RZ 3159 (# 93205) 09/16/99
Computer Science/Mathematics 98 pages

Research Report

Connection Handover in Wireless Mobile ATM Networks

Sonja Buchegger

IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland

Research
Almaden · Austin · Beijing · Haifa · T.J. Watson · Tokyo · Zurich

Connection Handover in Wireless Mobile ATM Networks

Sonja Buchegger

IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Abstract

When mobile wireless ATM (Asynchronous Transfer Mode) networks, for example in airplanes or ships, move from one access point in a coverage area to another, it is desirable to preserve the active connections with their respective negotiated quality of service parameters. In order to achieve this, the connections are handed over to the new access point instead of dropping and reestablishing them.

In this thesis the issue of connection handover in mobile ATM networks is discussed. First an overview of ATM, mobile communications and wireless mobile ATM is given. Then the concept of connection handover and its algorithms are outlined. A list of quality criteria is established. Message and data flow scenarios are given for multiple connection handover approaches. They are followed by a thorough investigation of connection handover performance according to the given quality criteria. A comparison of the different handover approaches in terms of performance is made. Next, the influence of the network topology on the performance of the connection handover is investigated. Some conclusions are drawn from the overall investigations concerning connection handover.

Contents

List of Figures

List of Tables

1 Introduction

Mobile ATM (Asynchronous Transfer Mode) networks in airplanes and ships provide on board communication infrastructures. Connectivity from such mobile networks to land-based fixed networks is essential in order to provide additional services such as Internet connectivity, telephony, and real-time exchange of operations data.

ATM over wireless media and enabling mobility is a fairly complex combination. The benefits of ATM, namely providing quality of service for different kinds of requirements, be it voice, image, video or data transfer, should also be preserved in the wireless environment. In addition to the problems caused by the media, the challenge of mobility has to be mastered. In order to set up connections, the location of the mobile network has to be known for routing purposes. Mobility also has an effect on connections that have already been established.

In the case where a mobile network has connectivity to a fixed network, its access point may change over time. That is the case when the mobile ATM network moves from one satellite coverage area onto the next. It is essential that existing connections be preserved following the change of access point. The connections are handed over to the new access point, hence the term "connection handover". This way connections do not have to be torn down and set up from scratch again. A complete re-establishment of the connections would introduce a substantial disruption of the connection adversely affecting any application in terms of delay and cell loss. The mobility should be transparent to the users and their applications. Handover algorithms aim at minimization of the impact of mobility providing for a seamless change of access points.

Algorithms for connection handover used for mobile end-user systems (e.g., mobile telephones, mobile PCs) may not apply directly in this environment due to the fact that in this case an entire network (switches plus local end systems) is mobile.

Figure 1.1 shows an airplane roaming between satellite coverage areas.

The rest of the thesis is outlined as follows:

Chapter 2 shows a short overview of mobile communications in general and explains the arising issues and challenges when parts of the network move. In contrast to communications in fixed networks, a permanent address can not automatically be mapped to a location, because the location can change over time. The impacts of mobility on addressing, routing, signaling, etc. are outlined. Also some examples of existing mobile communications systems are given.

Chapter 3 gives a short introduction to ATM (Asynchronous Transfer Mode). The concepts of cells, virtual connections and virtual paths are presented along with a de-

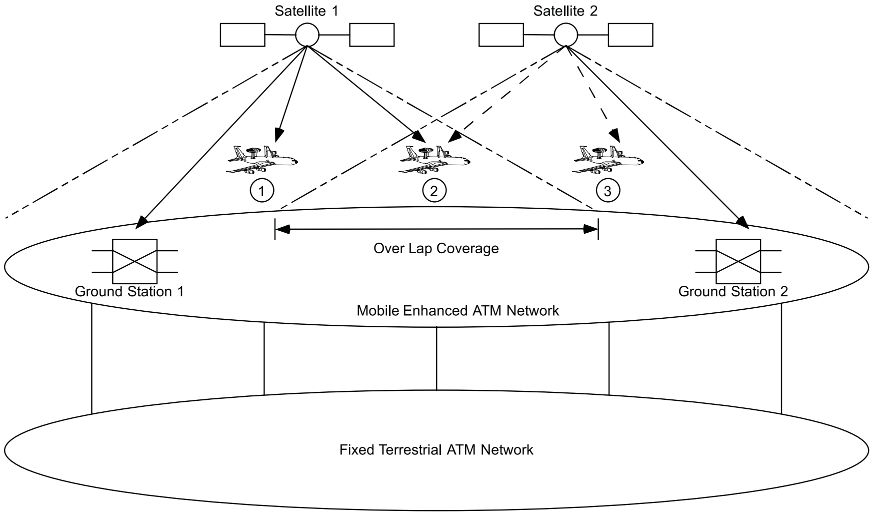


Figure 1.1: Mobile ATM Network Roaming
[Ray96a]

scription of the ATM protocol stack and architecture. A special emphasis is put on the explanation of the provision for quality of service in ATM.

Chapter 4 is about wireless and mobile ATM. The rationale for wireless and mobile ATM is looked at as well as its differences to 'regular' ATM. ATM has been designed for a fixed and wired network, therefore, enabling mobility and introducing the wireless medium requires changes. What these changes are and how they are made is explained in this chapter.

Chapter 5 describes connection handover in general and for wireless mobile ATM networks in particular. Several different algorithms for connection handover are explained and evaluated. In order to be able to compare the different approaches, a list of quality criteria for connection handover is established. Open issues remaining for further research are listed.

Chapter 6 talks about performance of connection handover algorithms. Based on message and data flow scenarios for different handover algorithms, the performance of these algorithms in terms of quality requirements is investigated and compared. Numerical results are shown.

Chapter 7 draws conclusions on the overall investigations in the thesis.

2 Mobile Communications

This chapter is a short overview on mobile communication systems, giving a classification according to the ITU (International Telecommunication Union) [Pap97]. Then, an example for mobile computing is outlined for a comparison with the system that is the subject of this thesis, wireless mobile ATM.

2.1 Definitions, Classification

Public mobile Systems :

- **Cellular Mobile Telephone** One example is the Global System for Mobile Communications (GSM), others are IS-41, IS-54 and IS-95 in the US, RCR-27 in Japan, and DCS-1800, a new version of the GSM system.
- **Paging Systems** A popular system is the Short Message Service (SMS) of the GSM.
- **Wireless Local Loops** These are special radio systems for extending the wired telephone.
- **Cordless Telephone Systems** Examples for cordless telephone systems are CT-1, CT-2 and DECT (Digital European Cordless Telephone).
- **Global Satellite Mobile Systems** There are various projects planned and some already implemented. Satellites can be placed at different heights respective to the earth, allowing a distinction between LEOs, MEOs and GEOs. LEO (Low Earth Orbit) are the satellites closest to earth and in contrast to GEO satellites, which are stationary with respect to the earth and placed in higher altitude, LEOs change their position relatively to the earth. An example for LEO satellites is IRIDIUM, one for GEOs is INMARSAT.
- **Cellular Digital Packet Data Systems** Examples are Mobitex and ARDIS.

Non-Public Mobile Radio Systems :

- **Professional Mobile Systems (PMR)** These systems include dispatcher services for corporate nets, military and emergency use.
- **Cordless microphones**
- **Wireless Local Area Networks (WLAN)**
- **Maritime and Air Systems**

Future Systems :

- **UPT (Universal Personal Telecommunication)** UPT enables access to telecommunication services while allowing personal mobility. It enables each UPT user to participate in a user-defined set of subscribed services and to initiate and receive calls on the basis of a unique personal, network transparent UPT number across multiple networks at any terminal, fixed or mobile, irrespective of geographical location, limited only by terminal and network capabilities and restrictions imposed by the network operator.

FPLMTS (Future Public Land Mobile Telecommunication Systems) FPLMTS are systems which will provide telecommunication services to mobile or stationary users by means of one ore more radio links. This mobility will be unrestricted in terms of location within the radio coverage area. They will extend the telecommunication services of the fixed network to those users over wide geographic areas, subject to constraints imposed by spectrum allocation and radio propagation and, in addition, will support a range of services particular to mobile radio systems.

PCS (Personal Communication Service) PCS is an extension and integration of existing and future wireless and wired communication networking features and capabilities, ultimately allowing communication with a person, regardless of his/her location. PCS features include: unified access, personalized service profiles, personal phone number, support for personal and terminal mobility, improved digital communication, etc.

2.1.1 Classification Based on Technology

According to the Position of Mobile Stations Land mobile systems

Maritime mobile systems

Air mobile systems

Special portable systems

According to the Position of Base Stations Land mobile systems

Satellite mobile systems

According to the Services Two directional voice (telephony)

Two directional data

Data broadcasting

Voice/video broadcasting

Paging

Navigation and positioning

According to the Signaling and Networking Analog and digital systems

Global and local (cellular) systems

According to the Frequency Band VHF systems (70 - 160 MHz)

UHF systems (e.g. 450 MHz)

Cellular standard systems (800 - 900, 1800 - 1900 MHz)

L-band satellite systems (e.g. 1500 MHz)

Microwave mobile networks (20 - 60 MHz)

2.2 Problems

Wireless Related Problems :

Fading Due to the limited coverage area, the signal strength decreases as the mobile moves towards the edge of the coverage area. When it falls below a threshold, a handover to another coverage area has to be done. When the signal is very weak, it can hardly be distinguished from noise.

Shadowing Buildings or Mountains get in the way between the mobile and the base station, preventing the signal from being propagated directly.

Frequency Band, Multiple Access The Frequency Band is limited and the number of users is increasing. Different algorithms have been developed and implemented to give access to multiple users at once. Amongst those algorithms are TDMA (Time Division Multiple Access), assigning time slots to the users, FDMA (Frequency Division Multiple Access), separating the users by frequencies, and CDMA (Code Division Multiple Access), giving a different code to every user to prevent interference.

Multipath Propagation The signal is sometimes reflected from surfaces resulting in the effect that the user may receive the same signal from different directions and at different times.

Mobility Related Problems :

Handover When the mobile moves from one coverage area to the next, the existing communications should be preserved. The roaming should be as seamless as possible and transparent to the far end party that communicates with the mobile.

Location Management When the mobile roams between coverage areas it needs mechanisms to be found for incoming communication.

2.3 Mobile Computing Example: Mobile IP

[Per97] An obstacle towards mobility in IP is the use of the IP address. Consider how IP addresses are used today in the Internet. In the first place, they are primarily used to identify a particular end system. IP addresses are also used to find a route between the endpoints. The route does not have to be the same in both directions. Putting these two uses together results in a situation fraught with contradiction for mobile computing. On one hand, a mobile computer needs to have a stable IP address in order to be stably identifiable to other Internet computers. On the other hand, if the address is stable, the routing to the mobile computer is stable, and the datagrams always go essentially to the same place. Mobile IP extends IP by allowing the mobile computer to effectively utilize two IP addresses, one for identification, the other for routing.

Mobile IP introduces the following new functional entities:

Mobile Node A host or router that changes its point of attachment from one network to another, without changing its IP address. A mobile node can continue to communicate with other Internet nodes at any location using its constant IP address.

Home Agent A router on a mobile node's home network which delivers datagrams to departed mobile nodes, and maintains current location information for each.

Foreign Agent A router on a mobile node's visited network which cooperates with the home agent to complete the delivery of datagrams to the mobile node while it is away from home.

A mobile node has a home address, which is a long-term IP address on its home network. When away from home, a care-of address is associated with the mobile node and reflects the mobile node's current point of attachment. The mobile node uses its home address as the source address of all IP datagrams it sends.

Mobile IP is a way of performing three related functions:

Agent Discovery Mobility agents advertise their availability on each link for which they provide service.

Registration When the mobile node is away from home, it registers its care-of address with its home agent.

Tunneling In order for datagrams to be delivered to the mobile node when it is away from home, the home agent has to tunnel the datagrams to the care-of address.

The following will give a rough outline of operation of the mobile IP protocol, making use of the above-mentioned operations.

Mobility agents make themselves known by sending agent advertiement messages. An impatient mobile node may optionally solicit an agent advertisement message.

After receiving an agent advertisement, a mobile node determines whether it is on its home network or a foreign network. A mobile node basically works like any other node on its home network whet it is at home.

When a mobile node moves away from its home network, it obtains a care-of address on the foreign network, for instance, by soliciting or listening for agent advertisements or contacting Dynamic Host Configuration Protocol (DHCP) or Point-to-Point Protocol (PPP).

While away from home, the mobile node registers each new care-of address with its home agent, possibly by way of a foreign agent.

Datagrams sent to the mobile node's home address are intercepted by its home agent, tunneled by its home agent to the care-of address, received at the tunnel endpoint (at either a foreign agent or the mobile node itself), and finally delivered to the mobile node.

In the reverse direction, datagrams sent by the mobile node are generally delivered to their destination using standard IP routing mechanisms, not necessarily passing through the home agent.

3 Asynchronous Transfer Mode (ATM)

This chapter gives an overview of ATM focusing on the areas that are relevant for the thorough understanding of the subsequent chapters. For a more detailed description of ATM see [DL95a], [DL95b], [Tan96].

3.1 Rationale for ATM

The main reason for the invention of ATM is to have a single high-speed network that integrates all kinds of information transfer like voice, video which are real-time traffic and can tolerate some loss of data but no delay, as well as non-real-time traffic (usually computer data), which can tolerate some time delay but no loss. Some data types are quite bursty (such as video transmission, needing a high bandwidth but only for a short time), others are sent at constant bit rates. The integration offers a solution to the problem given by multiple networks for telephony and data transfer and makes it possible to offer a variety of new services. By integrating and multiplexing different kinds of data in a single network, otherwise unused capacities and bandwidth can be better exploited.

The two main standardization bodies are the ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) and the ATM Forum, an industry group set up in 1991.

3.2 The ATM Network

An ATM network consists of ATM switches that are connected to each other via an interface called the NNI (Network-Node Interface). Terminals are connected to the switches by the UNI (User-Network Interface). If the switches are privately owned, the NNI between them is called Private Network Node Interface (PNNI). They are connected to public ATM switches by Public UNI.

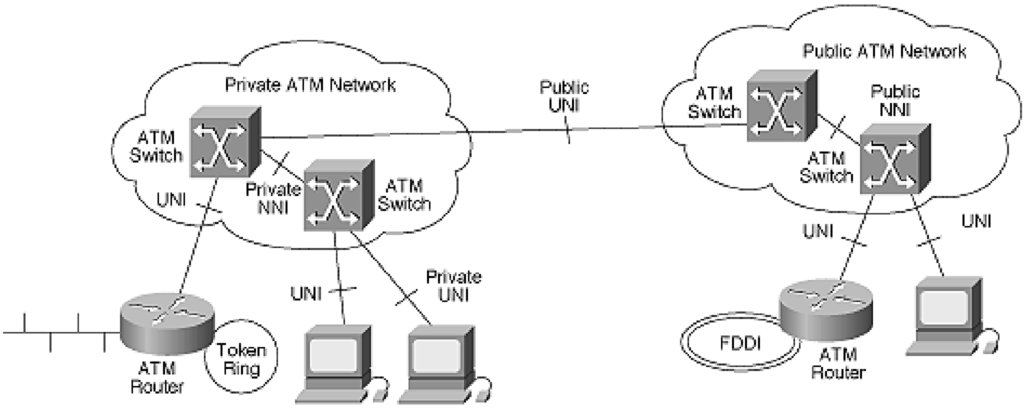


Figure 3.1: A Typical ATM Network

3.3 Basic Principles

3.3.1 Cells

ATM uses small, fixed sized packets, called cells, for data transmission. These cells consist of a 5 bytes header and 48 bytes for the payload. Since ATM has been designed to accommodate all different kinds of information, such as voice, video, images and data, the cell size is a compromise. Long packets have a better header/payload ratio reducing the overhead. Short packets have the advantage, that if one packet is lost or corrupt, not so much information is lost at once. Small packet sizes favor voice transmission for example, where occasional loss of small chunks of information can be tolerated. Data transmission would preferably use large packet sizes to efficiently move information without wasting much bandwidth for headers.

As opposed to packets, cells are fixed in size. The reason for not allowing for longer or shorter packages is mainly the so gained simplicity. ATM switching is done in hardware and fixed-sized cells speed up the switching process.

Figure 3.2 shows the content of a ATM cell. GFC stands for Generic Flow Control and is not used yet. The fields labeled VPI and VCI are small integers identifying the virtual path and virtual channel the cell takes to get to its destination. VPs and VCs are explained in the next section. The PTI field indicates the type of the payload. The CLP (Cell Loss Priority) bit can be set to distinguish between high and low priority cells. If congestion occurs and cells have to be discarded, the switch will attempt to discard cells that have the CLP bit set to one first. Finally, the HEC field is a checksum over the header.

Figure 3.3 illustrates the segmentation of packets into cells. The cells are then multiplexed in the switch and sent along the path to their destination.

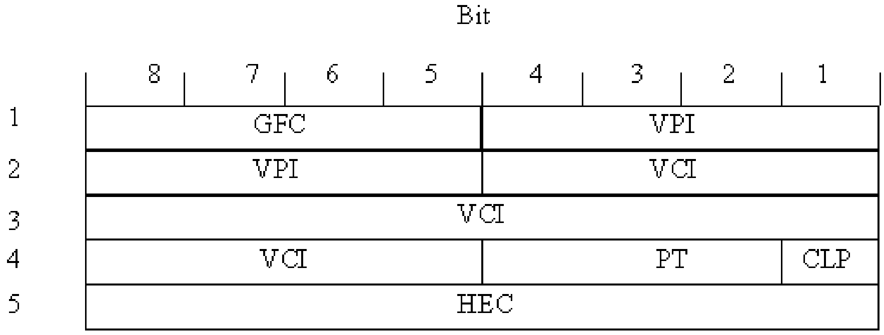


Figure 3.2: ATM Cell Header Format

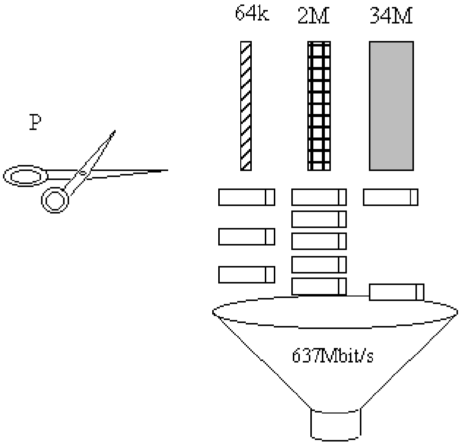


Figure 3.3: Segmentation and Multiplexing

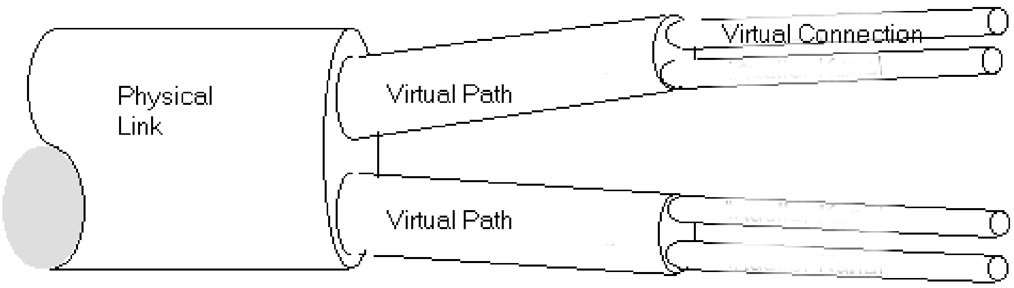


Figure 3.4: Physical Link/VPs/VCs

3.3.2 Connection Orientation

ATM is connection oriented. That means that before any cells are sent, a connection is set up between the source and the destination. Once the connection is established, all cells are sent along the same way through the network. A routing decision does only have to be made once at connection setup and not for each packet. As the cells all travel along the same path, they are guaranteed to be delivered in the same order as they were sent (in sequence). This is why the header of a cell does not need to contain any sequence number or even a destination address. When a connection is set up in one direction, a connection is also set up in the reverse direction allowing for bi-directional data transmission.

Two types of connections can be distinguished in ATM. Permanent and Switched Virtual Connections - PVCs and SVCs. PVCs are configured manually. These connections exist until they are manually released again. The other type of connection, SVCs, are set up using signaling. When one of the communication parties decide to terminate a connection, it is torn down and all reserved resources along the path are released.

ATM cells travel along so-called VCs (Virtual Channels) and VPs (Virtual Paths). VPs can contain several VCs. An SVC consists of all pairs of VCs and VPs between the calling and the called party.

Figure 3.4 illustrates the relation between the physical link, VPs and VCs.

3.3.3 Routing

In a switch, the traffic on a virtual connection is routed (switched) according to the VPI, VCI and incoming link of the data. A VPI/VCI is allocated for a virtual connection by the switch when the connection is set up, and it remains unchanged for the entire life time of the connection. It should be noted that the VPI/VCI values of a single connection are most likely to be different on different links. There is a routing table in each switch with routing information for all connections passing through the switch. The routing information includes the new VPI/VCI and the outgoing link for every incoming VC. The routing can also be done only according to the VPI. In this case the routing is called virtual path routing and the value of the VCI remains unchanged.

In order to find the correct entries for the routing table to alter the VPI and VCI in the cell header as the cells go through the switches, a path has to be computed from source to destination before the connection is set up. To choose a route through the network, the topology of the network together with information about its current state have to be known. The ATM routing protocol mainly considered in this thesis is PNNI [PNN96].

With PNNI a network can be divided into so called peer groups. Nodes within a peer group that have links to one or more other peer groups are called border nodes. Within each peer group a peer group leader is selected. This peer group leader represents the peer group to the rest of the network. PNNI provides a mechanism to build hierarchies of peer groups. The peer group leaders represent their groups as logical group nodes in higher layers of the hierarchy.

The advantage of such a hierarchy is that information of the network state and topology - and also updates of that information - is confined locally in peer groups. Only an aggregated representation of the network state of a peer group is given to the other peer groups of the same hierarchy. This mechanism limits the amount of information flooded through the entire network, thereby decreasing the overhead.

When a virtual connection is set up, a path from source to destination is calculated. Path selection in PNNI allows for different path computation methods such as shortest path or widest path. The result of the path selection is a so-called designated transit list or DTL, specifying the route to the destination.

3.3.4 Signaling

Signaling is the process used to set up new end-to-end VC or VP connections. When an end user makes a request for a connection to be set up, that request is sent from the end user to the switch to which the end user is connected using UNI signaling protocols. The end-user device communicates with the control point in the switch to which it is connected. PNNI signaling is used to set up the VC in each switch along the path of that connection.

The signals, that is to say the messages that are sent along the path from source to destination in order to set up a connection, follow a path specified by the so-called DTL (Designated Transit List) provided by PNNI.

To set up a connection, the calling party sends a *Setup* message to the switch it is attached to. The switch then sends back a *Call Proceeding* message to indicate that it has received the request. Between the switches along the path to the destination, the PNNI *Setup* and *Call Proceeding* messages are sent accordingly until the switch where the destination terminal is attached to sends a final UNI *Setup* to that terminal. If the destination agrees to establish the connection, it sends back a *Connect* message to its switch and gets a *Connect Ack* message back as acknowledgment. Between the switches on the way back to the source a PNNI *Connect* message is forwarded to the last switch, which in turn sends a final UNI *Connect* to the source terminal. After this process, the virtual connection is established and data can be sent.

Figure 3.5 shows the principle of how a connection is set up and how the parameters of the connection are negotiated.

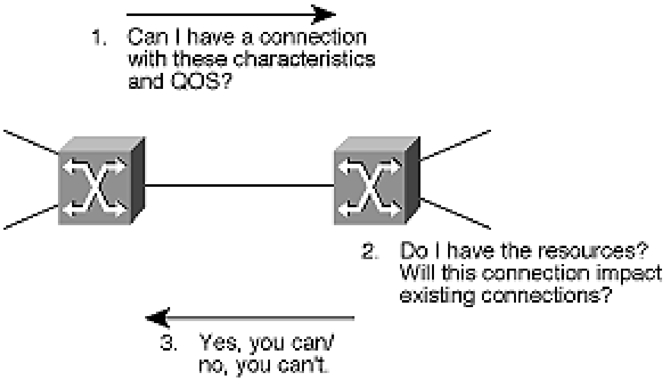


Figure 3.5: Connection Establishment

3.3.5 Traffic Classes/Service Categories

ATM is designed to transport different types of data.

Constant Bit Rate (CBR) The CBR connection type is used to carry constant bit rate traffic with a fixed timing relationship between data samples. In contrast to other network technologies, the bit rate can be specified in ATM.

Variable Bit Rate - real-time (rt-VBR) is intended for services that have variable bit rates combined with stringent real-time requirements, such as interactive compressed video (e. g. video-conferencing). Some compression schemes like MPEG for example send a complete base frame followed by a series of differences between the current frame and the base frame. This way, the transmission rates vary strongly in time. Despite this variation, it is very important that the ATM network does not introduce a delay variation of the cells (jitter) in order to prevent the video stream to be perceived in bad quality. In this case, both the average cell delay as well as the delay variation have to be tightly controlled. An occasional cell loss on the other hand can be tolerated and is best just ignored.

Variable Bit Rate - non-real-time (nrt-VBR) is for traffic where timely delivery is important, but a certain amount of jitter can be tolerated by the application.

Available Bit Rate (ABR) is a non-real-time traffic class like the nrt-VBR, but the bandwidth allocation is based on *best effort*. The network will give the connection as much bandwidth as possible, but it has to meet the minimum values of the traffic contract. The ABR introduces also flow control per virtual connection and is best suited for bursty traffic whose bandwidth range is roughly known.

Unspecified Bit Rate (UBR) does not offer any service guarantees. The user is free to send any amount of data up to a specified maximum, while the network does not promise any specific cell loss rate, delay or delay variation. This traffic class is best suited for IP packets, since IP does not make any promises about delivery either.

3.3.6 Quality of Service Parameters

When a virtual connection is set up with specified descriptors, these descriptors make up a traffic contract between the user and the network (or the customer and the carrier). The network will guarantee the type and quality of service agreed as long as the traffic conforms to the traffic contract.

To make it possible to have concrete traffic contracts, the ATM standard defines a number of QoS (Quality of Service) parameters whose values can be negotiated by the customer and the carrier. For each quality of service parameter, the worst case performance is specified, and the carrier is required to meet or exceed it. In some cases, the parameter is a minimum, in others it is a maximum. The quality of service is specified separately for each direction and consists of the following parameters:

PCR (Peak Cell Rate) is the maximum rate at which the sender is planning to send cells. This parameter may be lower than what the bandwidth of the line permits.

SCR (Sustained Cell Rate) is the expected or required cell rate averaged over a long time interval. For CBR traffic, SCR will be equal to PCR, but for all other service categories, it will be substantially lower. The PCR/SCR ratio is one measure of the burstiness of the traffic.

MCR (Minimum Cell Rate) is the minimum cell rate that the customer can accept. If ABR is specified, then the transferred cell rate can vary during the lifetime of the connection as long as it is always between the MCR and the PCR. If the minimum cell rate is set to zero, then ABR becomes similar to UBR traffic.

CDVT (Cell Delay Variation Tolerance) describes the maximal acceptable difference in arrival time deviation from the expected cell rate. Although the minimum, peak and sustained cell rates are specified, the cell still may not arrive uniformly according to the rates, but will deviate individually.

CLR (Cell Loss Ratio) measures the fraction of sent cells that are lost on the way or that arrive so late at the destination that they are considered useless, this is especially true for real-time data.

CTD (Cell Transfer Delay) is the average interval of time a cell needs to travel from source to destination. The terms average *end-to-end delay* and *transit time* are used synonymously as long as they are measured from source to destination.

CDV (Cell Delay Variation) describes how uniformly the cells are delivered.

CET (Cell Error Ratio) is the fraction of cells that arrive at the destination with one or more bits wrong.

SECBR (Severely-Errored Cell Block Ratio) is the fraction of N-cell blocks of which M or more cells contain an error.

CMR (Cell Misinsertion Rate) is the number of cells that are delivered to the wrong destination due to errors in the header of the cell.

It is important that no terminal violating its traffic contract can lower the QoS experienced by other terminals. It is necessary to ensure that a terminal is not sending more traffic than it is allowed to send. The function implementing this policy is called Usage

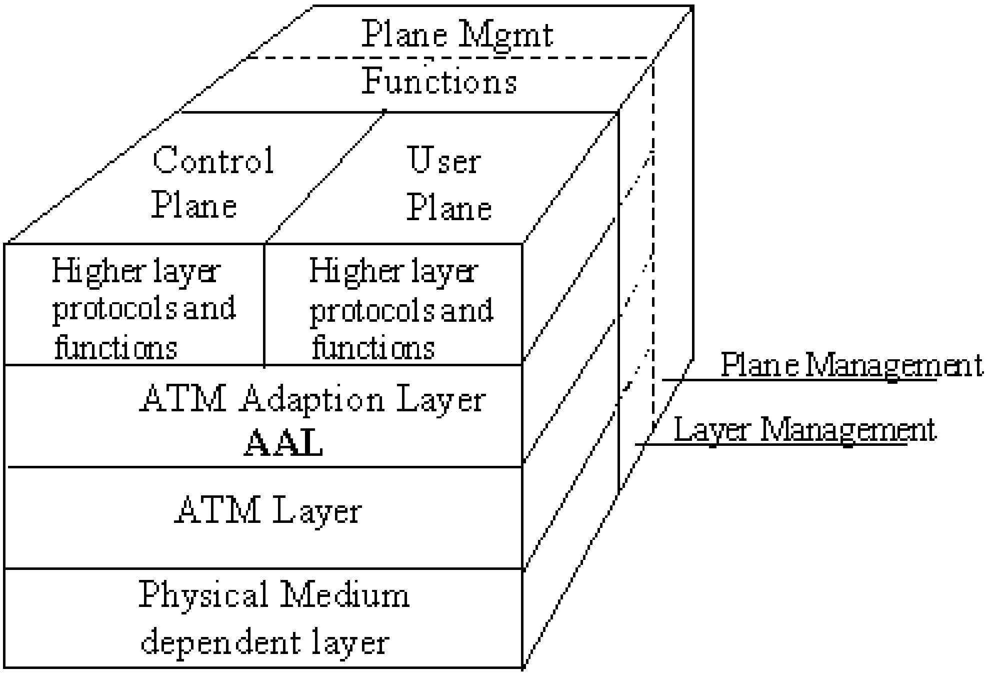


Figure 3.6: The ATM Reference Model

Parameter Control (UPC). The Generic Cell Rate Algorithm (GCRA) can be used to define conformance with respect to the traffic contract. For each cell arrival, the GCRA determines whether the cell conforms to the traffic contract of the connection. The GCRA is sometimes called the continuous-state leaky bucket algorithm.

3.4 Reference Model

The reference model 3.6 for ATM is different from the OSI and also from the TCP/IP model. It consists of three layers, the physical, ATM and ATM adaptation layers. The physical layer deals with the physical medium. ATM does not prescribe a particular set of rules, but instead says that ATM cells may be sent over the media by themselves, but they may also be packaged inside the payload of other carrier systems. ATM has been designed to be independent of the transmission medium. On the other hand, ATM has been designed with the reliability of fiber in mind, assuming low error probability.

The ATM layer deals with cells and cell transport. It defines the layout of a cell and what the header fields mean. It also deals with establishment and release of virtual circuits. Congestion control is also located here.

Because most applications do not want to work directly with cells, a layer above the ATM layer has been defined that allows users to send packets larger than a cell. The ATM interface segments these packets, transmits the cells individually, and reassembles them at the other end. This layer is the AAL (ATM Adaptation Layer).

Unlike the OSI reference model, the ATM model is defined as being three-dimensional. The user plane deals with data transport, flow control, error correction and other user functions. In contrast, the control plane is concerned with connection management. The layer and plane management functions relate to resource management and interlayer coordination.

The physical and AAL layers are each divided into two sublayers, one at the bottom to carry out the functionality and a convergence sublayer on top to provide the proper interface to the layer above it.

The PMD (Physical Medium Dependent) sublayer interfaces to the actual transmission medium. It moves bits on and off and handles the bit timing. For different carriers, this layer will be different.

The other sublayer of the physical layer is the TC (Transmission Convergence) sublayer. When cells are transmitted, the TC layer sends them as a string of bits to the PMD layer. At the other end, the TC sublayer gets a pure incoming bit stream form the PMD sublayer. It then has to convert this bit stream into a cell stream for the ATM layer.

The AAL layer is split into a SAR (Segmentation and Reassembly) sublayer and a CS (Convergence Sublayer). The lower sublayer breaks packets up into cells on the transmission side and puts them back together again at the destination. The upper sublayer makes it possible to have ATM systems offer different kinds of services to different applications (e.g., file transfer and video on demand have different requirements concerning error handling, timing, etc.).

4 Wireless Mobile ATM

4.1 Motivation

The WATM (wireless ATM) working group of the ATM forum, that was founded in 1996, has been working on a standard for wireless ATM. The standard is to be completed in 1999.

Broadband and mobile communications are presently the two major drivers in the telecommunications industry.[PPVM97] ATM is considered the most suitable transport technique for the future broadband integrated services digital network (B-ISDN), due to its ability to flexibly support a wide range of services with quality-of-service guarantees. On the other hand, wireless local area networks are becoming popular for indoor data communications because of their tetherlessness and increasing transmission speed. Wireless communications have been developed to a level where offered services can now be extended beyond voice and data. The combination of wireless communications and ATM can provide freedom of mobility with service advantages and quality-of-service guarantees. The main challenge of wireless ATM is to harmonize the development of broadband wireless systems with B-ISDN/ATM and ATM LANs, and offer similar advanced multimedia multiservice features for the support of time-sensitive voice communications, data traffic, video, and desktop multimedia applications.

An important motivation for mobile ATM [ALRR97],is to provide a high-speed backbone network that supports mobility and thereby provide a common infrastructure network to a diverse set of mobile technologies. The choice of ATM as the backbone network for a generic mobility supporting infrastructure is motivated by

- ATM's superior cost/performance when used as a switching technology for large traffic volumes.

-ATM's rich transport capabilities which allows different traffic types to be carried over the same network.

There is a homogeneity argument in favor of wireless and mobile ATM. ATM is a scalable technology and appears likely to be at the core of future multimedia networks. Therefore, extending the QoS-specifiable ATM VC model over the wireless hop leads to a homogeneous end-to-end network, simplifying the architecture. QoS specification can be used, for example, to give some connections higher priority during handover (see chapter 5), or to use rerouting policies tailored to the traffic characteristics. In general, VCs with QoS parameters in wireless and mobile ATM provide the ability to meaningfully distinguish the data packets being sent over the air, and not treat all of them according to one generic policy. [AHK+96]

4.2 Definitions

Communications between any set of devices can be visualized in relation to two sets of diametrically opposed characteristics: Fixed versus Mobile and Wireless versus Wired. These two sets can be used to separate the communications environment into four domains as shown in Table 4.1. [Gro98]

The four domains are: 1) Fixed and Wired, 2) Mobile and Wired, 3) Fixed and Wireless, and 4) Mobile and Wireless.

Mobile	Point of attachment to infrastructure changes over time (connections are halted first)	Point of attachment to infrastructure changes over time (connection is continuous)
Fixed	Fixed point of attachment to wired infrastructure	Wireless point(s) to point(s)
	Wired	Wireless

Table 4.1: Domains of Connectivity
[Gro98]

Mobile versus Fixed Wired and wireless relate to the connection method of the devices to the network. On the other hand, fixed and mobile relates to the logical connection of the component to the network. The difference between fixed and mobile communications lies in their use of two key network functions: the network's 1. assigning of unique identification (addresses) to devices, and 2. association of these identifiers to specific network access points.

In fixed communications, a permanent association exists between an end user device address and a single unique access point into the network. Although the relationship is typically taken to be that between an end user device and its access point this definition can also be applied to associations between devices within the network.

In mobile communications the end user device can connect to multiple access points and use network services. For mobility, the issue is the association between the end user device address and the various network access points. Both portable (plug and play) and continuously moving devices are considered as mobile.

Wireless versus Wired The dichotomy of wired and wireless relates to the media used for the physical connection to the network (or in a more generic sense: the other communicating device). Wired communication is the transmittal of information between two or more physically separate devices guided by conductors, such as copper twisted pair, fiber optics cable, or coaxial cable.

Wireless communications employs space as its media for information transmittal. The term "wireless" does not in and of itself imply mobility or any characteristic other than transmission of information, through space, between two or more physically separate devices.

4.3 Problems

There are several open issues in the development of wireless ATM. Most of them stem from the fact that ATM was designed with reliable fixed links in mind. More precisely, ATM assumes fixed users, plentiful and constant bandwidth allocated dynamically based on users' needs, full duplex and point-to-point transmission, very good transmission quality (which is why error detection and error correction techniques are limited), and low physical-layer overhead. On the other hand, in a wireless environment users can move inside the covered range. Existing connections have to change to at least partially different paths, when the mobile system moves to another access point, so that a handover has to take place. QoS parameters have to be re-negotiated for the 'new' connection to the mobile system, if they can not be accommodated on the new path. The available bandwidth in the radio interface is limited and can vary based on the quality of the channel, transmission is usually half duplex and point-to-multipoint due to the lack of available frequencies, transmission quality is usually poor because of the inherent unreliability of the wireless media requiring advanced error detection and error correction techniques. Moreover, the physical overhead is much higher than in fixed links, basically due to the synchronization delay between transmitter and receiver. [PPVM97]

The issues that need to be successfully addressed in making ATM work in a wireless mobile environment fall into two somewhat orthogonal categories: mobility related problems at the higher level and wireless related problems at the lower level. From a mobility perspective, the key ATM issue is that of VC management in the presence of mobility. Obviously, the VC route needs to be continually modified as the mobile systems move during the lifetime of a connection. From a wireless perspective, the key ATM issues are providing lower layer support for ATM QoS, and efficient transport of ATM over a slow, noisy air medium. [AHK+96]

Another issue is the integration of fixed and mobile parts of the network. Introducing the wireless media and mobility to fixed ATM networks should be transparent and seamless, allowing specialized mobility-aware switches to co-exist with other switches that implement only standard ATM. [Raj96] Communications should not be limited to either part of the entire network. Both kinds of networks, fixed or mobile, should be able to interwork. In order to achieve the necessary transparency to fixed parts of the network, the awareness of the presence of mobility has to be restricted to a preferably small part of the network. A first requirement is that the introduction of mobile parts of the ATM network does not affect the operation of the fixed part of the network that is not aware of the mobility. One step further, the two parts provide integrated communications facilities that remain the same with or without mobile systems involved. There has to be a smooth migration between these states, embedding support for mobility only in those parts of the network that need to be aware, keeping the compatibility to those that are not. [FDIS98]

4.4 Mobility Related Issues (ATM Layer)

Mobile systems are allowed to roam without encumbrance from one coverage area or cell to another within a designated service area. Mobility control refers to tracking the location of *idle* mobile systems as they move and updating the corresponding information in the network, so that incoming calls can be correctly delivered, and moving the connection and configuration parameters intact from one access point to another, as the *active* mobile system roams from one zone to another. [KA98]

4.4.1 Routing for New Connections

When a connection is set up the location of the mobile system has to be known in order to find a path for the connection. With moving systems the network topology constantly changes requiring an update of the topology information.

The integration of mobile ATM networks into a more global ATM networking infrastructure requires to extend the routing protocol and provide a management function for the location of the mobile resources [FDIS98]. The establishment of a connection between two end-systems of an ATM network requires to compute a path between these two end-systems. When introducing mobile networks, computing a path to reach an end-system belonging a mobile network depends on the current location of the mobile network, meaning to which access point the mobile network is currently connected. The PNNI protocol (see chapter 3) has the basic dynamic behavior required for the integration of mobile networks. The solution to provide location management of mobile resources is built on an extension of the PNNI protocol to support mobile networks. This provides a scalable solution for integrating mobile ATM networks into the ATM network of an access point provider.

It allows a mobile network to build its own PNNI hierarchy, and therefore supports mobile networks of any size, ranging from a single mobile ATM switch to a mobile network of hundreds of them. The most common scenario foreseen is however a mobile network structured as one peer group, with a logical group node representing that peer group. When a mobile network is granted connectivity to an access point switch of the network of its access point provider, the logical group node representing the mobile network integrates the PNNI hierarchy of the ground network. As soon as the mobile network has integrated the PNNI hierarchy of the network of its access point provider, the PNNI protocol advertises new routing information to reflect the appearance of the mobile network, and the later is now reachable from any end-system in a private network.

A first scenario of mobility, called intra peer group scenario, occurs when the mobile network moves between access points that belong to the same peer group. In this scenario, the motion of the mobile network is transparent to the PNNI hierarchy. The inter peer group scenario occurs when the mobile network roams between access points belonging to different peer groups. This scenario of mobility is handled by introducing the new concept of a mobile logical group node. Unlike a regular logical group node in the traditional PNNI, a mobile logical group node is allowed to move within the PNNI hierarchy to reflect the current location of its associated mobile network. The mobile logical group node dynamically joins an ancestor peer group of one of its current access points. When a mobile network joins the PNNI hierarchy of its access point provider network, or when its associated mobile logical group node changes its location within the PNNI hierarchy, the address prefix of the mobile network is advertised by the PNNI protocol.

Let us consider a mobile network with an address prefix 'A.A'. This address prefix is advertised by the PNNI protocol within the hierarchy of the access point provider network. The advertisement of the network prefix allows to locate the mobile network within the PNNI hierarchy of the access point provider, and thus to compute a path to reach that mobile network. Above a certain level of the PNNI hierarchy, the level at which the network of the access point provider is connected to private customer networks, the addresses of

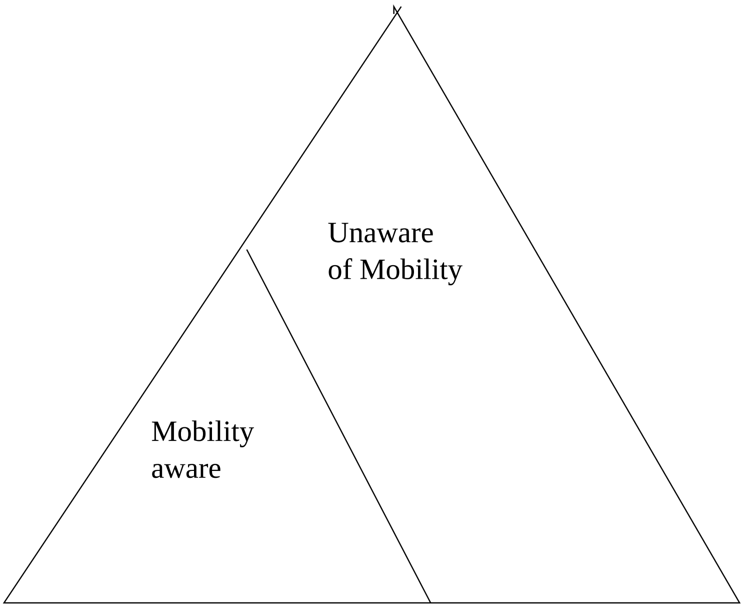


Figure 4.1: Mobility Awareness in a PNNI Hierarchy

the different mobile networks can be summarized into more global network prefixes, such as, in our example, 'A'. This summarization indicates that all the planes of 'A' can be reached through the network of the access point provider. The advertisement of the address prefixes of the different mobile networks can be confined to a specific part of the PNNI hierarchy. The other parts of the PNNI hierarchy (e.g. private networks) are not aware of the mobility and whenever a call is initiated from an end-system in a private network, the end-system does not have any knowledge of whether it is calling an end-system belonging to a mobile network or not. [FDIS98]

Figure 4.1 shows the awareness of mobility within a PNNI hierarchy. The mobile networks form their own hierarchy and are integrated into the hierarchy of the entire network at their top level.

4.4.2 Signaling and Re-routing for Existing Connections

The management of mobile networks within an ATM networks requires two types of supports : support from the routing to provide location management of mobile resources, and support from the signaling to enable the handover of existing connections. Handover is initiated when a mobile platform is moving away from an access point and is granted connectivity to the next access point. The subject of connection handover is discussed in detail in chapters 5 and 6.

4.5 Wireless Related Issues

ATM was originally designed to provide high-speed, highly reliable connection-oriented cell transfer over fiber optics with guaranteed quality of service negotiated prior to connection establishment. The default connection in ATM is point-to-point, although there can also be point-to-multipoint connections or vice versa.

A wireless environment, however, provides no means to support these properties. Usually wireless communication means lower bandwidth and is inherently unreliable. The connections are typically point-to-multipoint due to a lack of frequency.

While the ATM layer is concerned with challenges due to mobility, these layers below the ATM layer in the reference model have to cope with the fact that the mobile systems are wireless:

Data Link Layer

Physical Layer

Wireless links tend to introduce a higher error rate. Mechanisms like Forward Error Correction are used to cope with this problem. A high error rate increases the bandwidth required to get the data across the link. Since the bandwidth is an especially rare resource on wireless links, this aggravating effect of the higher error rate has a big impact.

Another undesirable property of wireless links is their low speed. Especially with GEO satellite links as considered in this thesis, the delay can be several hundreds of milliseconds. Every signal used therefore takes an unproportionally large period of time and has to be seriously evaluated in terms of its relevance for a protocol or whether it can be avoided.

5 Connection Handover in Wireless ATM Networks

This chapter gives an overview of connection handover in general and with respect to ATM networks. The overview is based on a variety of articles and contributions to the ATM Forum .

Connection handover (or by some authors referred to as hand-off) occurs when a mobile terminal, switch or entire network moves from one access point to another. The wireless link then is established to the new access point and the link to the old access point is released. Examples of reasons for handing over connections to another access point are decreasing signal strength, the mobile system moving out of the coverage area, optimization of traffic load or quality of communications in the network.

5.1 Components

The following glossary gives definitions of components involved in handover in wireless mobile ATM networks. A more fine grained selection of components can be found in the ATM Forum baseline text [Gro98]. For this thesis I have aggregated and generalized some of the components defined by the ATM Forum to a level of abstraction that is appropriate to understand the successive chapters. Other relevant terms and abbreviations can be found in appendix A.

Access Point (AP) Mobility enabled ATM switch equipped with a logical component that gives that switch the ability to communicate via a wireless communication link with wireless terminal adapters, or other wireless components.

Mobile ATM Terminal ATM terminal with the supplemental logical functionality of a mobile ATM terminal adapter and a wireless terminal adapter. Coupling the wireless and mobile functionality gives the terminal the ability to maintain all active VCs as the wireless link hands over from one wireless access point to another.

Mobile ATM Network ATM network with mobility functionality and therefore able to maintain active VCs as the wireless link moves from one wireless access point to another. Links to other networks (mobile or fixed) are wireless.

Mobile ATM Switch ATM switch with mobility functionality and therefore able to maintain active VCs as the wireless link moves from one wireless access point to another. Links to other networks (mobile or fixed) are wireless.

Mobile ATM System (MS) A mobile end-system, switch or entire network. Throughout this thesis this term is used when a distinction between them is irrelevant.

Base Station (BS) by some authors used synonymously to access point.

Far end Party (FEP) The terminal at the other end of the connection.

Crossover Switch (COS) An intermediate switch in the existing connection to a mobile terminal or network which acts as an anchor in rerouting a connection to the mobile terminal or network. Some switches may be statically designated as crossover switches (for example by pre-configuration or determined at setup time of a connection) or dynamically discovered. Usually the crossover switch will be a node in the fixed network. The case of mobile crossover switches is not addressed in the present work.

5.2 Quality Criteria for Handover

In order to be able to compare or design connection handover protocols, it is important to know what constitutes a 'good' handover. I have identified some criteria for the quality of connection handover. They vary in importance (according to priorities) and sometimes are tradeoffs to each other. Concerning performance criteria, the optimum is stated. The best handover protocol deviates the least from these criteria.

End-to-End Delay The end-to-end delay of the connections after handover should be minimal. The new path should be optimal.

Cell Loss, Buffer Requirements Cell Loss should be kept to a minimum. The amount of data lost equals to the required buffer size for lossless handover.

Sequencing Cells should reach their destination in the same order they were sent.

Cell Duplication Cells should not be duplicated.

Path Re-Use If the new path overlaps with the old one, it should reuse the resources reserved for the connection. Double reservation of resources on the same link should be avoided.

Total Handover Delay, Service Disruption Time The service disruption time should be minimal, for it affects the cell loss as well as the possibility of setting up connections. Also the total time needed to perform handover should be minimal.

Scaleability The protocol should be scaleable in terms of the number of VCs, the number of mobile systems, levels of mobility and the handover frequency.

Simplicity The protocol should be as simple as possible.

Robustness The protocol should be robust in case of failures.

Changes in Signaling As little as possible should be changed in UNI or PNNI signaling. (See chapter 3.) Only inevitable new messages should be introduced.

QoS The QoS should be maintained after the handover.

Data Looping, Zigzag Moving The protocol should be able to avoid data loops as a result from moving back and forth between access points.

Symmetry Performance should be similar on the uplink and downlink.

Transparency The terminal at the other end of the connection should not need to be aware of the handover or even of communicating with a mobile. In case of mobile networks not even the terminals within the mobile network should be aware of the mobility. Only those parts of the network that absolutely have to be aware of the mobility should be affected by the handover protocol. These parts should be kept to a minimum. Parts of the network that are not aware of the mobility should be able to participate in the communications.

5.3 Handover Categories

There are various possible scenarios for handover which leads to some distinctions:

According to the Co-Existence of Old and New Wireless Link :

In a ***soft handover***, the old wireless link will only be released after having successfully completed the setup of the new connection. Soft handover is only possible if the mobile system can be connected to more than one access point at a time, that is it is able to operate at more than one frequency simultaneously for instance.

Hard handover means that the old and the new wireless link do not overlap in time. The old link is torn down before the new one is active.

According to the Initiator :

Mobile-initiated handover means that the mobile system tells its access point when it wishes to have a handover . It is the responsibility of the mobile system to detect fading radio signals or other reasons like having reached a certain position (e.g. by using GPS, the global positioning system).

Network-initiated handover takes place when it is the responsibility of the network to tell the mobile to move to another access point.

Mobile-assisted handover is a combination of the two kinds described above. The mobile assists the network by giving information.

According to the Active Access Point :

Backward handover takes place, when the mobile is still linked to the old base station at the time of connection handover initiation. The major part of the signaling and other handover preparation is done before the actual switch of the wireless link to the new access point. After the radio link switch, the mobile only has to establish the connection to the new access point and resume the connections that have been previously rerouted.

Forward handover takes place, when the mobile is already linked to the new access point. This can be the case when the signal strength is decreasing too fast to maintain the wireless link to the old access point long enough for a handover. The signaling for the handover does not have to change except for one signal: The new access point merely informs the old access point about the radio link switch of the mobile, then the forward handover resorts to a backward handover.

According to the Scope : see also [Gro98].

Intra-AP handover occurs between different radio ports attached to the same access point.

Inter AP/intra-switch handover occurs between different radio ports attached to different access points on the same switch.

Inter-switch handover occurs between different radio ports attached to different access points on different switches. In chapter 6 concerning the performance of different connection handover schemes in mobile ATM networks, only inter-switch handover will be considered.

5.4 General Algorithm of Connection Handover

From all of the hard connection handover algorithms outlined in section 5.5 a kind of behavior pattern can be generalized and partitioned into the following steps:[1]

1 Connection handover initiation

2 Selection of a crossover switch

3 Establishment of a connection segment between the crossover switch and the new access point

4 Data flow along the new connection segment

5 Link from the mobile system to the new access point

6 Extension of the new connection segment from the new access point to the mobile system

7 Release of the unused part of the old path

5.5 Various Connection Handover Approaches

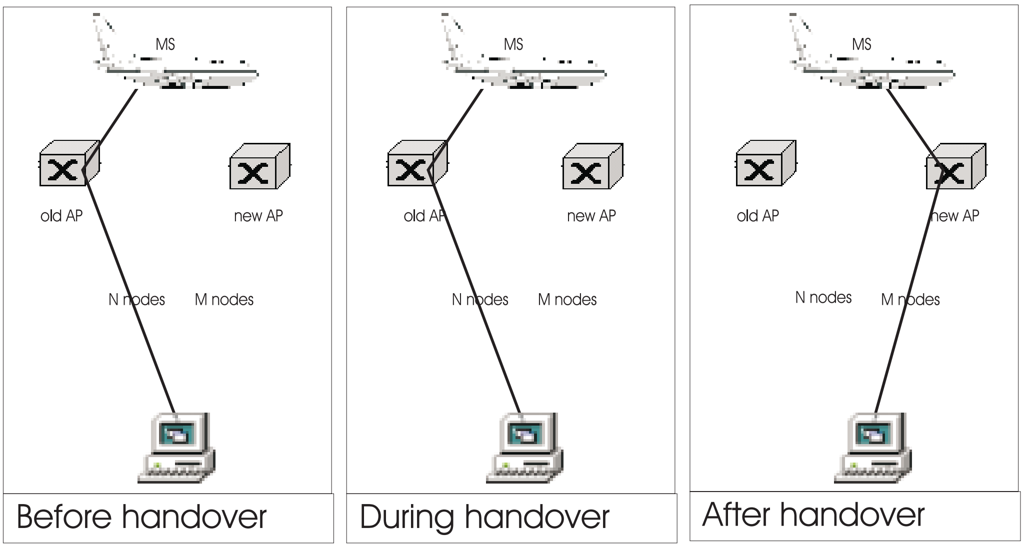
In the literature on connection handover (e.g.: [Toh97], [ALRR97], [MCCM97], [Raj96], [CRC$^+$97]) the following algorithms can be found.

Full Establishment (Also referred to as *connection reestablishment*). A straight forward approach is to establish a new connection from the moving terminal to the far end party of the connection existing at handover time. The old connection is torn down. See Figure 5.1 for visualization. This approach does not require any new signals. However, there are several drawbacks in this approach. Since the connection is completely reestablished, the far end party of the connection is affected by the handover of the mobile. Establishing an entire connection to the far end party and completely releasing the old connection also takes much more time than only doing this for a part of the path. This increases cell loss as well as the total handover delay.

Partial Path Rerouting , *incremental reestablishment*[Kee93] or *anchor rerouting* all refer to the following algorithm that can be seen in Figure 5.2 : A crossover switch is selected somewhere on the old path. The crossover switch can either be determined at the time when the connection is set up, preconfigured, or discovered dynamically during the handover. This crossover switch then serves as an anchor point for the

---

[1] Note that with soft handover the order of the steps can vary. For example step number 5 (wireless link establishment to the new access point) usually occurs earlier.



Before handover

During handover

After handover

Figure 5.1: Full Establishment

connection. That means that the connection remains unchanged between the far end party and the crossover switch. When the mobile moves to another access point only the partial path between the crossover switch and the new access point is rerouted. The data flow then is redirected to the new part of the path and the old partial path between the crossover switch and the old access point is released. An example for partial path rerouting is *nearest common node rerouting* proposed in [ACb]. This approach attempts to perform the rerouting for a handover at the closest ATM network node that is hierarchically above both of the access points in question or a parent of both access points in the network topology. This node then acts as an anchor point and crossover switch.

Path Extension This scheme (also referred to as *VC queuing*, *cell forwarding*[ACb] or *connection or VC extension* [MCCM97], [Raj96]) extends the path of the existing connection by adding the path from the old to the new access point as shown in Figure 5.3. The advantages of this approach are that it is fairly simple and fast. The old path is maximally reused. The cells all follow the same path, so no cells will arrive out of sequence. However, at each handover to a new access point, the path becomes longer, adversely affecting the end-to-end delay of the connection. If the mobile moves back and forth between base stations, some loop detection and elimination mechanism has to be introduced. The path extension scheme can also be viewed as a special case of anchor rerouting in the sense that the old access point acts as the crossover switch.

Path Splicing Proposed in [ALR98], this approach also has to select a crossover switch. The authors suggest that the crossover switch be discovered by the new access point targeting the search at the far end party of the connection. (See the description of the *loose select* crossover switch discovery 5.6.1.) Then a path from the crossover switch to the new access point is computed and *spliced* into the existing path. A scenario can be seen in Figure 5.4. After handover the partial path from the new access point via the crossover switch to the old access point is dropped. This approach entails double booking of resources along the same path but little cell loss, as can be seen in the section discussing handover performance.

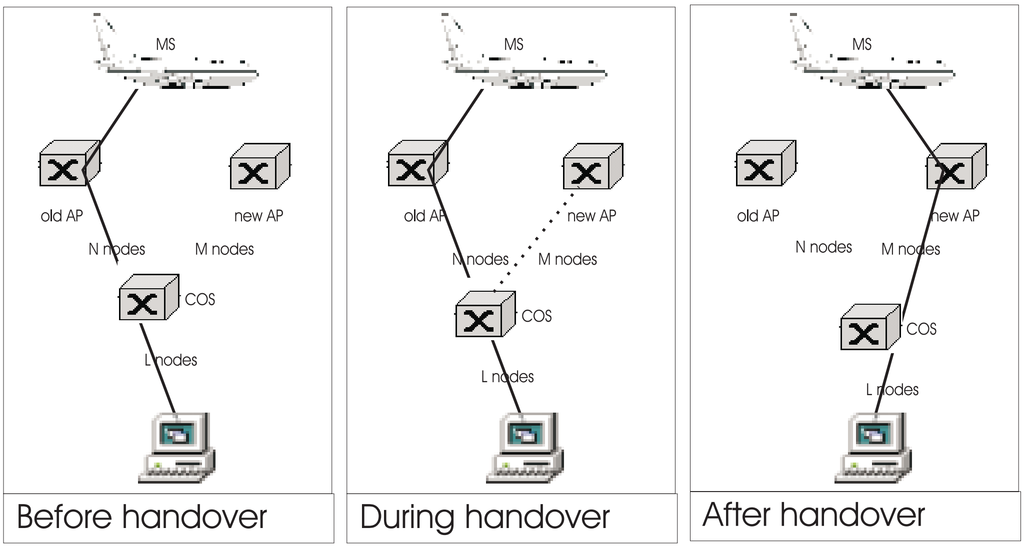


Figure 5.2: Partial Path Rerouting

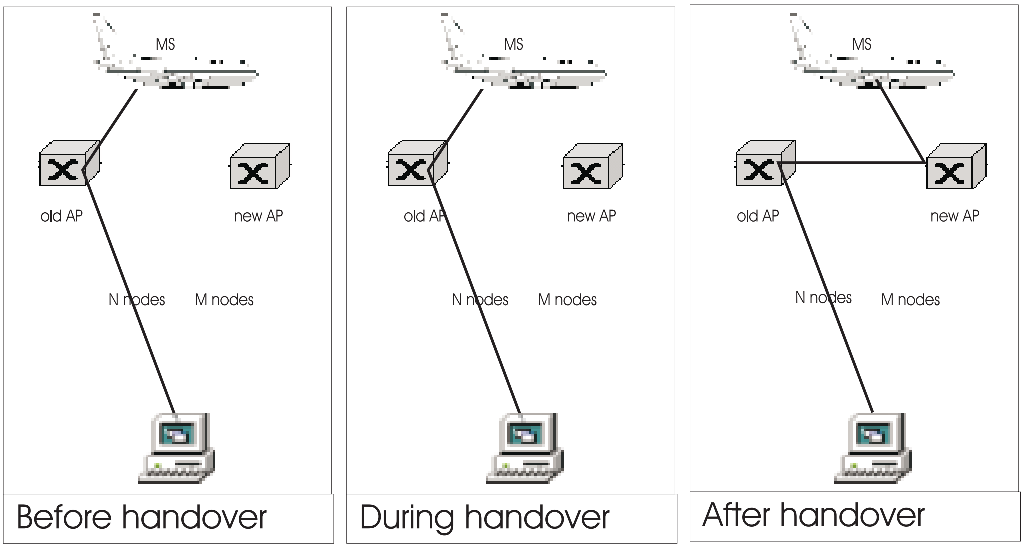


Figure 5.3: Path Extension

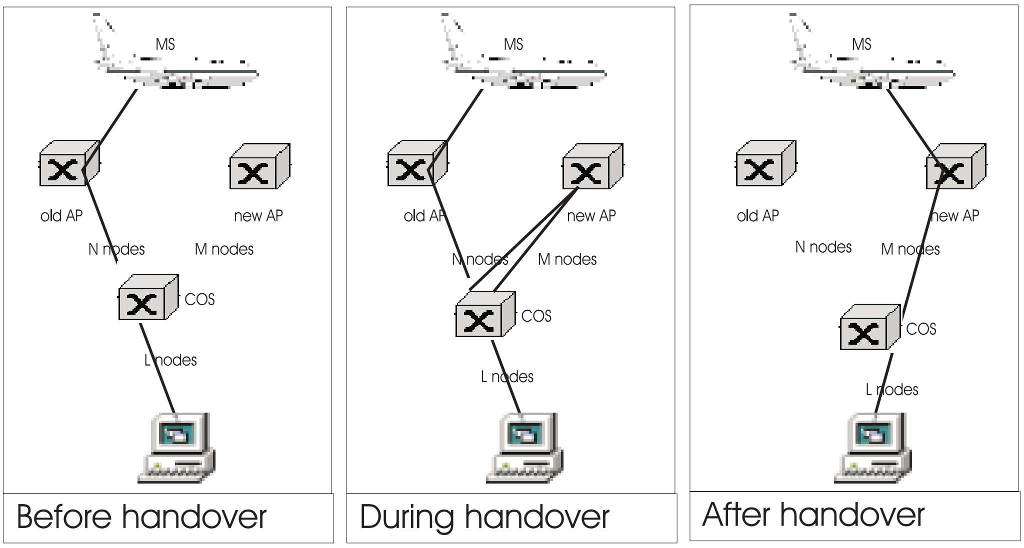


Figure 5.4: Path Splicing

Multicast Establishment Also referred to as *virtual tree based rerouting*. As described in [MCCM97] and shown in Figure 5.5, this approach preallocates resources in the network portion surrounding the location of the mobile. When a new connection involving a mobile is established, a set of virtual connections, named a *virtual connection tree*, is created, reaching all access points of the areas where the mobile might move to in the near future. Thus, the mobile can freely roam in the coverage area of the connection tree without invoking the network call acceptance capabilities during handover. The allocation of the virtual connection tree might be static or dynamic during the lifetime of the connection. This approach is fast and statistically guarantees the QoS contract also after handover, since the QoS is only negotiated once, at connection establishment, allocating resources in the entire area where the mobile is expected to roam. However, this approach is not at all efficient in terms of network bandwidth utilization, since it introduces the possibility of refusing a connection because of lack of resources that may never be needed, and high signaling overheads, especially in the case of dynamic tree allocation.

Two-Phase Handover proposed in [VKE96] combines the advantages of both path extension and partial path rerouting. The first phase consists of regular path extension. After the handover the path is optimized in the second phase by soft rerouting. The rationale behind this hybrid approach is the use of a fast procedure to handle the connection extension during handover, followed by the optimal VC reestablishment procedure, which is activated once the mobile is already connected to the new access point.

5.6 Crossover Switch (COS) Selection

The selection of the crossover switch is crucial for the connection handover performance. It has an influence on the end-to-end delay of the new connection, the path reuse, the signaling overhead, the cell loss, etc. The crossover switch can either be predetermined or dynamically discovered during handover.

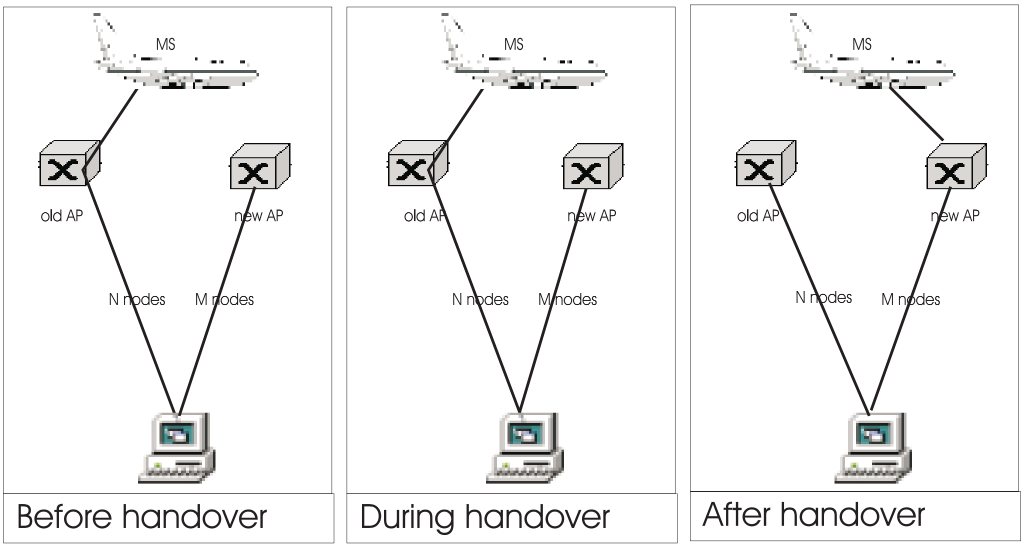


Figure 5.5: Multicast Establishment

Intra Switch Handover In the case of intra switch handover of a mobile terminal, when the wireless link only changes to another port, the crossover switch is the access point itself.

Inter Switch Handover For connection handover involving another access point or for mobile networks, there are multiple nodes that are possible candidates for the crossover switch:

- **Old Access Point** As in the path extension scheme, the old access point acts as the crossover switch for all the connections.
- **Predetermined Anchor Switch** The anchor rerouting scheme relies on a specified crossover switch before handover. Any crossover switch enabled node on the old path can be selected.
- **Crossover Switch Dynamically Discovered by the Old Access Point** With this approach the crossover switch is only discovered during the actual connection handover. The discovery is initiated by the old access point.
- **Crossover Switch Dynamically Discovered by the New Access Point** In this case the crossover switch discovery is initiated by the new access point. Almost all the dynamic crossover switch discovery algorithms use this approach.

5.6.1 Dynamic Crossover Switch Discovery Algorithms

The main purpose of the crossover switch discovery is to derive a minimum-hop path from the new access point to a crossover switch which lies in the original path of the connection [Toh97]. There are several algorithms presented in the literature. See [Toh97] for detailed description and implementation examples, [BG98] for a comparison and discussion.

Loose Select In this crossover switch discovery algorithm, the minimum-hop path from the new access point to the far end party of the connection is computed, regardless of the original path of the connection. However, if the chosen path from the new access point to the far end party has shared some part of the original path, then the crossover switch is the node where the two paths converge. [BG98] proposes a slight alteration to the algorithm by determining a candidate crossover switch at setup time that serves as a target for crossover switch discovery instead of the far end party or far end access point as suggested in [Toh97].

Distributed Hunt This algorithm requires every node to have information on its currently active connections. At the time of handover, a broadcast is sent out to discover which of the nodes in the network are possible crossover switches for a particular connection. Within this set of switches, the one with the least hops to the new access point is chosen. In the case of several possible crossover switches having the same number of hops to the new base station, an arbitrary switch is taken for the crossover switch. This approach is distributed so that there is no central register for locations for example. The new access point does not need to know the far end party of the connection, since once a crossover switch is found, there already is a connection to the target.

Prior Path Knowledge Much like in the distributed hunt algorithm, this scheme requires prior knowledge about the path of the old connection. But the information is not kept in each node but in a central communication server. The new access point computes paths to all the nodes in the existing path of the connection, selecting the node closest to itself. In the case of multiple possible crossover switches having the same number of hops to the new base station, the node closest to the old access point is chosen as the crossover switch.

Prior Path Optimal Resultant Derived from the prior path knowledge crossover switch discovery algorithm, this scheme enhances the algorithm by making sure that the new path is either shorter or equally long as the old path. If no equal or shorter paths are possible, this scheme falls back into prior path knowledge discovery. Like in prior path knowledge discovery, the paths from the new access point to the nodes in the old connection are computed. For each of the derived paths a comparison to the old path, that is the path from that node to the old access point, is made. Only those convergence paths that are equal or shorter than their counterparts of the old connections are chosen. Out of these valid paths, the shortest is selected as crossover switch. If multiple valid paths have the same minimum-hop count, again the node nearest to the old access point will be selected.

Backward Tracking In this algorithm suggested by [Kee93] the new base station triggers the old access point to invoke a distributed crossover switch location process. The algorithm backtracks along the old path starting from the old access point ending at the original far end party. Each node on that path will check its routing table to ascertain if it is the crossover switch. If the node uses the same port to reach the old and the new access point, then it is obviously on the part of the path, where the old and the new path overlap. Therefore, the switch located one hop ahead on the path towards the old access point is the diverging point and is chosen to be the crossover switch.

Route Optimization Proposed in [DVS97], this algorithm is initiated by the old access point, which computes a path between the far end party and the new access point. Then the old and the new path are compared. The node at the divergence point of the two paths is selected as crossover switch.

In a network where parts of the network are not able to support mobility, not all of the switches are enabled to act as a crossover switch. The algorithms described above have to be modified to allow for non crossover switch enabled switches to merely pass messages on to their neighbors on the connection or to be excluded from the set of valid crossover switch candidates generated by the algorithms.

5.7 Open Issues

The quality criteria for handover protocols in section 5.2 as well as the nature of wireless communications systems leave some issues for discussion and problems to solve. Some of these issues will be addressed in chapter 6.

QoS When handover to another access point is done and the best access point in terms of ability to accommodate all VCs is chosen, there is no guarantee that all QoS requirements can be fulfilled, but only that it will be the most suitable access point (the one most likely to meet the expectations) of the candidates provided by the mobile system. The question is, whether it is better to drop low-priority connections or to degrade the quality for all the connections, when the negotiated QoS can not be met after the handover. Another question arising is, of course, how the QoS policy should be realized once it is adopted.

VC Prioritizing In case of decreased QoS available on the new link or in order to minimize cell loss for certain connections, one would like to give priority to some VCs to be handed over.

Network Problems The handover protocol has to cope with:

> **Radio Link Failure** In case of a radio link failure the mobile has to establish another link, probably to a different access point, and has to resort to forward handover
>
> **Network Link Failure** If a link in the network fails during a handover measures have to be taken in order to prevent the failure from having an impact on the whole handover process.
>
> **Message Loss During Handover** The handover protocol should be robust also in case of message loss during handover. Time-out and re-transmission mechanisms have to be built in.

Buffering Questions on where and how much buffering should be built in arise. There are some tradeoffs between minimizing the buffer size and QoS maintenance as well as between the buffer size and its introduced additional delay.

Network Topology The influence of the topology of the network on the performance of the handover has to be taken into account. The information gained by investigating the relationship between topologies and handover algorithms can be used twofold. For existing networks the best handover algorithm can be chosen or, given the theoretically best algorithm, the information can be used to design a new optimal network for instance for an access provider. For a comparison of dynamic crossover switch discovery schemes, some research has been done in [Toh96c].

5.8 From Terminal to Network Mobility

Research on connection handover has almost exclusively focused on terminal mobility. In the rest of this document the mobility of entire ATM networks is being looked at. Research has been done on connection handover for mobile ATM terminals. I found the applicability of the results to mobile ATM networks limited by the following factors:

Number of Connections Since several terminals can be in a mobile ATM network, the number of connections to be handed over is likely to be much higher than for a single mobile terminal. Therefore, timing constraints like the maximum total handover delay or the impact on cell loss is different.

Bandwidth The same reasoning as above is valid in terms of bandwidth, which is likely to be much higher for a whole mobile network.

PNNI Instead of UNI Between a mobile terminal and the fixed network there is UNI. Between a mobile and a fixed ATM network, PNNI is more likely to be deployed than UNI. UNI and PNNI (see chapter 3 for protocol descriptions) differ in many ways, for example when a connection is set up, only UNI sends Connect_Ack messages. PNNI will provide features like edge-to-edge rerouting, etc. Many of the differences between UNI and PNNI matter when it comes to connection handover protocol design.

Frequency of Handover Compared to mobile terminals the coverage area is much bigger for mobile ATM networks, so connection handover is not expected to occur as frequently as for mobile terminals.

Levels of Mobility Mobile networks can be directly interconnected or supporting terminal mobility. Constructing mobile networks with several levels of mobility adds complexity to connection handover. If mobility is restricted to terminals, the worst case complexity will be that the mobile terminal has a connection to another mobile terminal via a fixed network. A random scenario for mobile inter-networking would be a small ship linked to a big ship that maintains a link to a fixed network. The small ship then could for example have connections to an airplane via the big ship, the fixed network and a satellite.

6 WATM Networks Handover Performance

6.1 Model of the Network

For the time being, only one mobile switch will be taken into consideration. The mobile network therefore consists of a mobile switch with terminals attached to it. The terminals are fixed with respect to the mobile switch. Therefore, the partial path of the VCs, that terminate at the terminal, between the terminal and the mobile switch is not affected by the handover and only the mobile switch will be considered (see figure 6.2). In figure 6.1, showing the mobile and the fixed network, the full line represents the path of a connection before handover, the dotted line shows the new path after handover. The two paths only diverge between the mobile switch and the crossover switch. The curve in the figure marks the border between the mobile network above and the fixed network below.

The mobile switch is connected to an access point via a wireless link. For the model two access points are assumed. For hard handover, only one radio link at the time can be active. This restriction does not apply to soft handover, where it is required to have two radio links active at the same time. The third node in the fixed network that is taken into account is the crossover switch (in the figures referred to as COS), marking the node at which the old and the new path to the mobile switch diverge.

Between the crossover switch and the old or the new access point, respectively, a number of n nodes can be assumed. As can be seen at the section describing the various approaches, n can be variable and potentially different for each VC. It is fixed for all VCs in case they share the same crossover switch.

6.2 Assumptions

In this section some assumptions regarding the requirements, system, network and delays are made.

Buffering No buffering of cells is assumed. Moreover, it is assumed that the data flow is not slowed down or stopped for the handover. This results in the fact that if there is a time gap between the old and the new connection, i.e. some idle time during which data is not sent or received on the new connection yet and not sent or received on the old one anymore, the data is just lost.

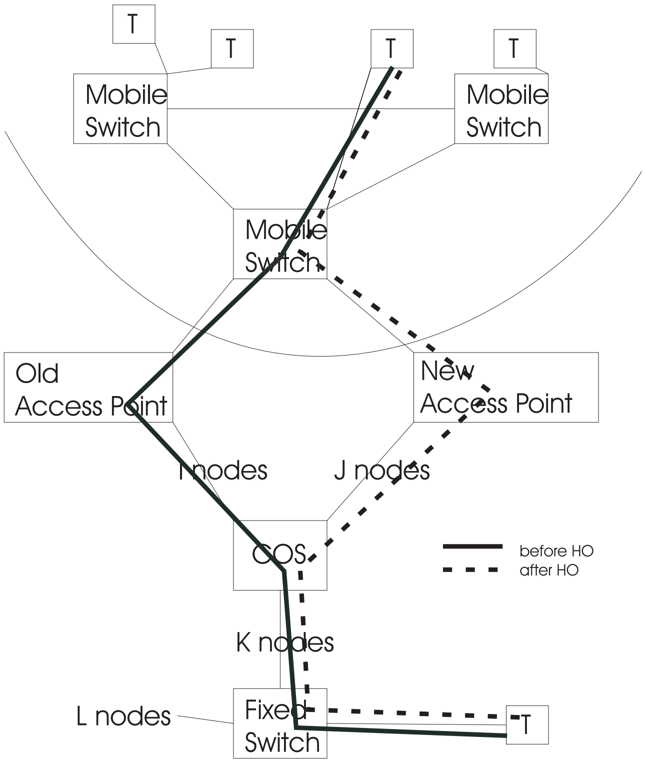


Figure 6.1: Mobile and Fixed Network

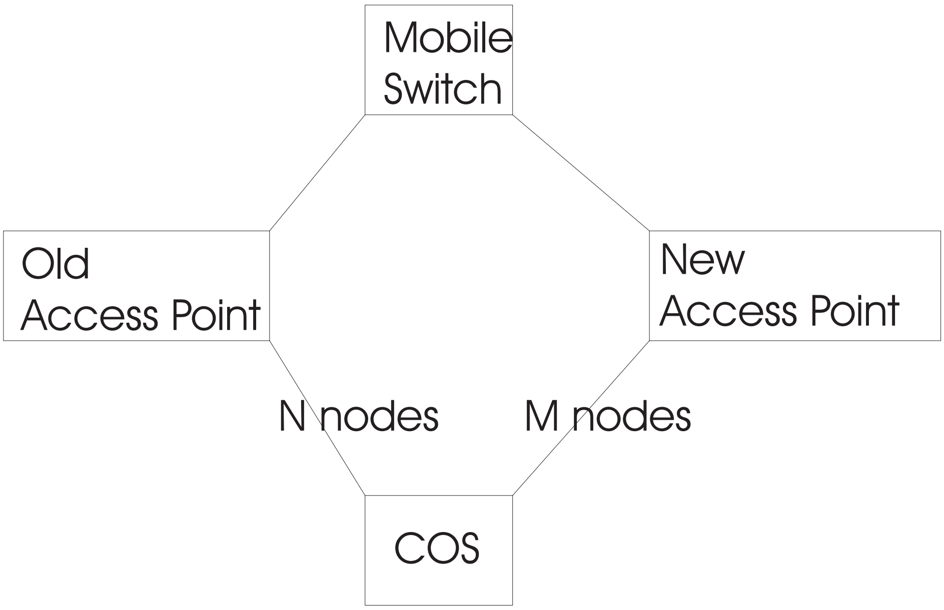


Figure 6.2: Model of the Network for handover

Requirements, Hardware The requirement of hard handover is assumed, unless it is explicitly stated otherwise. Furthermore, only the scenarios for forward handover are shown and discussed. The handover is initiated by the mobile switch which in turn gets an impetus from outside. Different scenarios on how handover is triggered can be imagined. One of these scenarios is that the mobile switch reaches a particular location where the handover is triggered. Another could be the relative signal strength. If the mobile switch does not initiate the handover when necessary, it can also be forced by the fixed network to perform a handover triggered by the signal HO_Force sent from the old access point to the MS. This signal is left out of the flow charts shown below, because it is optional and only occurs if the handover is triggered by the fixed network. It does not have an impact on the comparative performance of the different handover approaches described below.

It is assumed to be possible to separate the switch of the uplink cells flow (switch from the old to the new path) from the one of the downlink cells flow and perform them at different times.

Time Several assumptions on time such as computation, processing, transmission, propagation time are made and will be explained in more detail later.

For simplicity only two classes of signals are assumed. Signals that take longer to be processed (e.g. setup, release) and other signals that do not need path computation or any other delay in processing other than the most basic (e.g. a notify message). They will be referred to as *heavy* and *light* signals, respectively.

6.3 Approaches

In the previous chapter a number of algorithms on how to do connection handover were presented. Some of these algorithms can be combined and merged, resulting in a long list of approaches:

1 Full establishment

2 VP-based path extension

3 VC-based path extension

4 Hybrid path extension

5 VP-based anchor rerouting

6 VC-based anchor rerouting

7 VC-based anchor rerouting, shared crossover switch

8 Hybrid anchor rerouting, shared crossover switch

9 Dynamic crossover switch discovery, $< discoveryalgorithm >$

10 VP-based path splicing

11 VC-based path splicing, shared crossover switch

12 VC-based path splicing, dynamic crossover switch discovery

13 VC-based path splicing

14 Hybrid path splicing

Considering combinations with either hard or soft, forward or backward handover, this list would be even longer. Therefore, to look at performance issues, the following criteria are relevant:

Is the approach based on a VP or VCs? If it is VC-based, is the crossover switch shared, potentially different for each VC, but predetermined, or discovered dynamically? If it is discovered dynamically, which discovery algorithm is used? Is hard or soft handover considered? Forward or backward handover? Is the new partial path spliced into, added to or replacing part of the old path? Considering the various algorithms for dynamic cross over switch discovery, this would result in a very large number of distinguishable kinds of connection handover. In this thesis, only a few different kinds of handover with expected distinguishable performance differences are selected and described in detail.

6.3.1 VP-Based Anchor Rerouting

The crossover switch is the same for all connections and does not change over time. Before handover, a VP with fixed bandwidth is established between the crossover switch and the new access point. All the VCs in the old VP from the crossover switch to the old access point are rerouted at once. The bandwidth of the VP is determined ahead of time by an estimation or by simply taking the bandwidth the satellite link offers. The bandwidth is reserved assuming CBR (alternative: ABR VP).

Depending on the relationship between the bandwidth of the VP and the actual bandwidth used by the total of the VCs, there is a potentially unused bandwidth reserved or, in the contrasting case, not enough bandwidth reserved to accommodate all established VCs in the VP that is handed over.

A distinction can be made between those signals that only occur once per handover and those that have to be sent for each VC. Using a VP-based approach only requires the corresponding messages to be sent once, thus saving a substantial number of signals, the number depending on the number of VCs. This is due to the fact that all VCs are within a single VP, so they can all be handed over at once, not requiring extra signaling for each VC.

Data Flow:

Uplink The term *uplink* refers to the link from the mobile switch to the fixed network. This may seem counterintuitive for example for mobile switches in airplanes, but since in the ATM world the term *uplink* is traditionally used to refer to the link from user or edge devices to the ATM switch, which is also the case for wireless mobile end terminals, the notion is kept accordingly for mobile switches and entire mobile networks.

The periods marked *time* in the graphs will be discussed in the section on cell loss.

$t_{Radio(VP)}$ marks the point in time, when the uplink cells flow stops on the old path. This occurs just before the radio link is switched.

$t_{Up(VP)}$ is the point in time, when the uplink on the new path is activated and data sent.

Downlink	The term *downlink* refers to the link from the fixed network to the mobile switch. If $t_{Loss(VP)}$ occurs before $t_{Down(VP)}$, which it will in almost all cases, the following applies: $t_{Loss(VP)}$ is the point in time, when cells start getting lost on the downlink. This happens after the downlink is switched to the new path and cells arrive at the new access point. There, no connection to the mobile switch is established yet, so cells are lost. $t_{Down(VP)}$ is the point in time, when the downlink is initialized and data are sent from the new access point to the mobile switch. Otherwise, no data is lost.

Signal Flow:

HO_Request	This message is sent from the mobile switch to the old access point requesting for handover and therefore starting the whole process of the handover procedure. Along with the signal a list of one or several candidate new access points is provided by the mobile switch. This list can be based on the relative radio signal strength or on the relative vicinity in location of the new access points.
HO_Request_Query	On reception of the HO_Request, the old access point sends out this message to all the access points contained in the candidate list provided by the mobile switch. This message contains the necessary bandwidth and QoS parameters of the VP to be handed over.
HO_Request_Response	Every access point that received the HO_Request_Query checks the availability of resources to accommodate the request and tells the old access point by this message if it is able to do so or not.
HO_Response	As soon as the old access point receives the first HO_Request_Response from an access point, the sender of the message is chosen to be the new access point. This algorithm assumes that the one that is able to respond first is the closest and therefore best suited as a handover target. Contained in the HO_Response subsequently sent to the mobile switch is the information on which access point was chosen. If a candidate access point is not able to accommodate the request it sends back an HO_Failure. If the old access point does not receive any positive HO_Request_Response within a certain period of time or only receives HO_Failure messages, an HO_Failure message is sent to the mobile switch instead of the HO_Response, indicating that none of the access points in the candidate list were able to accommodate the handover. Having received that, the mobile switch has to trigger the handover procedure again trying other candidates.
VP_HO_Request	Since the new access point has already been determined, this message is sent to the (previously determined) crossover switch asking to set up a VP with the required bandwidth and QoS parameters for the partial path between itself and the new access point.
Setup	The crossover switch then sets up the partial path VP to the new access point accordingly with this message.
Connect	In response, the new access point sends this message back to the crossover switch.
Down_Ready	When the crossover switch received the Connect message it sends this (inband) signal to the mobile switch along the old path indicating that the new connection to the

	new access point has been established and that therefore it is now safe to switch the radio link.
	In contrast to the signal flows contained in some contributions, e. g.: [RLA97a], and also the baseline document of the ATM Forum, no Up_Ready is sent from the mobile switch to the network to indicate that the uplink is ready by having the last cell travel along the old path. This message is only necessary if cells are buffered and skipped here according to the assumption of no buffering, which renders it useless since data will be lost anyway, no matter at what node in the network.
	One could argue that the Down_Ready could also be skipped following the same reasoning, for it primarily serves the same purposes as the Up_Ready signal. In the scenarios shown in the flow charts, the Down_Ready nevertheless is needed as an indication as to when the radio link can be switched. It could be replaced by another signal that triggers the actual radio link switch. The only difference would be that the signal would have a different name.
Conn_Activate	Right after the radio link is switched, the mobile switch sends this message to the new access point to activate the connection. This message is only sent once during the entire handover procedure.
Conn_Active	On reception of the Conn_Activate the new access point sends this message back to the mobile switch indicating that the VP is ready and initializing the downlink.
Release	Some time after the radio link switch, the crossover switch releases the path to the old access point. In Figure 6.3, this message is shown after the Conn_Activate, but it could be an arbitrary time after the radio link switch.
Release_Ack	The old access point acknowledges the release of the old partial path by sending this message.

The signaling sequence can be seen in figure 6.3

The downlink data flow is shown in figure 6.4. The data flow chart is the same for all the hard handover approaches mentioned in this chapter. In the special case of path extension (explained below), the crossover switch and access point$_{old}$ are identical. The full line in the graph shows the data flow of a connection before the downlink switch. On reception of the Connect message at the crossover switch, the downlink is switched and the data flow is represented by the dotted line. After reception of the Conn_Activate message at the NMASE-E$_{new}$, the data flows along the broken line.

The uplink data flow can be seen in figure 6.5. Again, the data flow chart is the same for all the hard handover approaches mentioned in this chapter. In the special case of path extension, the crossover switch and access point$_{old}$ are united. The full line in the graph shows the data flow of a connection before the radio link switch. After the radio link switch the flow is represented by the dotted line. After reception of the Conn_Active message at the mobile switch, the data flows along the broken line.

6.3.2 VC-Based Anchor Rerouting, Shared Crossover Switch

Here the situation is similar to the one explained above, but each VC is handed over separately. Nevertheless, all VCs are handed over at the same crossover switch, provided that this is possible by routing accordingly. Some additional delay is introduced for each

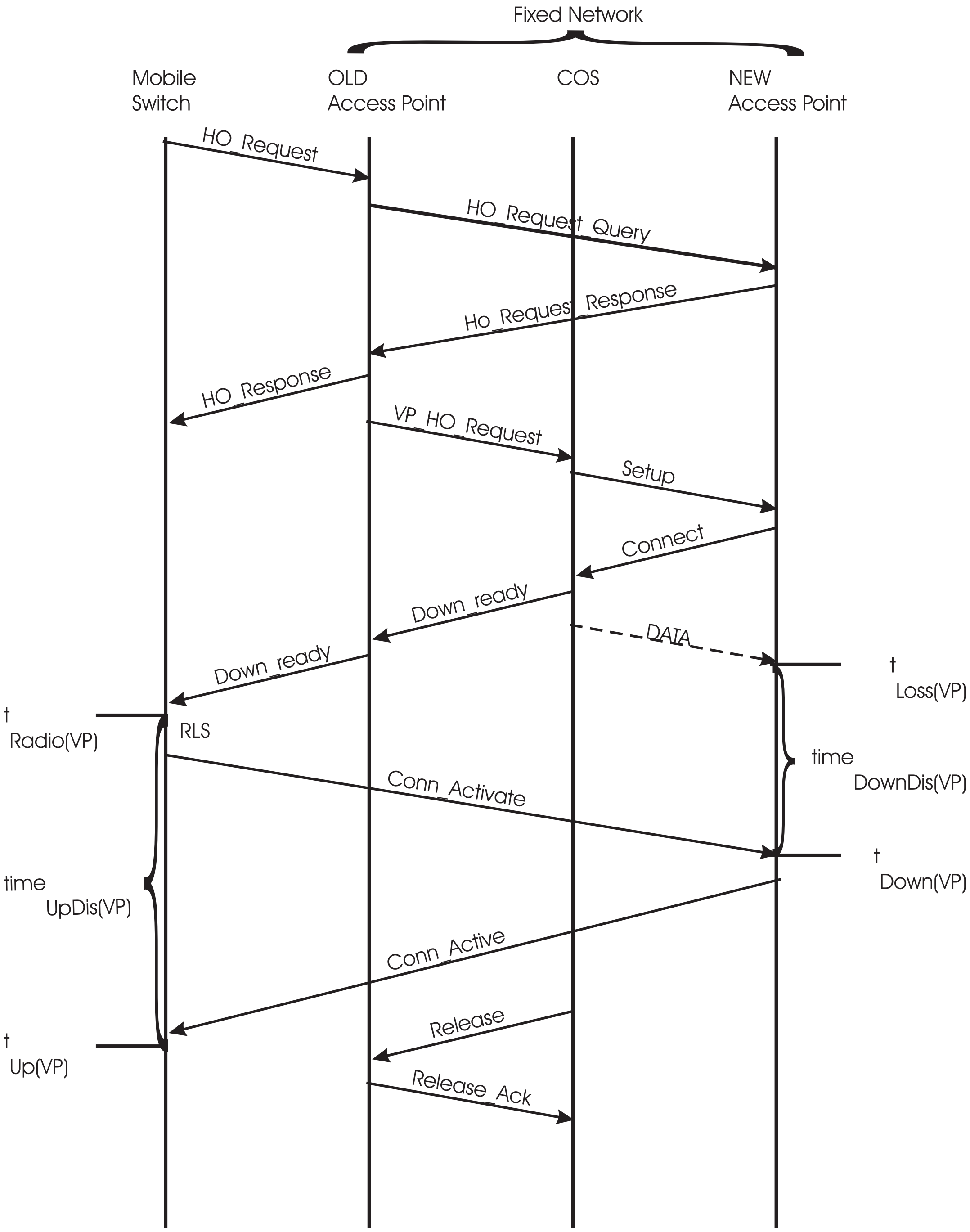


Figure 6.3: VP-based Anchor Rerouting

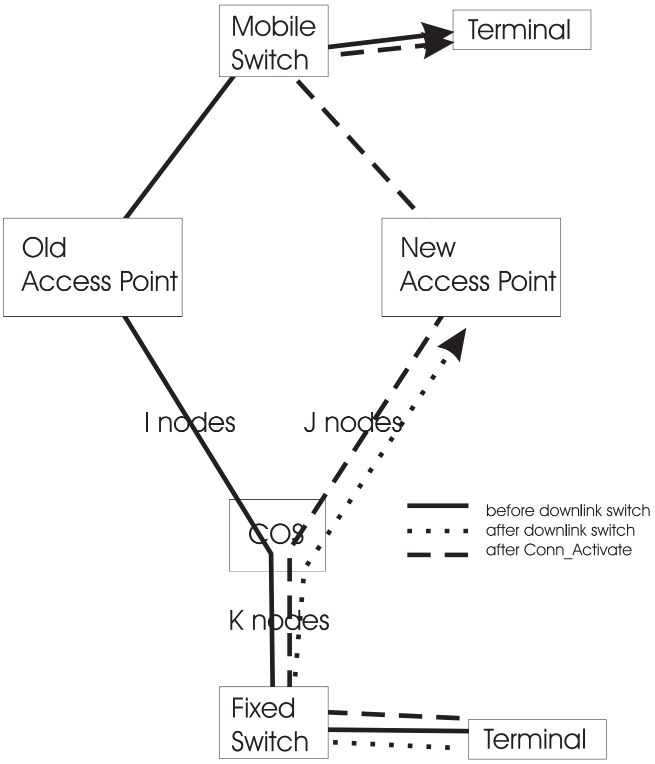


Figure 6.4: Downlink Data Flow for Hard Handover

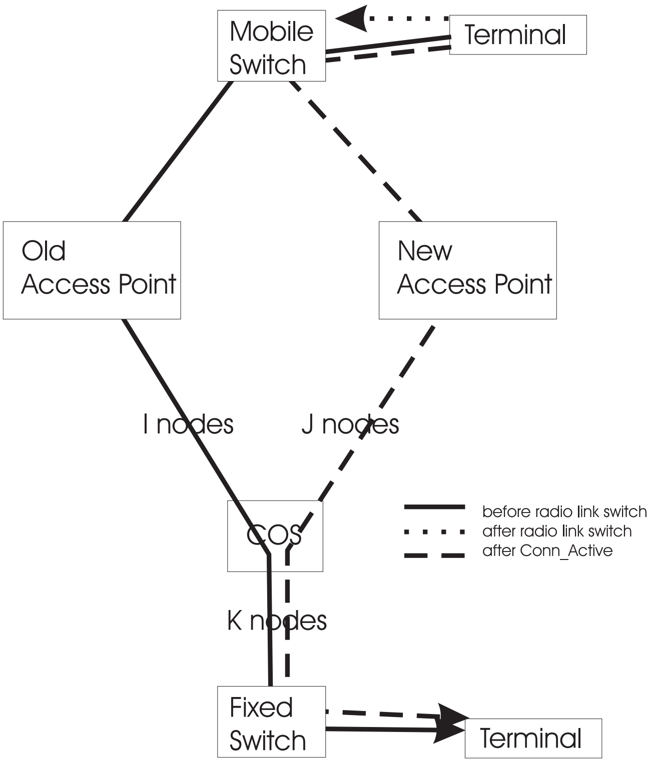


Figure 6.5: Uplink Data Flow for Hard Handover

VC, as the signals are sent sequentially and once per VC. The performance of this approach depends on the number of VCs to be handed over.

As far as the signaling is concerned, a distinction is made between signals that are sent once per handover and those that are sent for each VC to be handed over. In the flow chart, the signaling for three VCs is shown. Since the signals are processed sequentially, a constant time gap α is assumed as delay between subsequent signals due to processing. For performance measurement, one VC, henceforth referred to as VC$_j$, representing a random VC, is chosen. VC$_j$ is any VC in the sequence of VCs from VC$_1$ to VC$_n$.

Signal Flow:

The signals, as shown in figure 6.6 are essentially the same as with VP based anchor rerouting, except

VC_HO_Request

Setup

Connect

Down_Ready

Conn_Active

Release

These messages contain the corresponding information for the particular VC that is being handed over and therefore are sent separately and once for each VC.

Data Flow:

Uplink $t_Radio(VC)$ marks the point in time, where the uplink data flow stops on the old path. This occurs just before the radio link is switched. The radio link is switched, when all the VCs are ready, i. e. the Down_Ready signal for the last VC has reached the mobile switch.

$t_{Up(VC_j)}$ is the point in time, when the uplink on the new path is activated and data sent on VC$_j$, the typical VC.

Downlink $t_{Loss(VC_j)}$ is the point in time, when cells start getting lost. This happens after the downlink is switched to the new path and data on VC$_j$ arrive at the new access point. There, no connection to the mobile switch is established yet, so cells are lost.

$t_{Down(VC)}$ is the point in time, when the downlink is initialized and data are sent from the new access point to the mobile switch. This t-time is the same for all VCs.

According to the data flow described above, it can be seen that some points in time ($t_{<name>}$ in the graphs) have an increased delay due to the number of VCs handed over. However, other points in time remain the same for all the VCs. This can be explained by the fact that the radio link is switched only once per handover, thus requiring a point of synchronization of all VCs. The Conn_Activate signal is also sent only once, independent of the number of VCs, resulting in the same point in time to start the data flow on the downlink of the new path for all the VCs.

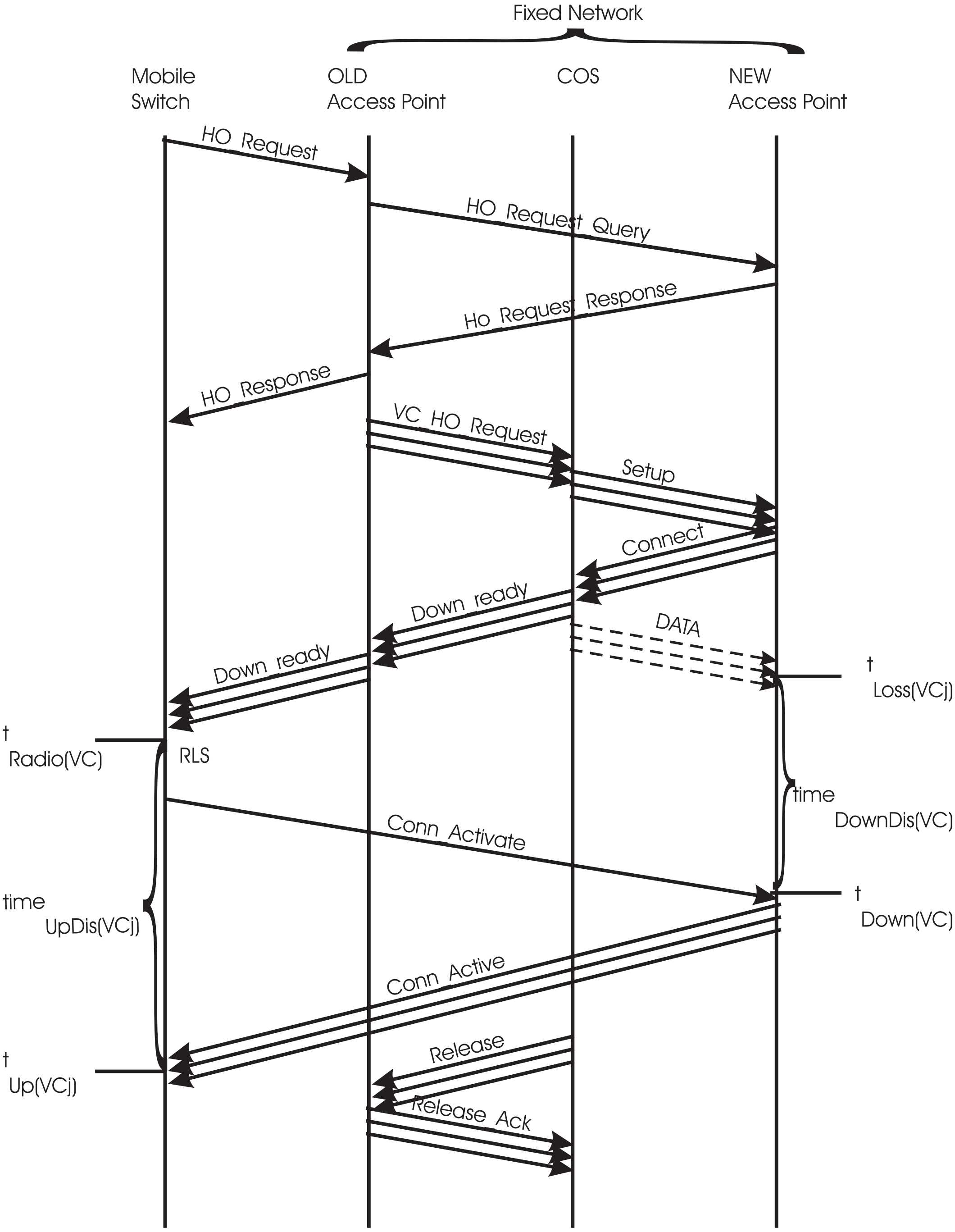


Figure 6.6: VC-based Anchor Rerouting, Shared Crossover Switch

6.3.3 VC-Based Anchor Rerouting

This approach is similar to the one explained above, as each VC is handed over separately. Nevertheless, each VC has its own crossover switch which can vary from VC to VC.

Signal Flow:

The signals are the same as with VC based anchor rerouting with a shared crossover switch with the only exception, that the VC_HO_Request messages may be targeted to a different crossover switch for each VC as can be seen in figure 6.7.

Since with this approach there is potentially a different crossover switch for each connection, this implies a different number of hops between the crossover switch and the old as well as the new access point for the VCs. This results in distributing the computation and signaling to several nodes, thus eliminating the sequential nature of the signaling in the case of a shared crossover switch.

The delay still depends on when the last VC is ready, but in the case of a shared crossover switch, this could be easily determined by counting the number of VCs and adding the constant delay. Now the situation changes and the factor that determine who the last VC is, are the rank of the VC in the handover sequence and the location of its particular crossover switch.

Now the '$t_{<name>}$-times' seen in the figure no longer depend on the number of VCs but also on delays comprised of propagation, processing and queuing delays.

Some distributions of the probability of number of hops between the crossover switch and the new (or the old, respectively) access point have to be assumed according to the topology of the network.

Data Flow: as in the previous section.

According to the data flow described above, it can be seen that some points in time ($t_{<name>}$ in the graphs) vary for different VCs. However, other points in time remain the same for all the VCs. This can again be explained by the fact that the radio link is switched only once per handover, thus requiring a point of synchronization of all VCs. The Conn_Activate signal is also sent only once, independent of the number of VCs, resulting in the same point in time to start the data flow on the downlink of the new path for all the VCs.

6.3.4 VP-Based Path Splicing

The signal flow can be seen in figure 6.8.

The VP-based alternative is chosen for the description of the handover approach. Therefore, the crossover switch is the same for all connections and does not change over time. Before handover, a VP with fixed bandwidth is established back and forth between the crossover switch and the new access point and spliced into the original path. All the VCs

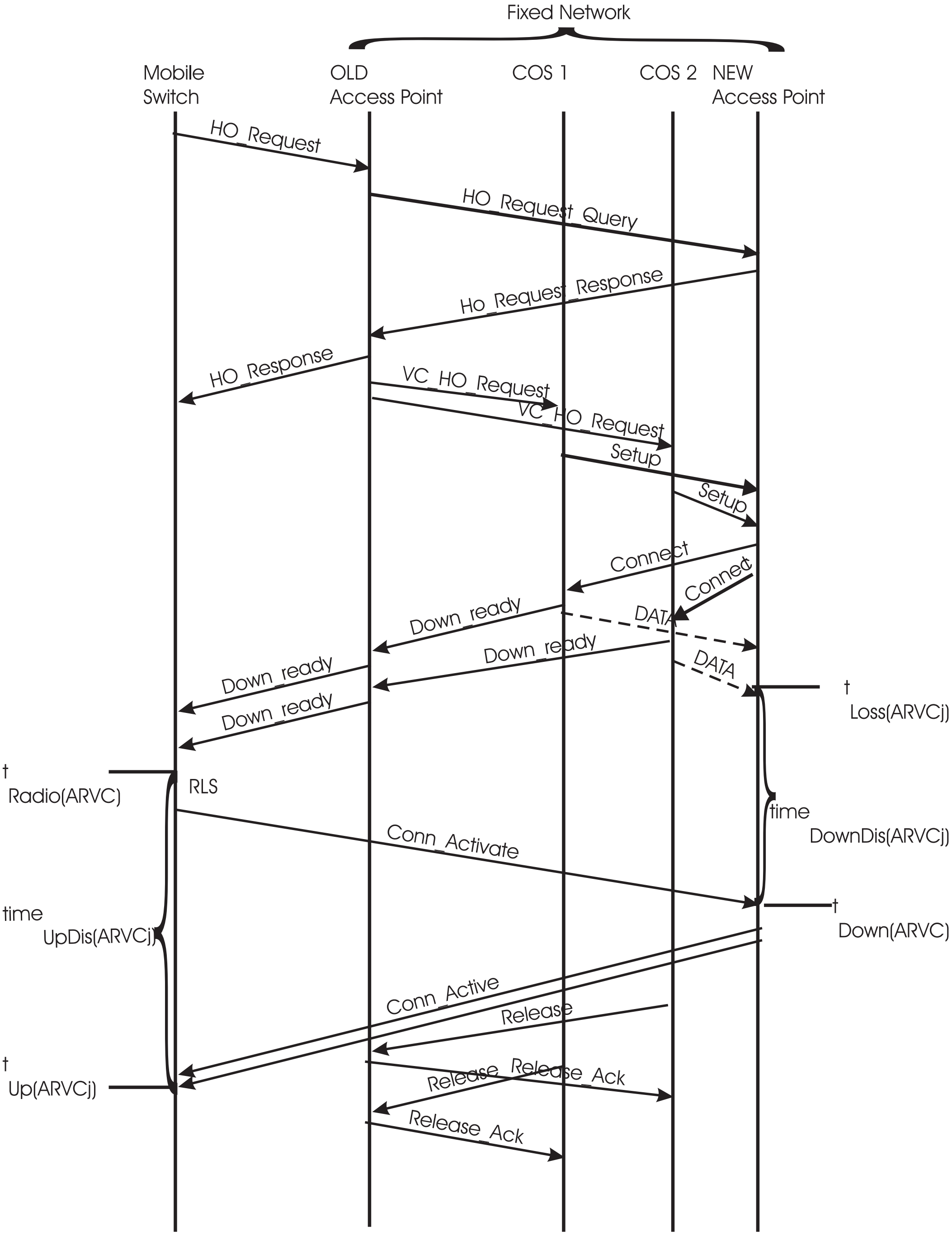


Figure 6.7: VC-based Anchor Rerouting

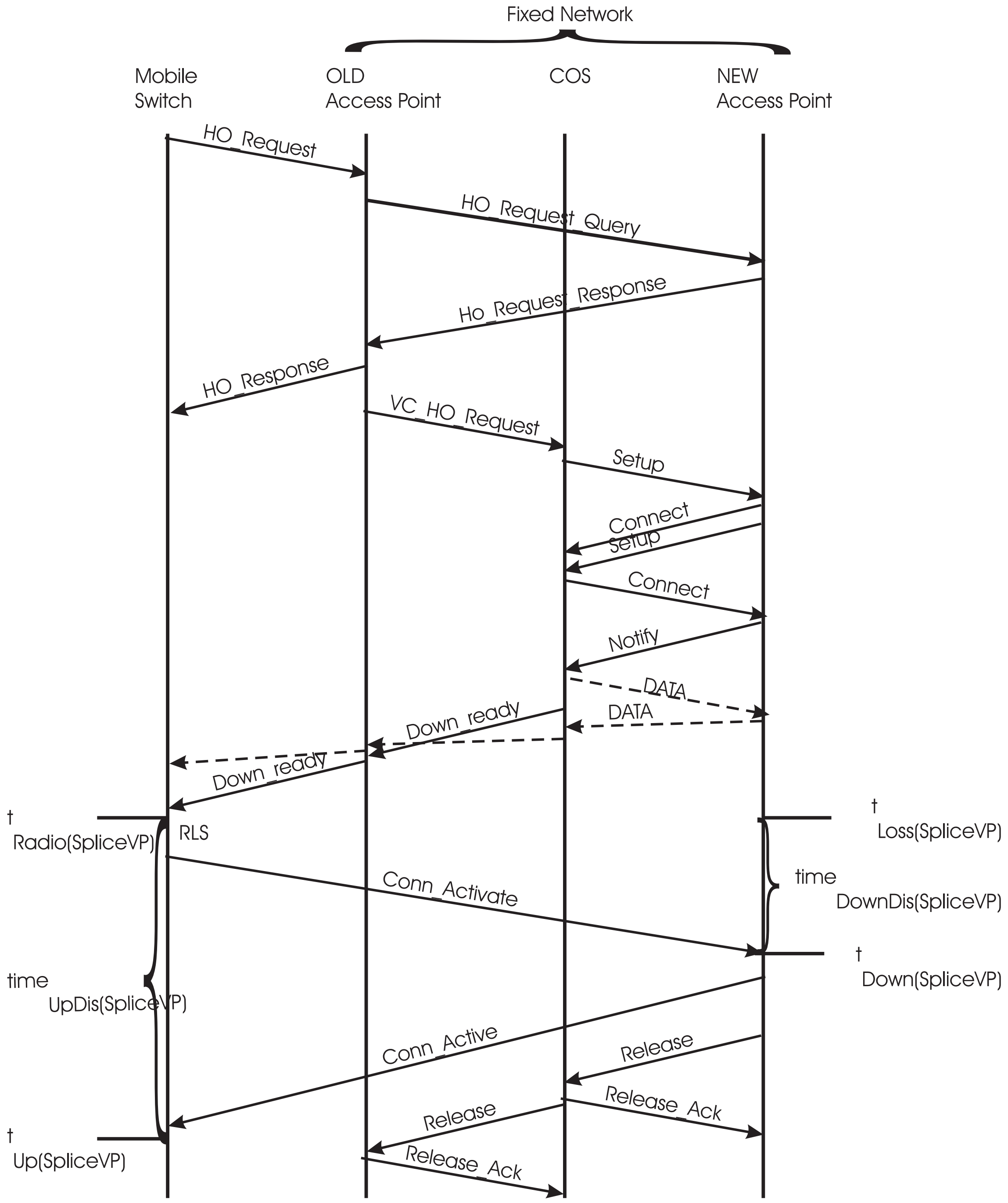


Figure 6.8: VP-Based Path Splicing

in the old VP from the crossover switch to the old access point, destined to the mobile switch, are rerouted at once to go to the new access point, return to the crossover switch and proceed to the mobile switch via the old access point. After handover, the partial path from the new access point via the crossover switch to the old access point is dropped. (See Figure 5.4.)

Data Flow:

Uplink	$t_{Radio(SpliceVP)}$ marks the point in time, when the uplink cell flow stops on the old path. This occurs just before the radio link is switched.
	$t_{Up(SpliceVP)}$ is the point in time, when the uplink on the new path is activated and data sent.
Downlink	$t_{Loss(SpliceVP)}$ is the point in time, when cells start getting lost on the downlink. This happens just before the radio link switch, when the mobile switch loses the link to the old access point.
	$t_{Down(VP)}$ is the point in time, when the downlink is initialized and data are sent from the new access point to the mobile switch. Since a part of the spliced in partial path, namely the path from the new access point to the mobile switch via the crossover switch and the old access point, is dropped, all the cells already underway on that path during the handover process are lost.

Signal Flow:

The signals correspond to those described at the VP-based anchor rerouting approach except for the modification of

Setup	The crossover switch sets up the partial path VP to the new access point. The new access point sets up the way back to the crossover switch.
Connect	In response, the new access point sends this message back to the crossover switch that also sends a Connect to the new access point.
Notify	This message serves to trigger the switch of the data flow to the path that was spliced in.
Release	Some time after the handover, the new access point releases the path, that was only spliced in, to the old access point via the crossover switch.
Release_Ack	The old access point acknowledges the release of the old partial path by sending this message.

6.3.5 Dynamic Crossover Switch Discovery (VC)

In the previously discussed approaches, the crossover switch has been a fixed switch chosen before the need for handover arises. With this approach, each crossover switch is discovered dynamically by the new access point.

When the connection is first set up, a candidate crossover switch is chosen. The setup message includes the ATM address of the first crossover enabled switch it encounters.

When the Setup reaches its destination, this address is stored there as the candidate crossover switch for the called party. On the way back, the address of the first encountered crossover enabled switch is included in the Connect message to serve as the candidate crossover switch for the calling party. Using this algorithm, an acceptable crossover switch is ensured for each VC, regardless if the mobile switch is the called or the calling party.

At handover time, the actual crossover switch, that may be different from the previously determined candidate, is dynamically discovered using the message VC_HO_Request. This message is sent from the new access point towards the candidate crossover switch and it is in search of the actual crossover switch. As soon as this message switch hits a crossover enabled switch belonging to the old path of the particular VC, this switch is chosen as crossover switch. In the worst case this search terminates at the candidate crossover switch.

The algorithm shown here corresponds to the 'loose select' algorithm presented earlier with the modification that the target for discovery is a crossover switch enabled node within or at the edge of the network.

With the approach of using a dynamically discovered crossover switch for each VC the single VCs still have different properties in terms of time. Whereas by using the same crossover switch for all VCs the gap α was constant and predictable, this is no longer true in the case of different crossover switches for different VCs. Depending on the time it takes the signals to travel to the particular crossover switch and back, i. e. the number of hops for simplicity, that gap can vary for each pair of VCs that are compared (see figure 6.9).

Depending on the number of hops to and from the crossover switch on the old and new path, the gap between the signals varies. It can be below an assumed unit of α in the fixed crossover switch case or well above. This fact renders the distribution of probabilities of number of hops to be encountered on the way to the crossover switch on either old and new path crucial.

These facts, however, are only relevant before the radio link switch occurs. After that only the communication between the mobile switch and the new access point has an impact on t-times, the access point new being the same for every VC.

Signal Flow:

With the exception of VC_HO_Request the signals are used as in the VC Anchor Rerouting Case. An additional signal, HO_Command is introduced.

Data Flow: same as in the section on VC-based anchor rerouting.

6.3.6 Hybrid VC/VP Anchor Rerouting

This approach is a mixture between VP and VC based anchor rerouting. When the old access point learns that the mobile switch requests a handover, it tells the crossover switch to establish a VP to the new access point with a bandwidth equal to the current total bandwidth used by the VCs of the mobile switch. This way only as much bandwidth as required is reserved. When there are new connections to be established that do not fit

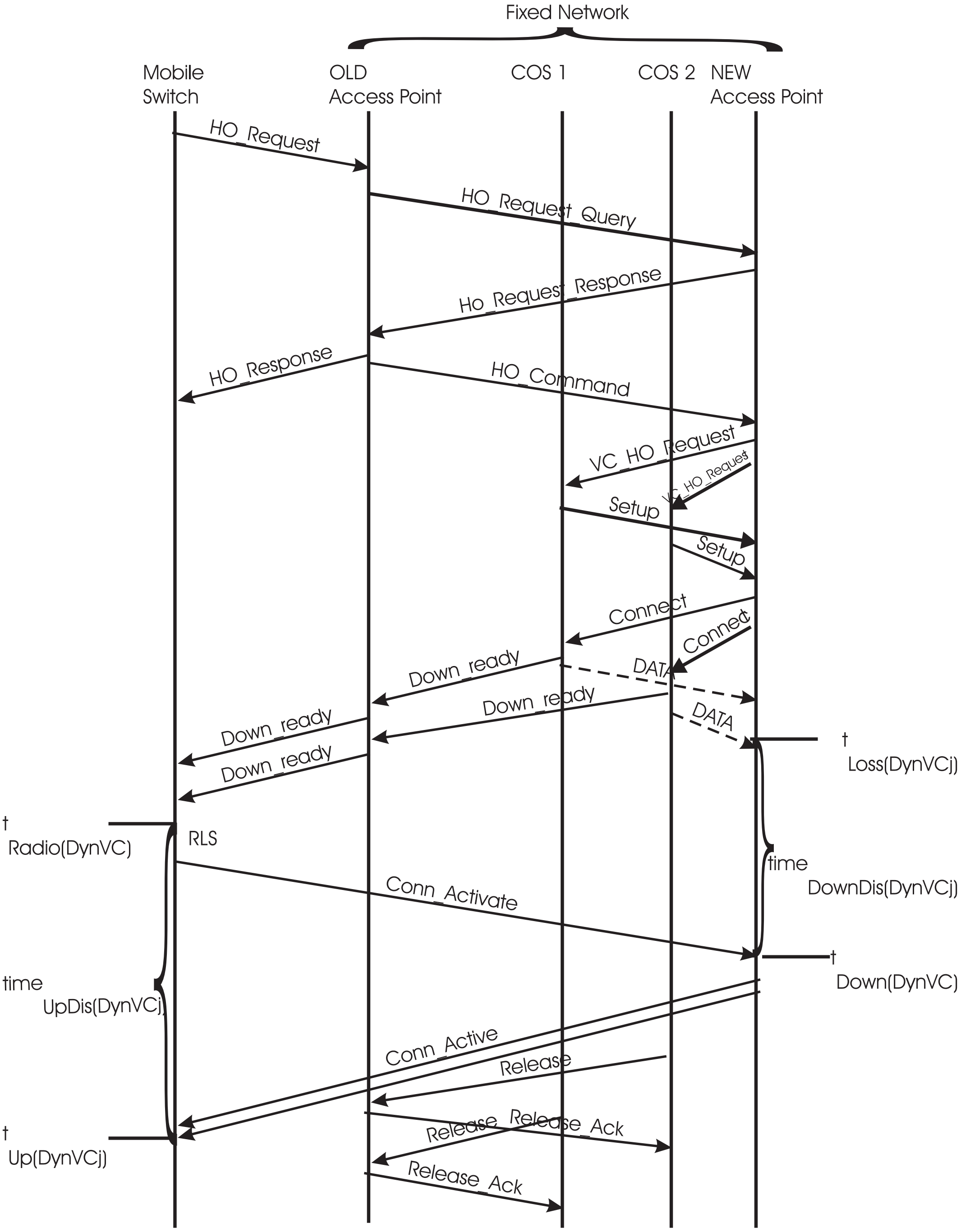


Figure 6.9: Dynamic Crossover Switch Discovery

into the allocated VP, extra VCs are established. At handover time, a VP according to the total bandwidth used (VCs in- and outside of the former VP) is set up.

Hereby, the advantages of the VP based anchor rerouting can be combined by the flexibility of the VC based anchor rerouting by the price of also having the adversity of sequential signaling for each extra VC. This has to be traded off against the potentially unused bandwidth of the VP approach.

The signaling corresponds to the chart shown for the VC based anchor rerouting. The only difference is that the signals drawn per VC only have to be sent for those VCs that did not 'fit' into the reserved VP bandwidth and QoS. So VC$_i$ is any of all the remaining VCs outside the VP, if there are any.

6.3.7 Path Extension

In this case (see figures 6.10 and 6.11), the crossover switch is the old access point. For this scheme all of the above approaches with fixed crossover switch can be looked at.

In terms of signaling, this approach can be considered being the simplest one. No crossover switch has to be determined and therefore no signaling between the old access point and the crossover switch is required, since they are exactly the same switch. The old path is maximally reused. In case of an air interface between the old and the new access point there is also exactly one hop between them, which in turn can nevertheless mean a long data propagation time due to the wireless medium.

Since the path is extended at each handover, the end-to-end delay also increases and more resources are allocated than necessary, because the path is suboptimal.

Another problem potentially arising by using this approach is the risk of having loops in the communication path when the mobile moves back and forth between the old and the new access point.

6.3.8 Soft Handover Combined with Dynamic Crossover Switch Discovery

This combination only works for the VC-based approach, since the VP-based approaches all share the same crossover switch. Using this approach, the assumption is that two radio links can be maintained at the same time. Whereas with hard handover, there is a point of synchronization for all VCs before the radio link is switched, thus losing cells until the new path for the last VC is set up, this is not the case with soft handover. With soft handover, each VC can be handed over individually and independently of other VCs. That way, the load due to handover signaling is distributed over time.

As with hard handover, here also two different scenarios can be considered as far as the crossover switch discovery is concerned. This can either be done by the known message VC_HO_Request or directly at the setup phase of the new partial path. That way one message less and only standard ATM signaling is used.

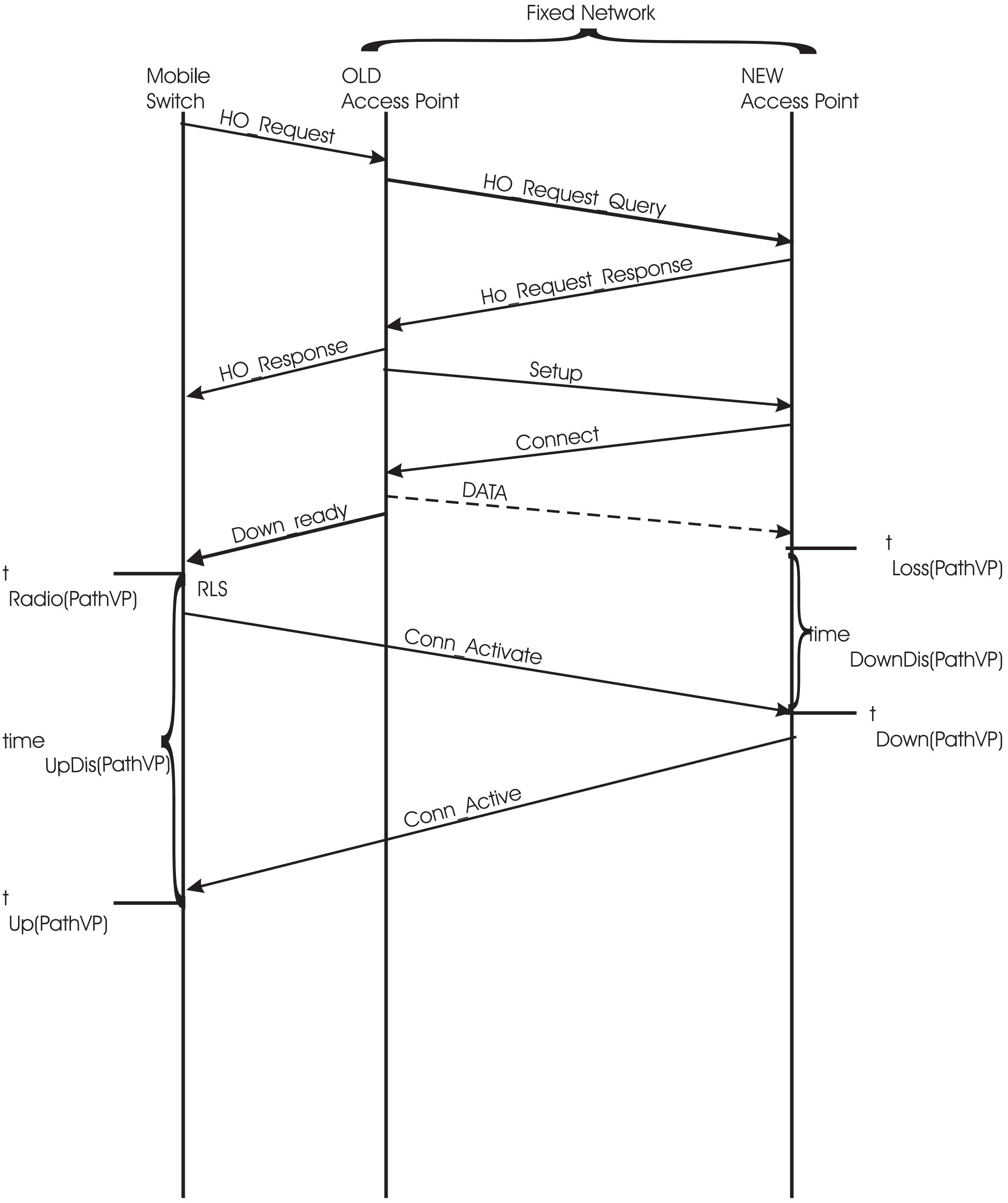


Figure 6.10: VP-based Path Extension

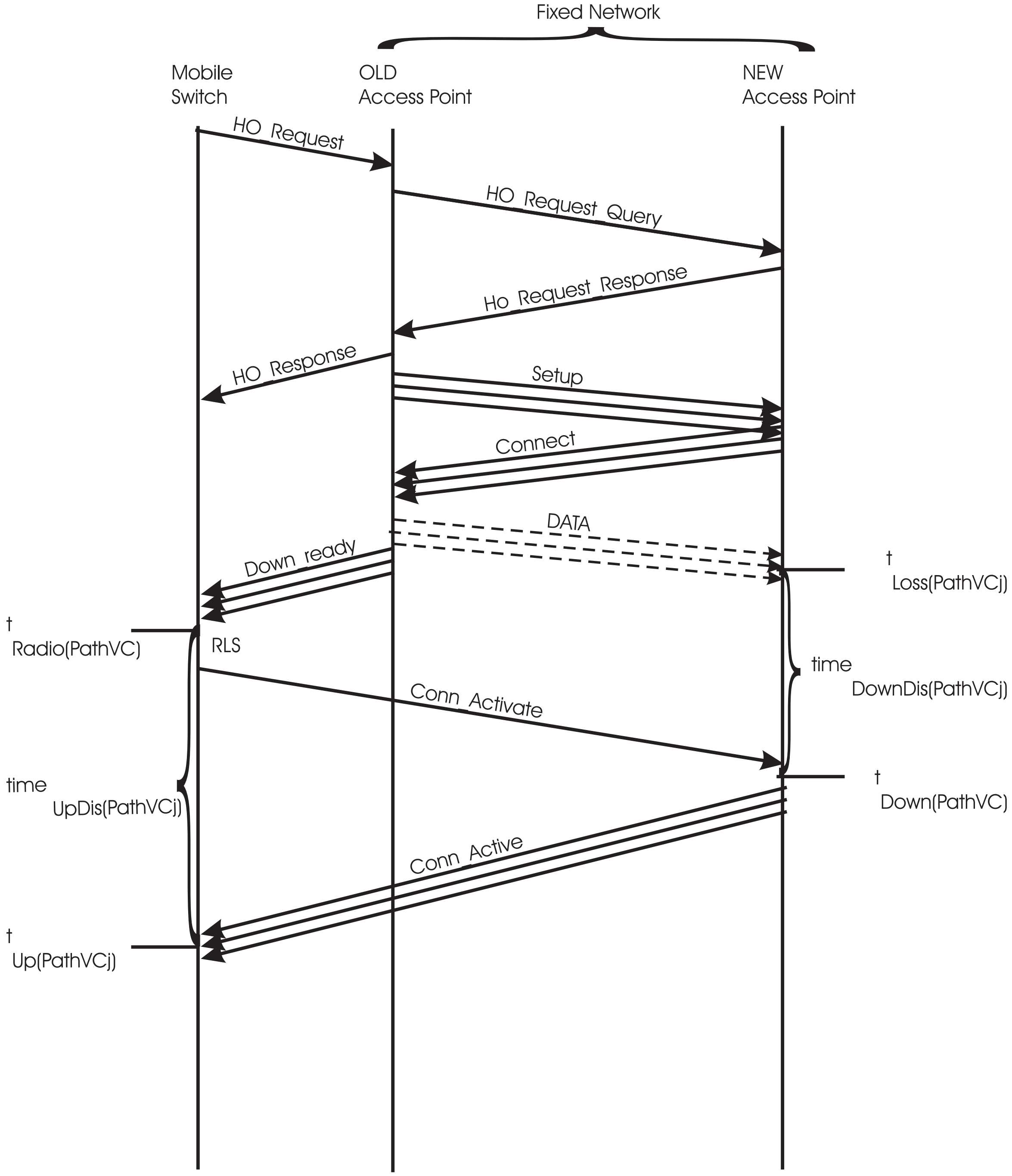


Figure 6.11: VC-based Path Extension

Signal Flow:

The messages shown in the 6.12 flow chart correspond to standard ATM signaling for connection establishment.

Although some messages are the same as the ones used for example in the dynamic crossover switch discovery algorithm for hard handover, the messages can be sent on direct paths to the destination. The fact that with soft handover two ATM links can be maintained at the same time can be exploited for signaling. That way, the mobile directly communicates with the new access point while at the same time it is still sending data on the old path via the old access point. This reduces the number of nodes affected by handover signaling, decreasing load impact and at the same time accelerating the handover process.

Data Flow:

Uplink The uplink data stream is switched as soon as the mobile system knows that the new part of the connection is fully established. It does so on reception of the Connect signal.

Downlink The crossover switch starts sending data down the link right after sending the Connect message.

6.3.9 Soft Handover Combined with VP-Based Anchor Rerouting

Also for soft handover, the alternative to establish a single VP containing all established VCs for the handover is considered (see figure 6.13) in order to minimize the signaling overhead and accelerate the handover process.

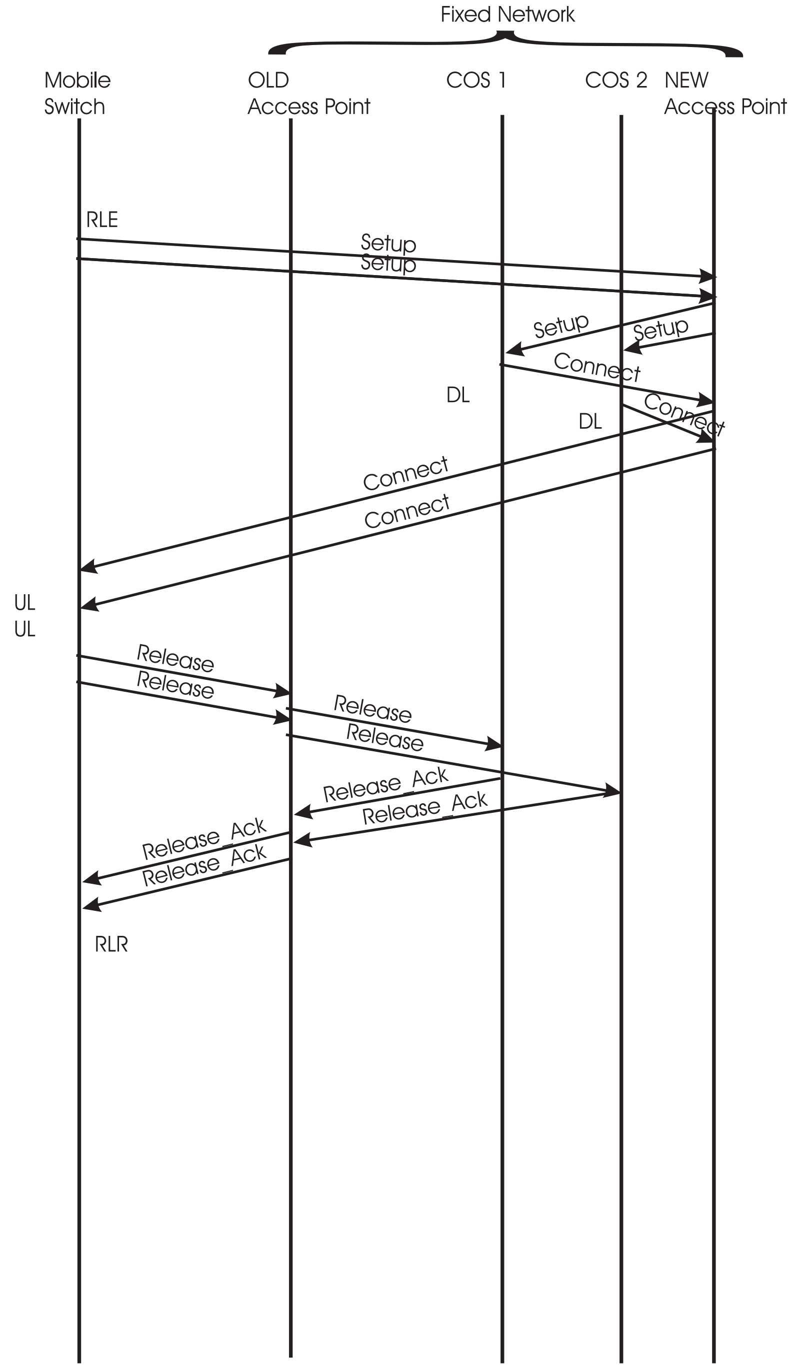


Figure 6.12: Soft Handover with Dynamic Crossover Switch Discovery

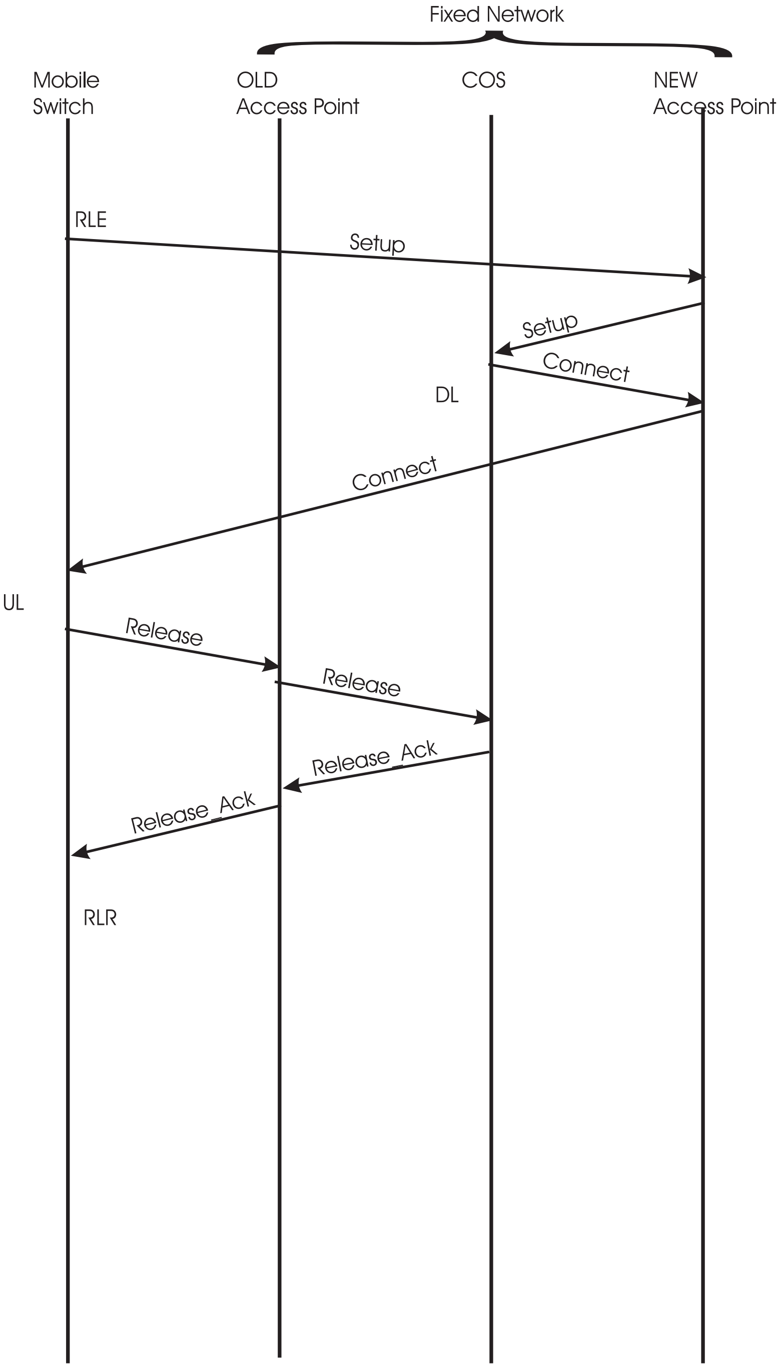


Figure 6.13: Soft Handover with VP-Based Anchor Rerouting

6.4 Performance Analysis

6.4.1 Cell Loss

Assuming no buffering, hard handover and moreover a steady stream of data sent up and down the link, cells are lost when sent during the disruption period due to the connection handover on the uplink and downlink, respectively. The data lost equals to the disruption period times the data rate.

$$data\ loss = time_{Dis} * rate$$

Therefore, the disruption period of the uplink and the downlink, respectively, have to be kept at a minimum in order to decrease the cell loss. The data rate is the sum of the rate of all the VCs that are active at handover time.

$$rate = \sum_{j=1}^{n} rate_{VC_j}$$

Since the data rate is the same for both the VP and VC-based approaches, provided there is sufficient bandwidth in the VP to accommodate all active VCs, the only relevant difference is the disruption period for the up- and downlinks.

$$rate = rate_{VP} = \sum_{j=1}^{n} rate_{VC_j}$$

For simplicity, it is also assumed that every VC has the same bandwidth. The disruption periods for the uplink (time_{UpDis}) and downlink ($\mathrm{time}_{DownDis}$), respectively, are calculated by subtracting $\mathrm{t}_{Radio(*)}$ (when the uplink data loss starts) from $\mathrm{t}_{Up(*)}$ (when the uplink data loss stops) and subtracting $\mathrm{t}_{Loss(*)}$ (when the downlink data loss starts) from $\mathrm{t}_{Down(*)}$ (when the downlink data loss stops). '*' stands for any of the strings added on the flow charts by using the different algorithms.

$$data\ loss\ on\ the\ uplink = time_{UpDis} * rate_{Uplink}$$

$$data\ loss\ on\ the\ downlink = time_{DownDis} * rate_{Downlink}$$

Uplink

The uplink disruption period is potentially different for each of the suggested approaches on handover, so a few examples are closely inspected and generalizations are derived.

VP-Based Anchor Rerouting

$$data\ loss\ on\ the\ uplink = time_{UpDis(VP)} * rate_{Uplink(VP)}$$

The period of time, when data is lost, starts at the switch of the radio link and ends on reception of the Conn_Active signal at the mobile system.

$$time_{UpDis(VP)} = t_{Up(VP)} - t_{Radio(VP)}$$

Within this period of time three events occur:

1 the radio link is switched

2 the Conn_Activate message is sent from the mobile to the new access point and processed there.

3 the Conn_Active message is sent from the new access point to the mobile and processed there.

These events can be represented by the following components (illustrated in Figure 6.14 along with those for the downlink disruption period.):

χ the time needed for the radio link switch.

ρ the propagation time for a signal over the wireless link

α the processing time for a signal like Down_Ready or Conn_Active. This consists of both receiving and generating a signal.

Therefore, the time$_{UpDis(VP)}$ consists of the corresponding periods of time:

$$\chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate) + \rho(Conn\_Active) + \alpha(Conn\_Active) = \chi + 2\rho + 2\alpha$$

VC-Based Anchor Rerouting, Shared Crossover Switch

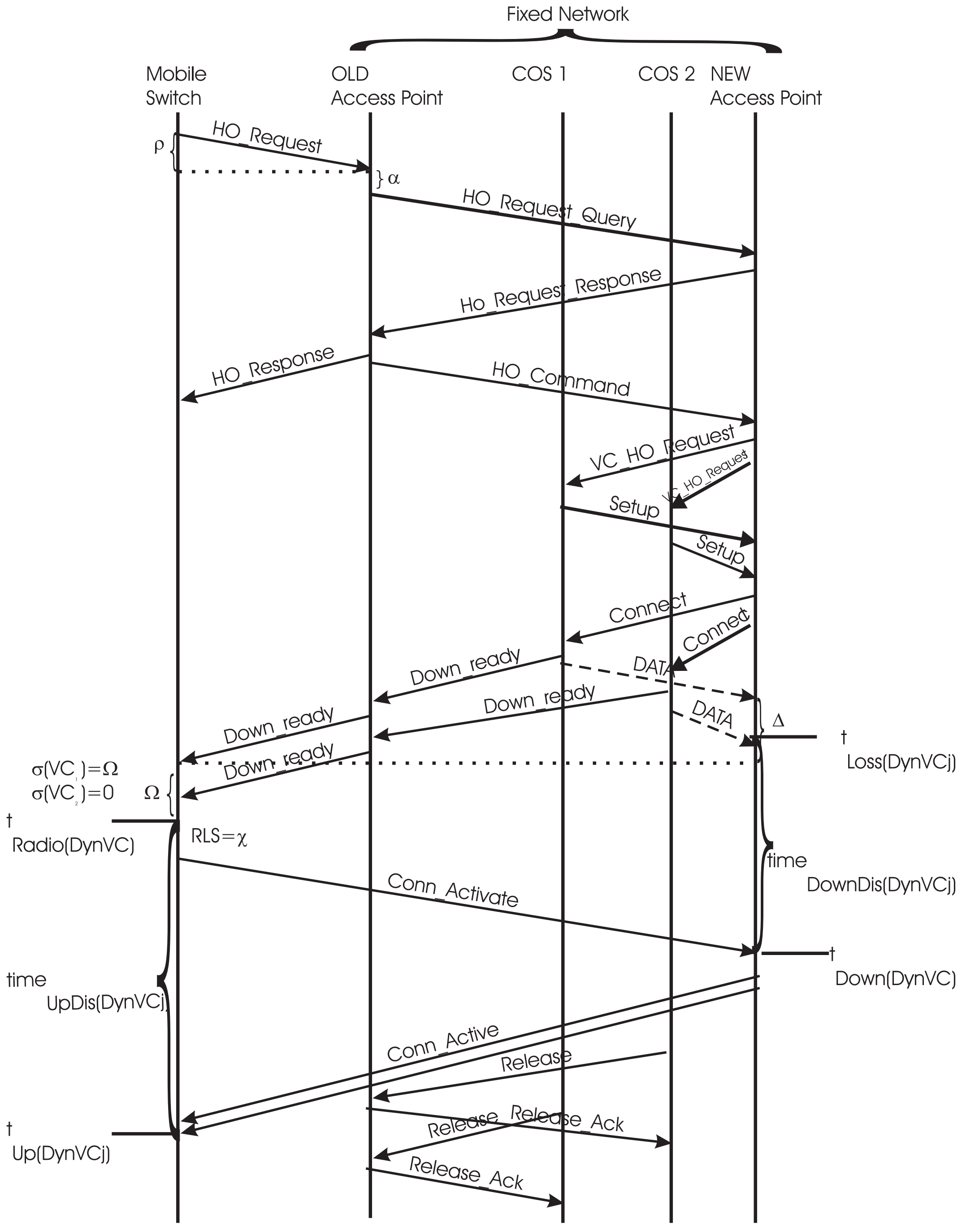


Figure 6.14: Time components

$$data\ loss\ on\ the\ uplink = time_{UpDis(VC)} * rate_{Uplink(VC)}$$

Since the uplink disruption period is multiplied by the total uplink data rate to get the amount of lost data, the average uplink disruption period for all VCs is being looked at. Therefore

$$\overline{time_{UpDis(VC)}} = \frac{\sum_{j=1}^{n} time_{UpDis(VC_j)}}{n}$$

VC$_j$ being any VC between VC$_1$ and VC$_n$ in the sequence of n VCs to be handed over.

The period of time, when data is lost, starts at the switch of the radio link and ends on reception of the Conn_Active signal for VC$_j$ at the mobile system. In the figures, this can be seen as:

$$time_{UpDis(VC_j)} = t_{Up(VC_j)} - t_{Radio(VC)}$$

Comparing to the VP equivalent for the time instances, both the radio link switch and the resumption of the uplink are delayed because of twice the additional signal processing α per VC between the signals, one before and one after the radio link switch (for processing Down_Ready and Conn_Active).

For the disruption period only the delay of processing the Conn_Active message matters. Because of this one additional α per VC for the Conn_Activate message,

$$\frac{(n-1)}{2} * \alpha$$

are part of the average time$_{UpDis(VC)}$, resulting in:

$$\overline{time_{UpDis(VC)}} = \chi + 2\rho + \frac{(n+3)}{2} * \alpha$$

So the average uplink disruption period for the VC-based anchor rerouting approach is the one for the VP approach plus the additional αs, if the number of VCs exceeds one.

$$\overline{time_{UpDis(VC)}} \geq time_{UpDis(VP)}$$

Hybrid Anchor Rerouting

To get the average data loss, the loss of all VCs in- and outside the VP has to be taken into account. The following calculation assumes that the VP is handed over first, followed by the VCs outside the VP:

$$\overline{time_{UpDis(hybrid)}} = \frac{time_{UpDis(VP)} \;*\; n \;+\; \sum_{i=1}^{k} i \;\alpha}{n}$$

(k being the number of VCs outside the VP, VC$_i$ between VC$_1$ and VC$_k$ in the sequence of k handled VCs outside of the VP.)

Similar to the VC-based anchor rerouting approach, the additional signals matter. In the case of the hybrid approach, these are the signals used for the k VCs outside the VP.

$$\overline{time_{UpDis(hybrid)}} = \chi + 2\rho + 2\alpha + \frac{\sum_{i=1}^{k} i}{n} * \alpha$$

$$= \chi + 2\rho + 2\alpha + \frac{(k+1) * k}{2n} * \alpha$$

Depending on the number of VCs outside the VP, the hybrid approach can be closer to the VP or to the VC approach as far as uplink cell loss is concerned.

$$time_{UpDis(VP)} \leq \overline{time_{UpDis(hybrid)}} \leq \overline{time_{UpDis(VC)}}$$

Other VC-Based Algorithms

The equation for VC-based anchor rerouting with shared crossover switch also applies to the dynamic crossover switch discovery approach, the VC-based anchor rerouting with potentially different crossover switches for each VC, as well as to the VC-based path splicing approaches and VC-based path extension.

This fact may seem strange at first, for one would expect the variety of crossover switches, with each having a relatively seen different distance to the access points, to have an impact on the disruption period for the approaches where the crossover switch is not shared. At least for the uplink this turns out to be false, which can be explained by the following reasons. According to the assumption that the uplink switch to the new path can be done independently from the downlink switch, the data only starts to be lost at the time the radio link is switched. It is therefore entirely irrelevant for the uplink disruption period, if the radio link switch is performed later because of the additional delay introduced by the dynamic crossover switch discovery approach. The time when the uplink stops is only relevant in relation to when it is started again, which is independent of the crossover switches, since it only requires communication between the mobile and the new access point.

As for all other approaches as well, the uplink disruption period for any given VC$_j$ lies between the radio link switch and the reception of the jth Conn_Activate message at the mobile.

Other VP-Based Algorithms

The equations for VP-based anchor rerouting also apply for the other VP-based approaches, namely path extension and path splicing.

Other Hybrid Algorithms

The equations for hybrid anchor rerouting also apply for the other hybrid approaches, namely path extension and path splicing.

Comparison of Uplink Cell Loss

Comparing all approaches with respect to the uplink disruption period, meaning actual data loss, the following applies:

$$time_{UpDis(VP-based)} \leq time_{UpDis(hybrid)} \leq time_{UpDis(VC-based)}$$

$$time_{UpDis(VP-based)} = \chi + 2\rho + 2\alpha$$

$$time_{UpDis(VC-based)} = \chi + 2\rho + \frac{(n+3)}{2} * \alpha$$

$$time_{UpDis(hybrid)} = \chi + 2\rho + 2\alpha + \frac{(k+1) * k}{2n} * \alpha$$

Base	Path Extension	Path Splicing	Anchor Rerouting
VC, shared COS	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$
VC, variable COS	n/a	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$
VC, dynamic COS	n/a	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$	$\chi+2\rho+\frac{(n+3)}{2}\ \alpha$
VP, shared COS	$\chi+2\rho+2\alpha$	$\chi+2\rho+2\alpha$	$\chi+2\rho+2\alpha$
hybrid, shared COS	$\chi+2\rho+2\alpha+\frac{(k+1)*k}{2n}\ \alpha$	$\chi+2\rho+2\alpha+\frac{(k+1)*k}{2n}\ \alpha$	$\chi+2\rho+2\alpha+\frac{(k+1)*k}{2n}\ \alpha$

Table 6.1: Uplink Disruption Periods

Downlink

Lost data on the downlink = downlink disruption time * downlink data rate. The data rate is the sum of the downlink rate of all the VCs that are active at handover time.

The downlink disruption time consists of several components. Some of these components are invariant for all the connection handover algorithms presented, others only appear with certain algorithms. This again depends on the categories that the algorithms are in, whether they constitute hard or soft, forward or backward, VC-, VP-based or hybrid handover and where the crossover switch is.

The components identified to have an influence on the downlink disruption time for at least some of the approaches, although some of them are valid for all, are the following:

Δ	the difference in time between the arrival of the Down_Ready signal at the mobile system and the arrival of the first cell at the new access point, sent down the link by the crossover switch. These events occur after the downlink has been switched at the crossover switch. The time difference depends on the location of the crossover switch, i. e. the number of hops to the old and to the new access point, respectively, as well as on the load of the nodes.
Ω	the time until the last VC is ready for the radio link switch. With hard handover this is a point of synchronization. Starting after the processing of the first Down_Ready signal at the mobile system, this time ends after the processing of the Down_Ready signal of the last VC at the mobile.
σ	the deviation of a particular VC from Ω.
χ	the time needed for the radio link switch.
ρ	the propagation time over the wireless link
τ	the propagation time over a wired link for one hop.
α	the processing time for a signal like Down_Ready or Conn_Active, consisting of both receiving and generating a signal.
α_{data}	the time data spends in an ATM switch (transmission delay, switching delay, etc.).
δ	the time data travels from the new access point to the mobile via the crossover switch.

Figure 6.14 illustrates the above mentioned time components.

As in the section covering the uplink, some relevant examples are taken and based on their behavior in terms of cell loss a generalization for handover categories is given.

VP-Based Anchor Rerouting

The amount of data lost on the downlink equals to the downlink disruption period times the data rate of the VP.

$$\mathit{data\ loss} \ = \mathit{time}_{DownDis(VP)} * \mathit{rate}_{Downlink(VP)}$$

The downlink disruption period lies between the point in time, when data is lost at the new access point because of the missing link to the mobile, henceforth referred to as $\mathrm{t}_{Loss(VP)}$, and the point in time, when the downlink is resumed ($\mathrm{t}_{Down(VP)}$).

$$\mathit{time}_{DownDis(VP)} = t_{Down(VP)} - t_{Loss(VP)}$$

Within these boundaries, there are the following components added up:

$$\mathit{time}_{DownDis(VP)} = \Delta + \alpha(\mathit{Down\_Ready}) + \chi + \rho(\mathit{Conn\_Activate}) + \alpha(\mathit{Conn\_Activate})$$

VC-Based Anchor Rerouting, Shared Crossover Switch

$$\mathit{data\ loss\ on\ the\ downlink} = \sum_{j=1}^{n} (\mathit{time}_{DownDis(VC_j)} * \mathit{rate}_{Downlink(VC_j)})$$

$$= \frac{\sum_{j=1}^{n} \mathit{time}_{DownDis(VC_j)}}{n} * \mathit{rate}_{Downlink(VC)}$$

Since the average downlink disruption period matters, VC_j is used as a reference:

$$\overline{\mathit{time}_{DownDis(VC)}} = \frac{\sum_{j=1}^{n} \mathit{time}_{DownDis(VC_j)}}{n}$$

$$time_{DownDis(VC_j)} = t_{Down(VC)} - t_{Loss(VC_j)}$$

Within these boundaries, there are the following components added up:

$$time_{DownDis(VC_j)} = \Delta_j + \alpha(Down\_Ready) + \sigma_j + \chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate)$$

Given that the crossover switch is shared, σ can be computed, as it results from the necessity to wait for the last Down_Ready signal to be processed at the mobile in order to switch the radio link. The radio link switch constitutes a point of synchronization for the VCs. When the data in VC$_j$ actually get lost, depends on the number of VCs.

σ therefore can be resolved as (n - j) * α to process the remaining Down_Ready signals.

$$time_{DownDis(VC_j)} = \Delta_j + \alpha(Down\_Ready) + (n-j) * \alpha(Down\_Ready) + \chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate)$$

resulting in the overall average of:

$$\overline{time_{DownDis(VC)}} = \overline{\Delta} + \frac{n+1}{2} * \alpha(Down\_Ready) + \chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate)$$

The simplified equation is:

$$\overline{time_{DownDis(VC)}} = \overline{\Delta} + \chi + \rho + \frac{n+3}{2} * \alpha$$

This leads to the following comparative equation:

$$\overline{time_{DownDis(VC)}} = time_{DownDis(VP)} + \frac{n-1}{2} * \alpha$$

Therefore, depending on the number of VCs,

$$time_{DownDis(VC)} \geq time_{DownDis(VP)}$$

Hybrid Approach

With the hybrid approach, some VCs are handed over outside of the VP that bundles several connections. Let k be the number of VCs outside the VP and VC$_i$ between VC$_1$ and VC$_k$ in the sequence of k handled VCs outside of the VP. In this scenario, all the VCs in the VP have to wait for additional k times the processing of the Down_Ready signal. Each VC$_i$ outside the VP has to wait for the rest of the VCs to have their Down_Ready signal processed, resulting in (k - i)α.

The following equation reflects this reasoning. The first part (n($\overline{\Delta}$ + α + χ + ρ + α)) shows the minimum of downlink disruption period for n VCs. This is the case, when k = 0, handing over all the VCs within one VP. As k grows, all the VCs within the VP (there are (n - k) of them) have to wait for all the k outside VCs (kα). Finally, all the outside VCs have to wait for their successors ($\sum(k - i)\ \alpha$). To get the average, the result is divided by n, the total number of VCs.

$$\overline{time_{DownDis(hybrid)}} = \frac{n(\overline{\Delta} + \chi + \rho + 2\alpha) + (n - k)k\alpha + \sum_{i=1}^{k}(k - i)\alpha}{n}$$

Simplified:

$$\overline{time_{DownDis(hybrid)}} = \frac{n(\overline{\Delta} + \chi + \rho + 2\alpha) + (n - k)k + \frac{k*(k-1)}{2}) * \alpha}{n}$$

Compared to the other approaches so far, the hybrid approach can be closer to the performance of the VP or VC-based anchor rerouting, depending on the number of VCs outside the VP.

$$\overline{time_{DownDis(hybrid)}} = time_{DownDis(VP)} + \frac{2nk - k^2 - k}{2n}\alpha$$

The cell loss can be reduced by handing the VP over after all the VCs have been handed over. This way, the VCs inside the VP do not need to wait for all the VCs outside, and the VCs outside only have to wait for one more, the VP. The equations above assumed the contrary, namely starting by the VP and then handing over the VCs outside the VP. Reversing the order would result in:

$$\overline{time_{DownDis(hybrid)}} = \overline{\Delta} + \chi + \rho + 2\alpha + \frac{\sum_{i=1}^{k}(k - i + 1)}{k}\alpha$$

$$= \overline{\Delta} + \chi + \rho + 2\alpha + \frac{\sum_{i=1}^{k} i}{k}\alpha$$

$$= \overline{\Delta} + \chi + \rho + 2\alpha + \frac{(k * (k-1))}{2k}\alpha$$

Regardless of the order,

$$time_{DownDis(VP)} \leq time_{DownDis(hybrid)} \leq time_{DownDis(VC)}$$

Dynamic Crossover Switch Discovery

$$\overline{time_{Down(DynVC)}} = \frac{\sum_{j=1}^{n} time_{Down(DynVC_j)}}{n}$$

$$time_{Down(DynVC)_j} = t_{Down(DynVC)} - t_{Loss(DynVC_j)}$$

The time between these two boundaries is composed of the following:

$$time_{Down(DynVC)_j} = \Delta_j + \alpha(Down\_Ready) + \sigma_j + \chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate)$$

In contrast to the VC-based handover with shared crossover switch, it is now more complicated to see what is behind σ, since the VCs no longer have a constant gap at the arrival of Down_Ready at the mobile.

The downlink disruption period, being the time between t$_{Loss(VC_j)}$ (i.e. when the first data sent via the new path on VC$_j$ arrives at the new access point) and the processing of the Conn_Activate at the new access point (t$_{Down(VC)}$), therefore can not be predicted knowing the number of VCs. How long this period is, depends strongly on when the radio link is switched, which in turn depends on when the last VC is ready. This time, Ω, as explained in the beginning of this section, depends on how long it takes the slowest VC to get all its signaling concerning the handover done. Assuming equal processing times at each node for simplicity, two factors determine who the last VC is. First, it depends on which rank the VC has in the order of the handover sequence and, second, on the location of the crossover switch for this VC.

Ω therefore is a function of the worst case VC's rank and number of hops between the crossover switch and the old access point (1 signal concerned) and more importantly the number of hops between the crossover switch and the new access point (3 signals involved) - all in relation to the properties of the best case VC (again, see the definition of Ω in the beginning of the section).

$$\begin{aligned}\Omega = & max(rank * \alpha + 3 * (\alpha + \tau) * (\#hopsToNew + \#hopsToOld)) \\ & -min(rank * \alpha + 3 * (\alpha + \tau) * (\#hopsToNew + \#hopsToOld))\end{aligned}$$

For the worst case VC (VC_{wc}) and the best case VC (VC_{wc}), the following holds:

max(rank*α+3*(α+τ)*(#hopsToNew+#hopsToOld)) =

rank$_{wc}$*α+3*(α+τ)*(#hopsToNew$_{wc}$+#hopsToOld$_{wc}$)

min(rank*α+3*(α+τ)*(#hopsToNew+#hopsToOld)) =

rank$_{bc}$*α+3*(α+τ)*(#hopsToNew$_{bc}$+#hopsToOld$_{bc}$)

Zooming into the components we get:

$$\Delta_j = (\alpha + \tau) * \#hopsToOld_j + \rho(Down\_Ready) - (\alpha_{data} + \tau) * \#hopsToNew_j$$

$$\begin{aligned}\sigma_j = & (rank_{wc} * \alpha + 3 * (\alpha + \tau) * (\#hopsToNew_{wc} + \#hopsToOld_{wc}) \\ & -(rank_j * \alpha + 3 * (\alpha + \tau) * (\#hopsToNew_j + \#hopsToOld_j))\end{aligned}$$

In order to have the values for the variables that the disruption period depends on, some assumptions on propagation, transfer, computation, processing time etc. have to be made along with a probability distribution for the number of hops likely to be between the access points and the crossover switch. Without taking into account some factors like number of VCs, distances, etc. no performance comparison to e.g. the VP-based anchor rerouting approach can be made.

$$\overline{time_{Down(DynVC)}} = \overline{\Delta} + \overline{\sigma} + \chi + \rho + 2\alpha$$

VC-Based Anchor Rerouting

All the results shown above for rerouting with dynamic crossover switch discovery are also valid for regular VC-based anchor rerouting, where every VC can potentially have a different crossover switch. The fact that with this approach the crossover switch is predetermined instead of dynamically discovered does not have an influence on the equations themselves but potentially on the location of the crossover switch, which in turn determines Δ, Ω and σ.

Path Splicing

This approach differs significantly from all the others in its behavior regarding lost data on the downlink. With path splicing, the data flow down the link can continue until the actual radio link switch takes place. The cell loss does not depend on other VCs. The components that constitute the uplink disruption time are:

$$\overline{time_{DownDis(PS)}} = \chi + \rho(Conn\_Activate) + \alpha(Conn\_Activate) + \overline{\delta}$$

δ is the time data is already traveling from the new access point to the mobile system via the crossover switch and the old access point. All the cells underway can not be redirected to the mobile and are lost.

The downlink disruption period is invariant to VC-, VP-based or hybrid approaches, when path splicing is applied.

Path Extension

Data takes longer to travel on a wireless link than on its wired equivalent. This has a positive effect on data loss on the downlink in VP as well as VC-based path extension, if the access points are connected by a wireless link themselves. Compared to their anchor rerouting counterparts, the path extension approaches do not need the Down_Ready message from crossover switch to old access point, that would increase the downlink disruption period. These two factors contribute to a minimization of Δ.

Comparison of Downlink Cell Loss

Base	Path Extension	Path Splicing	Anchor Rerouting
VC, shared COS	$\Delta+\chi+\rho+\frac{n+3}{2}\ \alpha$	$\chi\ +\rho+\alpha+\overline{\delta}$	$\overline{\Delta}+\chi+\rho+\frac{n+3}{2}\ \alpha$
VC, variable COS	n/a	$\chi\ +\rho+\alpha+\overline{\delta}$	$\overline{\Delta}+\overline{\sigma}+\chi+\rho+2\alpha$
VC, dynamic COS	n/a	$\chi\ +\rho+\alpha+\overline{\delta}$	$\overline{\Delta}+\overline{\sigma}+\chi+\rho+2\alpha$
VP	$\Delta+\chi+\rho+2\alpha$	$\chi+\rho+\ \alpha+\delta$	$\Delta+\chi+\rho+2\alpha$
hybrid	$\Delta+\chi+\rho+2\alpha\ + \frac{2nk-k^2-k}{2n}\alpha$	$\chi\ +\rho+\alpha\ +\delta$	$\Delta+\chi+\rho+2\alpha\ + \frac{nk-k^2-k}{2n}\alpha$

Table 6.2: Downlink Disruption Periods

Note that Δ, δ and σ can be different for each connection due to the variation of the crossover switch location. For path extension and other shared crossover switch schemes, Δ is at least the same for all the connections of a particular handover.

The other components are constants invariant to schemes and locations.

6.4.2 End-to-End Delay

Hereby the end-to-end delay on the new path is measured. Since the absolute number does not reveal much on the quality of the connection handover, it has to be related to another value. This other value can be the end-to-end delay of the connection on the old path or to the end-to-end delay on the new path achieved by other handover approaches.

The end-to-end delay on the new path for a particular VC is obviously the same for all approaches that are based on either anchor rerouting with a shared crossover switch for all connections, or path extension. The relation to the end-to-end delay of the old path depends on the selection of the crossover switch. In the worst case for the end-to-end delay on the new path, path extension, the end-to-end delay on the old path is increased by the delay between the old and the new access point at each connection handover process.

While the crossover switch is the same for all connections using the approaches mentioned above, determining the crossover switch for each connection potentially shortens the delay by allowing for a custom crossover switch for each connection. However, if the crossover switch is determined at setup time, the location of the new access point can not be taken into account for location optimization.

The dynamic crossover switch discovery approach promises a better performance as far as end-to-end delay is concerned, for by discovering the best crossover switch for each connection, the total path gets closer to being optimal between the sender and the receiver. Again, the total delay depends on the location of the crossover switch, which is in turn is a function of the selected dynamic crossover switch discovery algorithm. In this text only the proposed combination of electing a candidate crossover switch at connection establishment combined with a loose select algorithm at handover time has been considered. A comparison between this and other crossover switch discovery algorithms would be worth making.

Path splicing is a special case regarding the end-to-end delay. During the time when data is sent along the spliced-in part of the path, the delay increases because of the newly added delay of double the path between the crossover switch and the new access point. But this additional delay is only temporary until after the handover process.

6.4.3 Sequencing

For all the approaches described in this chapter, sequencing is guaranteed. Although data flow on the new path could be faster than on the old path, having cells that are sent later arrive earlier at the crossover switch, this does not mean that cells are received out of sequence. Since once the crossover switch switches to the new path, it does not return to the old path again, the cells can not be received out of sequence.

If buffering is used, which has not been the case in throughout this thesis, some mechanism, like an indication of the last cell, has to be introduced to avoid having cells arrive out of order. The only algorithm where sequencing is guaranteed up and down the link is path extension. Path splicing ensures sequencing on the downlink data flow, but not on the uplink.

6.4.4 Cell Duplication

Since only point-to-point connections are used, no multicast entailed, cells are not duplicated in any of the handover approaches presented in this chapter of the thesis.

6.4.5 Total Handover Delay

Depending on what is ultimately considered to be the ignition of the handover, assuming it to be the sending of the HO_Request message, and what to be the end, assuming the reception of the last Conn_Active message at the mobile switch, the total time required to hand over all active connections varies according to the different approaches.

The total delay of the handover process depends on several factors. First, the number of signals to be processed. The more signals are involved, the longer they take to be processed. Simple schemes are therefore preferable. VP-based approaches outperform the others by drastically reducing the signaling overhead. Soft handover also requires comparatively little signaling. Second, parallel processing of signals decreases the time spent on signal processing. Distributing the processing load over several crossover switches allows for parallelism. For VC-based approaches, sharing the crossover switch is therefore adverse. Third, the location of the crossover switch is crucial in terms of delay. The number of nodes between the crossover switch and the access points contribute to the delay by adding processing, transmission and propagation delay.

These factors contribute to the tendency that VP-based schemes have shorter total handover delays than VC-based schemes with varying crossover switches, which in turn are faster than VC-based schemes with a shared crossover switch.

6.4.6 Route Reuse

Because of the spatial locality in movement, it is very likely that the re-established path to the new access point shares most of the nodes and VPs in the old path. It is more efficient to reuse most of the VC segments of the old path instead of clearing them.

Obviously the path extension scheme offers the maximal reuse of the old path. For the approaches based on anchor rerouting with shared crossover switch, this parameter again remains the same for all connections handed over, regardless of the approach. As mentioned when talking about the end-to-end delay, here again the crossover switch is the crucial criterion. Its location determines the length of the rerouted partial path. The location depends on the network topology, the policy for crossover switch determination and discovery and on the availability of crossover enabled ATM switches in the network.

Double booking on the same path occurs with path splicing. Potentially this can also happen in the case of predetermined crossover switches. The only algorithm of those presented throughout the thesis to prevent double booking is dynamic crossover switch discovery, that makes sure to stop allocating resources as soon as it hits the old path.

6.4.7 Scalability

Again, the approaches using the least number of signaling and having the shortest total handover delay perform better in terms of scalability to a larger number of connections. To get a better grip on the overall scalability issues, the numerical results will be analyzed.

6.4.8 Robustness

Inherently, soft handover is the approach that performs best in terms of robustness for the second active radio link that allows to tolerate one radio link failure. The hard handover approaches have to resort to forward handover in case of a radio link failure, which takes longer because of the need of additional communication between the access points. This additional disruption period contributes to increased cell loss.

Regarding network link or node failures, all the approaches are equally robust provided that mechanisms like time-outs and fallback scenarios are built in. These mechanisms are approach invariant.

6.4.9 Changes to Signaling

{ATM!signalingSoft handover uses the least amount of changes to the signaling. With this approach, only standard ATM signals are used, only requiring the use of additional information elements within the Setup and Connect messages. The benefit of this is that only the ATM switches involved in mobility (the mobile, the access points and the crossover switch) need to be able to interpret and use the information elements. For all other switches, the messages appear as mere Setup and Connect messages and can be processed as usual.

All other approaches add new messages to the signaling protocol. This causes the need to update also all potential intermediate switches to enable them to process the additional messages, which adversely effects transparency and migration issues.

6.4.10 Quality of Service

Although some of the criteria for Quality of Service, such as cell loss and end-to-end delay, have already been discussed, many issues remain open for further investigation:

Bit Rates Contracts on CBR, ABR, UBR, negotiated at setup time of the connections, may no longer be fulfilled after connection handover.

Cumulated QoS Values Since the partial path to the crossover switch is always set up from the new access point, it can happen, that the old path was set up from the far end party, leaving the crossover switch with two partial paths set up in its direction. Cumulated values such as end-to-end delay increase with each hop on the path of a connection. A special calculation has to be carried out in order to cope with setups in different directions.

Prioritization The order in which the connections are handed over and a priority scheme for the case that the new access point can not accommodate all active VCs are a crucial part of Quality of Service ensuring. This requires two mechanisms: one to give priority to certain VCs based on a priority scheme, for example on ATM address or bit rate, and another one to actually comply with this priority scheme when it comes to dropping connections and ordering the remaining ones.

Delay Variation Jitter increases due to different path lengths of the old and new path.

6.4.11 Data Looping

The only connection handover algorithm, where data looping can occur as a result of zigzag moving between access points, is path extension. A mechanism to detect loops a soon as the mobile system attaches to the new access point has to be built in that algorithm. Path splicing has one inherent loop between the crossover switch and the new access point. But this loop is planned and only temporary. The data goes from the crossover switch to the new access point and back only once, there is no danger of several loops in that algorithm.

6.4.12 Symmetry

The only approach that is symmetric on up- and downlink is the one described as example for soft handover. All the hard connection handover algorithms behave differently for up- and downlink data flow.

6.4.13 Transparency

All the approaches that do not use standard ATM signals only, but introduce new ones (like HO_Request etc.) require all the nodes between and including the access points and the crossover switch to be able to process these extra signals and therefore be aware of the mobility. By restricting access points and crossover switches to a limited part of a network or to a access point provider network, also the awareness of the mobility is restricted and transparent to the rest of the network. If only standard ATM signals are used (like Setup and Connect), then only the access points and the crossover switch have to be aware of the mobility. This is only the case with the scenario shown for soft handover.

6.5 Parameters

Number of VCs Varying the number of VCs concerns all approaches that do not use a VP, since the signals sent out for each VC have to be processed sequentially when

Quality Criterion	Variables	Relation to Topology
Cell Loss	Δ, σ, δ	probabilities of number of hops between the crossover switch and the access points, number of VCs handed over one crossover switch, load
End-to-End Delay	end-to-end delay on the new vs. the old path	probabilities of number of hops from the crossover switch to the new vs. to the old access point
Path Reuse	distance between the crossover switch and the old access point	probabilities of number of hops between the crossover switch and the old access point
Total Handover Delay	location of the crossover switch	probabilities of number of hops between the crossover switch and the access points, number of VCs handed over one crossover switch, load

Table 6.3: Impact of Topology on Performance

sent from the same source.

Data Rate of VCs In order to determine data loss, the data rate has to be known. The total bandwidth used is of interest.

Times for signal processing and data transmission and propagation. Included in the time parameters could be assumptions on load in the various nodes.

Network Topologies The reason for investigating the impact of different network topologies on the performance of connection handover in wireless mobile ATM networks is twofold. Knowing the relationship between a network topology and a handover algorithm and given one of them, an optimization of the other part of the relationship can be found. Given a network, one can determine the best handover algorithm to optimize some of the quality criteria. In the reverse direction, the research results can give design guidelines for a new network of an access provider. Heuristics can be used to find a suitable topology. In addition to that, the number of necessary crossover enabled ATM switches and the best location to place them within the network of the access point provider can be found based on the information on the topology impact. The quality criteria, for which the location of the crossover switch is crucial, are cell loss, end-to-end delay, route reuse and the total handover delay.

The information on the network topology is used to get a probability distribution of the number of hops between the crossover switch and the two access points, which has a major impact on connection handover performance.

The performance of the handover algorithms at the quality criteria mentioned above varies according to the parameters shown in table 6.3.

Mobile as well as private fixed networks are linked to the access point provider's network through access points. These access points serve as wired or wireless access points. Since the partial paths from the access points to the terminals within

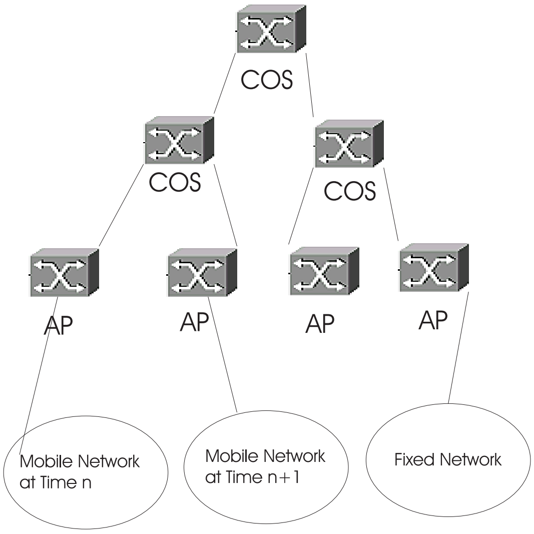


Figure 6.15: Model of an Access Point Provider Network with a Tree Topology

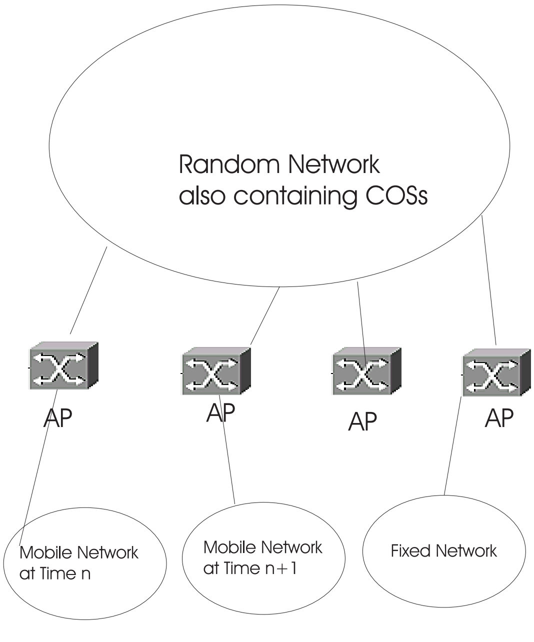


Figure 6.16: Model of an Access Point Provider Network with Random Topology

a mobile or private network do not change because of the handover, VCs start and end at access points. The case that a VC ends within the network of the access point provider is not considered. The figures 6.15and 6.16 show examples of access point provider networks with a tree and a random topology, respectively. Given values for the variables of the topology along with a number of VCs, the result should include probability distributions of:

Number of Hops between the Crossover Switch and the Old Access Point This parameter has an influence on the downlink disruption period as well as on the end-to-end delay. For anchor rerouting approaches, the crossover switch should be close to the old access point to minimize downlink data loss. The optimum is achieved in path extension approaches, where the old access point acts as crossover switch. For the dynamic crossover switch discovery approach the worst case crossover switch (see section on cell loss) should be close to the old access point as well. Uplink disruption times are independent of the crossover switch location for all the scenarios shown here.

Number of Hops between the Crossover Switch and the New Access Point This parameter has a major influence on the downlink disruption period in case of varying crossover switches per VC, an impact on all downlink disruption times with shared crossover switch and also on the end-to-end delay. For varying dynamic crossover switches this number should ideally be the same for each VC in order to minimize the downlink disruption period. Any deviation means cell loss. For the individual VC it is best to have many hops between the crossover switch and the new access point. Uplink disruption periods as well as downlink disruption periods for path extension are independent of this parameter.

Number of VCs per Crossover Switch The number of VCs increases the number of signals to be processed, which in turn adds to the load and therefore to the delay.

Number of Hops on the Old Path for end-to-end delay determination.

Number of Hops on the New Path for end-to-end delay determination.

Number of Hops Reused for reuse determination. This value can also be calculated by subtracting the hops from the crossover switch to the new access point from the number of hops on the new path.

Another influence on the location of the crossover switch is the principle of locality.

Locality of Movement Mobile systems are very likely to move from one access point to another in vicinity. Geographic and topologic vicinity are often correlated, so that the old and the new access point are likely to be fairly close to each other in most cases. Since the crossover switch is the anchor point, where the old and the new path diverge, it is therefore likely to also be located in the vicinity.

Locality of Communication Considering the case where a company owns a corporate fixed network and a mobile network on an airplane, both parts of the network connected to the provider's network via access points, a great number of connections will be between the mobile and the fixed corporate network. All these connections would have their endpoints within the provider's network in common, increasing the likelihood of a shared crossover switch.

The following topologies will be investigated and used as parameters with the topology specific stated assumptions for simplicity:

Tree specifying the degree, number of nodes, leafs or levels. All leafs are potential access points. All nodes are crossover switch enabled. VCs start and end at leafs only. Probability distributions for the location of the chosen crossover switch should be given for two random access points to roam and one access point where the VC ends. Assuming a tree with all leafs on the same level, the variables are degree, height, location of the old access point, location of the new access point and location of the far end party. The number of access points equals to $degree^{levels}$.

The principles of locality are also taken into account.

Three variations of crossover switch determination are considered:

A one node is declared as fixed shared crossover switch - the same crossover switch for all VCs or for a VP. This node could be the root of the tree. The hops from the root to the access points equals to the levels. For load balancing reasons, an alternative would be to assign a shared crossover switch to each mobile network.

B determination of the crossover switch as the ingress switch of the provider's network at setup time. This way, the crossover switch can either be the access point of the moving network or the access point of the far end party, depending on who is the calling party.

C dynamic crossover switch discovery using loose select targeted to the access point of the far end party.

Meshed giving the number of nodes. Every node will be considered able to act as an access point as well as crossover switch. As with the tree topology, three variations (A, B and C) of crossover switch determination are considered.

Random indicating the number of nodes and links. Edge nodes are considered access points, while the inner nodes are assumed to have crossover switch capabilities. As with the tree topology, three variations of crossover switch determination (A, B and C) are considered. The path to all three alternatives has to be computed. In this case, although the random network can be viewed as an n-ary partial tree, the number of hops to the root (as the chosen crossover switch in A) is not given by height as with the tree topology.

6.6 Simulation Process Model

Figure 6.17 consists of input parameters, boxes and output parameters. The input parameters mainly consist of the parameters listed in the previous section or thereof derived ones. The output parameters mostly conform to the performance criteria mentioned previously and give information for other quality criteria.

The box marked *perform HO* should be iterated using different connection handover protocols but with the same network to compare the output parameters.

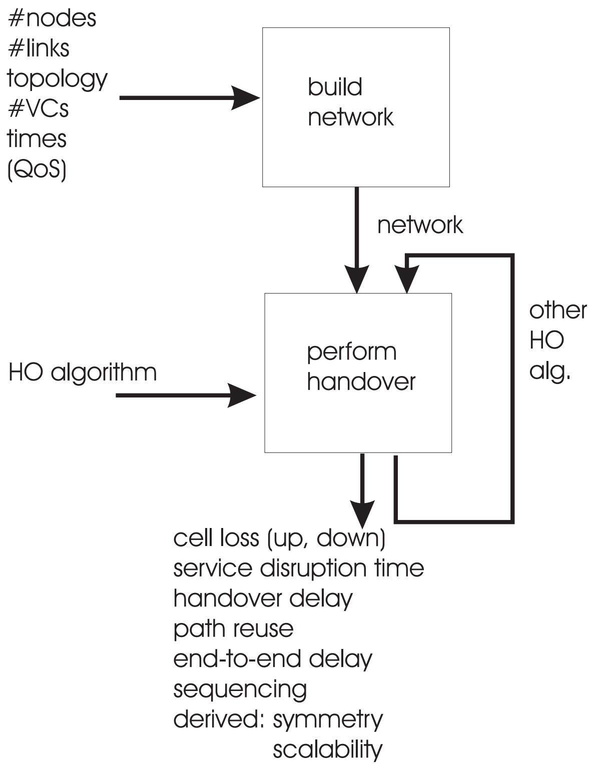


Figure 6.17: Simulation Process Model

6.7 Numerical Results

Figure 6.15 shows a simple model of an access point provider network with tree topology, which can be used as a visual reference for the following results.

The results have been obtained by building different networks with tree topology, varying with height and degree parameters. Given the access points of the mobile network before and after the handover (access point i and access point i+1 with (i+1) $\text{mod}_{\#accesspoints}$ equaling to zero for local movement) and the access point of the fixed network, the crossover switch has been found employing the dynamic crossover switch discovery algorithm used throughout this thesis. A code example in the appendix shows how this was implemented.

The reasons for choosing tree topologies for the the simulation, despite being a quite unrealistic topology for an access point provider network, are the following:

Simplicity The properties of a tree shaped graph are well agreed upon. A tree can be characterized by its degree and height, if it is full and balanced, which renders its presentation quite simple.

Generalizability I hoped to find formulas for full binary trees that could be generalized to n-ary trees and at the next level to partial n-ary trees, approximating random graphs without circles. Some of the obtained results have fulfilled this expectation and can be applied to various types of trees.

Comparability By choosing a random topology, one would have to stick with this specific topology in order to be able to compare the different approaches. Only punctual results for specific networks could be obtained. If a tree topology is chosen, the graphs are divided into classes of equivalence by specifying the degree and heigt, so that results can be obtained by using generic formulas, thereby avoiding to do the simulation all over again, if only one node or link in the topology changes.

Extension to Graph Theory There are various formulas and algorithms for searching, ordering and path computation in trees. As far as I scanned the literature, these take the root of the tree as a starting point. During the simuluation I found formulas (here applied to the dynamic discovery of the crossover switch, but other applications could be imagined) that reflect search algorithms that start from leafs rather than from the root and that are generic enough to scale to different types of trees. This would not have been possible with a random topology, because it would only have been valid for a very limited set of graphs.

The graphs in this section show some of the numerical results obtained based on the following time assumptions:

χ the time needed for the radio link switch = 5 ms.

ρ the propagation time over the wireless link = 100 ms.

τ the propagation time over a wired link for one hop = 1 ms.

α the processing time for a signal like Down_Ready or Conn_Active, consisting of both receiving and generating a signal = 10 ms.

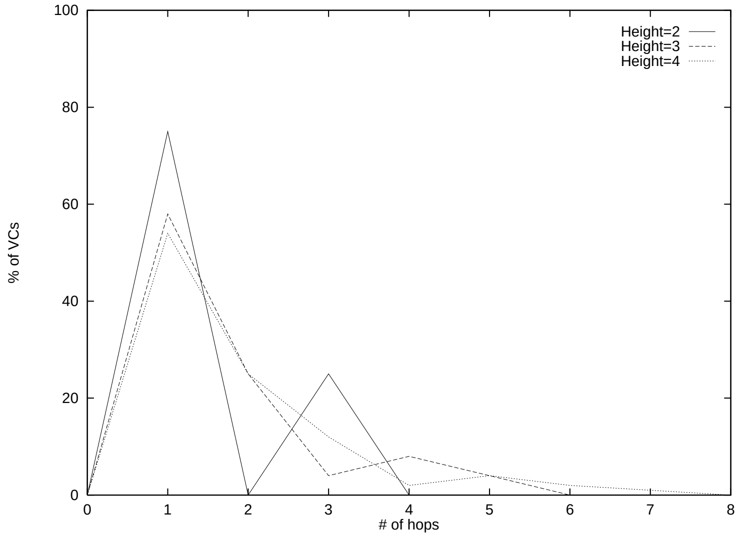


Figure 6.18: Distribution of Number of Hops between APs and COS (Local Movement)

α_{data} the time data spends in an ATM switch (transmission delay, switching delay, etc.) = 1 ms.

δ the time data travels from the new access point to the mobile via the crossover switch = g * #hops.

Inspection of the results of the simulations and analysis have led to the following result:

For local movement, the average number of hops from the access points to the crossover switch can be calculated by the following formula:

$$\overline{\#hopsToOld} = \overline{\#hopsToNew} = degree * \frac{\#COSs}{\#APs}$$

example:

binary tree (degree = 2), 16 access points (leaves), 15 crossover switches (inner nodes), 31 nodes, height 3.

$$\overline{\#hopsToOld} = \overline{\#hopsToNew} = 2 * \frac{15}{16} = 1.875$$

Figure 6.18 illustrates the distribution of the number of hops between the access points and the crossover switch for local movement.

The distribution changes, when random movement between all the access points is considered (figure 6.19).

Figure 6.20 shows the distribution of the height of the crossover switch in binary trees with

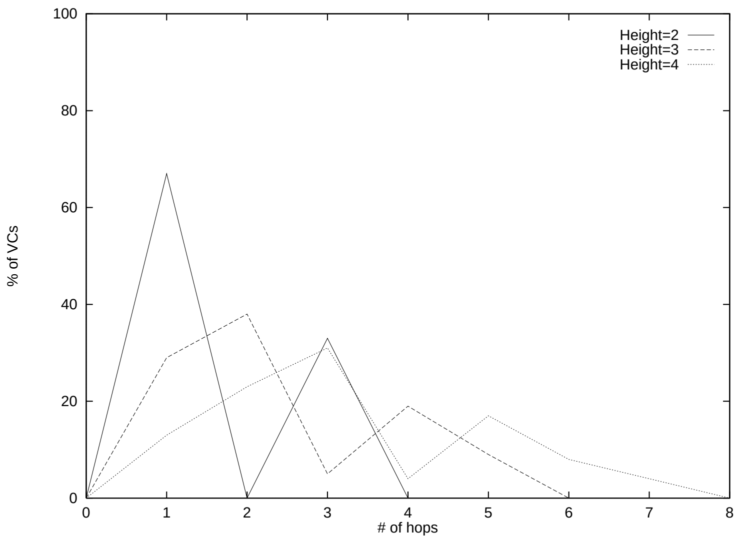


Figure 6.19: Distribution of Number of Hops between APs and COS (Random Movement)

local roaming between the access points as opposed to random movement in figure 6.21. The height of the crossover switch increases faster, when random movement is allowed.

In Figures 6.22 and 6.23, the average number of hops from the access points to the crossover switches is plotted in relation to the height of the network tree. Whereas with random movement the average number of hops increases steadily with height, it only varies slightly with local movement.

Applying the obtained numerical results on relative locations of the crossover switch and the access points as distributions for connection handover performance analysis, comparisons like in figure 6.24 can be made, where the uplink disruption period (potential cell loss period) for the VP, VC and a hybrid approach are shown.

If anchor rerouting using a shared crossover switch is employed, then the downlink disruption periods increase linearly with the number of VCs to be handed over. Since all connections have to be handed over by the same crossover switch, their signaling has to be done sequentially in the same switch. The topology, as can be seen from figures 6.25 and 6.7, has a minor influence. The main influence, the load, is determined by the number of VCs.

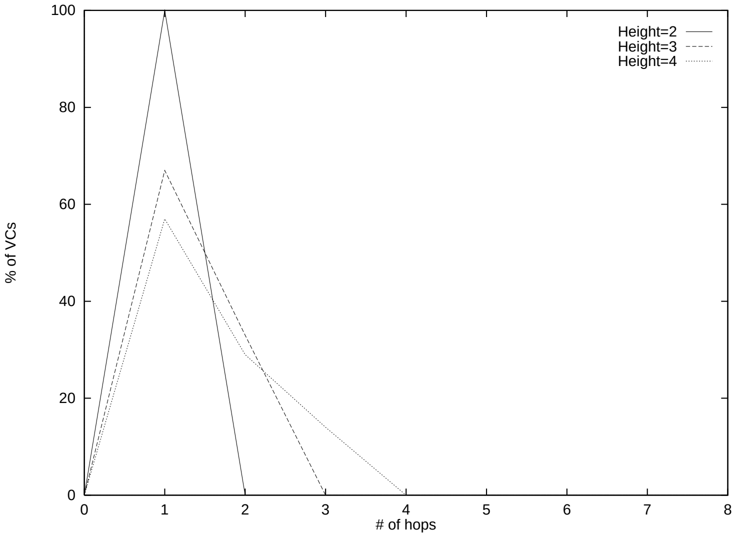


Figure 6.20: Distribution of COS Height (Local Movement)

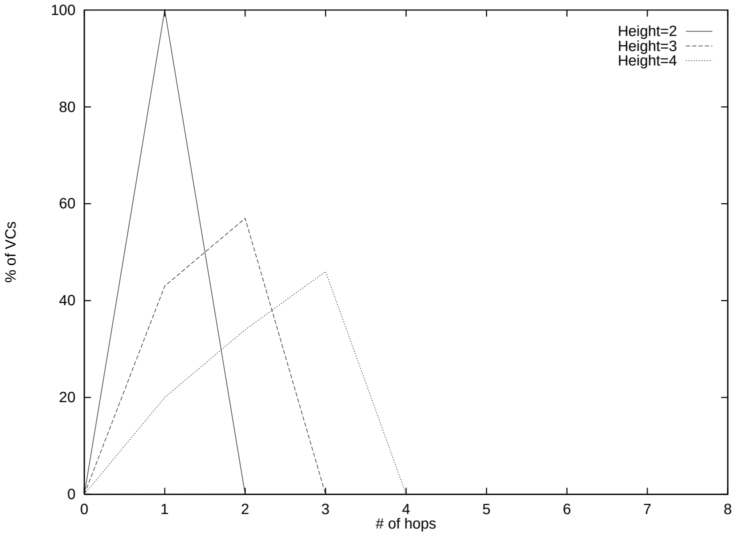


Figure 6.21: Distribution of COS Height (Random Movement)

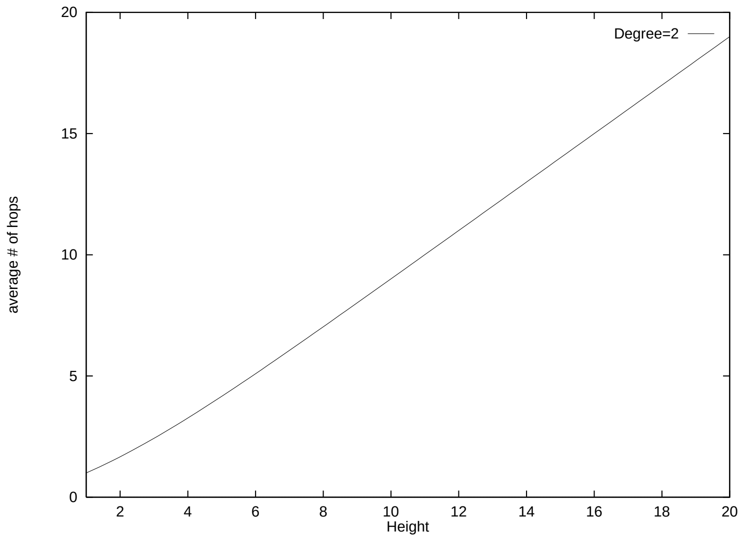


Figure 6.22: Average Number of Hops between APs and COS (Random Movement)

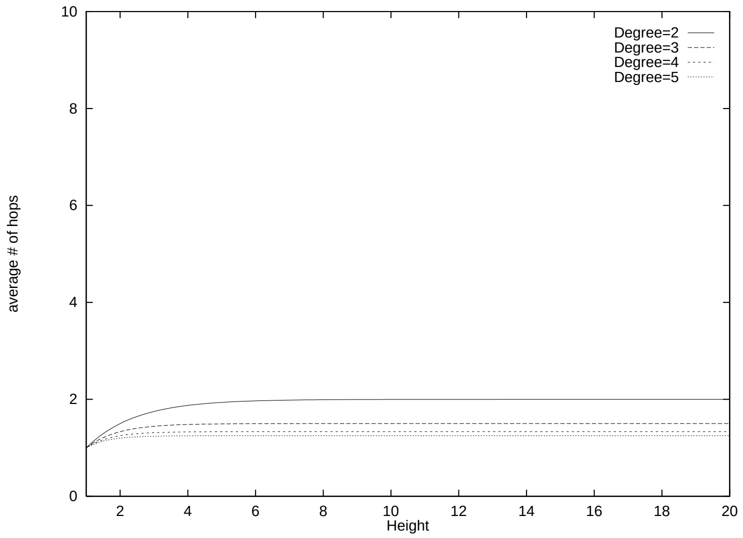


Figure 6.23: Average Number of Hops between APs and COS (Local Movement)

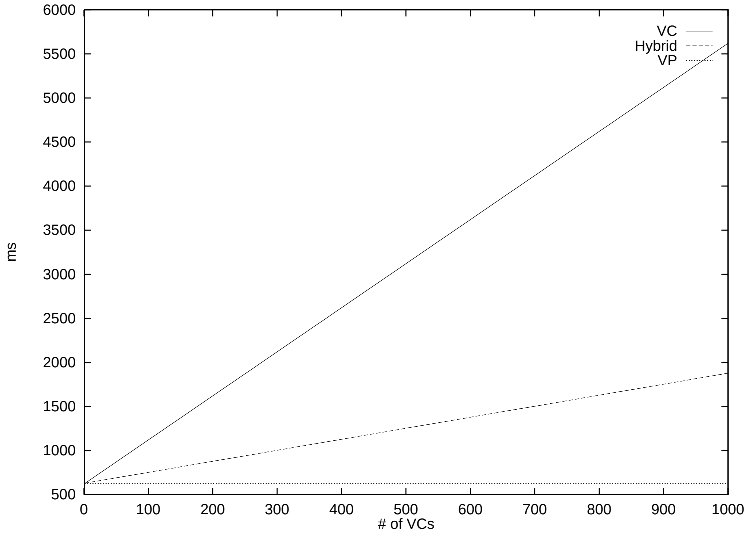


Figure 6.24: Uplink Disruption Period

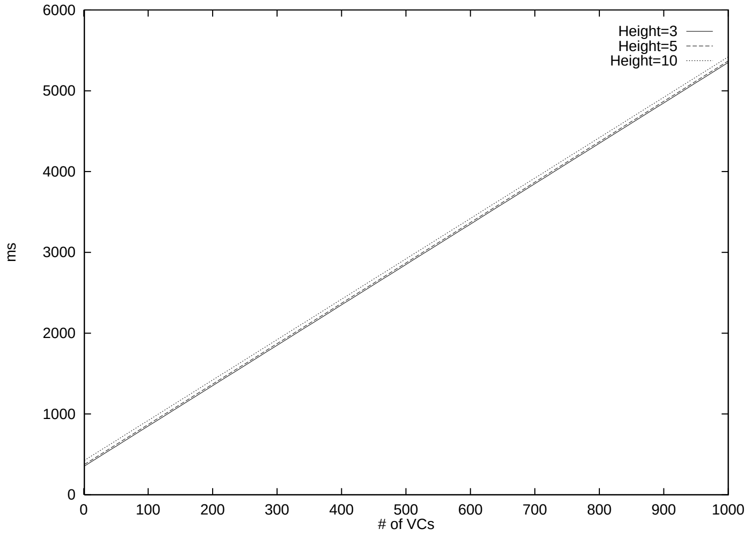


Figure 6.25: Downlink Disruption Period, Shared COS, Binary Tree

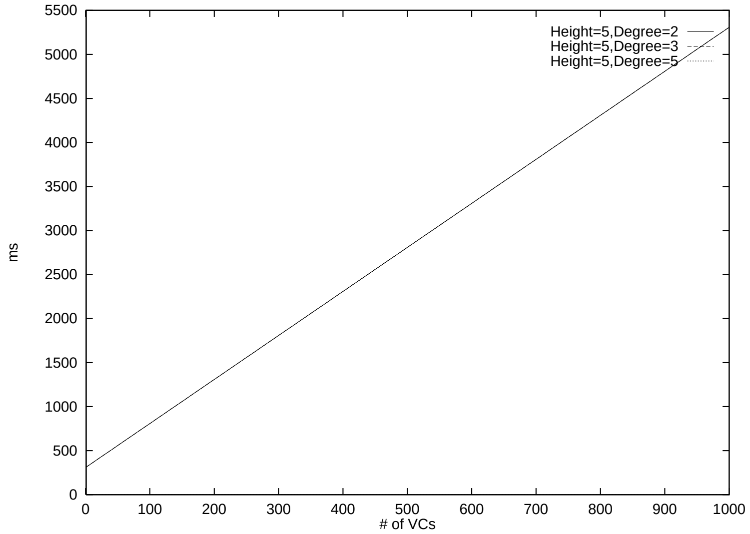


Figure 6.26: Downlink Disruption Period, Shared COS, N-ary Tree

7 Conclusions

Since some of the quality criteria that have been applied to measure the performance of different handover approaches are trade-offs to some others, unfortunately there is no such thing as the *Perfect Handover Algorithm* under all circumstances. Therefore, conclusions can only be drawn based on a specified set of circumstances. To decide, which handover approach is best suited for a specific set of circumstances remains up to the owner of the mobile network and depends on quality priorities. Different applications prefer different quality criteria. Data transfer will focus more on cell loss, while voice applications will prioritize end-to-end delay, to state a simple example. The results of this thesis can be used as a basis for deciding which approach to use and is therefore highlighting the behavior patterns for different handover approaches with respect to different quality criteria.

But now to some detailed conclusions: With hard, backward connection handover, cell loss on the uplink is connection handover algorithm invariant. The only influence on uplink cell loss, at a given total uplink data rate at handover time, is the number of VCs active at handover time. With this respect, a VP can be treated as one VC.

For the cell loss on the downlink with hard, backward connection handover, the following can be concluded:

A major component of the downlink disruption time of a VC is the difference in time between the arrival of the Down_Ready message at the mobile and the arrival of data at the new access point, sent on the new part of the path. This time difference depends on the location of the crossover switch, which in turn is determined by the handover algorithm and the topology of the network. Heuristics to determine the location of the crossover switch based on topology information have been given in the previous section on numerical results.

Since the dynamic crossover switch discovery approach strongly depends on a distribution of the number of hops between the crossover switch and the access points, the probabilities have to be determined or assumed (again, see the previous section on numerical results for heuristics), before a comparison with the other approaches really makes sense. So far, especially regarding cell loss, the VP based anchor rerouting has been proven to be preferred to the VC based anchor rerouting and the dynamic crossover switch discovery (only uplink) as well.

The dynamic crossover switch discovery scheme as it is presented here, performs best for worst case VCs, i.e. the VCs whose crossover switch is most distant to the access points. It could be worth thinking of an alteration to the algorithms as to favor the average or even the VCs having the least distant crossover switch. The overall cell loss behaviour of the dynamic crossover switch discovery scheme is minimized, when the individual VCs deviation from each other in terms of number of hops to the crossover switch is minimal.

In terms of end-to-end delay, inherently the dynamic crossover switch discovery algorithm performs best, path extension worst.

Soft handover always outperforms its hard counterpart, since the handover can be prepared by simultaniously using both links to the old and to the new access point, respectively. Therefore, the disruption periods are minimized, as there is no need for synchronization points (e.g. waiting for other connections to be prepared for handover in order to switch the radio link) during the handover procedure.

A Glossary

A

AAL ATM Adaptation Layer

ABR Available Bit Rate

AMPS Advanced Mobile Phone Service.

AP Access Point

APCP Access Point Control Protocol]

ATM Asynchronous Transfer Mode

B

B-ISDN Broadband-Integrated Services Digital Network User Part

BRAN Broadband Radio Access Network

BS Base Station

C

CAC Call Admission Control

CBR Constant Bit Rate

CDMA Code Division Multiple Access: a mechanism to share a communications link among several users.

COS Crossover Switch, the switch where the old path (before connection handover) and the new path diverge.

CS Cellular Systems. A form of two way radio in which signals are broadcast into a network of interlocking cells. Digital transmission using different frequency spectrum is beginning to replace the older analog technology on cellular networks and communications satellites.

D

DECT Digital Enhanced Cordless Telecommunications stations, usually via FFS satellites, to home antennas.

DLC Data Link Control

DTL Designated Transit List

E

ESI End System Identifier

ETSI European Telecommunication Standards Institute

F

FB Frequency Bandwidth. The range of frequencies required for specific uses (analog or digital). Television, for example, involves transmitting a far greater volume of information (pictures elements) than voice signals (6 MHz vs. 4 kHz bandwidth). Voice, in turn, requires more bandwidth than low speed data. In some cases, high speed data requires more bandwidth (¿64 kbps) than voice.

FDMA Frequency Division Multiple Access system for cellular radio.

FEC Forward Error Correction

FEP Far-End Party

FN Foreign Node

FPLMTS Future Public Land Mobile Telecommunications System

G

GEOS Geostationary Earth Orbiting Satellites are satellites that remain at a distance of 36, 000 kilometers from earth and can always be directed toward the same geographic area. Also called 'geosynchronous satellites', their footprint - the area that its signals cover - may be as large as an entire continent.

GSM Groupe Speciale Mobile or sometimes referred to as Global System for Mobile communication (or derivatives), are used all over the world and in the US where it is referred to as TDMA.

H

HA Home Address

HN Home Node

HO Hand Off/HandOver: a process of communications when a user moves from one base station to another base station.

L

LEO Low Earth Orbiting Satellites are small satellites orbiting 780 kilometers from the earth. Because they take less power to receive their signals or send signals to they are well suited to hand-held mobile uses. Their foot print is small and they provide global coverage by linking with one another in constellations.

LLC Logical Link Control

LM Location Management.

LMCS Local Multipoint Communications Systems. Recently licensed and yet-to-be developed, these are broadband wireless systems organized on the same principle as cellular and PCS but capable of transmitting multi-media and broadcasting services.

LR Location Register

M

MAC Medium Access Control

ME Mobility Enhanced.

MEO Medium Earth Orbiting Satellites. They occupy an orbit in between LEOs and GEOs.

MSS Mobile Satellite Services use a new generation of satellites to direct signals to handheld terminals or small antennas mounted on moving ships or vehicles.

MT Mobile Terminal

N

NCCI Network Call Correlation Identifier

NNI Network-Network Interface

nrt-VBR Non-real time Variable Bit Rate

O

OAM Operation And Management

OPFDMA Orthogonal Public Frequency Division Multiple Access for Future Public Land Mobile Telephone Service (FPLMTS) is a third generation wireless standard which was developed by ITU-T.

P

PCS Personal Communications Services is a new form of digital radio communications that encompass mobile and ancillary fixed communications that provide services to individuals and businesses and can be integrated with a variety of competing networks. It includes second generation high-tier PCS (digital cellular systems) operating in the 800 MHz, 900 MHz, 1800 MHz, and 1900 MHz frequency bands, low-tier PCS operating in the 1900 MHz frequency band, and third generation PCS that are evolutions of second generation PCS.

PDU Protocol Data Unit

PHY Physical layer

PNNI Private Network-Network Interface

PNNI+M Mobility enabled Network-Network Interface

Q

QoS Quality of Service

R

RP Radio Port

RRM Radio Resource Management.

rt-VBR Real time Variable Bit Rate

S

SAAL Signaling ATM Adaptation Layer

Spectrum A range of frequencies; the radio frequency spectrum is considered as a reusable natural resource which is managed by national governments under international agreements. The spectrum is segmented into bands reserved for different types of uses including ham radio, paging, cellular, satellite, etc. Individual carriers are assigned specific frequencies within these bands.

T

TDD Time Domain Duplex

TDMA Time Division Multiple Access is a technique to multiplex channels using permanent time slots in the transmission. TDMA uses digital technology to create two or more separate channels on one previously analog channel.

U

UBR Unspecified Bit Rate

UNI User-Network Interface

UNI+M Mobility enabled User Network Interface

V

VC Virtual Channel - throughout the thesis also used to refer to virtual channel connections -

VP Virtual Path - throughout the thesis also used to refer to virtual path connections

VCI Virtual Channel Identifier

VPI Virtual Path Identifier

Bibliography

[ACa] Bora A. Akyol and Donald C. Cox. Handling mobility in a wireless ATM network. Stanford University.

[ACb] Bora A. Akyol and Donald C. Cox. Rerouting for handoff in a wireless ATM network. Stanford University.

[AHK+96] Prathima Agrawal, Eoin Hyden, Paul Krzyzanowski, Partho Mishra, Mani B. Srivastava, and John A. Trotter. SWAN: A mobile multimedia wireless network. *IEEE Personal Communications*, pages 18 – 33, April 1996.

[AK96] Geert A. Awater and Jan Kruys. Wireless ATM - an overview. *Mobile Networks and Applications*, 1(3):235 – 244, 1996. J.C. Baltzer AG, Science Publishers.

[ALBR97] Arup Acharya, Jun Li, Ajay Bakre, and Dipankar Raychaudhuri. Design and prototyping of location management and handoff protocols for wireless ATM networks. In *Proceedings of ICUPC '97*, 1997.

[ALR98] Arup Acharya, Jun Li, and Dipankar Raychaudhuri. Mobile ATM: Network architeture and protocols for location management and handoff. In *Proceedings of INFOCOM '98*, 1998.

[ALRR97] Arup Acharya, Jun Li, Bala Rajagopalan, and Dipankar Raychaudhuri. Mobility management in wireless ATM networks. *IEEE Communications Magazine*, pages 100 – 109, November 1997.

[BG98] Ravi Raj Bhat and Rajeev Gupta. Framework for dynamic COS discovery in wireless ATM. Contribution to the ATM Forum WATM 98-0005, 1998.

[Bha98a] Ravi Raj Bhat. Architectural details for wireless access in wireless ATM. Contribution to the ATM Forum WATM 98-0196, 1998.

[Bha98b] Ravi Raj Bhat. Extension to backward COS discovery (BCD) approach. Contribution to the ATM Forum WATM 98-0231, 1998.

[Bha98c] Ravi Raj Bhat. Handover signaling scenarios in WATM. Contribution to the ATM Forum WATM 98-0195, 1998.

[Bis97] Subir Kumar Biswas. An agent-based signaling architecture for supporting mobility in radio ATM networks. *International Journal of Communication Systems*, 10:87 – 101, 1997. John Wiley & Sons, Ltd.

[CRC+97] Matthew Cheng, Subhashini Rajagopalan, Li Fung Chang, Gregory P. Pollini, and Melbourne Barton. PCS mobility support over fixed ATM networks. *IEEE Communications Magazine*, pages 82 – 92, November 1997.

[Dea98] John Deane. A case for wireless soft QoS. Contribution to the ATM Forum WATM 98-330, 1998.

[DL95a] Harry J. R. Dutton and Peter Lenhard. *Asynchronous Transfer Mode ATM - Technical Overview.* Prentice Hall PTR, 2 edition, 1995. Communications IBM.

[DL95b] Harry J. R. Dutton and Peter Lenhard. *High-Speed Networking Technology: An Introductory Survey.* Prentice Hall PTR, 3 edition, 1995. Communications IBM.

[DPD98] I. Dravopoulos, N. Pronios, and T. Dagiuklas. A reason for using HO_Force signal. Contribution to the ATM Forum WATM 98-0289, 1998.

[DVS97] Gopal Dommety, Malathi Veeraraghavan, and Mukesh Singhal. Route optimization in mobile ATM networks. In *Proceedings of MOBICOM 97*, pages 43 – 54, 1997.

[Dyk97] Doug Dykeman. Personal communication. IBM Zurich Research Laboratory, 1997.

[FDIS97] Laurent Freléchoux, Doug Dykeman, Ilias Iliadis, and Paolo Scotton. PNNI routing support for mobile networks, 1997.

[FDIS98] Laurent Freléchoux, Doug Dykeman, Ilias Iliadis, and Paolo Scotton. Resource location in mobile ATM networks. In *Proceedings of ICATM '98*, 1998.

[Fre98] Laurent Freléchoux. Personal communication. IBM Zurich Research Laboratory, 1998.

[Gro98] WATM Working Group. WATM baseline text. ATM Forum btd-watm-01.07.ps, 1998.

[Han96] Harri Hansén. Connection management functions of a private wireless ATM network. Master's thesis, Helsinki University of Technology, 1996.

[HIK98] Harri Hansén, Jukka Immonen, and A. Kaloxylos. Description of the handover algortithm for the wireless ATM network demonstrator (WAND). VTT Information Technology, 1998.

[KA98] Bhumip Khasnabish and Majid Ahmadi. Integrated mobility and QoS control in cellular wireless ATM networks. *Journal of Network and Systems Management*, 6(1):71 – 89, 1998. Plenum Publishing Corporation.

[Kee93] Kimberly Keeton. Providing connection oriented network services to mobile hosts. In *Proceedings of USENIX Symposium on Mobile & Location Independent Computing*, pages 83 – 103, 1993.

[MCCM97] Marco Ajmone Marsan, Carla-Fabiana Chiasserini, Renato Lo Cigno, and Maurizio Munafò. Local and global handovers for mobility management in wireless ATM networks. *IEEE Personal Communications*, pages 16 – 24, October 1997.

[MHIV96] Håkan Mitts, Harri Hansén, Jukka Immonen, and Simo Veikkolainen. Lossless handover for wireless ATM. *Mobile Networks and Applications*, 1(3):299 – 312, 1996. J.C. Baltzer AG, Science Publishers.

[MRRB97] Håkan Mitts, Kristian Rauhala, Bala Rjagopalan, and Gregor Bautz. Proposed handover signaling architecture for release 1.0 WATM baseline. Contribution to the ATM Forum WATM 97-0845, 1997.

[Pap97] Laszlo Pap. Mobile communication systems. Lecture at University of Klagenfurt, 1997.

[Per97] Charles E. Perkins. Mobile IP. *IEEE Communications Magazine*, pages 84 – 99, May 1997.

[PNN96] PNNI SWG. Private network-network interface specification version 1.0 (PNNI 1.0), 1996.

[PPVM97] Nikos Passas, Sarantis Paskalis, Ditra Vali, and Lazaros Merakos. Quality-of-Service-oriented medium acces control for wireless ATM networks. *IEEE Communications Magazine*, pages 42 – 50, November 1997.

[PR97] Richard H. Paine and Sankar Ray. Boeing service scenario. Contribution to the ATM Forum WATM 97-0902, 1997.

[PZK97] Kaveh Pahlavan, Ali Zahedi, and Prashant Krisnamurthy. Wideband local access: Wireless LAN and wireless ATM. *IEEE Communications Magazine*, pages 34 – 40, November 1997.

[Raj96] Bala Rajagopalan. Mobility management in integrated wireless ATM networks. *Mobile Networks and Applications*, 1(3):273 – 285, 1996. J.C. Baltzer AG, Science Publishers.

[Raj98] Bala Rajagopalan. Rationale for ATM layer control signaling for handover at the UNI. Contribution to the ATM Forum WATM 98-0301, 1998.

[Ray96a] Sankar Ray. WATM requirements for airborne platforms communicating via satellite links. Contribution to the ATM Forum WATM 96-0062, 1996.

[Ray96b] D. Raychaudhuri. Wireless ATM: An enabling technology for multimedia personal communication. *Wireless Networks*, 2:163 – 170, 1996. J.C. Baltzer AG, Science Publishers.

[RL98] Bala Rajagopalan and Jun Li. Incorporating failure recovery and support for multiple vcs in the baseline handover procedure. Contribution to the ATM Forum WATM 98-0255, 1998.

[RLA97a] Bala Rajagopalan, Jun Li, and A. Acharya. Signaling for connection rerouting for handoff control in wireless ATM. Contribution to the ATM Forum WATM 97-0338, 1997.

[RLA97b] Bala Rajagopalan, Jun Li, and A. Acharya. Signaling requirements for handoff control. Contribution to the ATM Forum WATM 97-0638, 1997.

[Tan96] Andrew S. Tanenbaum. *Computer Networks.* Prentice Hall International Inc., 3 edition, 1996. International Edition.

[Toh96a] Chai-Keong Toh. A handover paradigm for wireless ATM LANs. *ACM Symposium on Applied Computing (SAC'96)*, 1996.

[Toh96b] Chai-Keong Toh. A hybrid handover protocol for local area wireless ATM networks. *Mobile Networks and Applications*, 1(3):313 – 334, 1996. J.C. Baltzer AG, Science Publishers.

[Toh96c] Chai-Keong Toh. Performance evaluation of crossover switch discovery algorithms for wireless ATM LANs. In *Proceedings of INFOCOM 1996, Volume 3*, pages 1380 – 1387, 1996.

[Toh97] Chai-Keong Toh. *Wireless ATM and Ad-hoc Networks.* Kluwer Academic Publishers, 1997.

[UYA97] Hüseyin Uzunalioğlu, Wei Yen, and Ian F. Akyildiz. A connection handover protocol for LEO satellite ATM networks. In *Proceedings of MOBICOM 97*, pages 204 – 214, 1997.

[VKE96] M. Veeraraghavan, M. Karol, and K. Y. Eng. A combined handoff scheme for mobile ATM networks. Contribution to the ATM Forum WATM 96-1700, 1996.

[WCK98] Indra Widjaja, Anna Cui, and Jeong Geun Kim. Buffering locations for backward handoff. Contribution to the ATM Forum WATM 98-0278, 1998.

[WK98] Indra Widjaja and Jeong Geun Kim. Performance issues in backward handoff. Contribution to the ATM Forum WATM 98-0091, 1998.

[WL97] Daniel Wong and Teng Joon Lim. Soft handoffs in CDMA mobile systems. *IEEE Personal Communications*, pages 6 – 17, December 1997.

[YBF$^+$96] R. Yuan, S. K. Biswas, L. J. French, J. Li, and D. Raychaudhuri. A signaling and control architecture for mobility support in wireless ATM networks. *Mobile Networks and Applications*, 1(3):287 – 298, 1996. J.C. Baltzer AG, Science Publishers.

Index