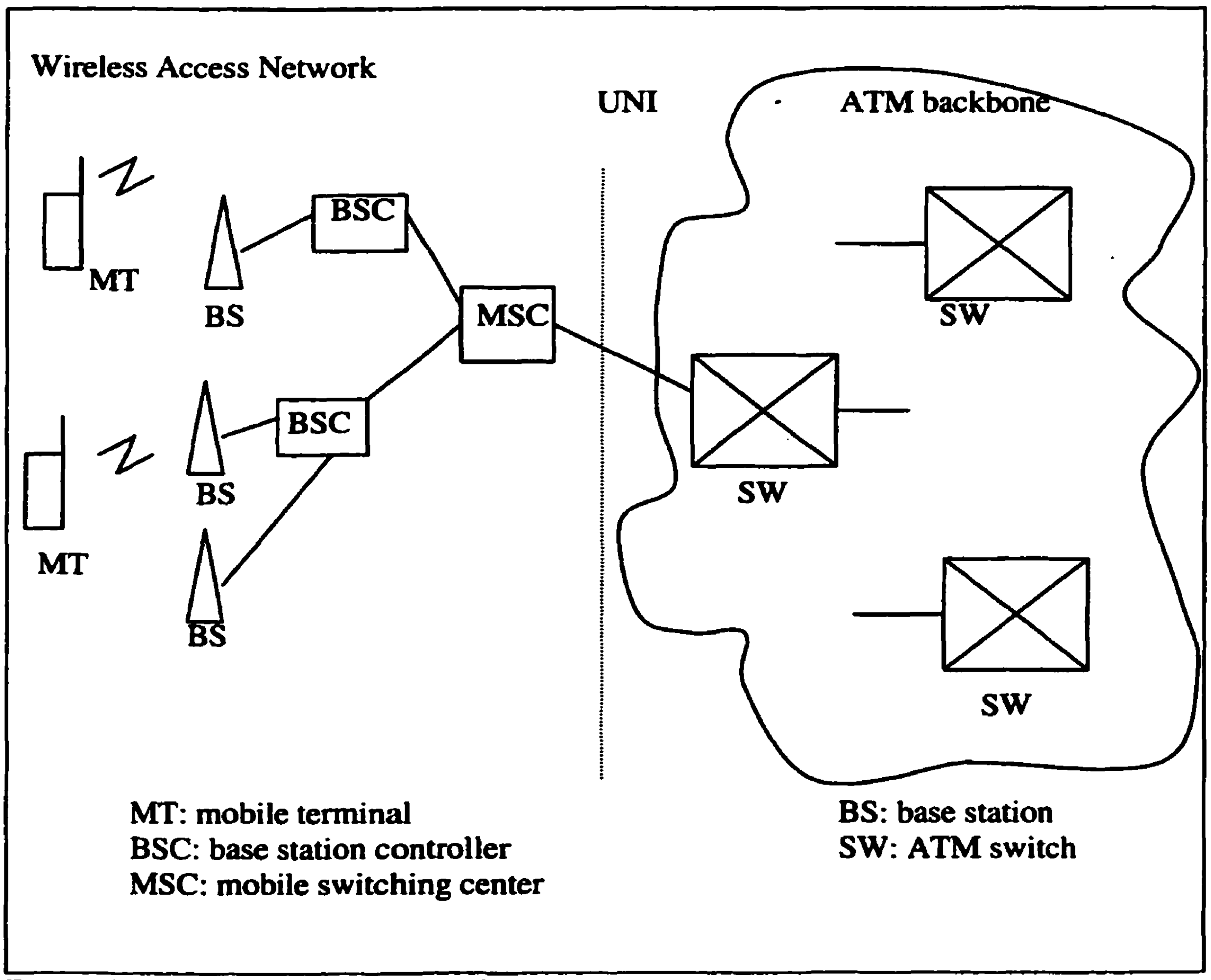

Figure 2.4 W-ATM architecture 1

Architecture 2 – The MSCs are omitted and mobility is supported by mobility-enhancement (ME) ATM switches -- BWSC switches. This architecture requires modifications in the base station controllers (BSCs), which can be modified to accommodate an ATM interface only with the PCS protocols remaining basically unchanged, and in the ATM switches (only the access switches). The second architecture should be more cost-effective than the first thanks to the low cost of ATM switches. (A typical cost of a telephone switch is about 300-500 cents/kb/s, while that of an ATM switch is about 2 cents/kb/s) [18].

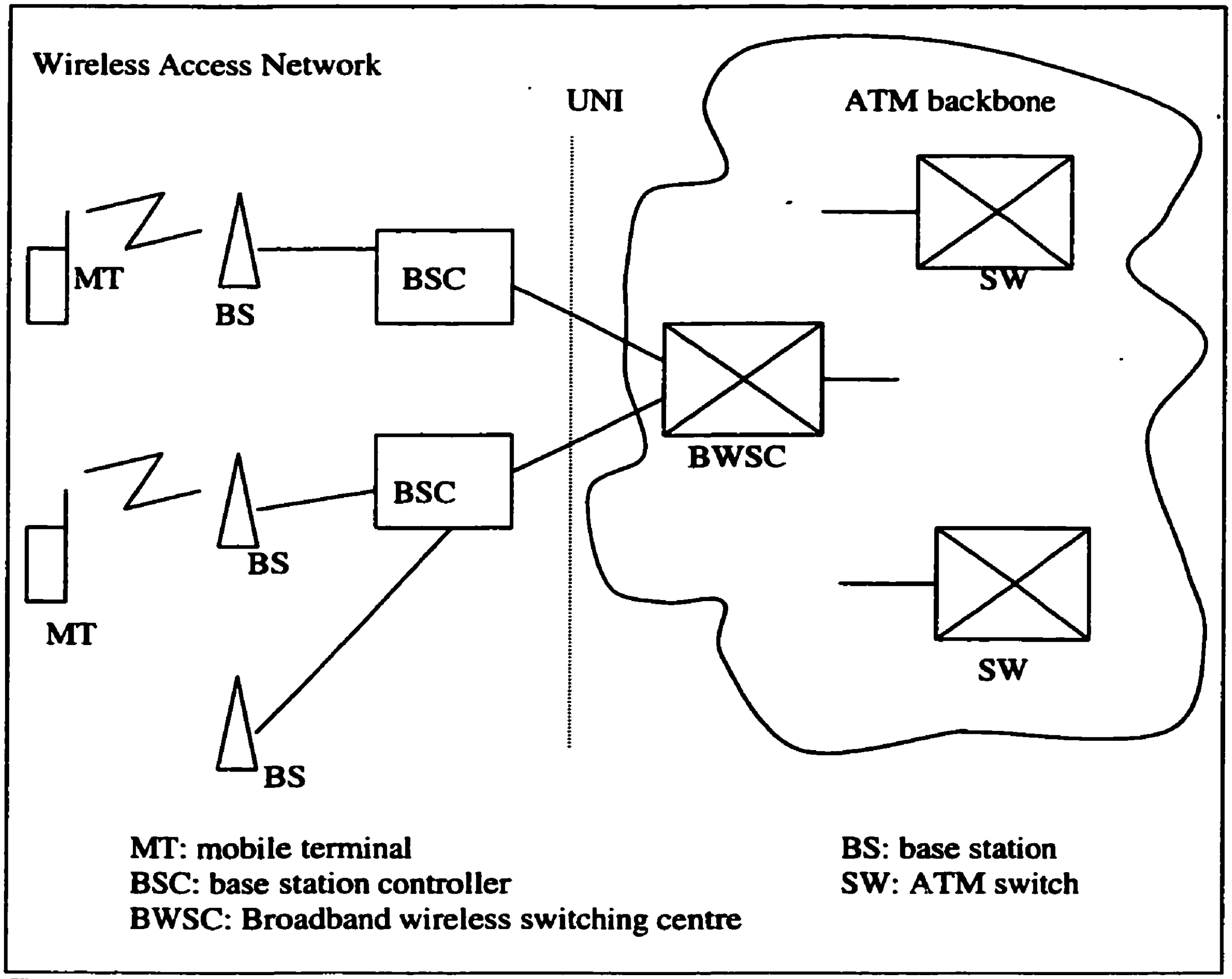

Figure 2.5 W-ATM Architecture 2

Architecture 3 – The BSCs are omitted, and the base stations (BSs) are connected directly to the ME ATM switches. The mobility functions supported by the BSCs are moved to the BSs and/or the ME ATM switches. This may complicate the design of the BSs and ME ATM switches. However, keeping "one integrated network" as the ultimate goal in mind, a mobility enhanced ATM switch is the only solution. Backbone ATM switches can be gradually enhanced to support mobility in a smooth way. With the mass production of ME ATM switches, network with ATM switches can be operated in a cost-effective way. So architecture 3 is adopted in this thesis.

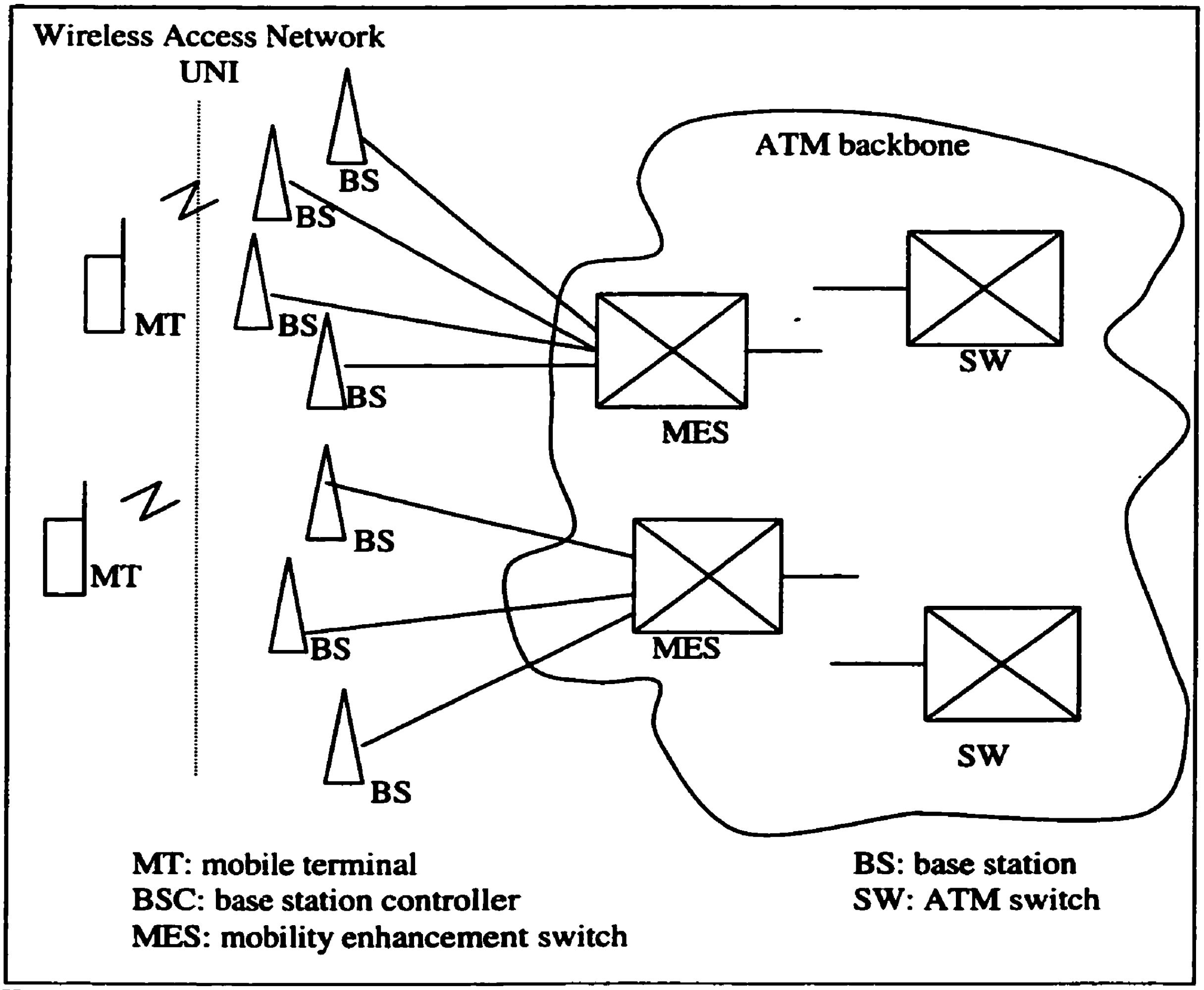

Figure 2.6 W-ATM Architecture 3

2.4 Time Varying Quality of the Two-way Radio Links Between Mobile and Base Station

Unlike the wireline network, wherein the physical link between a remote user terminal and the end-office switch or remote concentrator are of time-invariant high quality, in W-ATM, the mobile radio channel places fundamental limitations on the performance of wireless communication systems [1]. The transmission path between the transmitter and the receiver can vary from simple line-of-sight to one that is severely obstructed by buildings, mountains, and foliage. The radio link is subject to several time-varying impairments arising from inherent user mobility and unavoidable changes due to motion of the surrounding environment (closing doors and passing trucks are two examples of the latter). These impairments are manifested in a time-varying bit error rate (BER) performance of the radio link, with the BER often too low to meet the needs of the application.

In the following sections a brief insight of the root cause of propagation channel impairment is provided by introducing log-normal shadowing and fading propagation models.

Propagation models have traditionally focused on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location [29]. Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver (T-R) separation distance are useful in estimating the radio coverage area of a transmitter and are called *large-scale propagation models*, since they characterize signal strength over large T-R separation distances (several hundreds or thousands of meters). On the other hand, propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances (a few wavelengths) or short time duration (on the order of seconds) are called *small-scale or fading models*. The free space propagation model is used to predict received signal strength when the transmitter and receiver have a clear, unobstructed line-of-sight path between them.

2.4.1 Log-normal Shadowing Propagation Model

Log-normal shadowing belongs to large-scale propagation models. Both theoretical and measurement-based propagation models indicate that average received signal power decreases logarithmically with distance, whether in outdoor or indoor radio channels [29]. Such models have been used extensively in the literature. The average large-scale path loss for an arbitrary T-R separation is expressed as a function of distance *d* [29].

$$\overline{PL}(dB) = \overline{PL}(d_0) + 10n \log \frac{d}{d_0} \qquad (2.1)$$

where n is the path loss exponent that indicates the rate at which the path loss increases with distance, and d_0 is the close-in reference distance, which is determined from measurements close to the transmitter.

The model in equation (2.1) does not consider the fact that the surrounding environmental clutter may be vastly different at two different locations having the same T-R separation. This leads to measured signals that are vastly different than the average value predicted by equation (2.1). Measurements have shown that at any value of *d*, the path loss *PL(d)* at a particular location is random and distributed log-normally (normal in dB) about the mean distance dependent value. That is [29]

$$\begin{aligned} PL(d) &= \overline{PL}(d) + X_\sigma \\ &= \overline{PL}(d) + 10n \log\left(\frac{d}{d_0}\right) + X_\sigma \end{aligned} \qquad (2.2)$$

and

$$P_r(d) = P_t(d) - PL(d) \qquad (2.3)$$

where X_σ is a zero-mean Gaussian distributed random variable (in dB) with standard deviation σ (also in dB). P_r is the received power and P_t is the transmit power.

The log-normal distribution describes the random *shadowing* effects, which occur over a large number of measurement locations, having the same T-R separation, but have different levels of clutter on the propagation path. Simply put, log-normal shadowing implies that measured signal levels at a specific T-R separation have a Gaussian (normal) distribution about the distance-dependent mean of (2.1), where the measured signal levels have values in dB units. The standard deviation of the Gaussian distribution that describes the shadowing also has units in dB. Thus, the random effects of shadowing are accounted for using the Gaussian distribution that lends itself readily to analysis.

The close-in reference distance d_0, the path loss exponent n, and the standard deviation σ, statistically describe the path loss model for an arbitrary location having a specific T-R separation, and this model may be used in computer simulation to provide received power levels for random locations in communication system design and analysis. This model is used in the simulation models developed to evaluate the performance of the new handoff scheme proposed in this thesis.

Shadowing, due to the presence of large physical objects (buildings, walls, etc.) which preclude direct line-of-sight between a radio transmitter and receiver, also gives rise to strong signal power attenuation, but varies at a much lower rate than does multipath fading [1].

2.4.2 Small-scale Fading and Multipath

Many physical factors in the radio propagation channel influence small-scale fading. These include the following:

- Multipath propagation – The presence of reflecting objects and scatterers in the channel create a constantly changing environment that dissipates the signal energy in amplitude, phase, and time. Multipath propagation, due to the superposition of radio waves reflected from surrounding objects, gives rise to frequency-selective fading which, in turn, causes frequent and deep reduction in received power for narrowband signals and rapidly changing time dispersion of broadband signals. (By way of example, in a vehicle travelling at 60 mph, the received power for a narrowband signal might fade 10 dB or more several tens of times per second) [1].
- Speed of the mobile – The relative motion between the base station and the mobile results in random frequency modulation due to different Doppler shifts on each of the multipath components.
- Speed of surrounding objects – if objects in the radio channel are in motion, they induce a time varying Doppler shift on multipath components.

- Transmission bandwidth of the signal –If the transmitted radio signal bandwidth is greater than the "bandwidth" of the multipath channel, the received signal will be distorted.

In mobile radio channels, the Rayleigh propagation model is commonly used to describe the small-scale fading.

The above propagation models show root cause of these time-varying impairments is multipath propagation, shadowing, distance-dependent signal power path loss, etc. Path loss refers to the reduction in the fraction of transmitted power that is actually received as the distance between transmitter and receiver increases, and also produces a slowly time-varying effect. These propagation related phenomena (multipath, shadowing, and path loss) reduce immunity to radio receiver noise and, perhaps more important, increase the susceptibility to co-channel interference arising in other cells which happen to be using the same frequency. In addition, the level of co-channel interference encountered by a given base-mobile link is strongly influenced by the time-varying number of currently active users in any other geographical cells sharing a common frequency, and by the proximity of a mobile to its serving geographical cell's edge of coverage.

The radio link characteristics mentioned in this section are factors that impact the operation of W-ATM networks.

2.5 Technical Issues for Wireless ATM Networks

There are several technical issues to be addressed in wireless ATM network design which can be seen from the rest of this section. One of the issue is the gaps between wireless and wired transmission systems. This issue is basically due to the fact that the ATM specifications have been developed based on high-speed, reliable optical fibre transmission links [1]. On the contrary, the transmission speed of wireless systems is relatively low, and wireless links sometimes suffer high bit error rates (BERs). Another incompatibility is limited capacity due to the finite frequency spectrum resource.

The advantage of wireless systems is definitely mobility resulting from its tetherless feature, which cannot be supported by wired systems. Because of fading, a trade-off between transmission speed and mobility is necessary. The comparison of wireless access schemes in terms of mobility and information rate is shown in Figure 2.7.

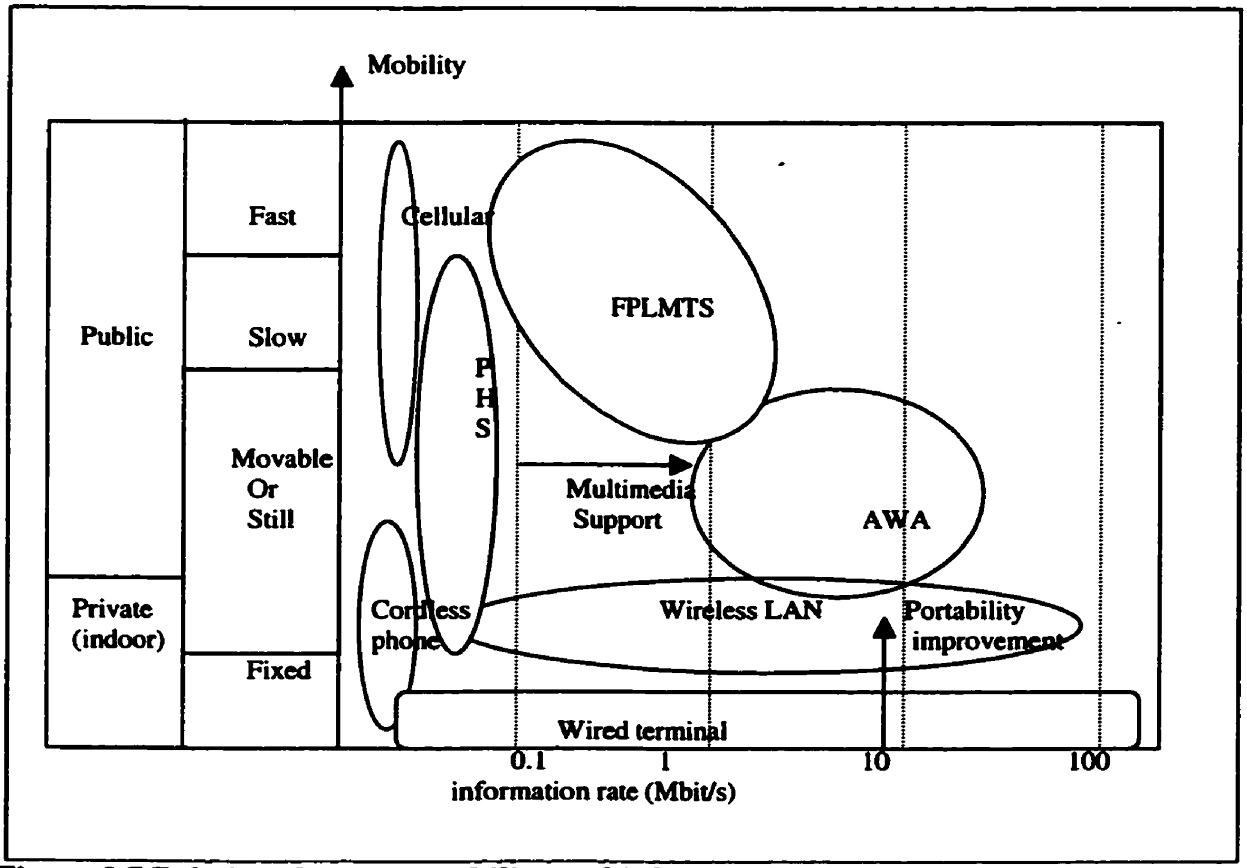

Figure 2.7 Relations between mobility and information rate

In Figure 2.7, AWA is ATM wireless access, PHS is personal handy phone systems and FPLMTS is future public land mobile telecommunication systems.

To connect ATM-based wireless access systems to ATM-based networks, the above-mentioned incompatibility and the trade of between data rate and mobility speed issues need to be solved. Besides, to enable nomadic access to ATM networks, additional functions are required in ATM networks for mobility support. They include handoff control, location management, mobile user authentication and registration, routing of mobile connections. To support handoff, signalling must be extended for dynamic rerouting of a set of virtual connections (VCs) from one to another. Location management is the capability of mapping from "mobile terminal-id" to "routing-id" used to locate the current ATM end point. In conventional mobile systems, the end point is the mobile user, while the wireless ATM endpoint can be either the fixed user or the mobile user. The routing function is the capability to map "mobile terminal routing-ids" to "paths" in the network, and route identification for handoff. ATM routing algorithms need to be upgraded to support these additional mobility-related functions.

Handoff becomes a dominant system issue as cell size diminishes and the rate of handoff rises. Handoff becomes an even greater issue in a multimedia setting wherein QoS guarantees must be maintained. There are lots of proposals that address handoff issues in Wireless ATM networks. The next section provides a summary of existing handoff protocols.

2.6 Classification of Existing Handoff Schemes

Handoff procedure in an ATM network is performed, without an interruption of the user communication, to transfer user's wireless access between base stations and probably reroute user's connection between ATM switches.

Figure 2.6 shows a wireless ATM network consisting of base stations, mobile terminals, and network interface equipment. A mobile terminal might have a few simultaneous connections in the wireless ATM network. When a handoff occurs these connections may need to be rerouted.

As indicated in reference [8], the different approaches proposed in literature to handle handoffs in wireless ATM networks can be broadly subdivided into four categories, which have completely different routing strategy and hence characteristics, performance and impact on the ATM standards:

1. Whole connection set-up rerouting
2. Connection added-on rerouting
3. Virtual-tree based rerouting

4. Anchor rerouting

2.6.1 Whole connection set-up rerouting

The whole connection set-up rerouting requires a completely new connection to be set-up between the fixed end user and the mobile terminals. This is one of the earliest proposals, from the network point of view. This handoff approach is almost the same as setting up a connection when a call is initiated. Although it has a minor impact on the fixed network architecture, this procedure may not be fast enough to guarantee that handoffs do not cause timeouts to expire and connections to be abruptly terminated. In addition, both fixed end user and the mobile terminals must be involved in the set up of a whole new path operation. Figure 2.8 shows the VC evolution for a mobile terminal that roams through three geographical cells in a network adopting whole connection set-up rerouting scheme. LS represents local switch. The first connection is LS_1 -> LS_2 -> LS_4 -> LS_5 -> BS_1. The second connection is LS_1 -> LS_2 -> LS_4 -> LS_5 -> BS_2. The third connection is LS_1 -> LS_3 -> LS_4 -> BS_3.

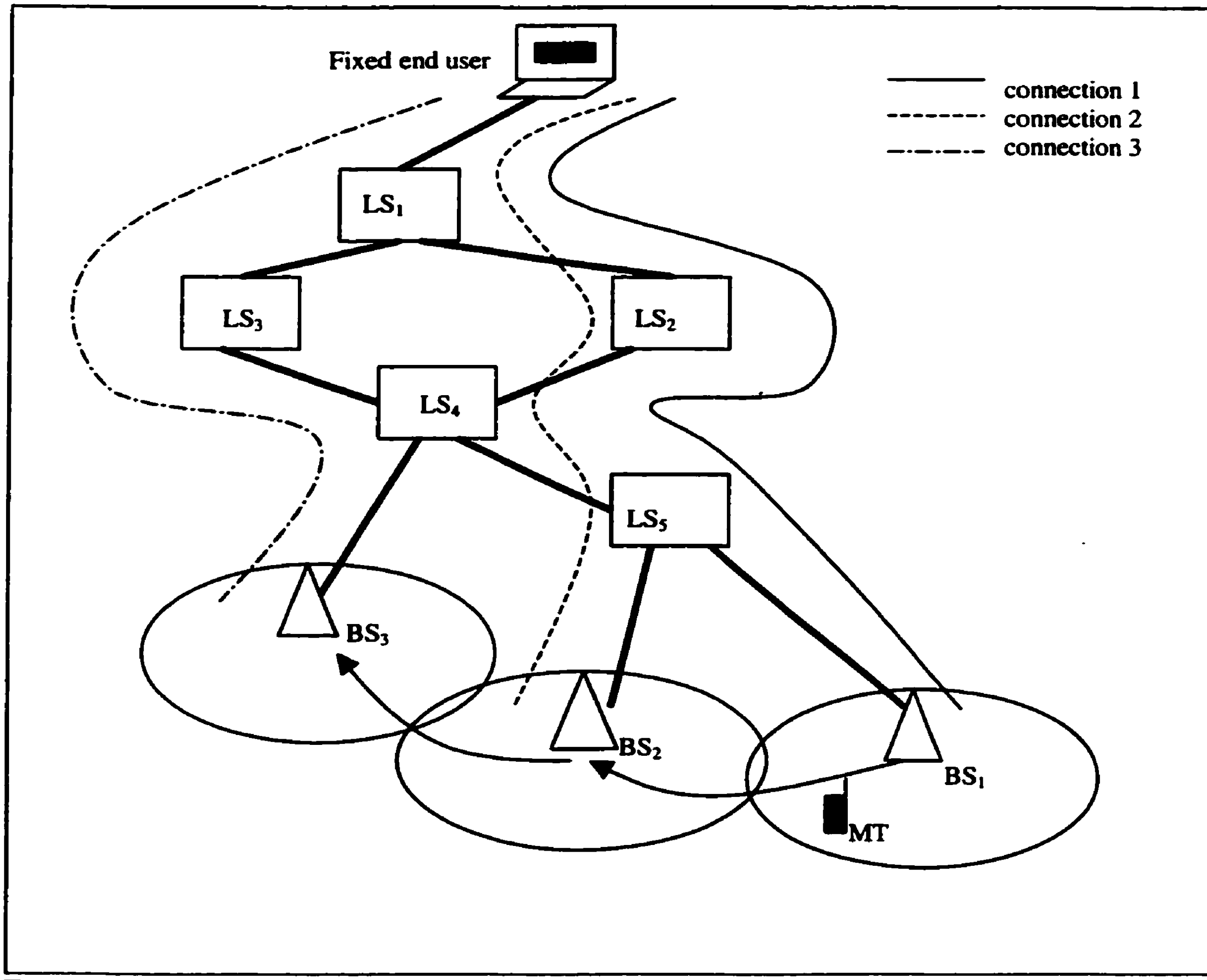

Figure 2.8 Whole connection set-up rerouting

2.6.2 Connection added-on rerouting

The connection added-on establishment algorithm extends the VC between the fixed end users and mobile terminal. The connection between the new ATM switch that the mobile terminal roams in and the old ATM switch the mobile associated with previously is added up to the existing connection. As proposed in [34] for the BAHAMA algorithm, this path extension can be performed by the source base station or as proposed in [35], by the LS. The advantage of connection added-on approach is as follows: simple and reasonably fast execution and inherently preservation of ATM cell sequence. However, there is no rerouting performed and this may cause wastage of bandwidth. When the mobile terminal roams around in a limited area, the mobile may possibly return to a previously visited base station. This can cause close loop connection. Figure 2.9 shows the VC modification needed to connect to the mobile terminal. The original connection of the MT is LS_1-> LS_2->LS_4 -> LS_5 -> BS_1. The second connection with add on path when mobile moves to BS2 coverage area is LS_1-> LS_2->LS_4 -> LS_5 -> BS_1->LS_5 -> BS_2. The third connection with add on path when mobile moves to BS3 coverage area is LS_1-> LS_2->LS_4 -> LS_5 -> BS_1->LS_5 -> BS_2 ->LS_5 ->LS_4 -> BS_3.

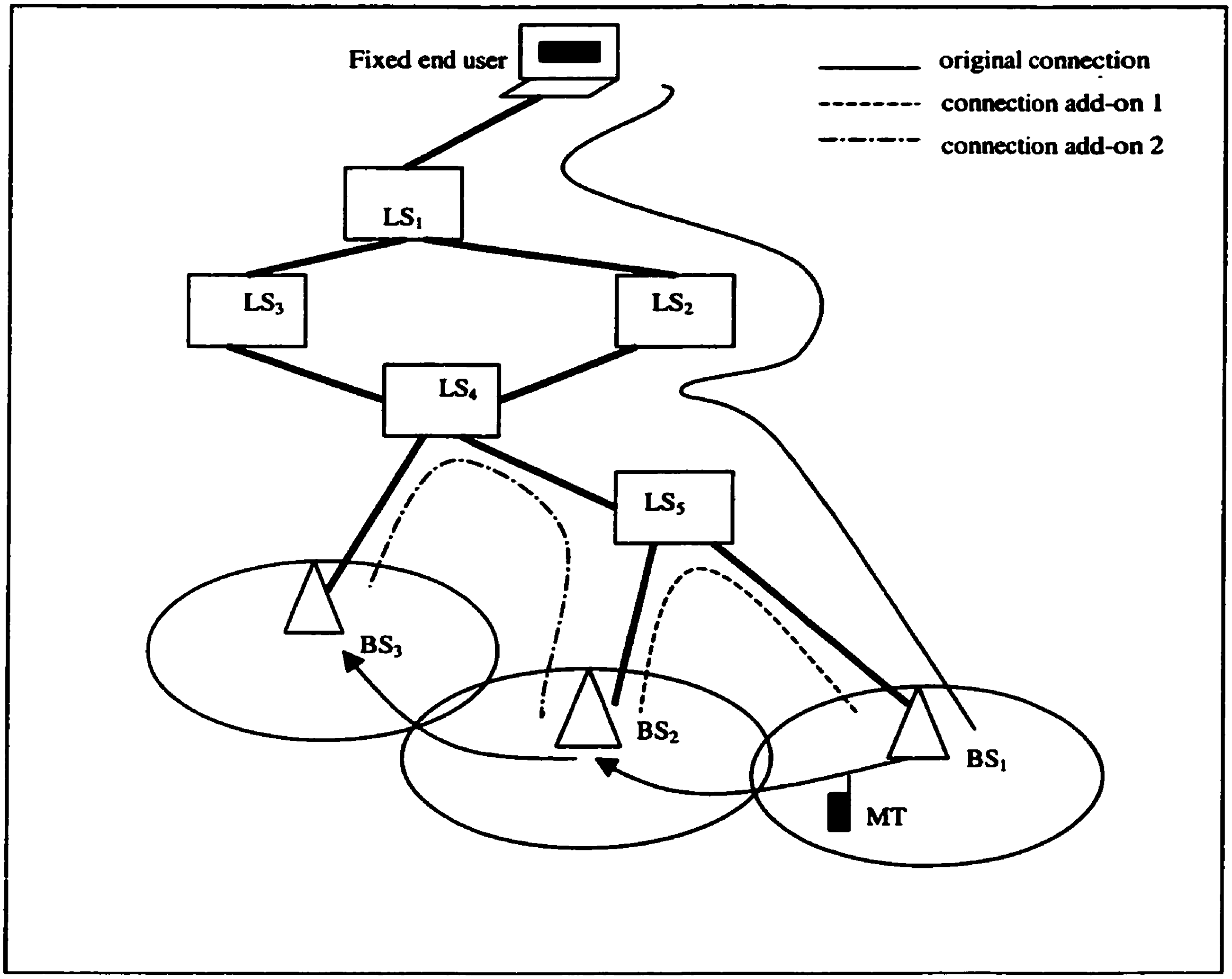

Figure 2.9 Connection add-on rerouting

2.6.3 Virtual-tree based rerouting

The virtual-tree establishment approach uses the concept of a virtual connection tree. The network predicts the mobile's movement and pre-allocates resources in the network portion surrounding the geographical cell where the mobile is located. When a mobile terminal establishes a wireless ATM connection, a connection tree is formed from the root ATM switch to the leaves in the tree effectively producing a point-to-multipoint connection. Thus, the mobile user can freely roam in the tree-covered area without invoking the tearing down and reestablishing connection procedure during handoff. The allocation of the virtual connection tree may be static [31] or dynamic [36] during the connection lifetime. The advantage of this scheme is that it is fast and statistically guarantees the QoS requirements during the execution of the handoff procedure. However, the mobile terminal is utilizing only one leaf switch at a time, whereas the rest of the connection is not utilized. Another disadvantage is high signaling overheads are involved, especially in the case of dynamic tree allocation. Figure 2.10 shows a multicast establishment, assuming that the mobile moves within the three geographical cells.

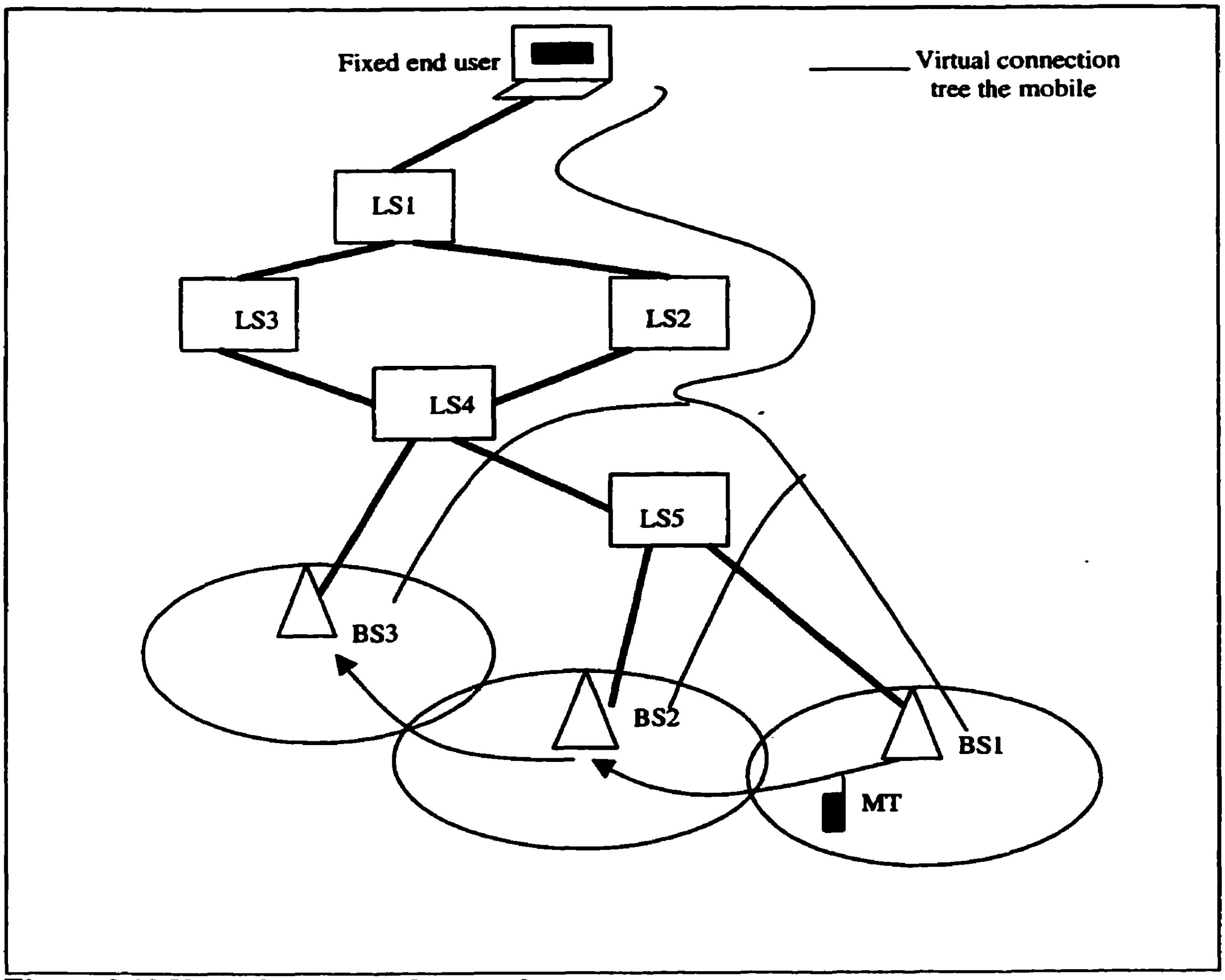

Figure 2.10 Virtual-tree based rerouting

2.6.4 Anchor rerouting

The anchor rerouting technique reuses part of the original path from the fixed end user to the ATM switch along the path. This is called an anchor switch. The path from this switch down to the mobile is different before and after handoff. A new path from this switch to the new base station is set up for the roaming mobile. The whole connection to the fixed end user and the mobile are the consolidation of these two paths. Because of the spatial locality in movement, it is very likely that the reestablished path to the new location of the mobile terminal shares most of the virtual paths (VPs) in the original path. As a consequence, this technique is expected to be fast, efficient, and transparent, so it can be imagined that the fixed end users do not perceive the network handoff as a service interruption. At each handoff, the new path of the connection is optimally established, thus avoiding resource wastage. Figure 2.11 also indicates the pivot node (PN), that is, the anchor ATM switch that connects the original path to the incremental path for the handoff occurring from BS_2 to BS_3. Reference [8] provides an example of the anchor rerouting scheme. Figure 2.11 shows a mobile's original connection as LS_1-> LS_2->LS_4 ->LS_5 ->BS_1. When the mobile moves from BS_1 to BS_2, the new connection is LS_1-> LS_2->LS_4 ->LS_5 ->BS_2, where LS_5 acting as an anchor switch. The path for mobile to communicate with the fixed end user while in BS_3 coverage area is LS_1-> LS_2->LS_4 ->BS_3, where LS_4 acting as an anchor switch.

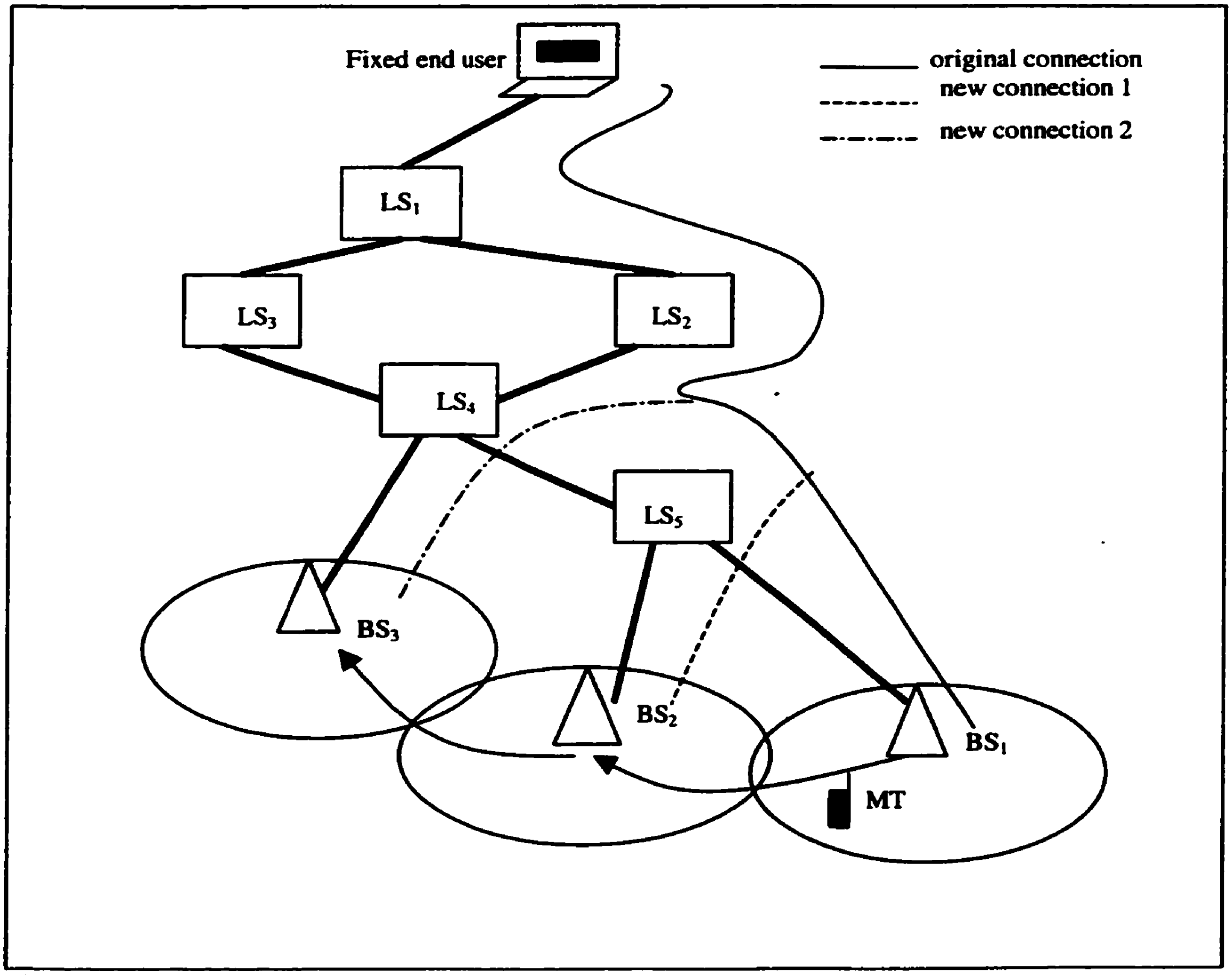

Figure 2.11 Anchor rerouting

The scheme proposed in this thesis is classified under the anchor rerouting category. The proposed scheme is called soft handoff like rerouting (SHLR) scheme, which is described in detail in the next chapter.

Chapter 3

Soft Handoff Like Rerouting Scheme

In this chapter, a novel handoff scheme called Soft Handoff Like Rerouting (SHLR) is presented. The goals of this handoff protocol design are first listed, followed by the introduction to the novel signalling protocol for handoffs in wireless ATM networks. Signalling message diagram during handoff procedure for successful handoff is also shown and explained in this chapter. In user connection rerouting, ATM cells along the connection path may need to be buffered, which can cause loss and delay of ATM cells and inefficient utilisation of network resource. So in this chapter buffer requirements of the network for both the SHLR and a previously proposed scheme [8] are addressed. The chapter is concluded with a comparison of the SHLR scheme with several proposed schemes in the literature.

3.1 Handoff Scheme Design Goals

The design goals for the handoff signalling protocol are:

- Present a new signalling protocol to support mobile's handoff in wireless ATM network.
- Reroute ATM connections seamlessly for users requiring reestablishment.
- Minimise user data loss and delay due to handoff.
- Improve radio link quality during handoff procedure.

3.2 Approaches in Designing Signalling Protocol

Two approaches that may be used for devising a signalling protocol for a wireless ATM network are identified. The first approach, called the *overlay* approach, keeps the existing ATM signalling protocol intact, and functions as an overlay network [6]. The second approach, which is referred to as the *migratory* approach, modifies the existing ATM signalling protocols to accommodate mobile users together with the fixed users and aims to minimise the overhead incurred in supporting mobile users.

Migratory approach is used in the signalling protocol design in the thesis. The motivation for this migratory design is the inefficiency problems associated with the overlay signalling approach. Specifically the inefficient problem of overlay approach to wireless ATM network signalling becomes more prominent for large scale network implementation due to the large number of ATM connection reestablishment attempts for support of mobile users. However, the overlay signalling approach does not require any modifications to the existing ATM signalling protocols. Although the migratory approach can lead to implementation complexity and backward complexity issues, it relaxes the constraint of not modifying the ATM protocols and aims to make the wireless network implementation more efficient.

3.3 The Architecture and Protocol Stack to Support Mobility in an ATM Network

The discussion of this section refers to W-ATM network architecture Type 3 that was presented in Chapter 2.

The W-ATM scenario shown in Figure 3.1 comprises mobile terminals (MTs), base stations (BSs), and ATM switches (ATM nodes). ATM switches and BSs belong to the fixed network segment. MTs directly communicate only with BSs, which provide the interface between the wired and wireless portions of the network. A number of BSs can be connected to the same ATM switch. The ATM switch connected to one or more BSs is termed the local switch (LS), with respect to the MTs that communicate with the BSs.

According to several proposals [8, 17, 3], during VC reestablishment, one ATM switch must perform special functions to manage the rerouting of the connection; this switch is called pivot node (PN) in this thesis; other authors call such an ATM node the cross-over switch (COS) [8]. Figure 3.1 shows the wireless ATM network architecture that is adopted in this thesis. LS and PN are ATM switches that are enhanced to support mobility.

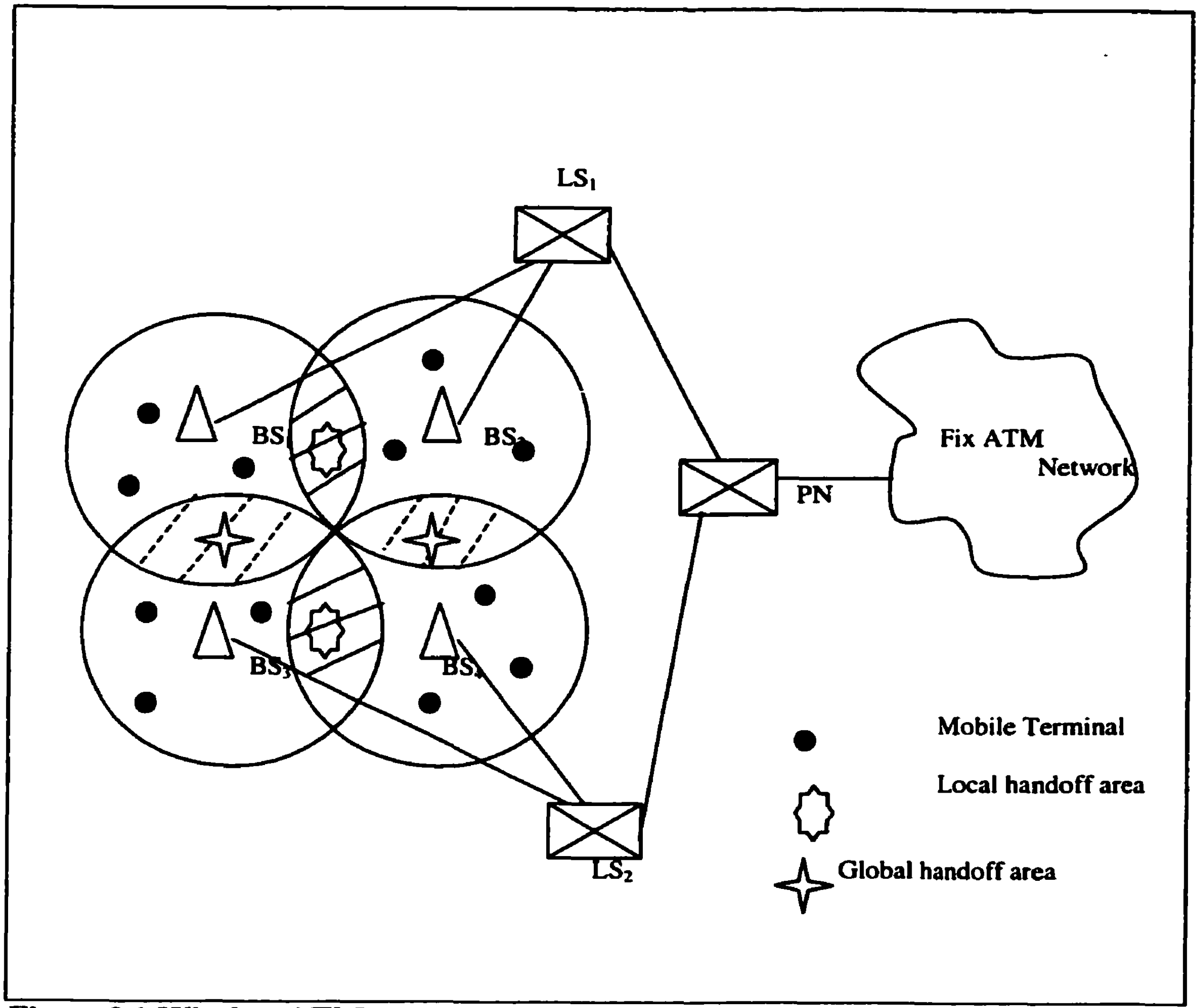

Figure 3.1 Wireless ATM network architecture

In Figure 3.1, there are four geographical cells, each under the control of one base station. BS_1 and BS_2 connect to LS_1, BS_3 and BS_4 connect to LS_2. Another ATM switch marked as PN connects LS_1 and LS_2 to the fixed ATM network. Both local and global network handoffs can happen in this network [8], as described below.

Figure 3.2 shows a global handoff case. A handoff is global when an MT moves either between BS_1 and BS_3 or between BS_2 and BS_4. In this type of handoff, both local switch and pivot node are involved in rerouting signalling messages associated with handoff.

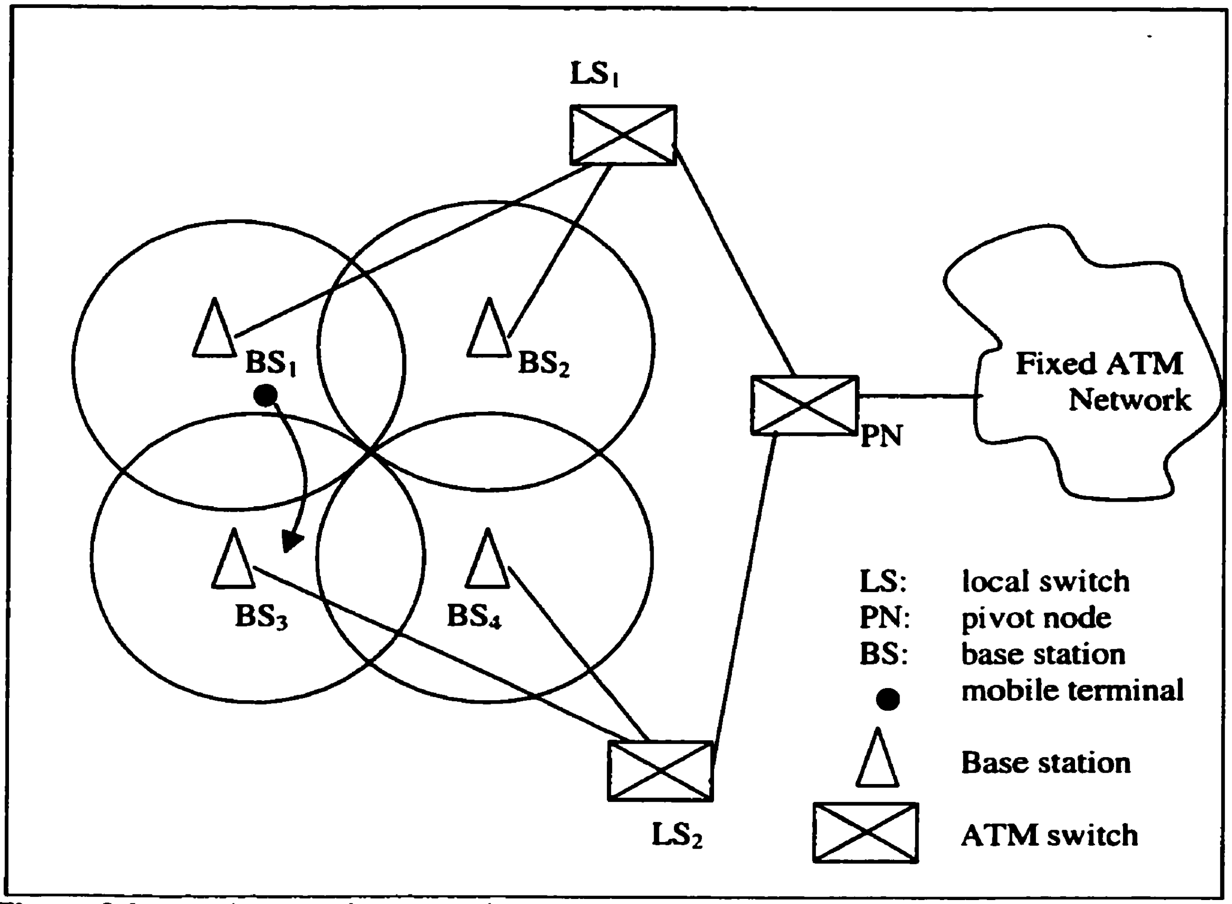

Figure 3.2 A mobile performs global handoff

As illustrated by Figure 3.3, a handoff is local when an MT moves either between BS_1 and BS_2 or between BS_3 and BS_4. The local switch is responsible for the correct update of ATM virtual circuit translation tables. This type of handoff requires minimum handoff switch involvement, that is, the pivot node is not affected.

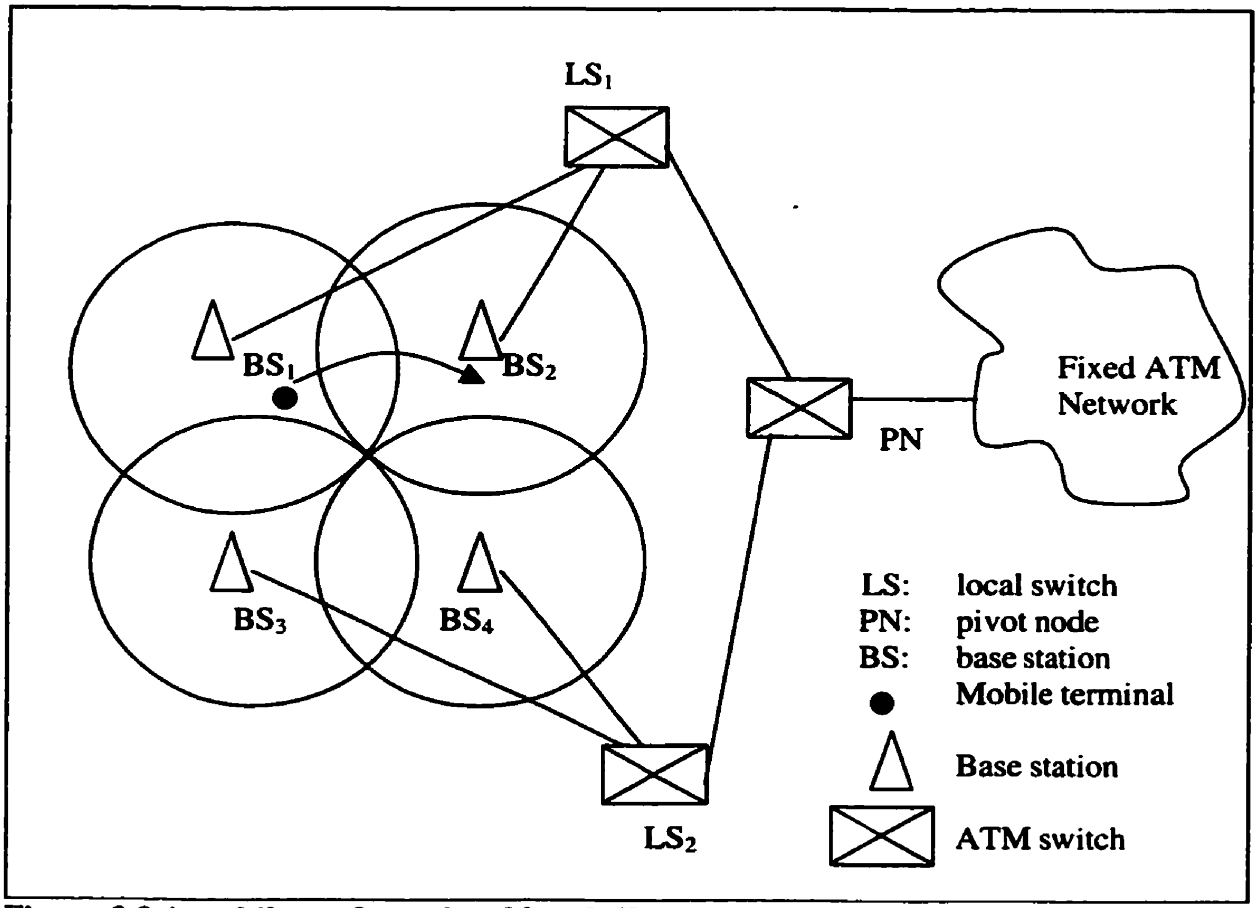

Figure 3.3 A mobile performs local handoff

The protocol stack within the fixed and wireless network segments is illustrated in Figure 3.4 [8]. Communication between an MT and a BS is performed using a radio interface that comprises the radio physical layer (RPHY), radio medium access control layer (RMAC), and radio logical link control layer (RLLC), which are globally referred to as the radio access layer (RAL). RPHY layer is used to transmit wireless data through the air. RMAC layer is used to enhance the physical layer transport capability. RMAC layer controls the wireless link according to the bandwidth explicitly required by the user as well as the amount of ATM cells input. RLLC layer is introduced to reduce the throughput performance degradation due to the bit error in a wireless link. The RAL (including RPHY, RLLC and RMAC) corresponds to physical layer (PHY) of native ATM protocol architecture. Similar to fixed ATM networks, end-to-end connections between mobile/fixed terminals are provided by the ATM adaptation layer (AAL) present in the end-user protocol stack. Conversely, the protocol stacks at BSs and ATM switches reach only the ATM layer.

Observe that, in the protocol stack, ATM cells are exchanged all the way between end-user terminals, thus implying that ATM cells are carried over the radio interface. This approach yields simple and homogeneous network architecture with end-to-end ATM cell delivery through standard ATM service access points.

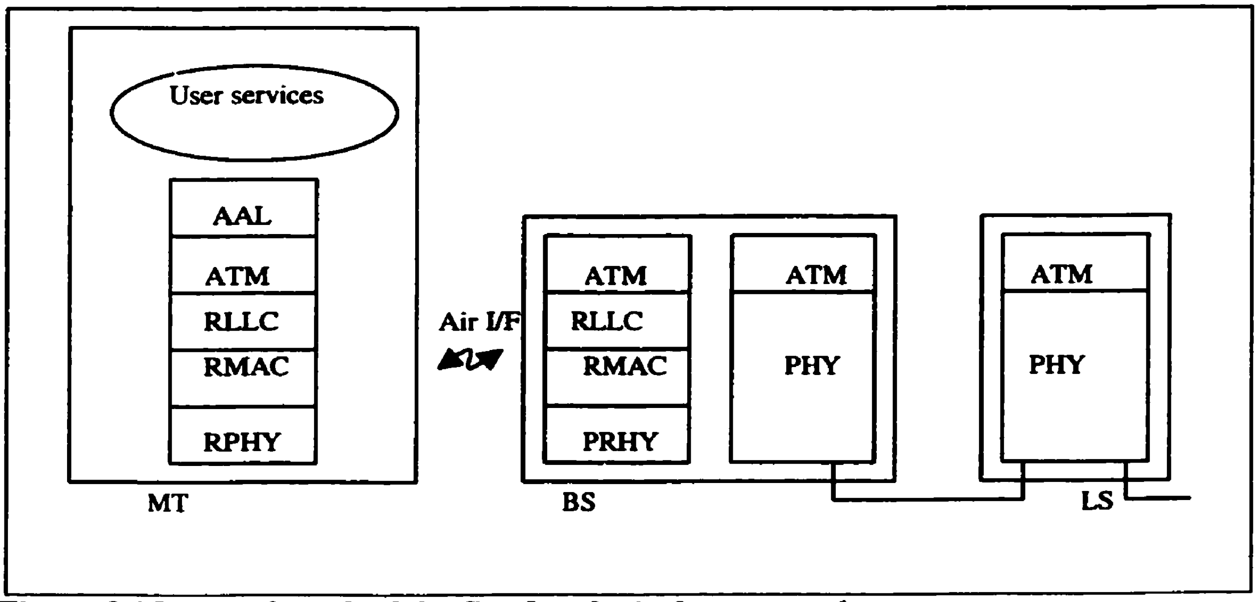

Figure 3.4 Protocol stack of the fixed and wireless network segments

3.4 Soft Handoff Like Rerouting Scheme

Handoff procedure is performed to ensure the integrity of a radio connection and to minimise interference to the users in the coverage area of neighbouring cells. In this section a new handoff scheme called Soft Handoff Like Rerouting Scheme (SHLR) is presented.

There are two levels of handoff event [2]: network and radio. The radio-level handoff is the actual transfer of the radio link between two base stations; the network-level handoff supports the radio-level handoff by performing rerouting and buffering. The radio-level handoff determines some of the procedures used in the network-level handoff, as shall be seen later. The network level handoff is performed first then followed by radio level handoff, which is described by steps 4 to 6 of the SHLR scheme step-by-step operation presented below.

The following assumptions are made:

- The handoff of each connection is carried out separately by executing same handoff procedure for each connection. Only the handoff procedure for a single connection is described here.
- Handoff is mobile initiated. The mobile monitors the link quality in terms of downlink data signal strength from the home base station it is communicating with. If the signal is lower than a pre-specified threshold, the mobile selects a destination base station that has the strongest signal strength if there is any and sends a handoff request message to the BS.
- The mobile uses the existing link to the home base station to initiate the handoff request message.

The above assumptions serve as basis for the following step-by-step operation of the proposed SHLR scheme. Refer to Figure 3.2 for global handoff scenario and Figure 3.3 for local handoff scenario that will be described here.

1. A handoff procedure starts between BS_1 and BS_3 (as illustrated in Figure 3.2). Assume BS_3 is the destination base station and BS_1 is the original base station.

2. The LS_1 first checks to see if there is a direct physical link from itself to both BS_1 and BS_3 exists. Direct physical link means there are no other ATM switches between LS_1 and BS_1 or BS_3. There are two possible outcomes to satisfy LS_1's check:

i). If there were a network connection between BS_3 and this LS_1, a new connection is established without any further network switches involvement. This is the case where local handoff is performed between BS_1 and BS_2 (as illustrated in Figure 3.3). LS_1 acts as an anchor for the connection reestablishment. Both BS_1 and BS_2 act as network connection elements for the mobile. This process is explained in detail later in step 6 of the SHLR procedure. Once the radio-level handoff is finished, BS_2 is the only end point of the mobile connection.

ii). If LS_1 and BS_3 are not connected by a direct physical link (as illustrated in Figure 3.2), then the LS_1 will forward the *handoff initiation* message to the next ATM switch along the connection path.

3. The *handoff initiation* message is forwarded along the mobile terminal's network connection. The ATM switch that receives this message uses an algorithm to check if both the LS_1 and the destination base station can be routed through its ports. The switch that meets this requirement is called the pivot node (PN). The PN uses the algorithm to check if there is any resource available at the destination base station. The resource includes the fixed part of the connection and the radio link. This thesis focuses on handoff rerouting scheme design, so the discussion of the algorithms by which ATM switches use to search the route and that the PN uses to check BS resource availability is out of the scope of the thesis. Once the pivot node is found, the handoff initiation message is not forwarded to any further switches after this point. If no resource is available, the handoff request fails and the handoff procedure ends.

4. Once the PN is found and the destination BS resource is allocated, the PN then forwards a *handoff request acknowledgement* message to LS_1 notifying the LS_1 of successful network connection rerouting. This *handoff request acknowledgement* message includes the available resource information to set-up a radio link connection to BS_3. This message will be forwarded by LS_1 to BS_1 and finally to the mobile. PN also forwards a *reroute connection request* message along the newly established path to the destination base station. Upon receiving this message, the ATM switches along the new connection set up the new connection table and forward the message all the way to the destination base station BS_3.

5. When BS_3 receives the *reroute connection request* message, it allocates a radio link for the mobile. A new connection reestablishment is complete at this point.

6. At the same time PN sets up the connection, it updates its connection table by adding the new connection without tearing down the old connection. This is a make-before-break approach. Any downlink ATM cells from the fixed ATM network along the mobile connection to this PN are duplicated to both BS_1 and BS_3. PN is also ready to receive any uplink ATM cells from the mobile through both BS_1 and BS_3 by allocating a PN buffer. The mobile, upon receiving the *handoff request acknowledgement* message, extracts the radio link information and starts to send ATM cells to both BS_1 and BS_3. At the same time, the mobile allocates a MT buffer ready to receive downlink data. For downlink data, since they are sent by PN to both BS_1 and BS_3, mobile receives duplicate data from both old connection and new connection during this period. Also, arriving ATM cells are buffered until the arrival of the corresponding duplicate cells. As long as the duplicate cell arrives, the ATM cell is ready to be removed from the buffer. On the other hand, uplink data that are sent from mobile arrive at PN via both BS_1 and BS_3. The PN buffer stores the ATM cell until its duplicate arrives. The ATM cells are then ready to be removed from the buffer and forwarded to the end user through the network connection. The difference between PN buffer and MT buffer is that in PN buffer, multiple handoff ATM cells can arrive at the same time.

This two-way multipath cell-forwarding between PN and the mobile simulates the soft handoff, where the mobile talks to both original BS and destination BS during the radio link transition phase.

7. The multipath cell-forwarding phase lasts until the new radio link is stabilised. This thesis focuses on handoff rerouting scheme and hence the discussion of the determination of the stability of the new radio link is out of the scope of this thesis. When the radio link is stabilised, the old connection is torn down by sending a *handoff complete* message through both BS_1 and BS_3 to PN from MT. The old path between $BS_1$and PN is cleared and only the new path is used for data transmission. Any data buffered at the PN buffer and MT buffer is sent out through the new path only.

3.5 Signalling Diagram for Successful Global Handoff

Section 3.4 provides a step-by-step version of handoff processing in a W-ATM network. This section discusses the signalling messages required for the handoff procedure. All the signalling messages proposed here are new messages created in the thesis. Figure 3.5 shows the signalling messages that transverse the network during handoff procedure. BS_1 represents original BS and BS_3 represents destination BS. Other nodes marked in Figure 3.5 represent possible ATM switches along the user connection. The signalling messages are used as follows:

HOI: Handoff request initiation. This message is sent by mobile to start the handoff process. The handoff process is always mobile initiated.

HRA: Handoff request acknowledgement. When PN receives a handoff request message and allocates resource at the destination base station, it sends a handoff request acknowledgement (HRA) message to the mobile through the original base station and starts duplicating downstream data to both original base station and destination base station. The HRA message tells the mobile which radio link to use at the destination base station. When the mobile receives HRA message, it sets up another radio channel to the destination base station and starts radio-level handoff by duplicating data to both original and destination base stations.

RCR: Reroute connection request. This is a message that PN sends to the destination base station. It informs destination base station to allocate a radio channel to the mobile. At the same time, PN duplicates downstream data to the destination base station. Upon reception of this message, destination base station sets up the radio link, forwards any downstream data to the mobile and receives any upstream data from mobile to the network. If the mobile receives HRA (where mobile can get the radio link information about the BS_3) later than RCR reaches BS_3, the duplicate data forwarded by BS_3 is not received by mobile.

HOC: Handoff complete. After the radio level handoff is stabilised, the mobile sends a HOC message to PN through both the new and old connection paths. At the same time it stops transmission of information data to the original base station. Mobile tears down the radio link connection to the original base station and transmits data only to destination base station. Each switch along the old connection path discards the downstream data to the mobile when forwarding this message to PN. Upon receiving this message, PN tears down both uplink and downlink connection to the original base station and forwards data through the new connection only. This message implies the end of handoff procedure.

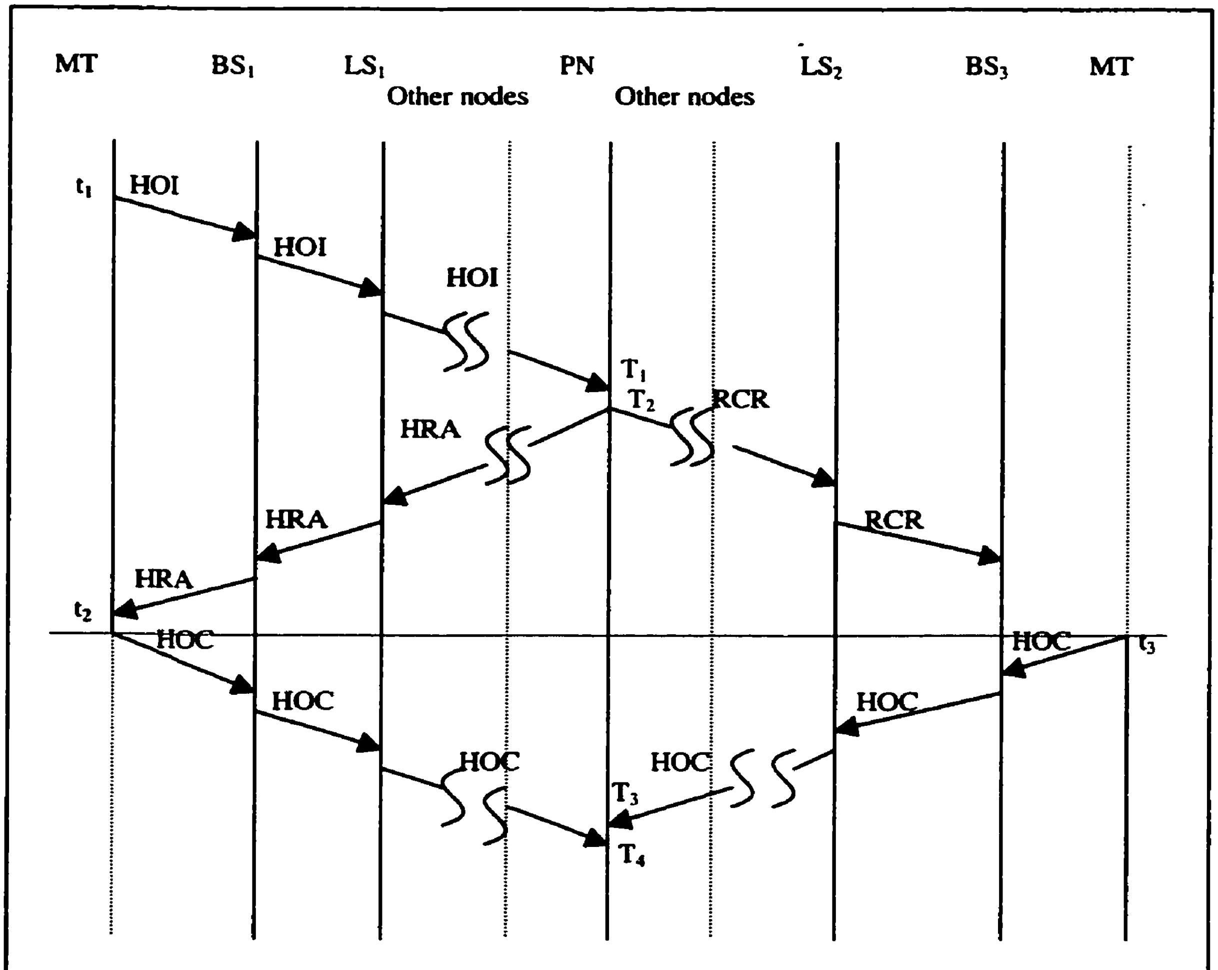

Figure 3.5 SHLR signalling messages for successful handoff

3.6 SHLR Buffer Requirements

Buffers are needed to store duplicate ATM cells at both PN and MT for uplink data and downlink data respectively. PN buffer is implemented in PN and an MT buffer is implemented in each mobile.

Since PN sends ATM cells to the mobile through both original base station and destination base station in a point-to-multiparty manner during handoff procedure, it ensures timely delivery of downstream data. This is very important for time-sensitive traffic such as voice traffic. Since the duplicate ATM cells traverse through different paths, they may arrive at the mobile at different times. In order to guarantee in-sequence delivery of ATM cells, a handoff buffer is allocated in the mobile and PN respectively. When PN duplicates data, a sequence number is added, so when mobile receives the two ATM cells, it can match them to each other. When one of the two ATM cells arrives, the buffer at the mobile stores the data and waits for the second one to arrive. It also starts a timer to wait for the duplicate cell. Time-sensitive traffic and time-insensitive traffic have different timeout limits, so that the delay requirement for time-sensitive traffic can be guaranteed. A duplicate ATM cell that arrives later than its time limit is discarded.

For the radio link part of the path, ATM cells may experience fading, shadowing and path loss. This can cause ATM cell arriving at the handoff buffer with a SNR value lower than the QoS requirement. The scheme implemented in the buffer to deal with this issue is: if the duplicate ATM cell arrives prior to the expiry of the timeout, the cell with better quality in the sense of higher signal-to-noise ratio is chosen for processing and the other one is discarded. As a result, there is a higher chance for an ATM cell to be above the QoS required SNR threshold with duplicate cell than that for a single ATM cell.

In the same way, a mobile sends duplicate ATM cells to both the original base station and destination base station. Since PN can receive data from both old path and new path, PN is responsible for combining the two streams into one. A PN handoff buffer is allocated at the PN for receiving duplicate data during handoff procedure. Duplicate ATM cells are treated the same way as in downlink handoff buffer.

From the above buffer management scheme, it can be found that the handoff procedure induced buffer occupancy is due to:

1. Radio link stabilisation duration. It is during this time that mobile and PN duplicate data are stored in the PN buffer and MT buffer respectively.

2. The different propagation time between the old path and the new path. This can be referred to as the path difference caused delay. The ATM cells are temporarily stored in the buffer before the duplicate cell arrives over the other path. PN supports ATM cells associated with multiple simultaneous handoffs. The longer the path differences the more likely the multiple handoffs data will be buffered before the duplicate cell arrives. Hence, more buffer size needed. For MT buffer, at most ATM cells of one connection need to be buffered, so the buffer size needed is relatively small.

3. Timeout. If one of the duplicate ATM cells never arrives in the buffer, then when timeout timer expires, the ATM cell waiting in the buffer is sent. This can happen when the mobile sends HOC message to the original BS to disconnect the old path. Since PN may duplicate downlink ATM cells to the original BS before the HOC message from the original BS arriving PN, the ATM cells sent through this path is discarded and can never reach the mobile.

The total buffer delay is the result of a) the above mentioned handoff procedure caused delay. ATM cells need to wait for the duplicate even though the network transmitter is ready; b) and queuing delay which is due to the busy status of the network transmitter even though the ATM cells are ready.

Soft handoff can cause buffer delay because of path difference during radio link handoff. However, it is important to note that handoff is executed when mobile is moving from one geographical cell into another. As a consequence, the original link can become poorer. With this fact considered in SHLR protocol, the duplicating data scheme that is specified in step 6 of section 3.4 and the buffer management scheme that is stated in this section are implemented to improve the radio link quality.

In any ATM network, the sequence of individual ATM cells must be preserved for correct reassembly of encapsulated user information. Therefore, during a handoff, the preservation for cell sequence is a primary concern in the network. Since in SHLR data stream pass through both old connection and new connection and converge at the network buffer, cells are delivered through the network in sequence for both mobile and the end user.

In the following two sections a handoff scheme proposed by Ajmone-Marsan et al. [8] is introduced and buffer requirement for the scheme is discussed. The goal is to compare the previously proposed algorithm with what is proposed in this thesis.

3.7 Hard Handoff Like Rerouting Scheme

The handoff protocol proposed by Ajmone-Marsan et al. [8] uses break-before-make idea for radio link handoff which is similar to hard handoff, hence the scheme is thus referred to as Hard Handoff Like Rerouting (HHLR) scheme in this thesis.

The following is the signalling diagram and signalling messages of HHLR scheme provided by reference [8]. Other nodes marked in Figure 3.6 represent possible ATM switches along the user connection.

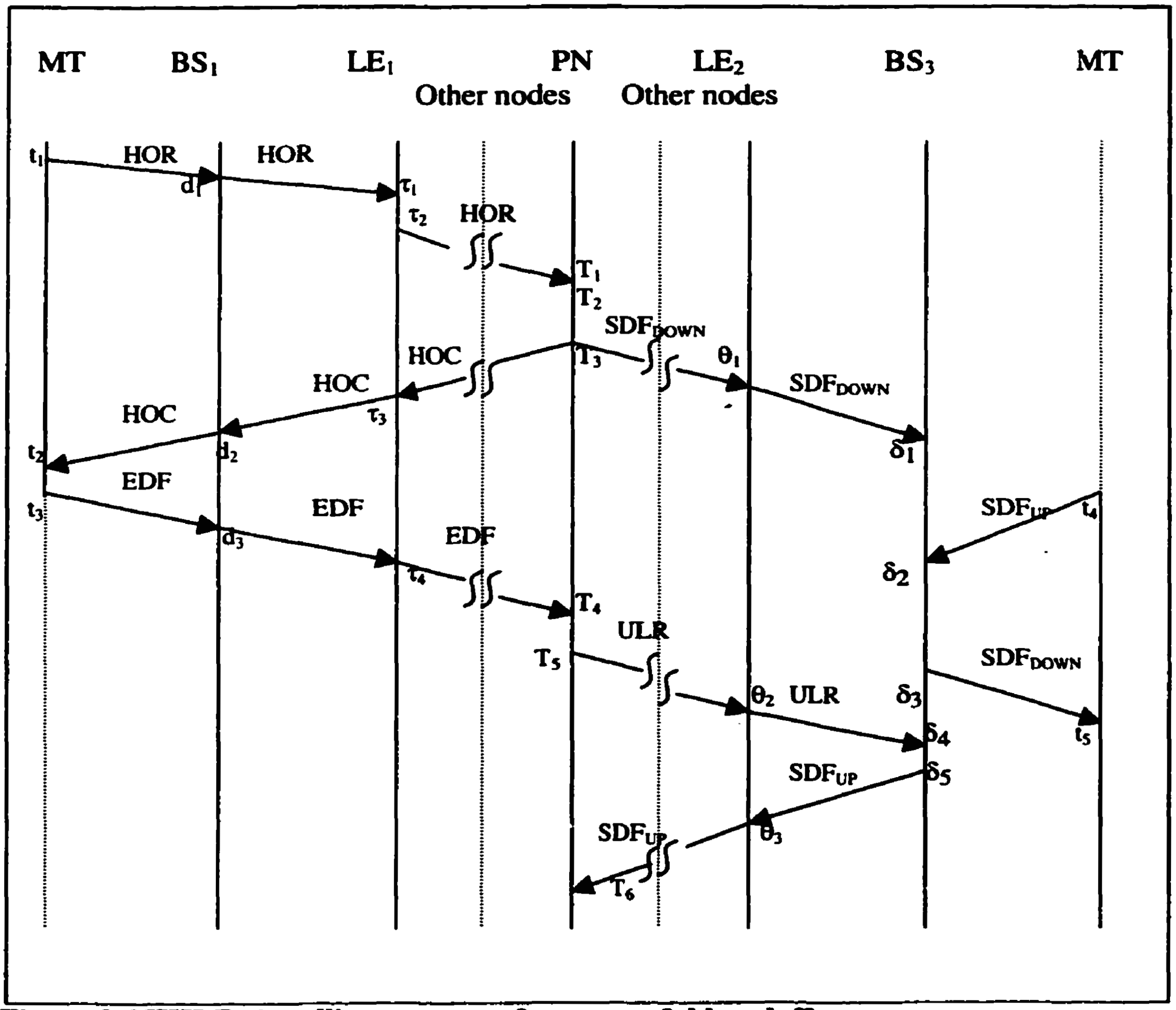

Figure 3.6 HHLR signalling messages for successful handoff

HOR: Handoff request. Handoff is mobile terminal initiated. The mobile terminal monitors the wireless link quality in terms of received signal power. If the source base station's signal is less than the handoff threshold, the mobile terminal searches for the base station that has strongest signal power as destination base station. The mobile terminal sends the handoff request to the network with the selected base station identifier. While waiting for the network to respond, the mobile terminal is still sending data stream to the end user through the original base station.

HOC: Handoff confirm. This message acknowledges that the network is ready to switch the connection to the mobile terminal from the source base station to the destination base station. The handoff request message traverses the network from the nearest ATM switch to the end user for the mobile user connection. Each of the network switches on this path, upon receiving this message, checks to see if there is a connection to the destination base station. If there is a connection to that base station and if the resource (radio channel) is available at the destination base station, this network switch establishes a new path to the destination base station BS_3 for both uplink and downlink connections. The network switch then sends a HOC· message to notify the mobile terminal to switch to BS_3. Any downlink data sent by the end user arriving at this switch are forwarded through the new path to BS_3.

SDF_{DOWN}: Start data flow on the new downlink connection. This message is sent by the PN on the incremental connection established to BS_3 before any downstream data. When the SDF_{DOWN} message reaches BS_3, if the mobile terminal has not completed the radio handoff yet, all downlink data are stored at the BS_3 in the downstream handoff buffer.

EDF: End of data flow on the uplink connection through the source base station. This is the message sent from mobile to the network telling the radio link handoff is taking place. When mobile receives the HOC message, mobile terminates the upstream transmission to BS_1 by sending the EDF message. EDF is sent after the last ATM cell sent from the mobile. This message informs the PN that no more upstream data cells will arrive through the original base station so that PN can update uplink data stream routing table.

ULR: New uplink connection ready. This message is sent by the PN to enable the uplink connection of the incremental part. Upon receiving EDF sent from mobile, PN updates the routing table and sends the ULR message to BS_3 indicating that the upstream transmission through BS_3 is enabled.

SDF_{UP}: Start data flow on the new uplink connection. This message is originated from the mobile after the completion of radio link handoff. After radio link handoff is completed, mobile terminal will resume sending uplink data through BS_3. When BS_3

receives this message, it forwards the downstream data stored in the downstream handoff buffer to the mobile with the SDF_{DOWN} message as the very first prior to sending any data ATM cells. While waiting for the ULR message from PN indicating the uplink connection ready, BS_3 buffers all uplink data messages sent from the mobile. Upon receiving of the ULR message from the PN, BS_3 forwards the upstream data stored in the upstream handoff buffer to the network with the SDF_{UP} message as the very first of any data ATM cells. This message also means the completion of the handoff procedure.

3.8 HHLR Scheme Buffer Requirements

Handoff procedure induced buffer delay (or handoff delay) for a mobile connection starts from the time ATM cells are stored in the buffer and ends when the ATM cells are emptied from the buffer for HHLR scheme [8].

A time sequence diagram of the handoff events can be drawn to analyse the buffer delay. Figure 3.7 shows the time sequence diagram for downlink buffer delay analysis. Each of the time stamp shown in Figure 3.7 and Figure 3.8 can be referenced to time stamps marked in Figure 3.6.

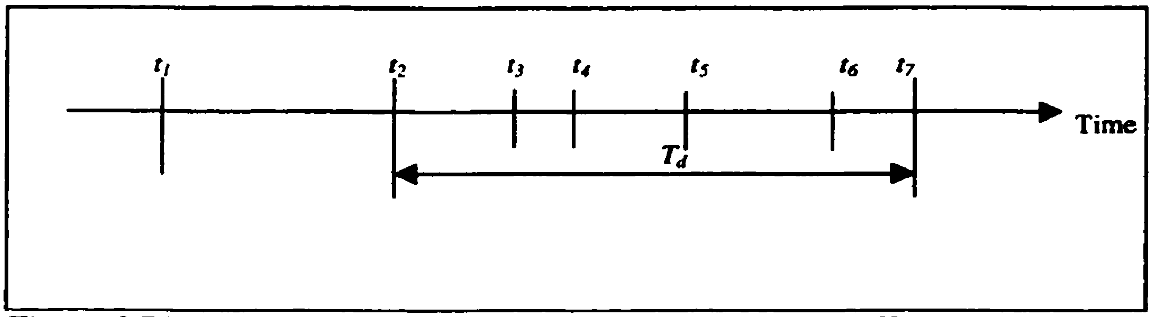

Figure 3.7 Time sequence diagram for HHLR downlink buffer analysis

Figure 3.7 illustrates the times when handoff messages are received or sent through the network for downlink buffer analysis. T_d represents downlink buffer duration.

t_1 is the time when PN sends HOC to originating BS and SDF_{DOWN} to destination BS. t_2 is the time when SDF_{DOWN} message is received by BS_3 and t_3 is the time when BS_1 receives HOC message. t_4 is the time when the mobile sends a EDF message through BS_1 and t_5 is the time when the mobile sends a SDF_{UP} to BS_3. t_6 is the time when BS_2 receives the SDF_{UP} message and t_7 is the time when BS_3 ready to forward downlink ATM cells for the mobile.

Since downlink data are sent through new path from PN to BS_3 right after SDF_{DOWN} message, BS_3 gets this message at time t_2. Downlink buffer at BS_3 starts to store downlink data for the mobile in handoff at this time. BS_3 then waits for the SDF_{UP} message sent from mobile after radio link handoff is completed. After receiving SDF_{UP} at time t_6, BS_3 updates the mobile state to ready-to-send-downlink-data and starts to empty buffer for this mobile at t_7. Time elapsed from t_2 to t_7 is the buffer delay of downlink buffer.

Equation (3.1) shows the relation between the handoff process time intervals and downlink buffer duration. Let $[x_1, x_2]$ represents the time interval between x_1 and x_2. Downlink buffer delay can be expressed as:

$$T_d = [t_2, t_7] = [t_1, t_3] + [t_3, t_4] + [t_4, t_5] + [t_5, t_6] + [t_6, t_7] - [t_1, t_2] \quad (3.1)$$

There are six time intervals on the right hand side of Equation (3.1). The first and the sixth terms are propagation delay or path delay experienced by the handoff messages. The second, third and forth terms are radio link handoff delay. The fifth terms is BS_3 processing delay. Downlink buffer duration due to handoff procedure is caused by

a) Difference of propagation delay (path difference) between old and new path: $[t_1, t_3]-[t_1, t_2]$.
b) Radio link handoff delay.
c) Processing delay.

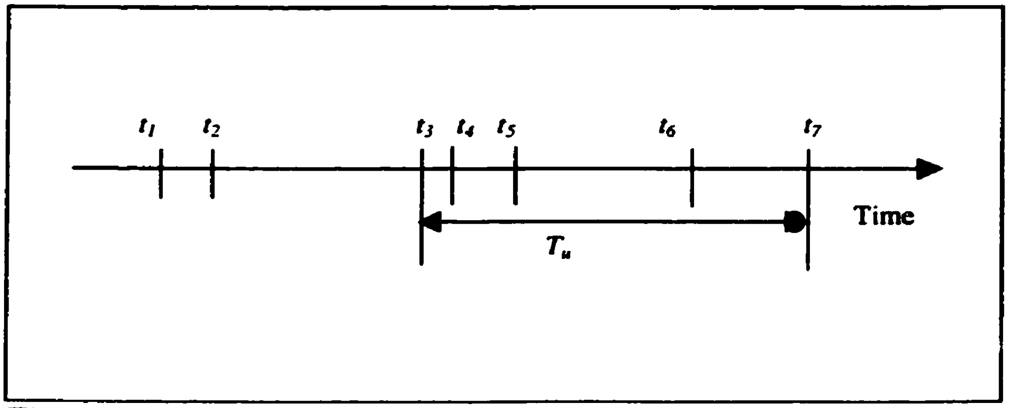

Figure 3.8 Time sequence diagram for HHLR uplink buffer analysis

Figure 3.8 illustrates the times when handoff messages are received or sent through the network for uplink buffer analysis. T_u represents the uplink buffer duration. t_1 is the time when mobile sends EDF message to $BS_1$and t_2 is the time when mobile sends SDF_{UP} message to BS_3. t_4 is the time when PN receives EDF message and t_5 is the time when PN sends ULR message to BS_3 after the uplink routing table is updated. t_3 is the time when BS_3 receives the SDF_{UP} message. Upon receiving the ULR message at time t_6, BS_3 updates the MT status and forwards the SDF_{UP} message to PN at time t_7.

If the uplink is not ready, any data sent from mobile after SDF_{UP} (t_3) are stored in uplink buffer until the ULR message is received at time t_6. BS_3 updates the mobile state to ready-to-send-uplink-data and starts to empty uplink data from uplink buffer at t_7. So, time elapsed from t_3 to t_7 is the buffer delay of uplink buffer.

Equation (3.2) shows relation between the handoff processing time intervals and uplink buffer. Let T_u denotes uplink buffer delay and $[x_1, x_2]$ represents time interval between x_1 and x_2. Uplink buffer delay can be expressed as:

$$T_u = [t_3, t_7]= [t_1, t_4] + [t_4, t_5] + [t_5, t_6]+ [t_6, t_7]-[t_1, t_2]-[t_2, t_3] \quad (3.2)$$

There are six time intervals on the right hand side of the equation (3.2). The first and the third terms are propagation delay or path delay that experienced by handoff messages. The fifth and sixth terms are radio link handoff delay. The second and forth terms are BS_3 processing delay.

Uplink buffer delay is mainly due to

a) Addition of propagation delay (path addition) of old and new path: $[t_1-t_4]+[t_5-t_6]$.

b) Difference between the delay of item a) and the radio link handoff delay.

c) BS_3 processing delay.

Since the radio link transmission speed is at the speed of light ($3*10^8$ m/sec) and the BS switch processing speed is in the range of hundreds of megabits, both b) and c) caused delay are negligible [8]. The total buffer delay is the result of a) The handoff procedure execution delay when the ATM cells have to wait even though the network transmitter is ready; b) and queuing delay which is due to the busy status of the network transmitter even though the ATM cells are ready. In HHLR scheme, ATM cells need to be buffered during handoff procedure, from t_2 to t_7 for downlink buffer and from t_3 to t_7 for uplink buffer. For the connections that have time-sensitive traffic, the requirement of timely delivery ATM cells is very stringent. How to minimise the buffer delay during handoff procedure becomes a key issue in the design handoff of protocol.

The simulation and analytical analysis results of the performance of SHLR and HHLR schemes are presented in chapter 5. In the following section, several previously proposed handoff schemes are overviewed and compared with the SHLR scheme.

3.9 Comparison of SHLR Scheme with Several Proposed Schemes in the Literature

SHLR scheme enhances the current ATM signalling protocols by adding support for rerouting an active connection. In the current version of ATM signalling protocols there is no provision for rerouting a connection once it is established. By using SHLR scheme, multiple connections between a mobile and end users may be rerouted seamlessly, without re-establishing the connection during handoff procedure.

3.9.1 Comparison of SHLR Scheme with HHLR scheme

SHLR scheme is an improvement over HHLR scheme. The key change is the way of establishing new route while handoff procedure is performed. In HHLR scheme, the mobile first tears down the radio link to the old base station, then performs radio link handoff and establishes the new radio link with the destination base station. In SHLR scheme, old connection and radio link between PN and old base station for both uplink and downlink data streams are kept to transmit data until the radio link handoff is stabilised.

There are two cases that can cause low signal to noise ratio and even data packet loss: i) in wireless environment, data signal may experience fading, path loss, etc; ii) during handoff, a mobile can be in a situation when its signal is getting poorer, and because of the mobility feature, mobile may move back and forth between the two base stations. The beauty of the SHLR handoff scheme is the improvement in radio link quality through duplication of data to both the old and new connections and by comparing them at the buffer to get better quality data, hence the better quality user connection is guaranteed.

From the buffer requirement analysis of HHLR scheme, the uplink buffer delay (τ_u) due to handoff procedure is mainly due to the addition of the propagation delay between the old and new connections. In SHLR scheme, buffer delay due to handoff procedure relates only to the propagation delay difference between the old and new connections. SHLR scheme lowers the buffer occupancy and hence the cell loss ratio and mean cell queuing delay.

3.9.2 Comparison of SHLR Scheme with NCNR Scheme

3.9.2.1 Introduction to NCNR

The Nearest Common Node Rerouting (NCNR) scheme was proposed by Akyol et al. in [2]. The authors of [2] have utilized the zone concept to illustrate the NCNR algorithm. NCNR is based on finding the ATM node that is a root of or common to both of the zones involved in the handoff transaction. The rerouting is then performed starting from this nearest common node (NCN). Time-sensitive traffic and time-insensitive traffic have different handoff procedures. For time-sensitive traffic, network establishes a new connection from NCN to the radio port (base station). Both old connection and new connection are kept to forward ATM cells during handoff procedure. The old connection is not cleared until the radio link is stabilized. For time-insensitive traffic (i.e., throughput-dependent traffic), during radio link handoff, downlink information ATM cells are buffered at the ATM node connecting the old radio port and new radio port for downlink data. Uplink ATM cell are buffered at the mobile. If old BS buffer is non-empty before the handoff is started, then old BS buffer is transmitted to the mobile if possible. No information data is transmitted in the downlink direction until the radio level handoff is completed. Upon completion, the ATM cells stored in the old BS buffer are transmitted to the new BS and go in front of all other cells buffered for transmission so as to preserve ATM cell sequence.

3.9.2.2 Comparison

The NCNR scheme uses different algorithms for time-sensitive and throughput dependent traffic to meet the QoS requirements for both traffic types. However, using different algorithms for different traffic types adds up complexity to the implementation of the protocol. In SHLR scheme, the only difference for time-sensitive and throughput-dependent traffic is the different time constraints at the buffer

As stated in Reference [2], throughput-dependent traffic cannot tolerate cell loss. In NCNR, before the radio link handoff, buffered ATM cells at the old BS are transmitted to the mobile. Since the radio link for the old connection is already poor, transmission of cells stored in old BS buffer may cause loss of ATM cells. In SHLR during handoff, throughput-dependent traffic is duplicated and a better cell is selected by the PN, resulting in less cell loss.

In NCNR, after radio level handoff, downlink cells stored in old BS buffer have to be forwarded back to NCN, then through the new connection to the mobile before those ATM cells buffered in the new BS can be transmitted. This cell back-forwarding procedure uses up network resource and can cause additional delay to the ATM cells that are buffered at the new BS. The back-forwarding procedure is eliminated in SHLR resulting in less cell queuing delay.

The good point of NCNR is that time-sensitive traffic is duplicated during radio link handoff to deliver ATM cells in a timely manner. Unfortunately, these duplicated data in NCN are not compared to combat poor radio link transmission. In contrast, SHLR makes use of these duplicated data so that there is a higher chance for SHLR to have more ATM cell above the required SNR threshold and hence improves the link quality.

3.9.3 Comparison of SHLR scheme to BAHAMA Scheme

3.9.3.1 Introduction to BAHAMA

The BAHAMA [34] handoff scheme is proposed for a wireless ATM local area network (LAN). The BAHAMA architecture consists of a flat network of base stations and user terminals. The base stations are interconnected with ATM links. The BAHAMA architecture is also based on forwarding of cells after a successful handoff. The initial base station (same as ATM node) acts as an anchor for the handoff connection and forwards the user cells to the destination base station. After the handoff is completed, the initial base station migrates the user connection to an optimal route in the network if the mobile stays within the coverage area of the destination base station for an extended period of time. Since the ATM cells are always routed through the initial ATM node during handoff, the cell sequence is preserved. The BAHAMA LAN uses the virtual path indicator of the ATM cell header for routing. This simplifies the rerouting since all that needs to be changed in the cell header is the virtual path indicator provided that the virtual connection indicator is available at the destination base station.

3.9.3.2 Comparison

BAHAMA rerouting scheme uses cell-forwarding scheme to keep the ATM cells transmission in sequence. When the ATM switch associated with the previous base station acts as an anchor for the handoff, all the ATM cells still go through the previously established connection before the extended connection is established. This guarantees the preservation of cell sequence. For a fast-moving user in a system that has small radio coverage areas, there will be a network trail left behind the mobile, so cell-forwarding through extended connection wastes the network bandwidth. In the SHLR scheme, the connection is simply rerouted from PN to the new ATM node of the wireless network. This minimises the bandwidth usage by eliminating the part of the forwarding connection between the PN and the old base station.

To perform BAHAMA, either the anchor ATM switch or the new added ATM switch can buffer the ATM cells during the handoff procedure. In SHLR, there is no buffering delay hence the total amount of ATM cells buffered can be decreased.

BAHAMA rerouting scheme does not support the transmission of uplink and downlink ATM cells through both old and new base stations during radio level handoff. This may cause a problem for time-sensitive traffic data streams in terms of delay experienced during handoff.

As a handoff algorithm for wireless access networks, BAHAMA does not consider shadowing loss. In SHLR scheme, algorithm is implemented to make use of duplicated packets and improve radio link quality during handoff.

BAHAMA scheme assumes a flat (or ring) network model where all neighbouring ATM switches will have direct connection in between. This approach, while suitable for a

LAN, will not scale well for a wider-area network. SHLR scheme is developed with wide area networks (WANs) in mind and does not have such a problem.

In this chapter, the SHLR scheme is presented in detail with the step-by-step operation provided. A related scheme called HHLR is also introduced in great detail. What is the performance of these two schemes in terms of cell loss ratio and mean cell queuing delay? In the SHLR scheme, PN buffer handles handoff of a larger geographical area than both uplink and downlink buffers in HHLR scheme. Does SHLR scheme outperform HHLR scheme? The next chapter provides the simulation analysis for both SHLR and HHLR schemes.

Chapter 4
W-ATM Network Simulation Models

This chapter presents models for analyzing SHLR performance using simulation technique. Analytical techniques, computer simulation, projections from existing experience, and experimentation are the available methods that can be used to evaluate and compare network designs and protocols. Unlike analytical models, which often require many assumptions and are too restrictive for most real-world systems, simulation modelling places few restrictions on the classes of systems under study. The execution of a computer simulation model is comparable to conducting an in-vitro experiment on the target system [22].

In this chapter the design of network simulation model is introduced. Performance analysis of SHLR scheme is built on this simulation model. The parameters of interest are cell loss ratio and mean cell queuing delay. The remainder of this chapter is organized as follows: in section 4.1 the simulation modelling tool is presented, followed by a discussion of network modelling in section 4.2. Simulation network design forms the topic of section 4.3 and followed by simulation network verification of section 4.4. Finally, the simulation scenarios are introduced in section 4.5.

4.1 Simulation Modelling Tool

From a high-level perspective, telecommunications networks can be seen as users who generate demands for network resources, and protocols that control the allocation of network resources to satisfy those demands. The generation of user demands and their satisfaction are encapsulated in simulation events, which are ordered by their time of occurrence. The action of the protocols depends on the state of the network at the time the demand was issued. This event-based processing is referred to as discrete-event simulation [22]. The SHLR simulation programs used in the thesis are implemented with OPtimized Network Engineering Tools (OPNET) developed by OPNET Technologies, Inc. [27]. Both behaviour and performance of modelled systems can be analysed by performing discrete event simulations. OPNET models are hierarchical: network models, node models and process models. Network models define the position and interconnection of communicating entities, or nodes. Network models are made up of subnets and node models. Node models are objects in a network model. Node models typically depict the interrelation of processes, protocols, and subsystems. Node models are made up of modules with process models. Process models control module behaviour. Process model combines the graphical power of a state-transition diagram with the flexibility of a standard programming language (e.g. C) and a broad library of pre-defined modelling functions.

4.2 Simulation Network Modelling

The purpose of the simulation is to evaluate the performance of SHLR scheme The SHLR results are then compared with those of HHLR scheme.

In SHLR scheme, uplink ATM cells of multiple MT connections may be buffered in the PN buffer at the same time, but on the downlink direction the MT buffer in the mobile terminal only stores ATM cells for one mobile terminal. Hence PN buffer or uplink buffer is more critical than MT buffer. To be comparable, uplink buffer is simulated for both SHLR and HHLR schemes. The rest of section 4.2 presents the idea of the network modelling by showing the simulation models that play key role in the simulation.

4.2.1 Network Level Modelling

Figure 4.1 shows the simulated wireless ATM network. There are 40 mobiles, 4 base stations and 1 pivot node. Each BS covers a hexagon shaped area with radius 4km. For the SHLR scheme a PN buffer is implemented in the PN. For the HHLR scheme, an uplink buffer is implemented in each BS. All 4 uplink buffers are observed during the simulation. The links between BSs and PN are optical links with speed 150 Mbps. The path difference in SHLR scheme for the duplicated ATM cells is 0.17 msec. The value 0.17 msec is chosen so that the 2 identical ATM cells from different paths arrive before next ATM cell arrives in buffer, which is one of the situations that can happen during handoffs. The peak data rate for each the mobile is designed to be 2.0 Mbps, this means the inter-arrival time between 2 consecutive ATM cells is about 0.192 msec. For HHLR scheme the uplink buffer duration (τ_u) is mainly due to the propagation delay of the sum of the old path and new path. This delay should be longer than that in SHLR scheme, and is set to 0.36 msec. Timeout duration in SHLR scheme is assumed to be 0.195 msec for all mobiles. The timeout value is set so that the duplicated ATM cells always arrive before the timeout timer expires and the two ATM cells can be used to improve radio link quality during handoff.

The simulation developed in this thesis runs local handoff scenarios. As far as the handoff is concerned, the difference between local handoff and global handoff is the processing delay of the switches and the propagation delay of the links. Global handoff experiences longer processing delay and propagation delay. Hence the results generated from different simulation scenarios in the thesis can be applied to global handoff scenario.

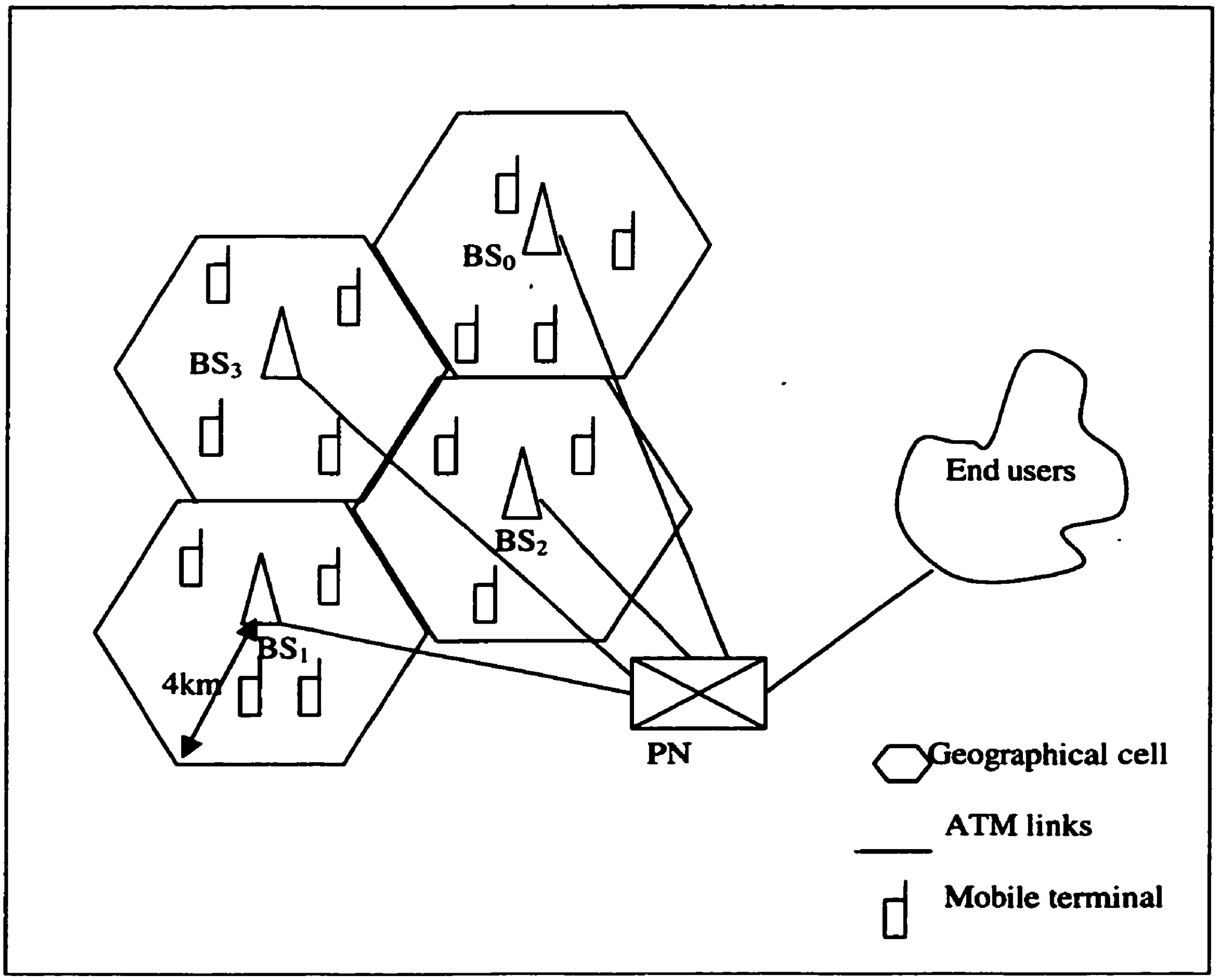

Figure 4.1 Network level modelling

4.2.2 Node Level Modelling

Figure 4.2 shows how PN is modelled. The nodes labelled *pt_x* and *pr_x* are point-to-point transmitters and receivers respectively and are linked to each base station. A pair of *pt_x* and *pr_x* represent duplex link between BS_x and PN, where $x = 0, 1, 2, 3$, the BS index. The node labelled *channel_mgr* represents PN functionality. The node *channel_mgr* manages data, signalling messages and channels. The node labelled *phone_sys* represents end users in the network. It receives data from and generates data to each of the mobile through *channel_mgr* node. End user data from *phone_sys* are switched to different links according to the mobile ID inserted in the header of each ATM cell. Node *channel_mgr* knows which base station the mobile is communicating with and it forwards the data to the BS. There are 3 kinds of data streams (shown as arrow lines) for data or messages communicating between *channel_mgr* and each transmitter or receiver: 1) stream1: information data steam; 2) stream 2: handoff message, location update and call connection clean up messages stream; and 3) stream 3: call connection set up messages stream. Data or message sent from one node triggers an interrupt to another node as shown in Figure 4.2.

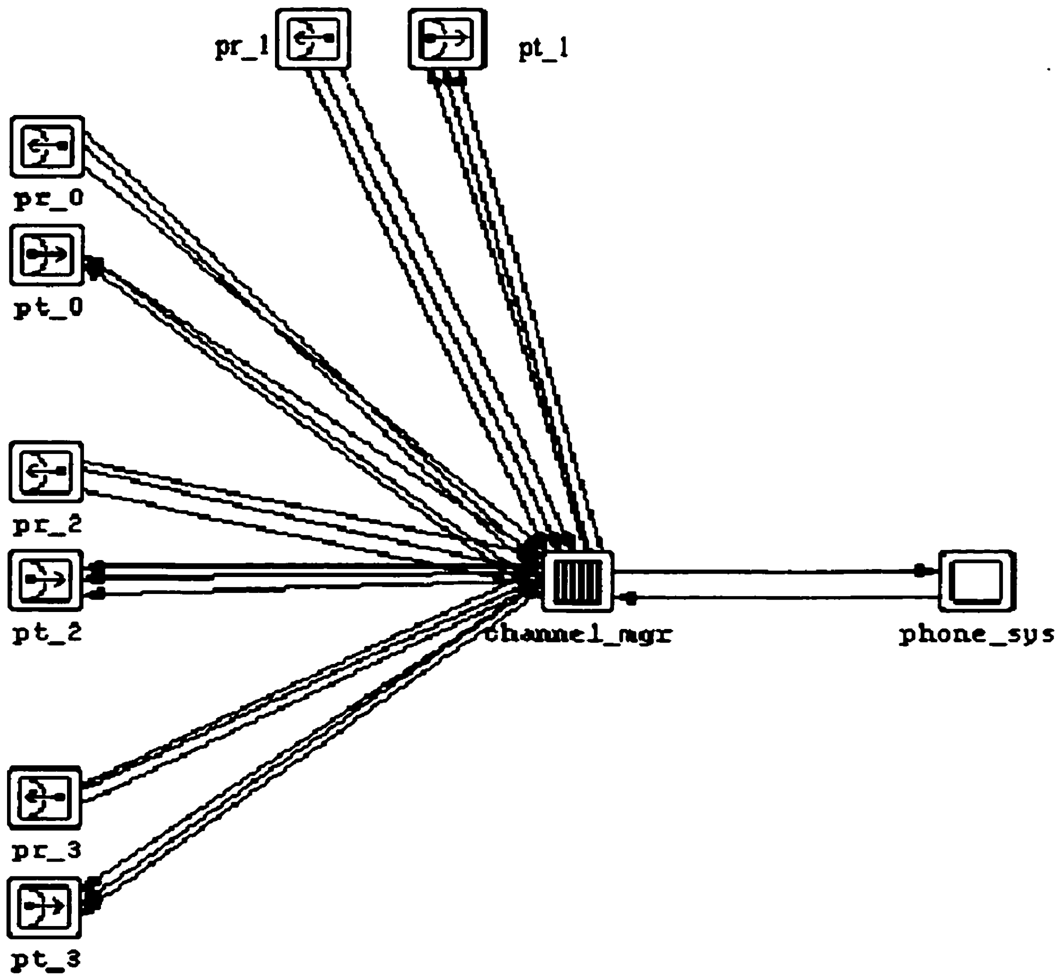

Figure 4.2 PN node level modelling

Figure 4.3 shows the base station node model. The node labelled *a_0* is an antenna through which a BS transmits and receives data from the radio link. The nodes labelled *rr_x* and *rt_x* are radio transmitters and receivers respectively, where $x = 0, 1, 2, 3$. The nodes *rr_0* and *rt_0* are used for receiving and transmitting data during MT call set up event. The node *rr_1* and *rt_1* are used for receiving and transmitting signalling messages associated with handoff, location update and call connection tear down events. The nodes *rr_2* and *rt_2* are used for receiving and transmitting information data generated from mobiles and end users respectively. Each arrow line between node *rr_2* and node *call_proc_up* represents on uplink channel and similarly each link between node *call_proc_down* and node *rt_2* represents one downlink channel respectively. The node labelled *bis_set* is used to probe the busy and idle state of the access channel. The node labelled *locator* takes care of generating location update messages that are broadcasted to the mobiles. Uplink and downlink data are processed in node labelled *call_proc_up* and *call_proc_down* respectively. Signalling messages including call connection set up and termination, location update and handoff are processed in *handoff* node. If necessary actions (e.g. add mobile to the mobile list after a channel is assigned to the mobile) need to be taken, the node *handoff* communicates with *call_proc_up* and *call_proc_down* as represented by arrow lines between handoff and *call_proc_up/ call_proc_down* nodes. Finally, nodes *pt_0* and *pr_0* are transmitter and receiver for BS to communicate with the PN respectively.

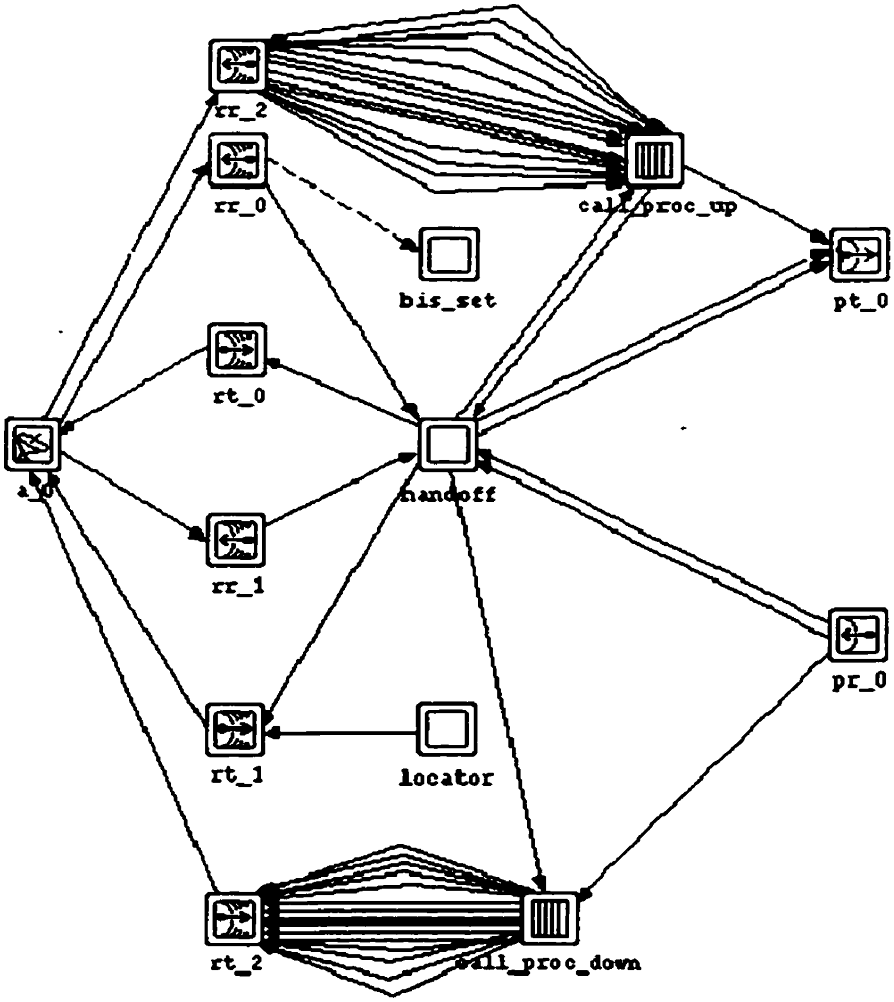

Figure 4.3 BS node level modelling

Figure 4.4 shows the node model for MT in the SHLR scheme. The node labelled *mob_gen* is responsible for mobility control. The node labelled *call_gen* is responsible for call generation. The node labelled *cbr_gen* takes care of traffic generation: CBR or VBR traffic. The node labelled *call_server* is responsible for signalling message handling. It talks to node *cbr_gen*, *call_proc_up*, *call_proc_down* corresponding to the messages received. The nodes *rt_0* and *rr_0* are transmitter and receiver for signalling messages respectively. One arrow line between node *call_server* and *rt_0* forwards call set up message and the other one forwards the remaining signalling messages. The nodes *rt_1* and *rr_1* are transmitter and receiver for information data processing. The node labelled *call_proc_up* takes care of ATM cells sent from traffic generator. It forwards data by using one of the arrow lines between node *call_proc_up* and *rt_1*, or duplicates (during handoff) data by using both of the arrow lines between node *call_proc_up* and *rt_1* to the radio transmitter *rt_1*. The node labelled *call_proc_down* takes care of ATM cells sent from end user. The node *sink* is the place where all received data are discarded, which simulates the end of the path for the downlink data. The node *resp_filter* receives and checks the signalling messages and forwards them to *call_server*. The node labelled *a_0* is an antenna through which a BS transmits and receives data from the radio link.

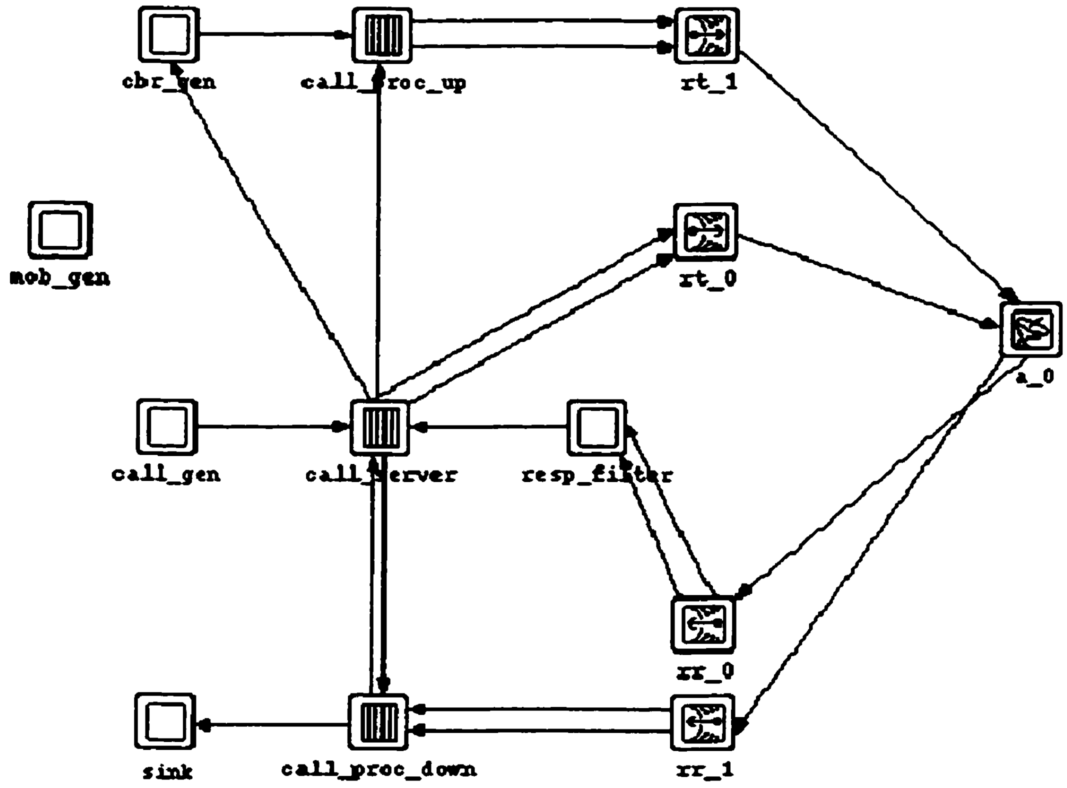

Figure 4.4 MT node level modelling

4.2.3 Process Level Modelling

There is a process model that defines the behaviour for each of the nodes (except transmitters, receivers and antennas) in node level modelling. To represent the generic functions of process level modelling, two of the process models are chosen from those designed for SHLR and HHLR schemes simulation.

Figure 4.5 shows the process model for node *channel_mgr* in PN node level modelling. Each circle (black or white) represents a state. An interrupt can trigger the execution of the state transition process. C code is embedded in each state. The black coloured circle represents a *forced* state, which upon entry, the code in it is executed and then the state is exited after the execution. The white coloured circle represents an *unforced* state. Upon an interrupt, the code in *unforced* state is first executed. If one of the transition conditions (dashed arrow lines to the *forced* states) is satisfied, the process transits to the *forced* state. The code in the satisfied *forced* state is executed and the process returns to the *unforced* state.

Different kinds of messages (call set up, call termination, handoff, data traffic) are processed in different states. State *idle* is an *unforced* state waiting for interrupts to happen. If the interrupt is from stream 1 (as stated under the description for PN node level design), meaning information data from the mobile are received (TRA_UP_IN), the process transits to *tra_up* state to forward data to the end user (node *phone_sys* in PN node level modelling) or buffer the data (during handoff). If the interrupt is from *phone_sys*, meaning information data from the end user are received, the process transits to *tra_down* (TRA_DOWN_IN) state to forward or duplicate (during handoff) data to the mobile. If the interrupt is from stream 2 (as stated under the description for PN node level design) and the message is a handoff message (VOICE_IN), the process transits to *voice* state. In *voice* state, the process allocates a new a channel if the message type is a handoff request message or releases a channel if the message type is a radio link ready message, etc. If the interrupt is from stream 2 and the message is for call connection tear down, the process transits to *call_te* (CALL_TERM_IN) state to release a channel. If no other transition condition meets, *default* transition is executed. If the interrupt is from stream 3 (as stated under the description for PN node level design), meaning the message is for call set up (DATA_IN), the process transits to *data* state and allocate a channel for the mobile user if there is a channel available. While in *idle* state, the simulation termination condition (END_SIM_IN) is checked and if true the interested simulation data are collected by executing function *record_data()*.

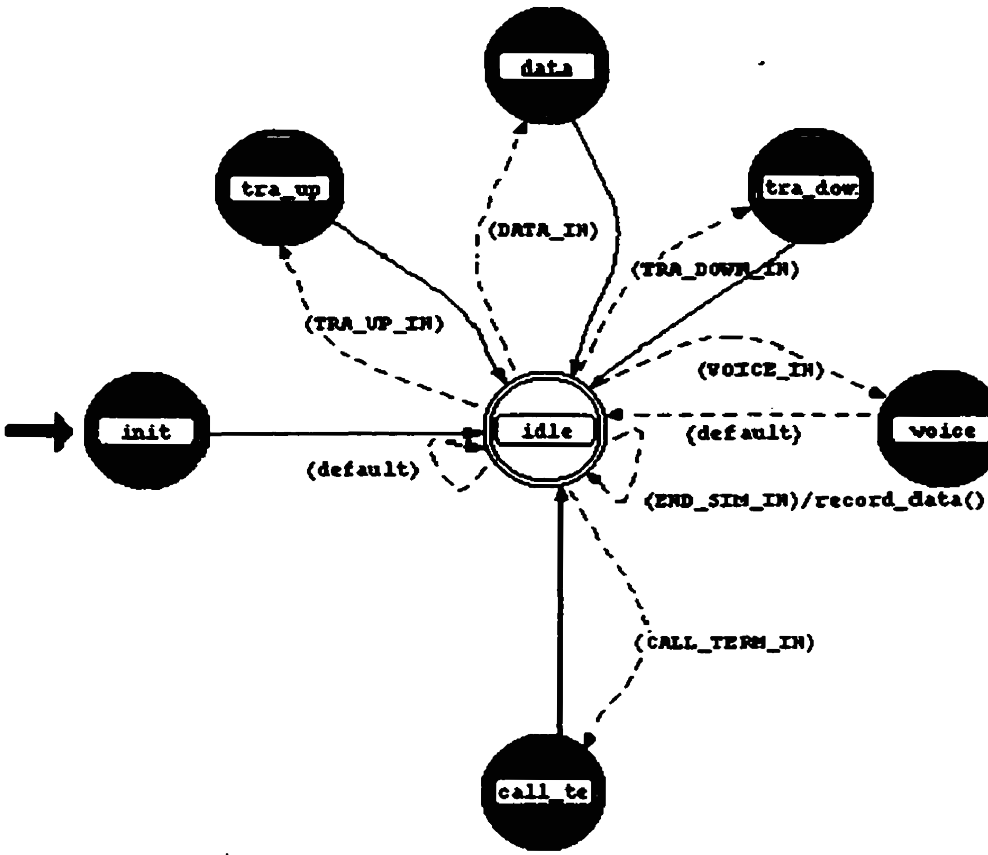

Figure 4.5 PN process level modelling

Figure 4.6 shows the process model implemented in mobile's node model *call_server* node. Call set up, call termination, handoff, location update functions are all handled in MT process model.

Call set up function is handled by states *call_in*, *await_r*, *check_res*, *xmit_ac*, *BLOCKED*, and *recheck*. The implementation of call set up procedure is borrowed from OPNET process model named *cell_call_svr_m* [27]. The mobile first sits in *idle* state, when a call starts (through receipt of an interrupt from *call_gen* node in MT node level modelling), the process transits to *call_in* state. In this system, there is a maximum number of attempts that a mobile can try to probe if the home BS is busy or idle, and also a maximum number of attempts that a mobile can try to send an access message. In *call_in* state, mobile resets access attempts and busy attempts counters. The access channel is shared by the mobile users in a geographical cell, so collision may happen when more than one user try to access the BS at the same time.

In *await_r* state, the mobile transmits an access packet, and starts a timer (timer 1) to wait for the response from the access channel. The mobile also sits in *await_r* for the expiration of another timer (timer 2) to recheck the busy-idle status. If a new call arrived (CALL_ARR) when the transmission of the current call is still in progress, this new call is queued.

The mobile process transits to *xmit_ac* state when timer 1 expires (ACCESS_TRY). The mobile tries to send an access message again if the total number of attempts is less than maximum access attempts, otherwise the access fails (ACCESS_DENIED) and the process goes back to *idle* state.

If timer 2 expires (RECHECK_BIS), the mobile process transits to *recheck* state. It checks the BS status (idle or busy). If the BS radio receiver is idle, the mobile starts timer 1 and re-sends access message. If the channel is busy, the mobile starts timer 2 again to wait for the receiver to be idle. If the number of attempts exceeds the maximum number of idle-busy attempts (BUSY), this call set-up is said to fail. The mobile process transits to *idle* state and waits for start of the next call.

If the mobile receives a response (RCVD_RESP), the mobile process transits to *chk_res* state, and checks the response message originated from the PN. If no channel is available (CALL_BLOCKED), the process transits to *BLOCKED* state from which it goes to *idle* waiting for next call set up event. If channels are allocated to the mobile (ASSD_CH), *start_voice()* function is first executed to set the channel attributes in nodes *rt_1* and *rr_1* and inform *cbr_gen* to start sending data (refer to Figure 4.4 for MT node level modelling). The mobile process then transits to an unforced state *call_p*. While in this state, if the simulation termination condition (SIM_END) is satisfied, the interested simulation data are collected by executing function *ho_report()*. If a handoff initiate timeout interrupt happens (HOWIN_END), the mobile process transits to *ask_ho* state to request a handoff if necessary. If a handoff message or location update message arrives (HANDOFF), the process transits to *handoff* state to handle the message accordingly. If the current call terminates (CALL_END), the process transits to *call_te* state and sends a message to PN to clean up the call connection and then transits to *idle* state. If another call arrives (CALL_ARR) while the mobile is on the phone (*call_p* state), the mobile queues the call (*q_call()*) until the current call is completed. The mobile checks queued call when in *idle* state. The mobile either starts another call connection set up request if there is a queued call or sits in *idle* state if no call is queued (SUBQ_EMPTY).

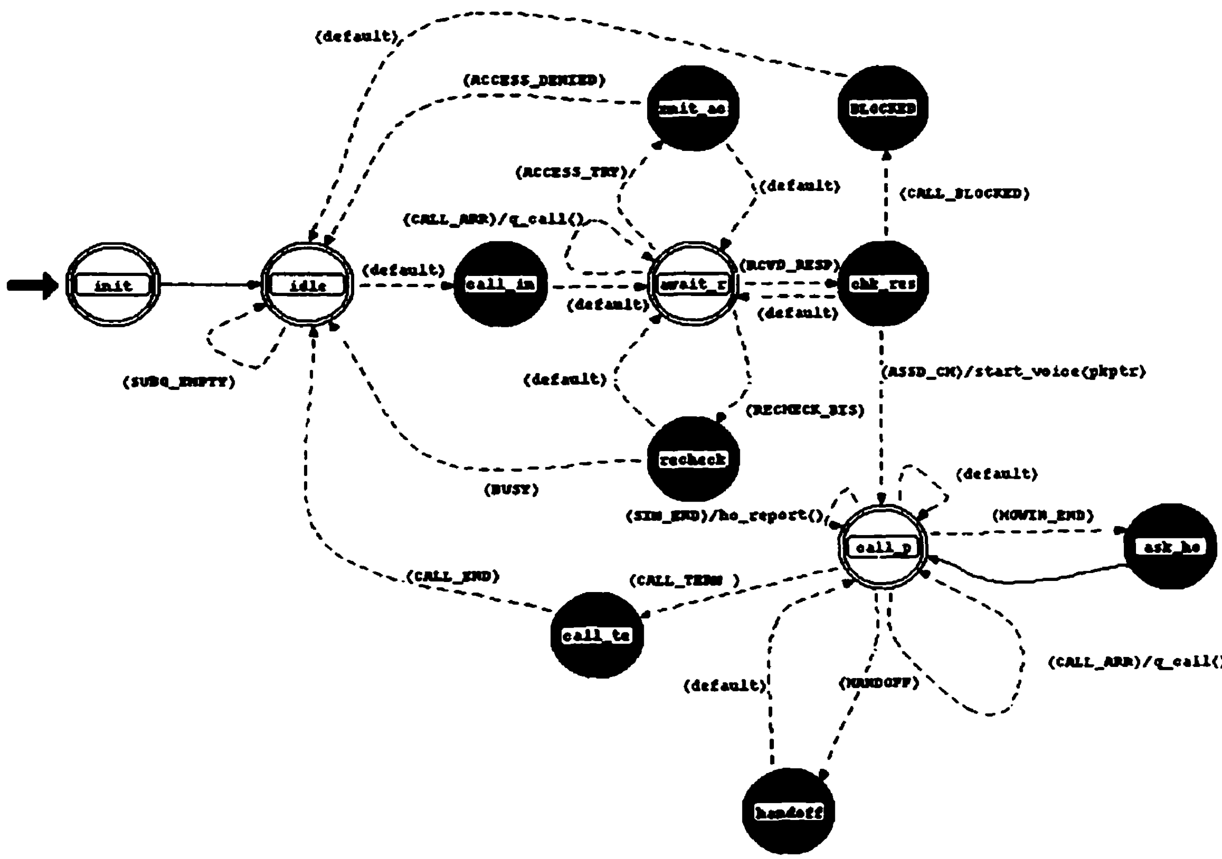

Figure 4.6 MT process level modelling

4.3 Simulation Network Design

In this section the selected criteria for wireless ATM network simulation design are presented.

4.3.1 Mobility Model

Each mobile in the network is moving with a constant speed but a random direction. Mobile's direction is changed periodically and uniformly to one of the four directions: North, South, East and West, as representative of micro-cellular network architecture in downtown core areas. If the mobile moves out of the area covered by the four geographical cells shown in Figure 4.1, the mobile is forced to move back the next time the mobile direction is updated.

4.3.2 Location Updating Procedure

The 4 base stations are sending location update messages periodically starting at the beginning of simulation time. The location update message is used by a mobile to locate its home BS, where it sends signalling messages and data. When the mobile receives the location update messages from 4 BSs, it measures the SNR value of each signalling message. The base station with the highest SNR is chosen as its home base station. Location is also updated after each handoff event. However location is not updated while a mobile is on a call, because the mobile keeps using the resource of the home base station where it originates the call. This home base station information and the resource information are used for handoff.

4.3.3 Channel Resource Management

Channel assignment strategies can be classified into fixed, flexible and dynamic [21]. The common underlying theme in all fixed assignment strategies is the permanent assignment of a fixed set of channels to each BS. Fixed channel assignment strategy is used in this simulation with 15 channels provided in each base station for simplicity. All channels of the 4 base stations are controlled by the PN, as PN has global view of the channel usage of all the base stations connected to it. A mobile asks for a channel to connect a new call through home BS or to perform handoff to destination BS. A mobile releases a channel when a call is terminated or when an old connection is torn down upon the handoff completion. For SHLR scheme, CDMA access scheme is used, where each mobile is assigned a spreading code generated from Walsh code when it asks for a channel. For HHLR scheme, FDMA access scheme is used, where each mobile is assigned a frequency. If there is no channel available in the home base station of the mobile, this call connection or handoff attempt fails.

4.3.4 Handoff Initiation

Each mobile keeps a list of SNR values for each of the 4 base stations. When the mobile receives the location update message from 4 base stations, the mobile extracts the SNR value from the messages and puts them in the BS SNR list, so that for each BS there is an updated SNR value. The mobile checks the SNR list right after the location update messages are received. If the MT's home BS's SNR value is lower than the threshold, the MT searches the SNR list for the highest SNR value [20]. If such a value exists, the mobile selects the BS with this SNR value as the destination BS and initiates a handoff request. The handoff initiation time is uniformly distributed between 1 msec and 1.2 msec, and between 1 msec and 5 msec for the purpose of running different simulation scenarios. The range of these two intervals are chosen so that during each interval, multiple mobiles can initiate handoffs closely enough, and hence ATM cells from those mobiles are stored in the same buffer during handoff. The above situation models the actual operation during network handoff. The radio link handoff latency is selected from a uniform distribution in the range of 7 to 11 ATM cells transmission time for simplicity.

4.3.5 Call Arrival Rate and Call Holding Time

Mobiles are the sources for generating calls. Call arrivals follow a Poisson process with mean call arrival rate 1/360 calls per second. Call holding time has exponential distribution with mean 288 seconds. New calls arriving before the current call is completed are queued. Another call starts right upon the current call's completion if there is a queued call. Call holding time is chosen to be less than the call inter-arrival time (360 seconds) so that there are fewer queued calls than the situation with a longer mean call holding time.

4.3.6 Mobile Source Traffic Type

Mobiles generate one of two traffic types: CBR, VBR. Figure A.1 and Figure A.2 (Appendix A) provides the definition for CBR and VBR traffic respectively. At the beginning of simulation, the mobile chooses a traffic type (CBR or VBR) with equal probability. The peak bit rate for both CBR and VBR is set to the same value 2 Mbps (Table 2.1). Both VBR and CBR traffic are modeled as ON-OFF sources. The ON period for VBR source is 352 msec and the OFF period is 650 msec [28], which apply to packet data. The transmitter at each of the BS empties data out of the buffer at a speed of 2.1 Mbps. This value is selected to empty the buffer at a faster rate than the rate the buffer is being filled.

4.3.7 Wireless Environment

Log-normal shadowing is assumed to simulate impaired radio link. For each ATM cell that needs to be processed through the buffer during handoff, the corresponding SNR value in decibel unit is generated. The received average SNR value is distance based. To this value is added a random variety selected from a normal distribution with standard deviation 3.0 dB. The deviation value is chosen for simplicity as long as the HHLR and SHLR schemes are comparable. For the HHLR scheme, if the SNR value is lower than the threshold, this ATM cell is counted as discarded due to shadowing. However this ATM cell is still further processed through the buffer (either stored or discarded if the buffer is full), since in HHLR scheme shadowing is not considered. For the SHLR scheme, a comparison of these values is done for duplicated ATM cells. Only when both of the values are lower than the threshold value is the ATM cell counted as discarded due to shadowing.

4.4 Verification of Simulation Models

The designed simulation network is verified before the final simulation run. The verification is done for each of the sub-module functions stated in section 4.3 and for the integrated function of the whole simulation network. For example, to verify handoff initiation and handoff signalling execution procedure of the mobiles, trace messages are printed out. The printed messages confirm that the mobile selects the highest SNR value if the received SNR value from the current BS is lower than the threshold. From then on, all the signalling messages presented in section 3.5 are then executed and printed out. The printed messages confirm the correct execution of the simulation program. Furthermore, information about the buffer size, the time each ATM cell arrives to the buffer, the time each ATM cell leaves or is discarded by the buffer are also printed out to verify the correct calculation of the simulation results: mean cell queuing delay and cell loss ratio.

Before the final simulation runs, the warm-up simulation runs for each scenario for each scheme are executed 10 times with different input seed. The warm-up simulation results showed the difference between simulation results were within 10% range. So the simulation results have 90% confidence. The results presented in this thesis are average values of 10 final runs for each single scenario.

4.5 Simulation Scenarios

To demonstrate the better performance of SHLR scheme over wireless environment, simulation scenarios both with and without log-normal shadowing need to be considered. In addition, because of the randomness of the handoff activity, ATM cells from only one mobile or multiple mobiles are processed by the buffer during network handoff. So the following simulation scenarios are selected:

1) Only 1 mobile is allowed to perform network handoff at one network node.

2) Maximum of 3 mobiles (M_{ho} =3) are allowed to perform network handoff simultaneously at one network node.

3) Maximum of 4 mobiles (M_{ho} =4) are allowed to perform network handoff simultaneously at one network node.

4) Log-normal shadowing considered with simulation scenario 3.

In the above scenarios, network node represents PN in SHLR scheme and BS in HHLR scheme respectively.

In this chapter, the simulation modelling is discussed in a great detail including network elements behaviour, network functionalities, etc. The well-defined simulation network provides solid background for the simulation results discussion of next chapter.

Chapter 5

SHLR Performance Analysis

This chapter presents performance analysis for SHLR scheme. The performance analysis includes results from both simulation approach and analytical approach. The results of SHLR scheme are compared with the results of HHLR scheme.

5.1 Simulated SHLR Performance Analysis

5.1.1 Input and Output Parameter List

Table 5.1 presents the parameters used in the simulation:

Table 5.1 Input parameter list

Parameter's Name	Value
Base Stations	4
Mobiles	40, evenly distributed across the 4 BSs
Call arrival rate	1/360 calls per seconds
Call holding time	288 seconds
Location update period	30 seconds
Handoff initiation window width	[1, 1.2] msec for scenario 3 [1, 5] msec for scenario 2
VBR source ON period duration	0.352 seconds
VBR source OFF period duration	0.650 seconds
Traffic source peak rate	2.0 Mbps
Transmitter rate	2.1 Mbps

The output results from the simulation are:

- Cell loss ratio
- Mean cell queuing delay (in msec)

Cell loss ratio and mean cell queuing delay are defined as follows:

Cell loss ratio due to lack of buffer space is defined as a ratio of discarded ATM cells (i.e. cells arriving when the buffer is full) to the total numbers of ATM cells processed by the buffer (i.e. cells that are both discarded and stored in the buffer).

ATM cell queuing delay is the difference between the time the ATM cell is sent out from the buffer and the time it is stored in the buffer

Mean cell queuing delay is defined as a ratio of total ATM cell queuing delay to total number of ATM cells processed by the buffer. The total ATM cells queuing delay is an accumulated delay experienced by all ATM cells that are stored in the buffer.

Cell loss ratio due to shadowing is defined as the ratio of the total number of ATM cells whose received SNRs are lower than the shadowing SNR threshold to the total number of ATM cells processed by the buffer.

5.1.2 Simulated Results and Discussion

The results obtained from the simulation are discussed in this section.

5.1.2.1 Cell Loss Ratio Simulation Results

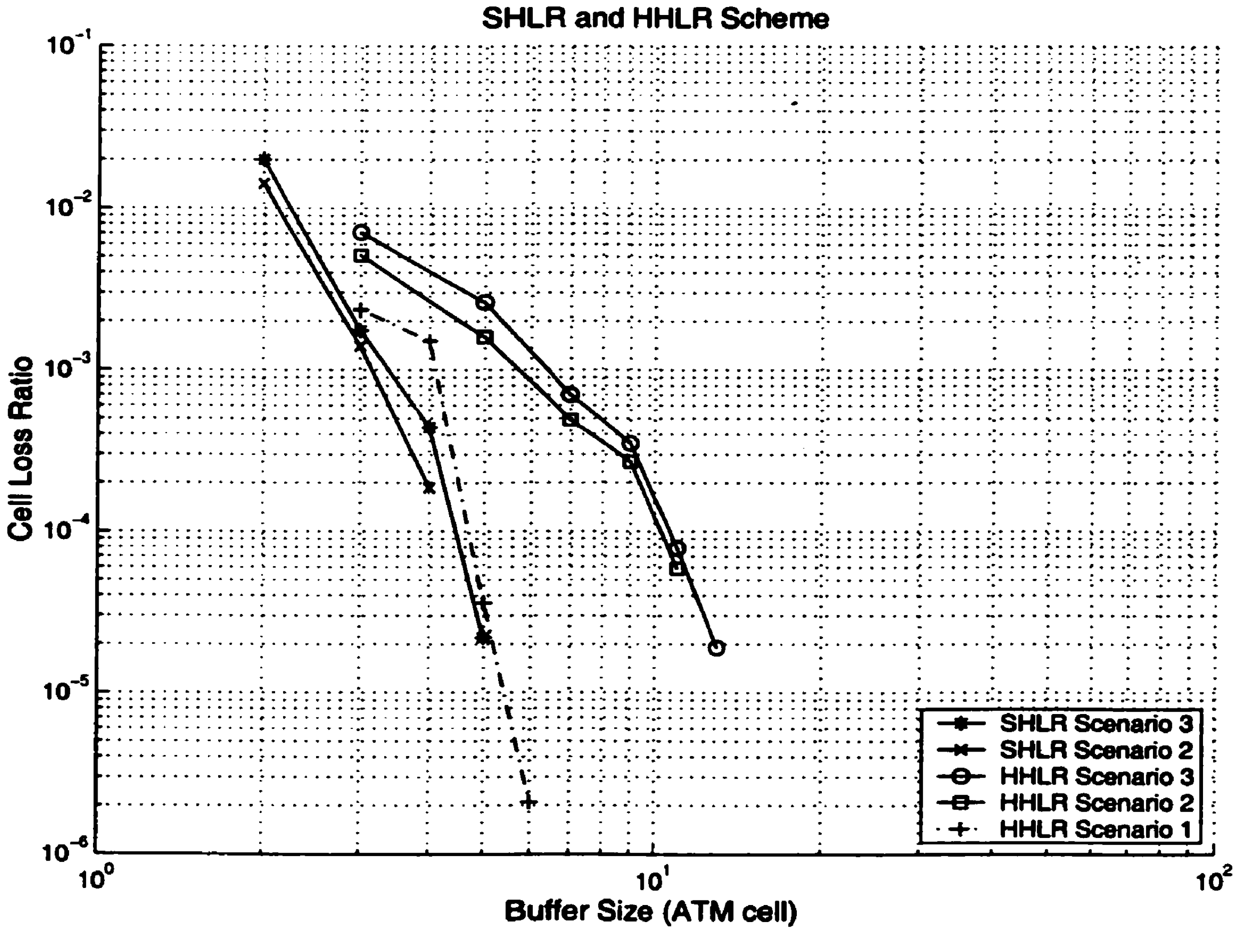

Figure 5.1 Comparison of cell loss ratio versus buffer size for SHLR and HHLR schemes

Figure 5.1 shows cell loss ratio versus buffer size for SHLR and HHLR schemes. In the simulation for SHLR scheme, the path difference (0.17 msec) for a pair of ATM cells is selected to be less than the time of 3 in sequence ATM cells that can arrive to the buffer (2*0.192 msec). It is found out that if only one mobile is allowed to perform network handoff at a time, for buffer size 3, the loss is zero. So results of scenario 1 are not shown in Figure 5.1 and the corresponding mean cell queuing delay results are not shown in Figure 5.5 as well. For scenarios 2 and 3, the larger the buffer size, the more the ATM cells that can be buffered and hence lower cell loss ratio.

The results for HHLR scheme scenarios 1, 2 and 3 are analysed as follows:

1. For all the 3 scenarios, the cell loss ratio decreases as the buffer size increases as expected. For example in scenario 3, at a buffer size of 3 the cell loss ratio is in the order of 10^{-3} which reduced to the order of 10^{-4} at a buffer size of 7.
2. Comparing the 3 scenarios, it is seen that scenario 3 has the largest cell loss ratio and scenario 1 has the smallest cell loss radio. The reason is that with larger M_{ho}, there is a higher chance to have more than one mobile doing network handoff simultaneously, causing ATM cells from multiple end users to arrive at the uplink buffer at the same time. As an example, for a buffer of size 5, the cell loss ratio of scenario 1 is $3.7*10^{-5}$, the cell loss ratio of scenario 2 is $1.7*10^{-3}$ and the cell loss ratio of scenario 3 is $0.8*10^{-3}$ respectively.
3. For the same buffer size, scenario 3 can cause more loss than scenario 2. Same reason can be applied to compare scenario 2 to scenario 1. In other words, to meet the same cell loss ratio requirement, a larger buffer size is needed for scenario 3 than scenario 2 (same for scenario 2 than that for scenario 1). This means to have a cell loss ratio say less than $7.0*10^{-5}$, the buffer size needed are 5, 11 and 13 for scenarios 1, 2, and 3 respectively.

The results for SHLR scheme scenarios 1, 2 and 3 are analysed and compared with HHLR scheme as follows:

For the same cell loss ratio, the buffer requirement is less for the SHLR scheme compared to that for HHLR scheme. Recall from Chapter 3 that for the SHLR scheme, path difference is the cause for buffering and hence cell loss. For the HHLR scheme, path addition (old path and new path) is the cause for buffering and hence cell loss. This is why a lower cell loss ratio is achieved for the SHLR scheme compared to that for the HHLR scheme.

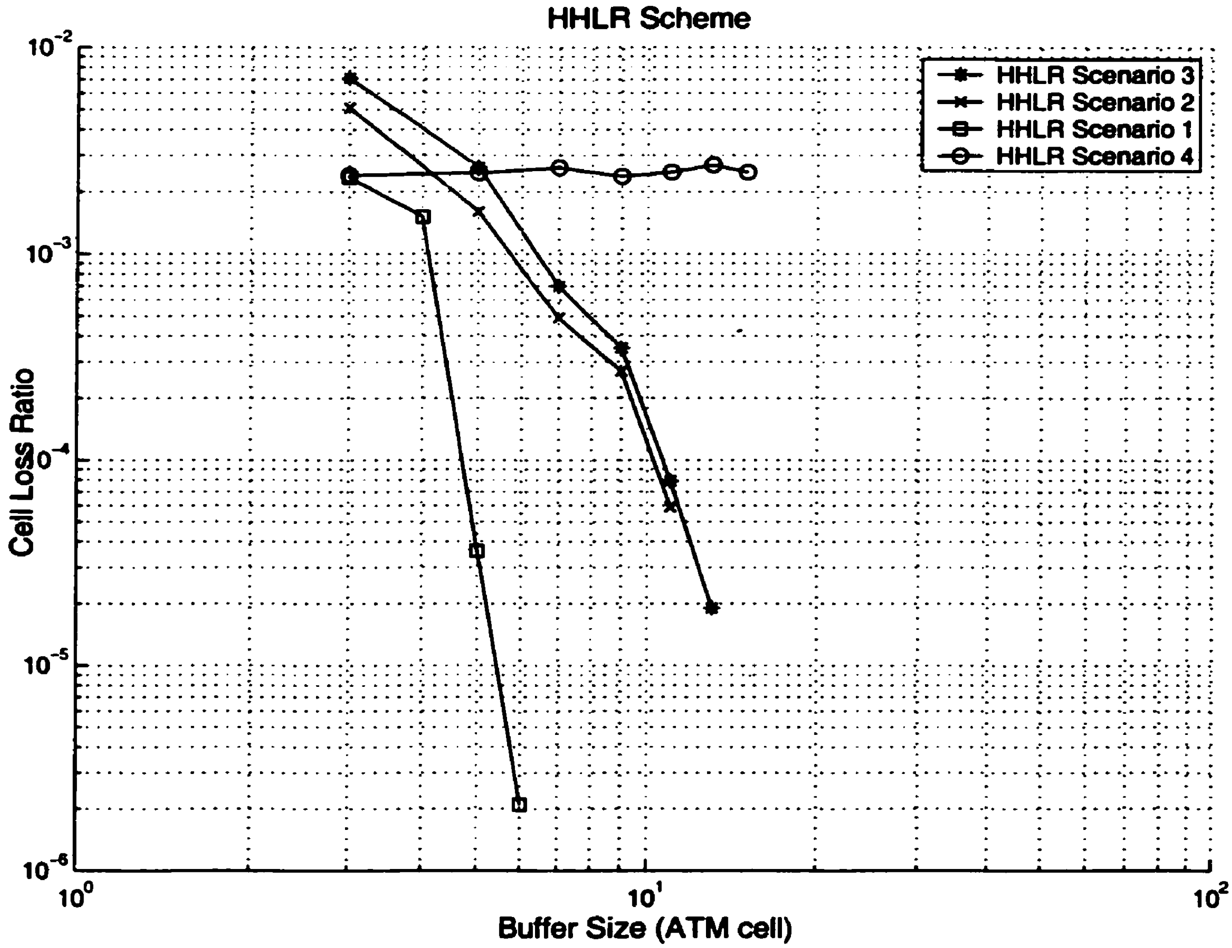

Figure 5.2 Cell loss ratio due to buffering compared with the cell loss due to shadowing for HHLR scheme

Figure 5.2 shows cell loss ratio due to buffering compared with the cell loss due to shadowing for HHLR scheme. The results are put together due to the fact that the total cell loss in the network is due to both the buffering and the shadowing. Unlike buffer size induced cell loss, it can be observed that cell loss due to shadowing is relatively constant and insensitive to the buffer size. The cell loss due to shadowing in the simulation is within the range of $2.0*10^{-3}$ to $3.0*10^{-3}$. Regardless of the buffer size value, the log-normal shadowing that ATM cells experience through radio link is not affected. It can be concluded from these results that finite buffer size is not the only cause of cell loss in the network. The results also imply that, if shadowing effect dominates, cell loss ratio cannot always be improved by adding more buffers. Comparing scenario 3 and 4, as shown if Figure 5.2, if the buffer size is less than 5, the overall cell loss ratio is mainly due to finite buffer size. For buffer size larger than 5, the cell loss ratio is mainly due to shadowing. In the latter case, increasing buffer size could not improve the cell loss ratio. It is concluded that two factors can affect the performance of a handoff protocol design for ATM networks with wireless extension: limited buffer size of fixed part of the network and the shadowing effect of the wireless part of the network.

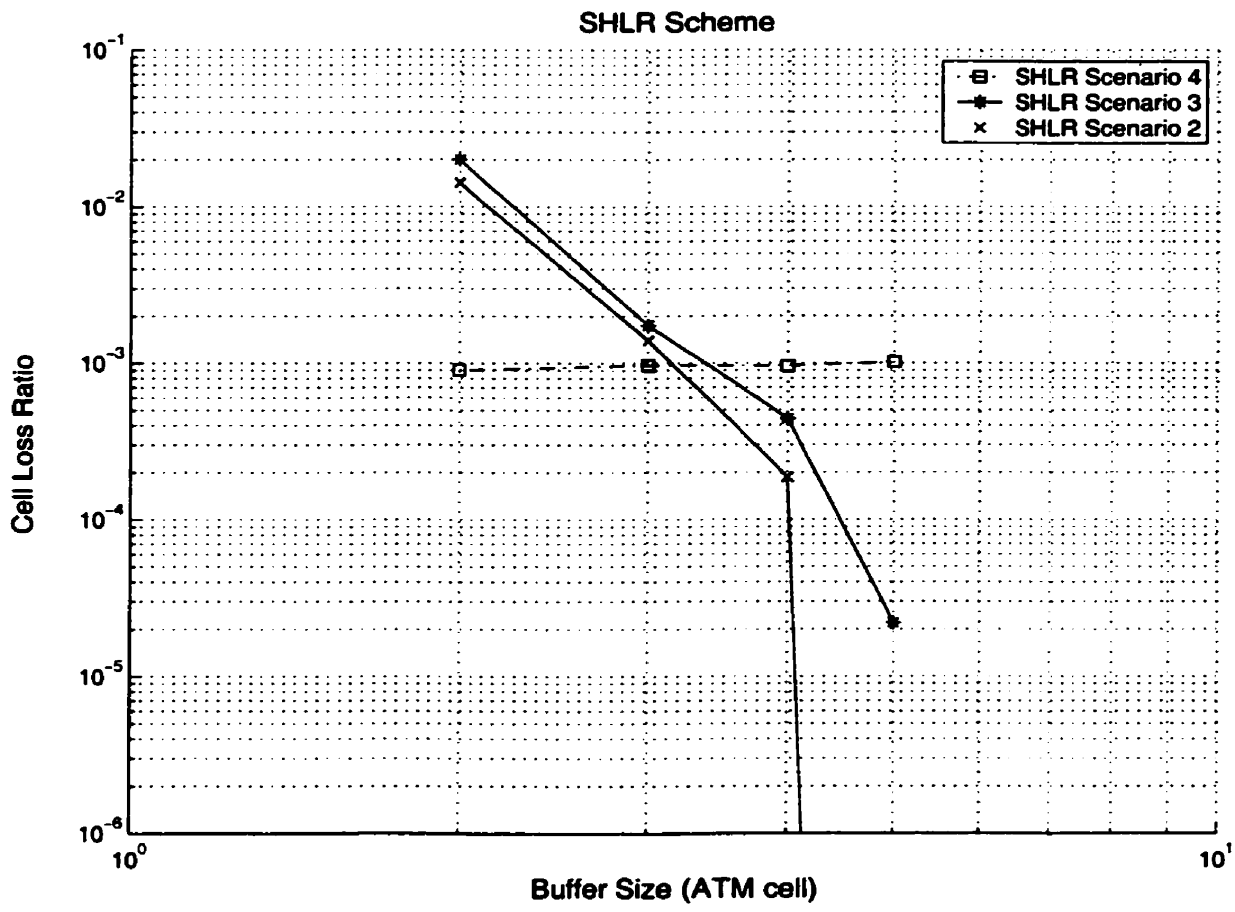

Figure 5.3 Cell loss ratio due to buffering compared with the cell loss due to shadowing for SHLR scheme

Figure 5.3 shows cell loss ratio due to buffering compared with the cell loss due to shadowing for SHLR scheme. The results shown in this figure validate the discussion for results shown in Figure 5.2. For buffer size 5 in scenario 2 with the simulated packets of 10^6 no cell loss observed, so the cell loss ratio falls to zero.

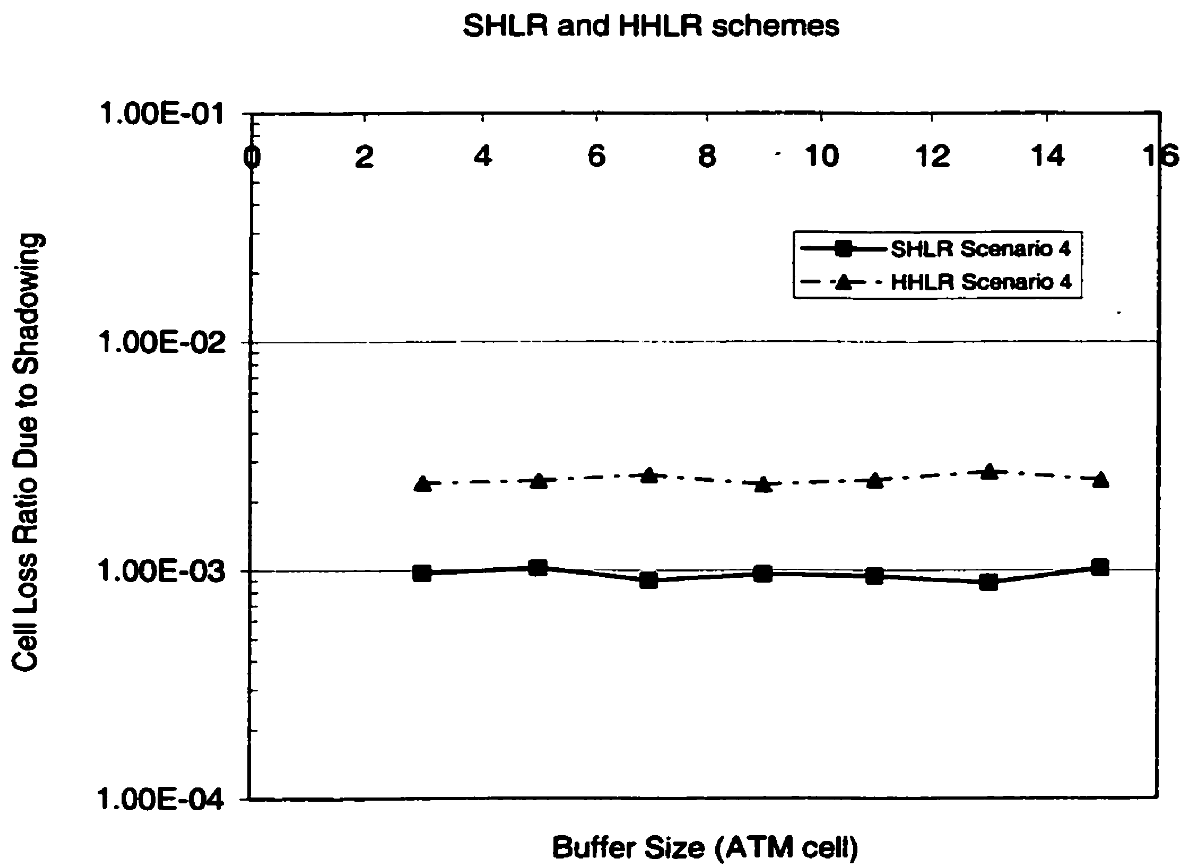

Figure 5.4 Comparison of cell loss ratio due to shadowing for both SHLR and HHLR schemes

Figure 5.4 shows the comparison of cell loss ratio due to shadowing for both HHLR and SHLR schemes. It is seen that, for the same wireless environment implemented, the cell loss due to shadowing in SHLR scheme is less than that in HHLR scheme. This demonstrates the beauty of SHLR scheme. Because of duplicating ATM cells in SHLR scheme, as long as one of the duplicated cells is above the SNR threshold, this ATM cell is valid and can be sent to the end user. Theoretically, the improvement by SHLR scheme should be 2 times better than that of HHLR scheme assuming the ATM cell transferred between the old path are independent from that of the new path. Since the received signal (in dB) is normally distributed, with the mean of the received signal due to distance, the chance that the signal is less than this distance-based mean is 1/2, which is the case for HHLR scheme. For duplicated ATM cells in SHLR scheme, four cases can happen in the PN buffer: i) the ATM cell from the old path is above and the one from the new is less; ii) the ATM cell from the new path is above and the one from the old is less; iii) both of them is above threshold; and iv) both of them are less than the threshold. The ATM cell cannot be transmitted only if case iv) happens. So the cell loss ratio in HHLR scheme should be twice as much as that of SHLR scheme, which can be seen in Figure 5.4. It is important to note that the performance improvement provided by SHLR is at the expense of increased cell delay due to waiting for the duplicate ATM cell.

5.1.2.2 Mean Cell Queuing Delay Simulation Results

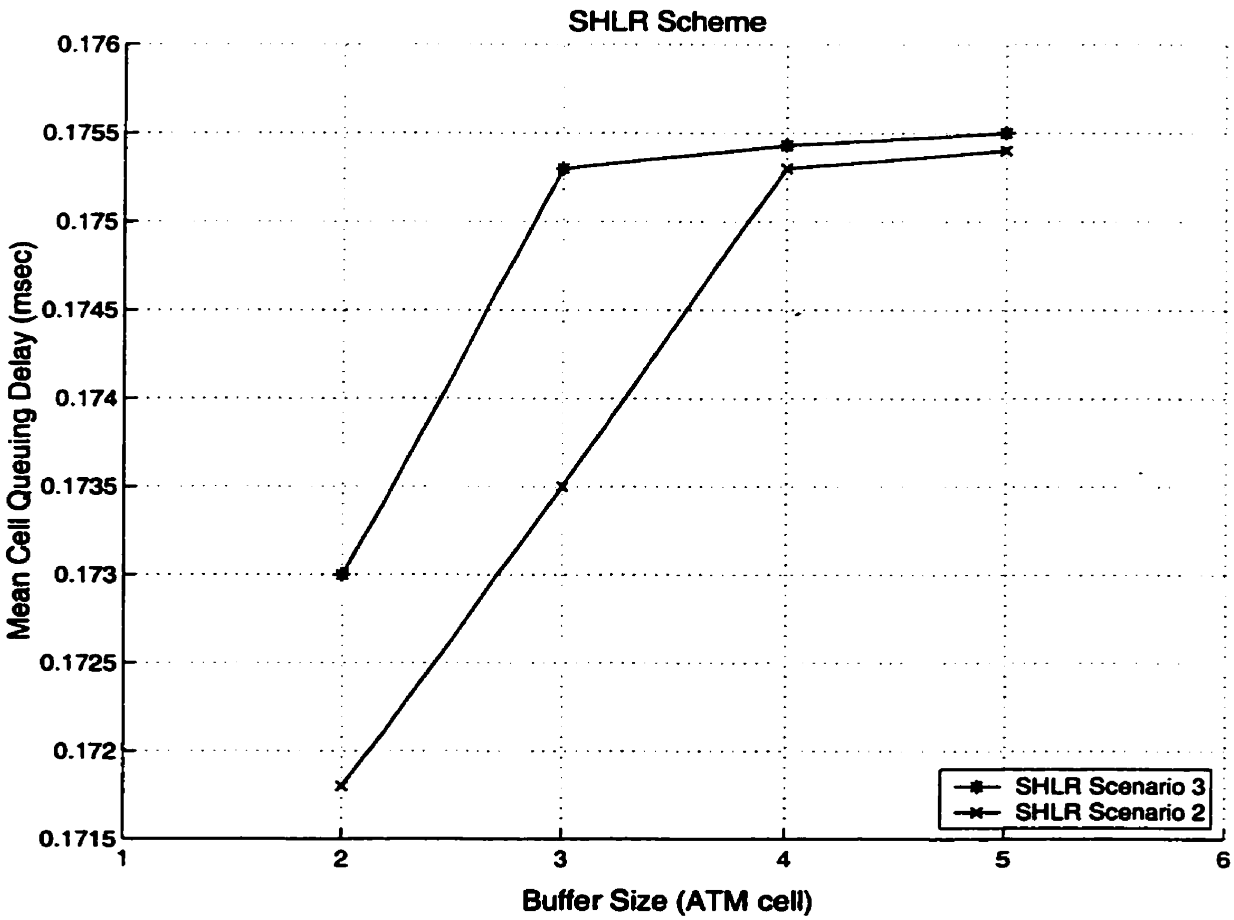

Figure 5.5 Mean cell queuing delay versus buffer size for SHLR scheme scenarios 2 and 3

Figure 5.5 shows mean cell queuing delay versus buffer size for the SHLR scheme scenarios 2 and 3. The cause of the delay which is specific to SHLR scheme is the path difference of the two duplicated ATM cells during radio link handoff. There are two possible cases during a simulation run: i) an ATM cell waits for its duplicate before it is sent; ii) if the duplicate cell never arrived, the ATM cell that arrives first waits in the buffer until a timeout expires and then it is ready to be sent. The above 2 cases are the causes of the cell queuing delay. Since in the simulation, the ATM cell inter-arrival time is 0.192 msec and the path difference is 0.17 msec, the duplicate ATM cell arrives before the next ATM cell arrives for one mobile. The constant value of path difference factor dominates the results of the mean cell queuing delay in the simulation. This can be seen from Figure 5.5, the mean cell queuing delay for both scenarios range from 0.1715 msec to 0.1755 msec. Had the value of path difference chosen to be larger than the ATM cell inter-arrival time, multiple ATM cells from one mobile may be queued in the buffer. From buffer analysis presented in chapter 3, queuing delay factor would then have affected the simulated results together with path difference. This simulation focus is on handoff procedure induced delay as shown in Figure 5.5.

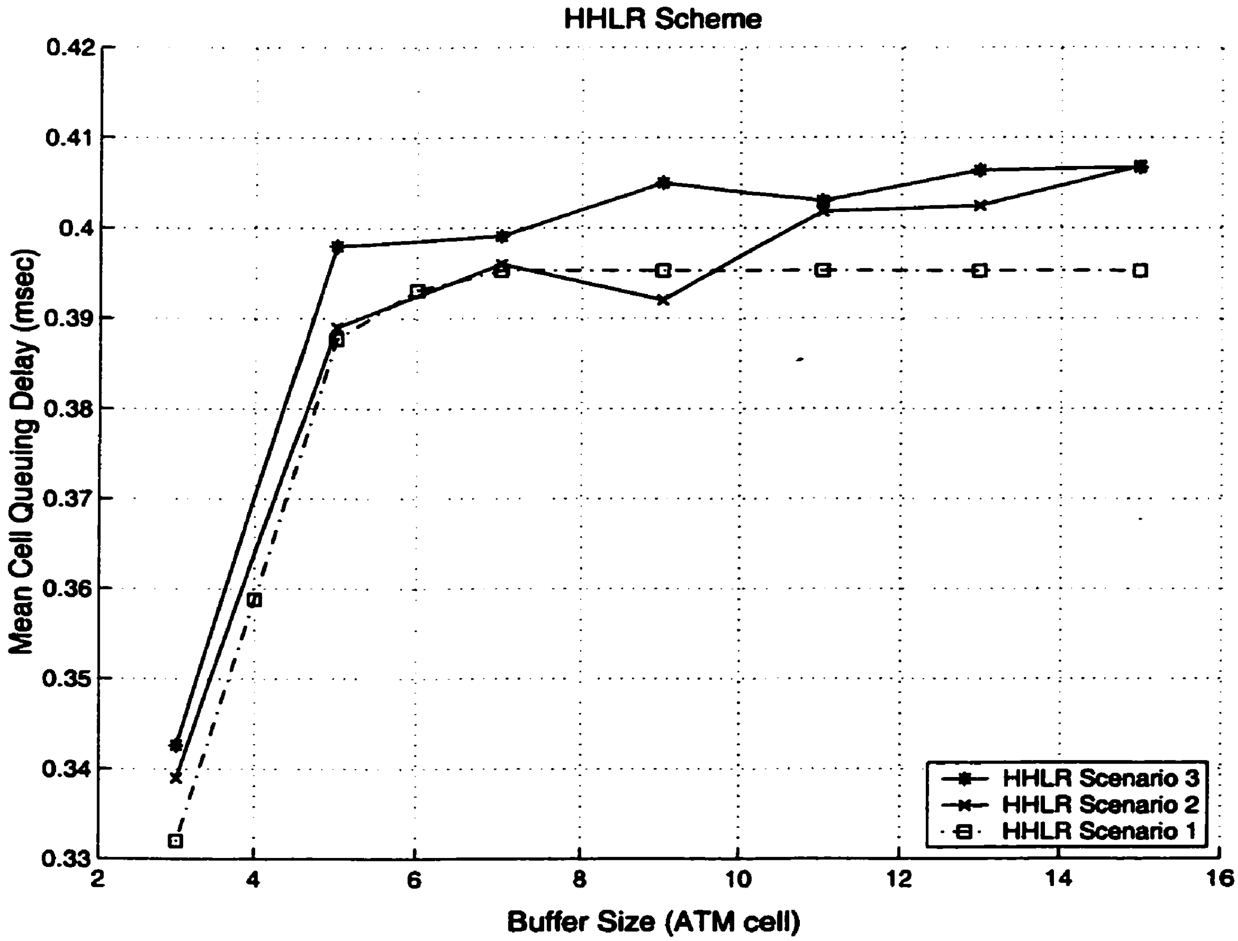

Figure 5.6 Mean cell queuing delay versus buffer size for scenarios 1, 2 and 3 for HHLR scheme

Figure 5.6 shows mean cell queuing delay versus buffer size for scenarios 1, 2 and 3 assuming HHLR scheme. Comparing the results of the three different scenarios, the mean cell queuing delay of scenario 3 is the highest while that of scenario 1 is the lowest. This is because scenario 3 has the highest M_{ho} value and scenario 1 has the lowest M_{ho} value. The higher the M_{ho}, the higher the chances that ATM cells from multiple mobiles stored in the buffer simultaneously. This causes higher delay for scenario 3 than that for scenario 2 and also for scenario 2 than that for scenario 1. At large buffer size (> 8 cells) the delays tend to a constant value, due to the finite buffer size.

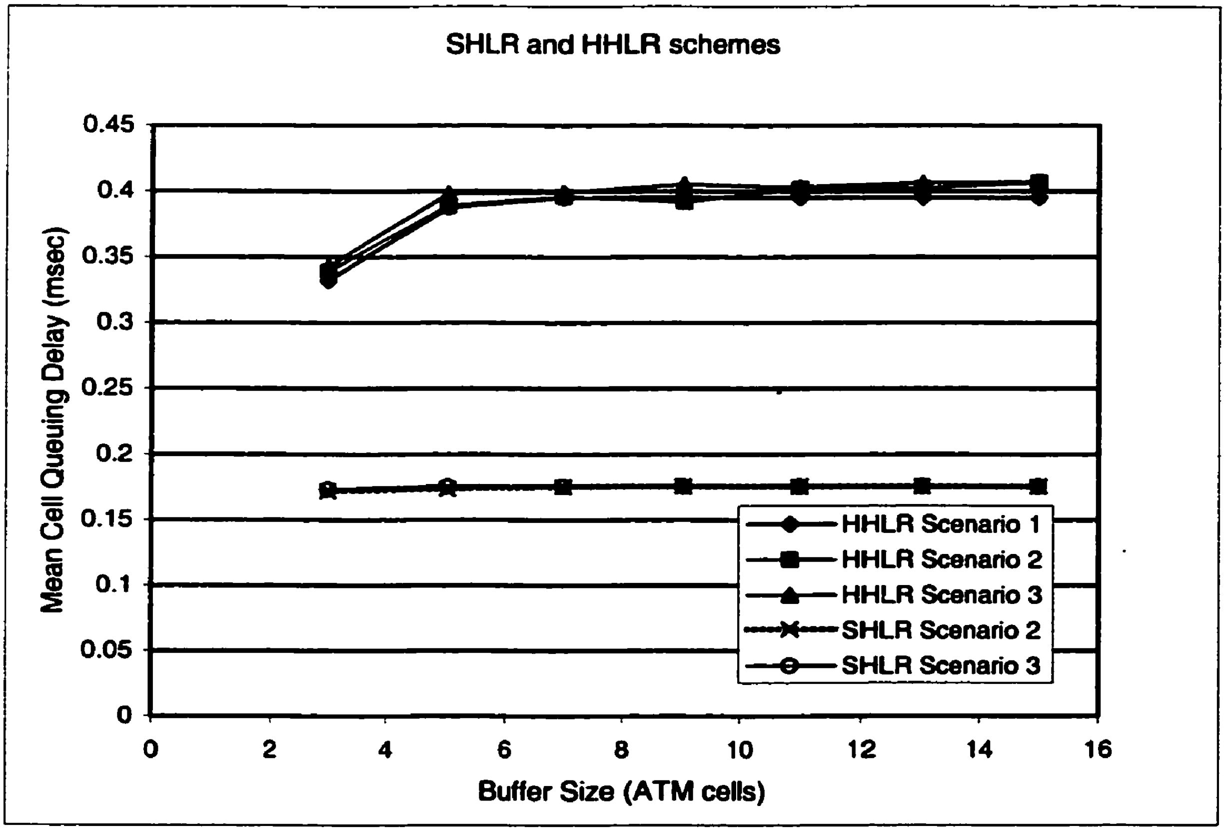

Figure 5.7 Comparison of mean cell queuing delay for SHLR and HHLR schemes

Figure 5.7 presents the comparison of mean cell queuing delay for SHLR and HHLR schemes. As the mean queuing delay in SHLR is mainly due to path difference (0.17 msec) and that for HHLR is due to handoff latency of the addition of the path propagation delay (0.36 msec), the results show that SHLR scheme has less delay compared to HHLR scheme. The fact is that time duration of the difference of the path propagation delay is always less than that of the addition of the path propagation distance under the same wireless ATM scenario simulated in this thesis. Hence, on average, there is always less number of ATM cells waiting in the buffer during the handoff procedure in SHLR scheme than that in HHLR scheme, resulting in less mean cell queuing delay.

5.2 Approximate Analytical Performance Analysis

5.2.1 Analytical Model

The W-ATM network has the following key elements:

- MTs: generate calls and send ATM cells to BS. MTs generate calls according to Poisson process. The traffic type may be CBR or VBR.
- PN buffer in SHLR or destination BS uplink buffer in HHLR: stores ATM cells for handoff MTs. The buffer has a finite length.
- Transmitter at the buffer: Empties the buffer if there are any ATM cells in it. The transmission rate is constant. Because the ATM cell has a size of fixed length (48 bytes), the transmission time is also constant or deterministic.

The above network elements can be represented in the queuing model by input source, buffer and server respectively.

The network scenario is modelled by M/D/1/N queuing model. In M/D/1/N, *M* represents Markov or Poisson input process with rate λ, the notation *D* represents the deterministic service time, as a result of the constant ATM cell size, the notation *1* represents single server (transmitter) in the system and *N* represents the buffer size. Figure 5.8 illustrates the M/D/1/N queuing model. The derivation of the M/D/1/N model is provided in Appendix B. The results of the queuing model analysis are the blocking probability which is same as cell loss ratio in the simulation, and mean queue waiting time which is same as mean cell queuing delay in the simulation.

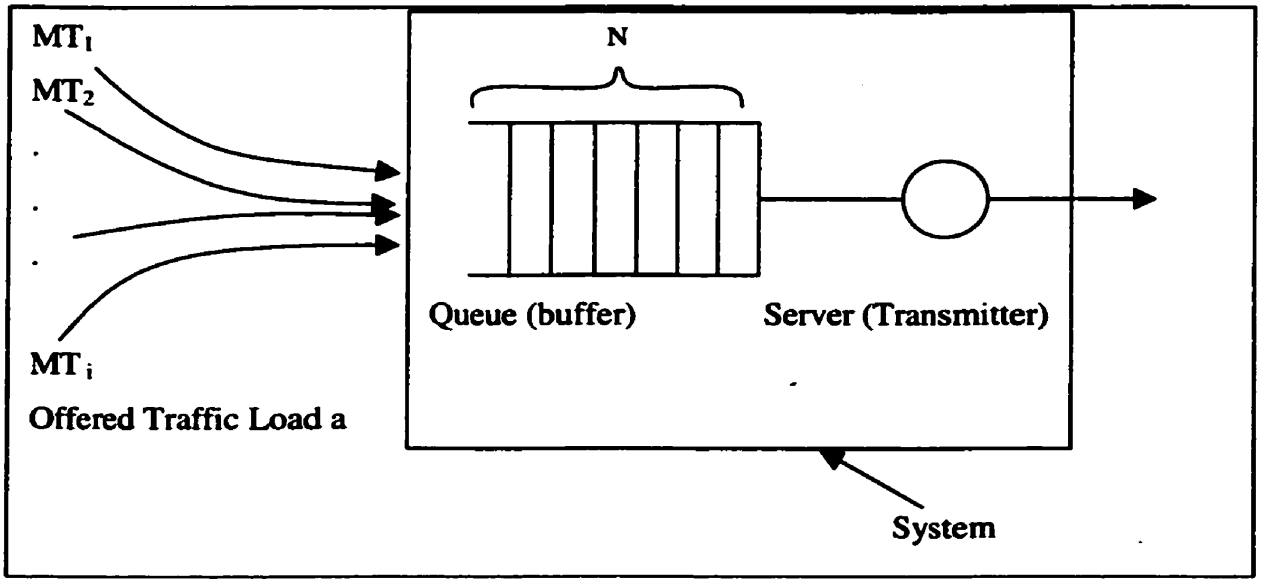

Figure 5.8 M/D/1/N queuing model

5.2.2 Input Parameters to Analytical Model

The analytical results for blocking probability and mean waiting time (derived in Appendix B) are dependent on traffic load. In both SHLR and HHLR schemes, where handoff behaviour is the focus, handoff buffer stores ATM cells only during the execution of the handoff procedure. During the time that the mobiles are not performing handoffs, the handoff buffer is not used. This implies that the actual input load seen by the handoff buffer is the traffic load due to handoff. This handoff-induced load is referred to as handoff traffic load in the thesis.

5.2.2.1 Offered Traffic Load

For all traffic classes $j \in J$, where J is an index set, with call arrival rate λ_j and call holding time h_j, the input traffic load is given by:

$$a = \sum_{j \in J} \lambda_j h_j \tag{5.1}$$

Considering i) ON-OFF VBR source data activity y_j, where y_j is a ratio expressed as $y_j = lon / (lon + loff)$. *lon* is the on period of the ON-OFF VBR source and the *loff* is the off period of the ON-OFF VBR source; ii) V_j (bps), the data speed of the *j*th source; iii) c (bps), the server transmission speed, the offered traffic load can be modified as [19]:

$$a = \frac{1}{c} \sum_{j \in J} \lambda_j h_j V_j y_j \tag{5.2}$$

The traffic load a in Equations (5.1) and (5.2) is dimensionless.

5.2.2.2 Handoff Traffic Load

Handoff traffic load depends on the carried traffic load and the MT handoff activities. The mobile's handoff activities are affected by the following factors:

a) Boundary crossing rate: how often the MT is moving across the boundary of geographical cell coverage area of each BS.

b) Handoff initiation rate: how often the MT initiates a handoff request. This is the time when MT finds a destination BS with better quality than the current BS and makes handoff request.

c) MT call activity: call arrival rate and call holding time.

The MT performs a handoff only when the mobile is crossing the boundary, is at the time of initiating a handoff and is during a call.

With the assumptions that the i) MT's moving direction is uniformly distributed; ii) the MT moving speed is constant in all moving directions; iii) the MT density in the coverage area is constant, which means the number of MTs moving out of the BS coverage area is equal to that of MTs moving in.

The MT boundary crossing rate is given by [7]:

$$BCR = \frac{M}{A}\frac{v}{\pi}L \tag{5.3}$$

where BCR is average boundary crossing rate.

M is the number of MTs in one geographical cell..

A is the area covered by one BS.

v is the MT moving speed.

L is the boundary length of the coverage area.

The probability that MT is crossing the boundary when mobile is on the phone P_{on} is:

$$P_{on} = \frac{1/h}{1/h + \lambda} \tag{5.4}$$

where h is the mean call holding time and λ is the mean call arrival rate. In the simulation, all mobiles are designed to have the same mean call holding time and mean call arrival rate.

The MT boundary crossing rate BCR_{on} when call is on is

$$BCR_{on} = BCR * P_{on} \tag{5.5}$$

Only those mobiles that are crossing the boundary and under another BS coverage area can be handed over to another BS. Assuming mobile moving direction is uniformly distributed, from Figure 4.1 it can be obtained that the chances that the mobiles in BS_0 and BS_1 can move to any other BS coverage area (e.g. from BS_0 to BS_2 or BS_3) is roughly 1/3. The chances that the mobiles in BS_2 and BS_3 can move to any other BS coverage area (e.g. from BS_2 to BS_0 or BS_1) is roughly 1/2:

$$BCR_{BS} = ((1/2 + 1/3) / 3) * BCR_{on} \tag{5.6}$$

The handoff request initiation rate P_{init} is decided by criteria set in the simulation: only when the MT's SNR is less than the threshold and one of the other BS has higher SNR. So the handoff rate

$$H = BCR_{BS} * P_{init} \tag{5.7}$$

Assuming MT's handoff process is Poisson. The probability P_h that the traffic load is due to handoff is:

$$P_h = \frac{H}{H + 1/h} \tag{5.8}$$

This means the input handoff traffic load a_h is given by

$$a_h = a * P_h. \tag{5.9}$$

5.2.3 Comparison of Simulated Results with Analytical Results

In this section, the generated results from M/D/1/N queuing model are compared with simulation results. For the analytic results, the handoff traffic is assumed as 0.56.

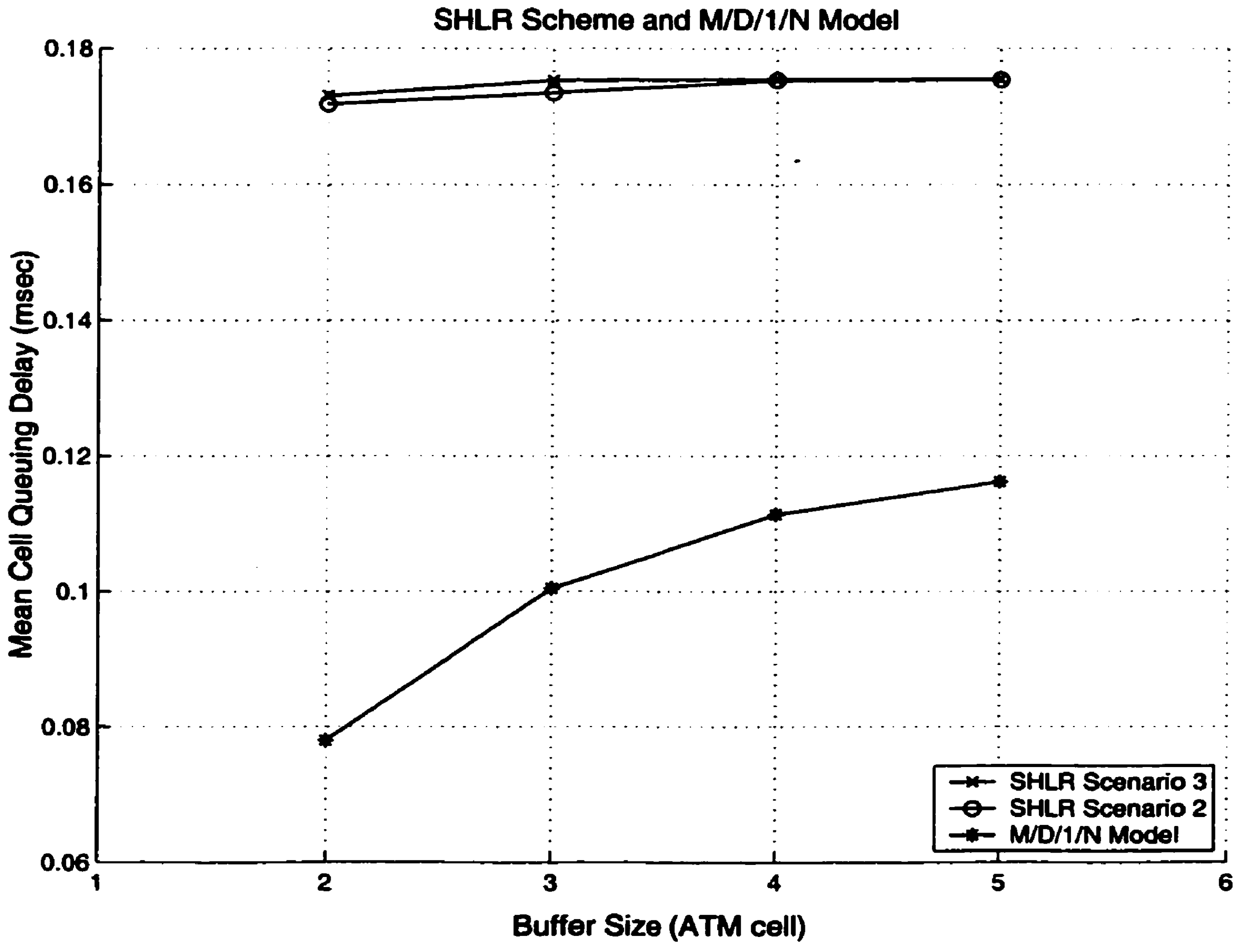

Figure 5.9 Comparison of the simulated mean cell queuing delay for SHLR scheme with analytical results

Figure 5.9 compares the simulated mean cell queuing delay for SHLR scheme with analytical results. It is found that, for all buffer sizes considered, the analytical results are better than the simulation results. This is because of non-inclusion of the handoff procedure delay due to path difference in the M/D/1/N model. Another fact observed in Figure 5.9 is that for the range of buffer size considered, the simulated queuing delay is relatively flat, contrast with that for M/D/1/N model which increases with buffer size. This behaviour is explained by recalling that the simulated mean cell queuing delay for the SHLR scheme is determined by a constant path difference. On the other hand, the main cause of the mean cell queuing delay for the M/D/1/N model results is queuing in the buffer prior to transmission.

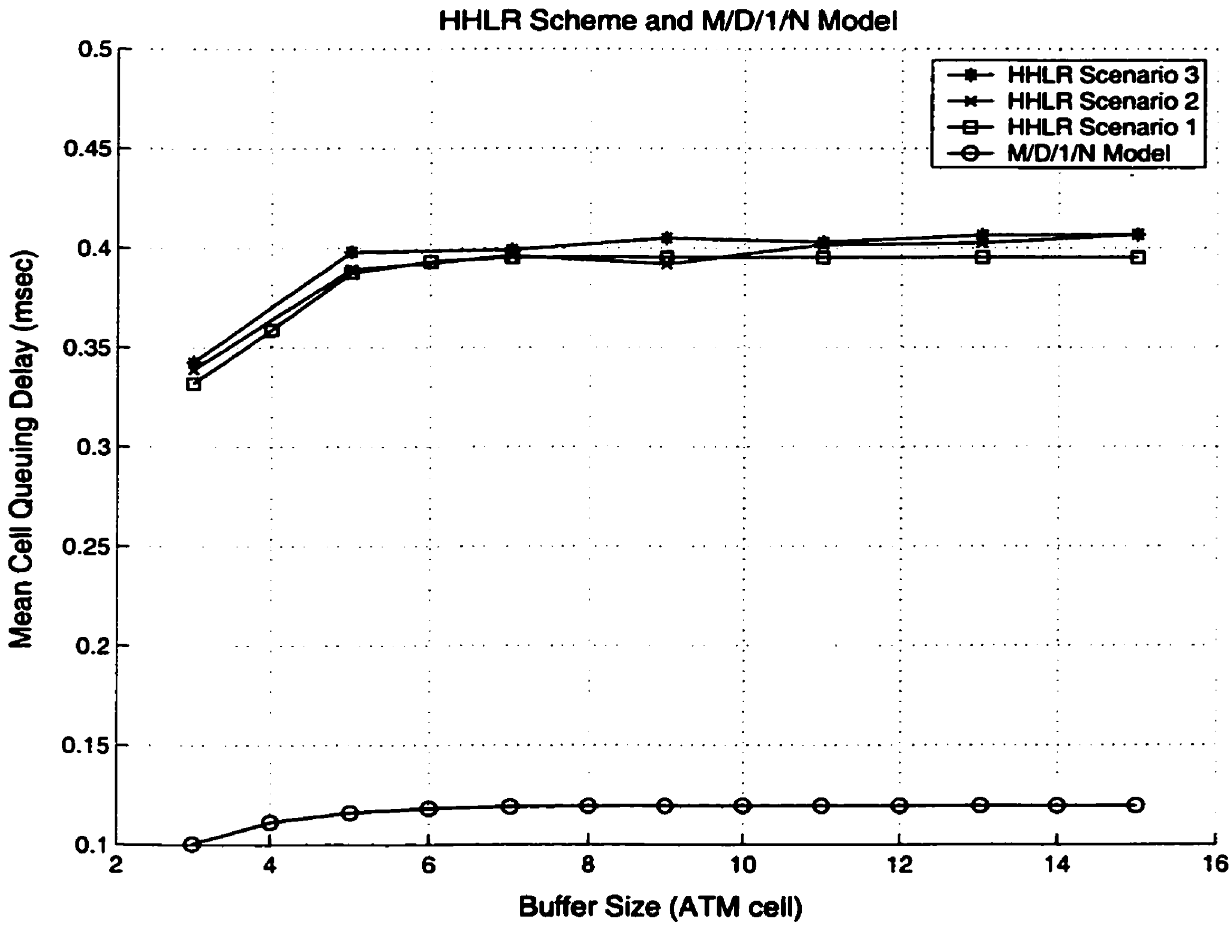

Figure 5.10 Comparison of simulated mean cell queuing delay for HHLR scheme with analytical results

Figure 5.10 compares the simulated mean cell queuing delay for HHLR scheme with analytical results. The non-agreement of the simulation results with the approximate analytical results is explained as follows. In HHLR, during handoff procedure, ATM cells are buffered and none of them can be sent out until the new connection path is ready. It is seen from Figure 5.10 that the analytical results are three times less than the simulation results. This means that the main cause of the delay is the handoff procedure delay when no ATM cell can be sent. This is why one of the objectives of SHLR scheme is to minimise the handoff latency and hence decrease the mean cell queuing delay.

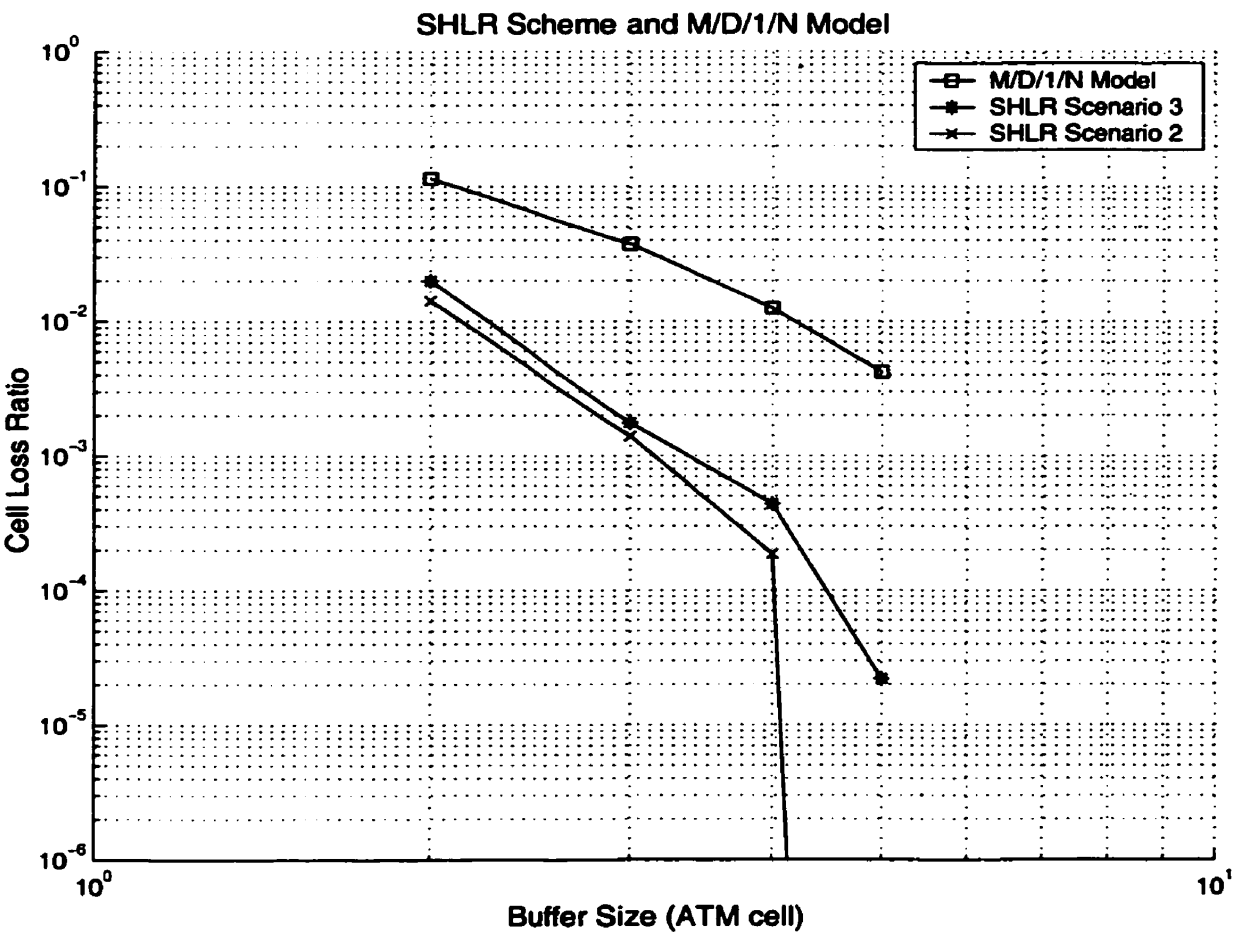

Figure 5.11 Comparison of simulated cell loss ratio for SHLR scheme with analytical results

Figure 5.11 compares simulated cell loss ratio for SHLR scheme with analytical results. The M/D/1/N model results are pessimistic compared to the simulation results. This is because the waiting period for duplicate ATM cell during the handoff procedure in SHLR scheme can cause buffer occupancy and hence loss for a finite buffer size. In M/D/1/N model, the cell loss is mainly due to buffering in a finite buffer.

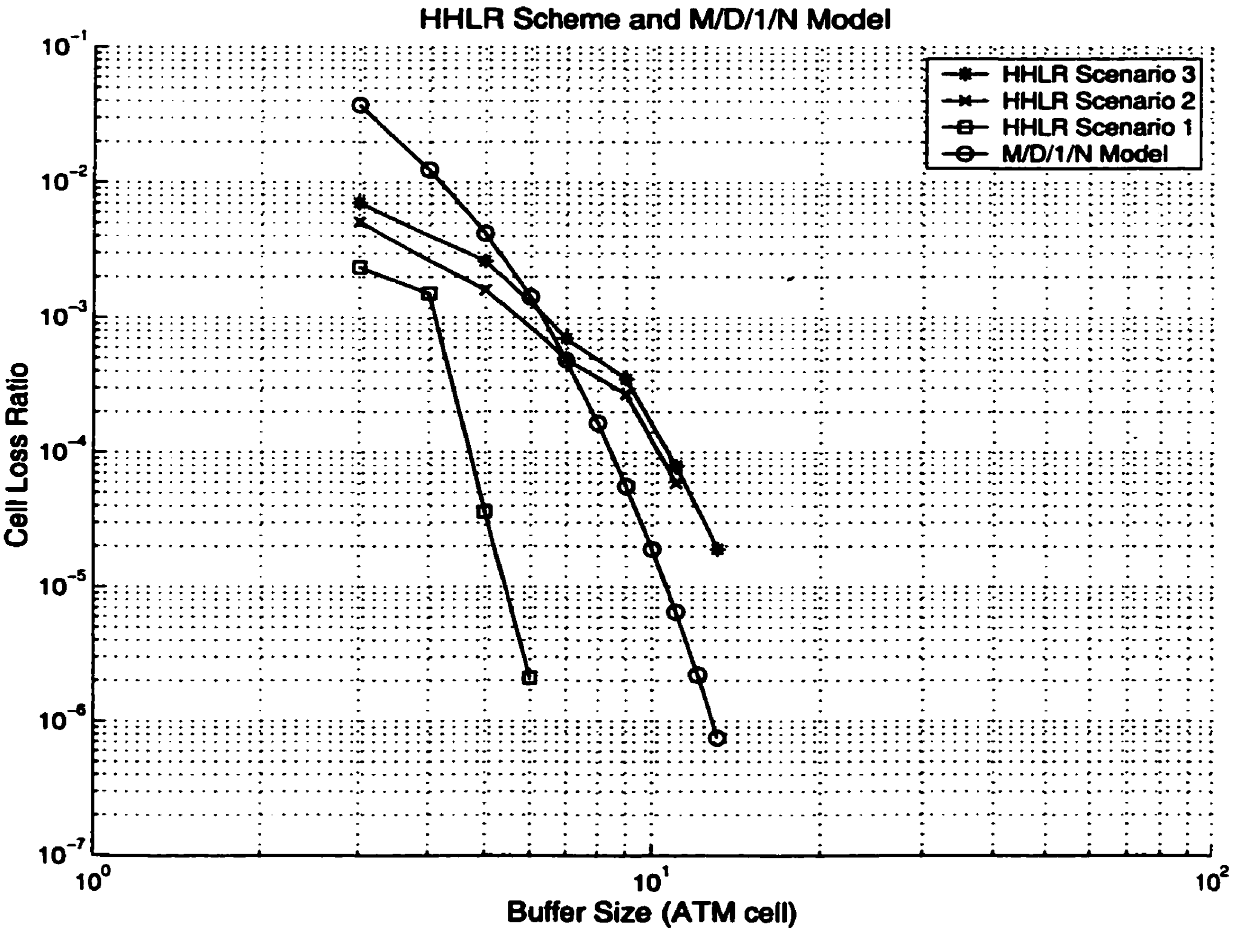

Figure 5.12 Comparison of simulated cell loss ratio for HHLR scheme with analytical results

Figure 5.12 compares simulated cell loss ratio for HHLR scheme with analytical results. Recall from the buffer requirement analysis for HHLR scheme, the uplink buffer delay during handoff procedure is mainly caused by the addition of propagation delay of old and new paths. This buffer occupancy during handoff is the main cause of the cell loss in the HHLR scheme. In M/D/1/N model, the cell loss is due to queuing in a buffer with finite size.

It can be seen from Figures 5.11 and 5.12 that the approximate analytical results for mean cell queuing delay are different from the simulated results. The reasons are as follows: analytical technique relies on several assumptions (e.g. assuming mobile's call activity follows Poisson process) in order to obtain a closed form solution. Analytical results are also obtained in a timely version, which is in contrast to long execution time of simulation runs.

Simulation technique enables the study of different scenarios of the SHLR and HHLR schemes as presented in Chapter 4. The approximate analytical approach does not provide this flexibility. Some of the analysis assumptions only approximate real-life situations and are made for analytical tractability. However compared to simulation analysis, analytical modelling is useful in evaluating the network performance because it enables necessary insights to be gained through sensitivity studies.

Chapter 6

Conclusion

This chapter concludes the thesis with a summary of the main contributions and suggestions for future extension of the research work on handoff schemes in wireless ATM networks.

6.1 The Main Contributions

6.1.1 Handoff Scheme Proposed

In this thesis a novel handoff rerouting scheme is proposed for mobiles moving from one base station to another base station in a W-ATM network. This handoff scheme is called soft handoff like rerouting (SHLR) scheme. SHLR is based on finding the ATM node that is a root of or common to both of the BSs involved in handoff. This common node is referred to as pivot node in the thesis, which serves as the anchor for ATM cell rerouting. The rerouting is then performed starting from this pivot node. Both radio link level handoff and network level handoff are studied. ATM cells are duplicated to both old and new path during the handoff procedure. Buffer is implemented in both pivot node and the mobile to temporarily store ATM cells during the handoff procedure. In particular, buffer is exploited to combat radio link destructive factors such as log-normal shadowing.

6.1.2 Methodologies Adopted

Both simulation and analytical approaches are used for evaluating the performance of the proposed SHLR scheme.

The performance evaluation considers simulation scenarios of three different traffic scenarios, combining CBR, VBR connections and shadowing environment are exploited. Furthermore, the performance of SHLR is compared with that of previously proposed HHLR scheme where the shadowing constraints are not taken into account.

The M/D/1/N queuing model is selected as an approximate analytical model to evaluate the performance of both SHLR and HHLR schemes.

6.1.3 Results Obtained

The performance metrics of interest are mean cell queuing delay and cell loss ratio. In this thesis, the performance of the SHLR is studied in a shadowing environment. Cell loss ratio is further classified to cell loss due to buffering and cell loss due to shadowing. The results for both SHLR and HHLR are presented as a function of buffer size.

6.1.4 Major Conclusions from the Results

The simulation results show that the proposed SHLR algorithm is promising and provides better loss and delay performance than those of the HHLR scheme.

For the same simulation scenario, the cell loss ratio and mean queue delay for SHLR scheme are less than those for HHLR scheme. For example, at buffer size 5, the cell loss ratio for SHLR scheme is on the order of 10^{-4}, compared to 10^{-2} for HHLR scheme. SHLR scheme provides two orders of magnitude improvement in cell loss performance over that for HHLR scheme. The mean cell queuing delay for HHLR tends to 0.4 msec, but for SHLR scheme, the mean cell queuing delay tends to 0.175 msec.

In addition, for all scenarios simulated, SHLR scheme gives about 2 times less cell loss due to shadowing than those results obtained for HHLR. The results confirm that making use of duplicated ATM cells in the buffer can improve radio link quality.

The approximate results obtained using the M/D/1/N model are somewhat optimistic compared to the simulated queuing delay results for both SHLR and HHLR schemes because the handoff delay is not considered in the analytical M/D/1/N model. The cell loss results obtained using the M/D/1/N model are pessimistic compared to the simulation results for both SHLR and HHLR schemes because the path difference is less than the queuing delay in the simulation scenarios.

6.1.5 Significance of the Results

The improvements that SHLR scheme offers over the HHLR scheme are quite significant: in terms of mean cell queuing delay and cell loss ratio due to shadowing, SHLR scheme is 2 times better than that of HHLR scheme; in terms of cell loss ratio due to buffering, SHLR scheme is at least 6 times better than that of HHLR scheme.

The results presented in this thesis are useful for dimensioning buffer requirements according to the QoS requirement. For example, with a network under a condition of simulation scenario 3, to have mean cell queuing delay less than 0.4 sec and cell loss ratio less than 10^{-4}, a buffer of size 7 should be allocated in the pivot node.

The results also help to dynamically allocate buffer space according to the number of simultaneous handoffs in progress. For example, to meet cell loss ratio of 10^{-5}, if only 1 mobile is performing handoff, a buffer of size 5 is good enough, whereas a buffer size of 13 is required when 4 mobiles are performing handoff simultaneously. An additional feature of SHLR scheme is the ability of recognising shadowing constraints associated with the handoff process.

The results in this thesis are useful for analyzing the network performance to gain an understanding on the major cause of cell loss: buffering or shadowing. In SHLR scheme scenario 3, if the buffer size is 4, the main cause of cell loss is shadowing and if the allocated buffer size is 2, the main cause of cell loss is due to lack of buffer space.

6.2 Future Work

SHLR scheme offers a practical solution to handoff issues in W-ATM networks. This thesis has demonstrated the advantages of the SHLR scheme in terms of low cell loss ratio, low mean queue delay and low shadowing loss. However, further evaluation is required in the following areas:

1. Work should be done to verify the better performance for different values of path differences (e.g. path difference larger than ATM cell inter-arrival time), which is the major cause for buffer occupancy of SHLR scheme.
2. Traffic with a larger burstiness (e.g. increase VBR source data activity) should be examined to compare with lower burstiness traffic. Efforts should be made in using other analytical models, e.g., MMPP/D/1/N model for bursty traffic source analysis.
3. Although SHLR scheme is designed to have a timeout set to meet delay requirement, it will be beneficial to investigate a scenario with different timeout limits for different delay requirement of ATM network applications (time-sensitive traffic and throughput-dependent traffic).
4. In a W-ATM network, mobiles may have multiple connections to different end users at the same time. Efforts including further simulation evaluation under this situation should therefore be made.

Reference

[1] Masahiro Umehira, Masamitsu Nakura, Hijin Sato, and Akira Hashimoto, "*ATM Wireless Access for Mobile Multimedia: Concept and Architecture*", IEEE Personal Communications, Oct. 1996, pp. 39-48.

[2] Bora A. Akyol and Donald C. Cox, "*Rerouting for Handoff in a Wireless ATM Network*", IEEE Personal Communications, Oct. 1996, pp. 26-32

[3] Bora A. Akyol and Donald C. Cox, "*Handling Mobility in a Wireless ATM Network*", Proc. INFOCOM'96, San Francisco, CA. March 24-28, 1996.

[4] Malathi Veeraraghavan, Mark J. Karol, and Kai Y. Eng, "*Mobility and Connection Management in a Wireless ATM LAN*", IEEE JSAC, Vol. 15, No. 1, January 1997, pp.50-68.

[5] Masaya Nishio, Noriteru Shinagawa, and Takehiko Kobayashi, *"A Lossless Handover Method for Video Transmission in Mobile ATM Network and Its Subjective Quality Assessment"*, IEEE Communications Magazine, November 1999, pp. 38 - 44.

[6] A. Akyol and Donald C. Cox, "*Signaling Alternatives in a Wireless ATM Network The Migratory Signaling Approach*", IEEE Journal on Selected Areas in Communications, Vol. 15, No. 1, January 1997, pp. 35 - 49.

[7] R. Thomas, H. Gilbert and G. Mazziotto, "*Influence of the moving of the mobile stations on the Performance of a Radio Mobile Cellular Network,*" Proc. 3rd Nordic Seminar, Paper 9.4, Copenhagen, Denmark, September 1988.

[8] Marco Ajmone-Marsan, Carla-Fabiana Chiasserini, Renato Lo Cigno, and Maurizio Munafo, "*Local and Global Handovers for Mobility Management in Wireless ATM Networks*", IEEE Personal Communications, Oct. 1997, pp. 16-24.

[9] David E. McDysan, Darren L. Spohn, *"ATM: theory and application"*, McGraw-Hill, 1994.

[10] Sheldon M. Ross, *"Introduction to Probability Models"*, Second Edition, Academic Press, 1981.

[11] G. Colombo, B. Jabbari, A. Nakajima, J. Kulkarni, "*Network Architecture and Traffic Issues for Wireless Personal Communications*", Dec. 1995, Volume XXIII n.6, pp. 863-873.

[12] Dipankar Raychaudhuri, "*WirelessATM Network: Architecture, System Design and Prototyping*", IEEE Personal Communications, Aug. 1996, pp. 42-49.

[13] Bernhard Walke, Dietmar Petras, and Dieter Plassmann, "*Wireless ATM: Air Interface and Network Protocols of the Mobile Broadband System*", IEEE Personal Commnuications, Aug. 1996, pp. 50-56.

[14] Anthony Acampora, "*Wireless ATM: A Perspective.on Issues and Prospects*", IEEE Personal Communications, Aug. 1996, pp. 8-17.

[15] Mischa Schwartz, "*Network Management and Control Issues in Multimedia Wireless Networks*", IEEE Personal Communications, June 1995, pp. 8-16.

[16] Ender Ayanoglu, Kai Y. Eng, and Mark J. Karol, "*Wireless ATM: Limits, Challenges, and Proposals*", IEEE Personal Communications, Aug. 1996, pp. 18-34.

[17] R. Yuan, S. Biswas and D. Raychaudhuri, "*Mobility Support in a Wireless ATM Network*", 5th Workshop on 3rd Generation Wireless Information Networks, 1995, Kluwer publishers, pp. 335-345.

[18] Matthew Cheng, Subhashini Rajagopalan, Li Fung Chang, Gregory P. Pollini, and Melbourne Barton, "*PCS Mobility Support Over Fixed ATM Networks*", IEEE Communications Magazine, November 1997, pp.82 - 92.

[19] Haruo Akimaru, Marion R. Finley, and Kyoto Yamori, "*A Practical Dimensioning Method for ATM Systems*", IEEE Transaction on Comm. Vol. 47, No. 2, February 1999, pp. 311-315

[20] Gregory P. Pollini, "*Trends in Handover Design*", IEEE Communications Magazine, March 1996, pp. 82- 90.

[21] Sirin Tekinay and Bijan Jabbari, "*Handover and Channel Assignment in Mobile Cellular Networks*", IEEE Communications Magazine, November 1991, pp. 42 - 46.

[22] Victor Frost and Benjamin Melamed, "*Traffic Modeling For Telecommunications Networks*", IEEE Communications Magazine, March 1994, pp.70-81.

[23] Yannis Viniotis, "*Probability and Random Processes for Electrical*

Engineers", WCB McGraw-Hill, 1998.

[24] Jerry Banks, John S. Carson, "*Discrete-Event System Simulation*", Prentice Hall, 1984.

[25] William H. Press, Brian P. Flannery, William T. Vetterling, "*Numerical Recipes in C*", Cambridge University Press, 1988.

[26] Haruo Akimaru and Konosuke Kawashima, "*Teletraffic -- Theory and Application*", 2nd Edition, 1999.

[27] OPNET Technologies, Inc., "*OPNET Modeler manual Version 4.0.A*", 1998

[28] H. Heffes and D. M. Lucantoni, " *A Markov-modulated characterization of packetized voice and data traffic and related statistical multiplexer performance*, " IEEE J. Select. Areas Commun., vol. 4, Sept. 1986, pp. 856-868.

[29] Theodore S. Rappaport, "*Wireless Communications -- Principles and Practice*", Prentice Hall, 1996.

[30] Daniel Wong, Teng Joon Lim, "*Soft Handoffs in CDMA Mobile Systems*", IEEE Personal Communications, December 1997, pp. 6 -17.

[31] Anthony S. Acampora, "*An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks*", IEEE JSAC, Vol. 12, No. 8, October 1994, pp. 1365-1375.

[32] Andrea Baiocchi, Nicola Lefari Melazzi, Marco Listanti, Aldo Roveri and Roberto Winkler, "*Loss Performance Analysis of an ATM Multiplexer Loaded with High-Speed ON_OFF Sources*", IEEE JSAC, Vol. 9, No. 3, April 1991, pp. 388 -392.

[33] Leandro Fernandes, "*Developing a System Concept and Technologies for Mobile Broadband Communications*", IEEE Personal Communications, February 1995, pp. 54-59.

[34] M.J. Karol etal., "*BAHAMA: A Broadband Ad-Hoc Wireless ATM Local Area Network*", Proc. ICC'95, June 1995, pp. 1216 - 23.

[35] T. La Porta, "*Distributed Processing for Mobility and Service Management in Mobile ATM Networks*", Wireless ATM Networking Wksp., New York, June 1996, pp. 86-92.

[36] O. Yu and V. Leung, "*Extending B-ISDN to Support User Terminal Mobility over an ATM-Based Personal Communication Network*", Proc. GLOBECOM'95, 1995, pp. 2289 - 2293.

[37] B. Ngo and H. Lee, "*Queueing Analyses of Traffic Access Control Strategies with Preemptive and Nonpreemptive Disciplines in Wideband Integrated Networks*", IEEE JSAC, Vol. 9, No. 7, September 1991, pp. 1093-1109.

[38] Hiroshi Saito, Masatoshi Kawarasaki, and Hiroshi Yamada, "*An Analysis of Statistical Multiplexing in an ATM Transport Network*", IEEE JSAC, Vol. 9, No. 3, April 1991, pp. 359-369.

[39] Sueng-Yong Park, Vaduvur Bharghavan, Sung-Mo Kang, "*Data Link Level Support for Handoff in Wireless ATM Network*", IEEE International Conference on Communications, Montreal, Canada, June 11-14, 1998, pp. 765-769.

[40] Chris Choi and K. F. Cheung, " *Sequential handoff Algorithm for Cellular Mobile Communication*", GLOBECOM'97, Phoenix, Arizona, November 3 - 8, 1997, pp. 716 - 719.

[41] Hiroshi Nakamura, Hisakazu Tsuboya, Masatomo Nakano, and Akihisa Nakajima, "*Applying ATM to Mobile Infrastructure Networks*", IEEE Communications Magazine, January 1998, pp. 66 - 73.

[42] Janny Hu, "*Applying IP over wmATM Technology to Third-Generation Wireless Communications*", IEEE Communications Magazine, November 1999, pp. 64 - 67.

[43] A.O. Fapojuwo and E. Nikiforov, "*Base Station Buffer Requriements for Handover Traffic Processing in ATM-based Wireless Networks*", IEEE VTC'2000, Tokyo, Japan, May 15-18, 2000, pp. 1737-1741.

APPENDIX A

Traffic Concepts and Markov Processes

A.1 Traffic Load

A *call* is defined as a demand for a connection to the telecommunication systems. The *holding time* is defined as the duration of the call. The *traffic load* is defined as the total holding time per unit time [26].

A.2 Call Arrival Process

The call origination process is assumed to follow a Poisson processing characterised by the following: as $\Delta t \rightarrow 0$,

1. The probability that a call originates in time interval $(t, t+\Delta t]$ tends to $\lambda\Delta t$ independent of t, where λ is a constant.

2. The probability that two or more calls originate in $(t, t+\Delta t]$ tends to zero.

3. Calls originate independently of each other.

In this process, the probability $p_k(t)$ that k calls originate in time (0, t] is given by [10]:

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad \text{(A.1)}$$

This is the *Poisson distribution* with mean λt, where λ is called the *call arrival rate*. This model is also known as *Poisson arrival* or *Poisson input*.

From (A.1), the probability that no calls originate in (0, t] is given by:

$$p_0(t) = e^{-\lambda t} \qquad \text{(A.2)}$$

Hence, the distribution function of the inter-arrival time (probability that the arrival time is no greater than *t*) is given by:

$$A(t) = 1 - e^{-\lambda t} \quad \text{(A.3)}$$

which is the *exponential distribution* with mean λ^{-1}.

This call arrival process is a pattern for voice traffic and it can also be used to data traffic activities. The assumption is made for analytic simplicity.

A.3 CBR Traffic

In CBR traffic, ATM cells are generated continuously with a constant bit rate (peak bit rate) during the whole call connection time. The mean bit rate (MBR) during the call is the same as the peak bit rate. See Figure A.1 for CBR traffic model.

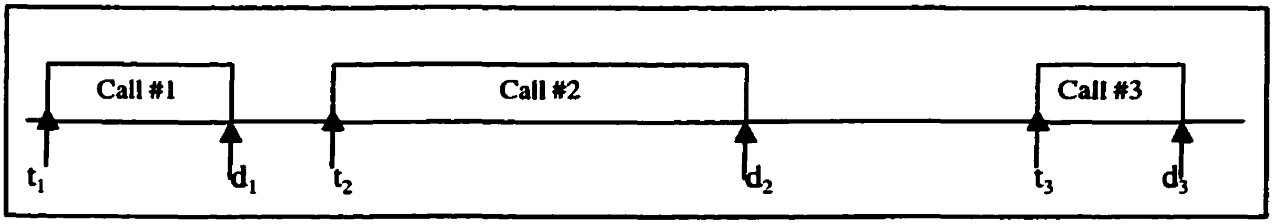

Figure A.1 CBR traffic model

The call arrival process of CBR traffic can be modelled by Poisson process with rate λ. The inter-arrival time [t_1 t_2] and [t_2 t_3], etc. between 2 calls has an exponential distribution with mean $1/\lambda$. The call duration or the call holding time [d_1 t_1] and [d_2 t_2], etc. also has an exponential distribution with mean *h*. A voice source is an example of a CBR source.

A.4 VBR Traffic

In VBR traffic, the calls are generated according to Poisson process. During the whole call connection time, there are ON periods and OFF periods. In ON periods, ATM cells are generated continuously with peak bit rate. During OFF periods, no ATM cell is generated. So, for VBR traffic, the mean bit rate (MBR) or the average bit rate during the call connection time is lower than the peak bit rate (PBR). The length of the ON period (Ion) and OFF period (Ioff) has exponential distribution. See Figure A.2 for VBR traffic model.

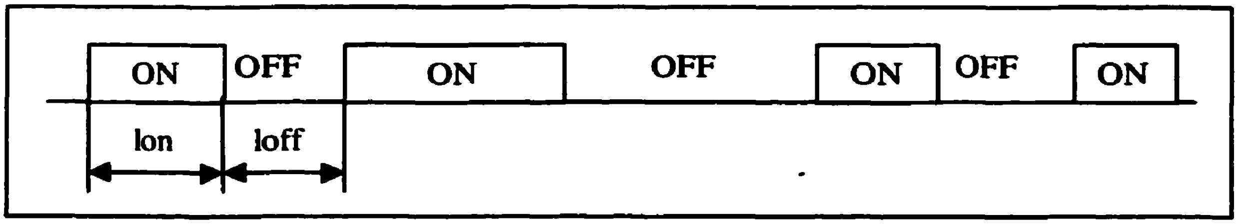

Figure A.2 VBR traffic model

The burstiness factor for the source is defined as β = Peak bit rate (PBR) / Mean bit rate (MBR), where MBR = PBR * $\frac{\text{Ion}}{\text{Ion} + \text{Ioff}}$. Data/Video source can be modelled by VBR traffic.

A.5 Markov Process and Renewal Process

A.5.1 Markov Process

A random process $\{X(t), t \in I\}$ is a Markov process if [23]

$$P[X(t_{n+1}) \le j \mid X(t_n) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1, X(t_0) = i_0]$$
$$= P[X(t_{n+1}) \le j \mid X(t_n) = i]$$

for any choice of $n, \{t_k\} \in I$. (A.4)

A Markov process with discrete state space is called a *Markov Chain*.

$$P[X(t_{n+1}) = j \mid X(t_n) = i, X(t_{n-1}) = i_{n-1}, \ldots, X(t_1) = i_1, X(t_0) = i_0]$$
$$= P[X(t_{n+1}) = j \mid X(t_n) = i]$$

for any choice of $n, \{t_k\} \in I$. (A.5)

A continuous-time Markov chain is a stochastic process that moves from state to state in accordance with a (discrete-time) Markov chain, but such that the amount of time it spends in each state, before proceeding to the next state, is exponentially distributed.

A.5.2 Renewal Process

If the interval X between consecutive occurrences of a certain event, say call arrival, is independently and identically distributed with a distribution function *F(x)*, the process $\{X\}$ is called a *renewal process*.

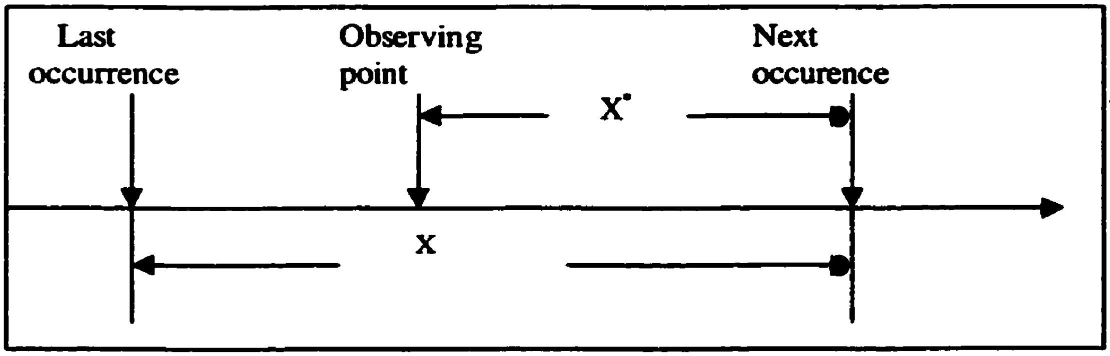

Figure A.3 Renewal process

In stochastic processes, a time instant at which the Markov property holds is called the *renewal point*. In a Markov chain, all time epochs at which the state changes, are renewal points.

A.6 Markov Property and PASTA

A random variable X is said to be memoryless or have Markov property, if

$P\{X > s+t \mid X > t\} = P\{ X > s\}$ for all X, t >0 (A.6)

Call inter-arrival time that is exponential distributed has Markov property.

Let P_j the probability that j calls exist at an arbitrary instant in steady state, and Π_j the corresponding probability just prior to call arrival epoch, as shown in Figure 5.4. For Poisson process with exponential inter-arrival time, they are equal:

$$P_j = \Pi_j \qquad (A.7)$$

This property is called Poisson Arrivals See Time Average (PASTA).

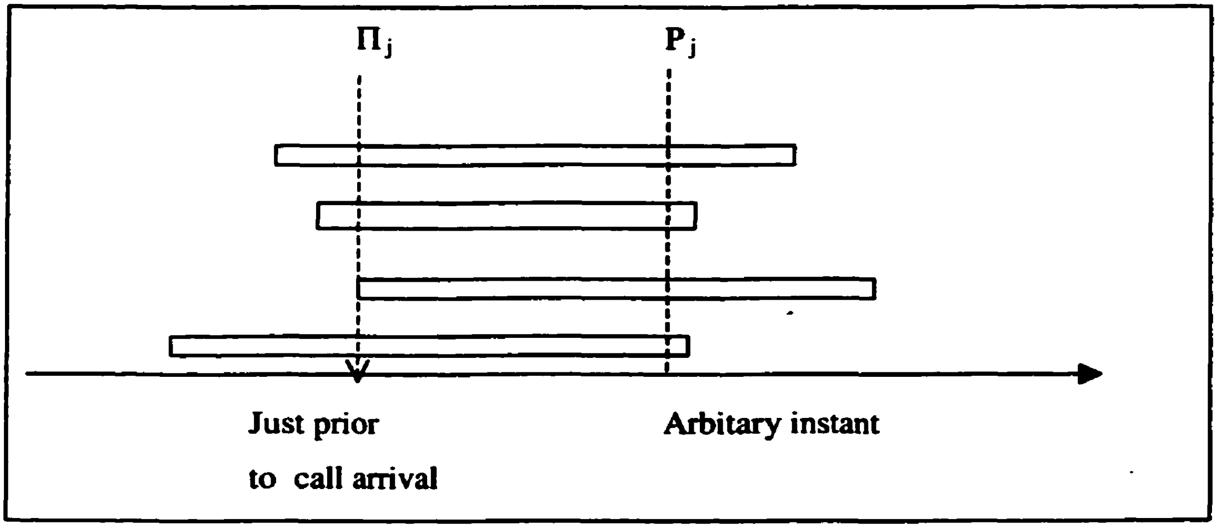

Figure A.4 Probabilities of system state

APPENDIX B

M/D/1/N Queuing Model Derivation

From M/G/1 queuing model, M/D/1/N is derived, hence in this section, M/G/1 system is first analysed. In M/G/1, *M* represents Markov process, *G* represents general distribution of the server service time and notation *1* represents one server. This is a case of infinite buffer size. [26]

B.1 M/G/1 Queuing Model

First of all, Little's formula is provided as follows to give the relationship between queue length and waiting time. This is an important property in queue model results analysis.

For a system in steady state, let λ be the arrival rate, *l* the mean number of waiting calls, and *W* the mean waiting time, which is equivalent to mean cell queuing delay, as shown in Figure 5.2. Then, The relation in general is:

$$l = \lambda W \qquad \text{(B.1)}$$

Equation (B.1) is known as the Little's formula.

Equation (B.1) has the meaning as follows: since *W* is the mean sojourn time of calls in the queue, if it is regarded as the mean holding time of the waiting ATM cells, the right hand side of (B.1) is the traffic load of the waiting calls. *l* is the average value of the buffer length seen by out side observers, while *W* is the delay experienced by arrivals.

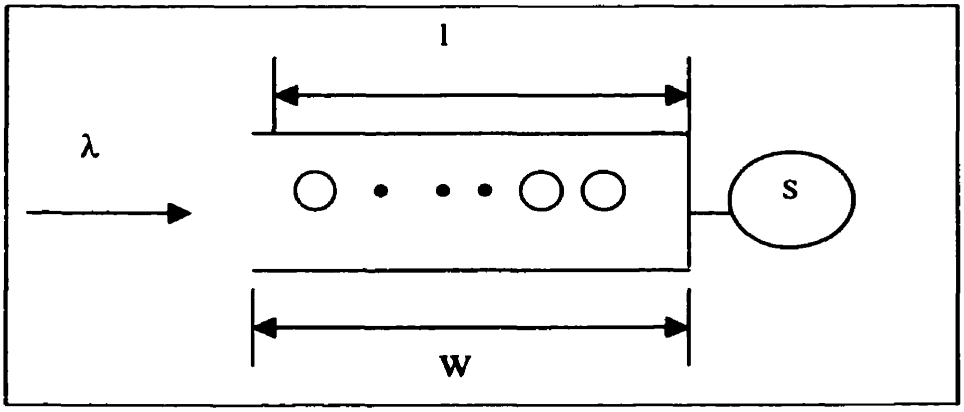

Figure B.1 Application of Little's formula

Consider the M/G/1 system with infinite buffer Poisson input at rate λ, and single server having a general service time distribution. The rest of this section gives M/G/1 and M/D/1/N queuing models analysis as in [26].

For the M/G/1 system, the departure epoch at which a call is terminated and leaves the system, becomes the renewal point. The reason is as follows:

Let k be the number of calls present in the system just after a call departure epoch in the steady state. Then,

(1) If $k = 0$, since the system is empty when the next call arrives, the call is immediately served, and the number of calls present just after its departure is equal to the number of arrivals during its service time.

(2) If $k >= 1$, since there are k calls waiting in the queue, the call at the top of the queue enters service, and the number of calls present at its departure is equal to k plus the number of arrivals during its service time.

In either case, because of the memory-less property of Poisson arrival, the number of calls present just after the next departure epoch is dependent only on the number k of calls existing just after the last departure epoch and independent of the previous progress, thus the Markov property holds (see Appendix A.5).

On the other hand, a stochastic process is called an *embedded Markov chain*, if the renewal points are embedded or hidden in particular time epochs such as call departure in M/G/1 systems.

In what follows, M/G/1 systems are analysed using the theory of embedded Markov chain.

Let Π^*_j be the probability that j calls exist just after a call departure in the steady state. Then, the relation can be given as,

$$\Pi_j^* = p_j\Pi_0^* + \sum_{k=1}^{j+1} p_{j-k+1}\Pi_k^*, j = 0,1,... \quad \text{(B.2)}$$

where p_j is the probability that j calls arrive during the service time.

Equation (B.2) is derived as follows: the first term in the right-hand side is the probability of j calls arriving when $k = 0$ corresponding condition (1) above. The second term is the probability of $(j-k+1)$ calls arriving when $1 <= k < j+1$ corresponding to condition (2) (one call is the departing call itself). Thus, the right-hand side probability is equal left-hand side Π^*_j. See Figure 5.3 for embedded Markov Chain.

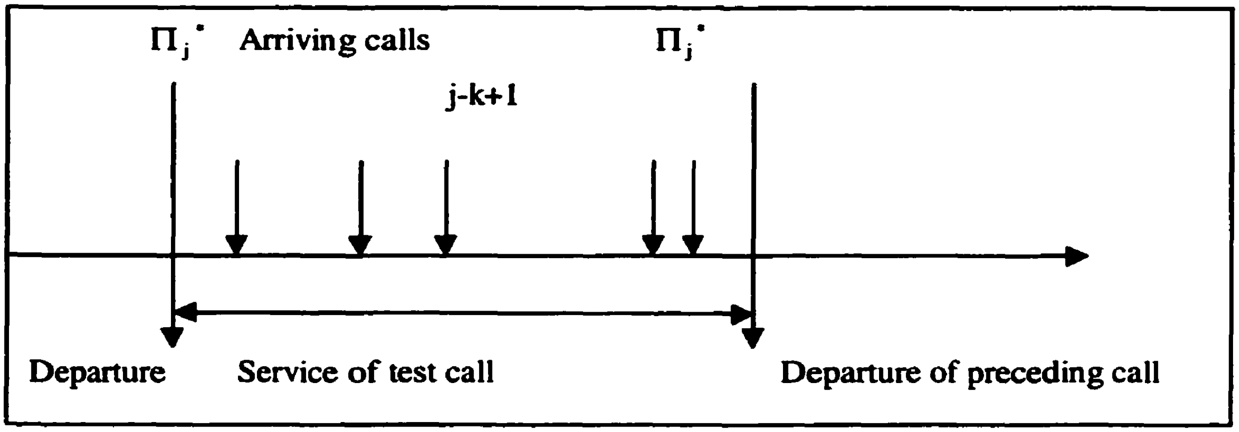

Figure B.2 Embedded Markov chain

Due to the Poisson process arrivals, the conditional probability that j calls arrive in the service time, given the service time is equal to x, is

$$p_j(x) = \frac{(\lambda x)^j}{j!}e^{-\lambda x} \quad \text{(B.3)}$$

From the total probability theorem, it can be derived

$$p_j = \int_0^{\infty} \frac{(\lambda x)^j}{j!} e^{-\lambda x} dB(x) \quad \text{(B.4)}$$

where B(t) is the service time distribution.

Define the probability generating function of Π^*_j by

$$g(z) = \sum_{j=0}^{\infty} z^j \Pi_j^* \quad \text{(B.5)}$$

multiplying (B.2) by z_j and summing, it can be derived

$$g(z) = \Pi^*_0 \sum_{j=0}^{\infty} z^j p_j + \sum_{j=1}^{\infty} \Pi^*_j \sum_{k=0}^{\infty} z^{j+k-1} p_k \quad \text{(B.6)}$$

Under a steady state condition

$$\lim_{z \to 1} g(z) = \sum_{j=0}^{\infty} z^j \Pi^*_j = 1, \quad \text{(B.7)}$$

and

$$\sum_{j=0}^{\infty} z^j p_j = \int_0^{\infty} \sum_{j=0}^{\infty} z^j \frac{(\lambda x)^j}{j!} e^{-\lambda x} dB(x) \quad \text{(B.8)}$$

$$= \int_0^{\infty} e^{z\lambda x} e^{-\lambda x} dB(X)$$

$$= \int_0^{\infty} e^{-(\lambda - z\lambda)x} dB(x)$$

$$= b^*(\lambda - z\lambda)$$

where $b^*(\theta)$ is the *Laplace-Stieltjes* transform of the service time distribution function $B(t)$. With the definition of *Laplace-Stieltjes* transform (LST)

$$f^*(\theta) = \int_0^{\infty} e^{-\theta x} dF(x) \quad \text{(B.9)}$$

It can be calculated Π^*_j, by using the inversion formula of probability generating function,

$$\Pi^*_j = \frac{1}{j!}\lim_{z\to 0}\frac{d^j}{dz^j}g(z) = \frac{1}{j!}g^{(j)}(0) \qquad \text{(B.10)}$$

Denoting the state probability at call arrival in the steady state by Π_j, it is known that $\Pi_j = \Pi^*_j$ in a stochastic process with discrete state space whose state transitions occur only by one step at a time, and this is the case for M/G/1 model under consideration. Mover, letting P_j be the state probability seen by an outside observer, from the PASTA the relation is:

$$P_j = \Pi^*_j = \Pi_j \qquad \text{(B.11)}$$

B.2 M/D/1/N Queuing Model

From now on, the case with finite buffer size N is considered. Using the same notation as before, and noting that $\Pi^*_j = 0$ for $j \geq N+1$ since a departing call can leave behind at most N calls, from (B.2) it can be derived

$$\Pi^*_j = p_j\Pi^*_0 + \sum_{k=1}^{j} p_{j-k+1}\Pi^*_k + p_0\Pi^*_{j+1} \qquad \text{(B.12)}$$

so

$$\Pi^*_{j+1} = \left(\Pi^*_j - p_j\Pi^*_0 - \sum_{k=1}^{j} p_{j-k+1}\Pi^*_k\right)p_0^{-1}, j = 0,1,\ldots,N-1 \qquad \text{(B.13)}$$

Equation (B.5) is the probability generating function of p_j, so p_j can be calculated by

$$p_j = \frac{1}{j!}\lim_{z\to 0}\frac{\partial^j}{\partial z^j}b^*([1-z]\lambda) \qquad \text{(B.14)}$$

For constant service time (D), the LST is given by:

$$b^*(\theta) = e^{-h\theta} \qquad \text{(B.15)}$$

where h is call holding time. From (B.7) it can be derived the probability that j calls arrive during the service time:

$$p_j = e^{-a}\frac{a^j}{j!}, \qquad j = 0,1,\ldots \qquad \text{(B.16)}$$

The state probability is:

$$P_j = \Pi_j = \Pi^*{}_j,\ j = 0,1,\ldots,N \tag{B.17}$$

$$P_{N+1} = \Pi_{N+1} \tag{B.18}$$

setting

$$C_j = \Pi^*{}_j / \Pi^*{}_0 \tag{B.19}$$

From Equation (B.13), it can be derived

$$C_{j+1} = (C_j - p_j + \sum_{k=1}^{j} p_{j-k+1} C_k) p^{-1}{}_0,\ j = 0,1,\ldots,N-1. \tag{B.20}$$

By recursively solving (B.20) for C_j, set

$$C = 1 + \sum_{j=1}^{N} C_j \tag{B.21}$$

noting (B.17), it can be derived

$$P_j = \frac{C_j}{1 + aC} \tag{B.22}$$

where a is the offered load

$$a = \lambda h \tag{B.23}$$

Using (B.22), the interesting result, probability of blocking (cell loss ratio), can be given:

$$P_B = P_{N+1} = 1 - \frac{C}{1 + aC} \tag{B.24}$$

The mean number of waiting customers in the system is given by:

$$L = \sum_{j=1}^{N+1} (j-1) p_j \tag{B.25}$$

For calculation convenience, it can be derived:

$$W = P_0 \left[NC - a^{-1} \sum_{k=1}^{N} (N - k + 1) C_k \right] h \tag{B.26}$$