PERFORMANCE EVALUATION OF END-TO-END
MOBILITY MANAGEMENT SOLUTIONS FOR TCP

A Thesis Presented to

Mohammad Ali Jinnah University

In Partial Fulfillment

of the requirement for the Degree

Master of Science
(Computer Science)

by

Peer Azmat Shah
Reg. # MS071017

August, 2009
THIS THESIS ENTITLED

“PERFORMANCE EVALUATION OF END-TO-END MOBILITY MANAGEMENT SOLUTIONS FOR TCP”

by Peer Azmat Shah

Reg # MS071017

has been approved

for Mohammad Ali Jinnah University

Supervisor: ______________________________________
Prof. Dr. Amir Qayyum
Mohammad Ali Jinnah University, Islamabad

Examiner 1: ______________________________________
Prof. Dr. Muddassar Farooq
FAST (NU), Islamabad

Examiner 2: ______________________________________
Prof. Dr. Sohail Asghar
Mohammad Ali Jinnah University, Islamabad

________________________________________________
Prof. Dr. Muhammad Abdul Qadir
Dean (Faculty of Engineering and Applied Sciences)
Abstract

Mobile wireless access to Internet, having rapidly growing demand, is currently being offered through many diverse networks. Most of the data centric applications use end-to-end reliable transport protocols such as TCP. Generally, the TCP connection is active until the user is connected to a single network. However, mobile user may not restrict himself to a single network and TCP has not been designed to cater for user mobility, specifically across heterogeneous networks. Hence the connection is broken if user can no longer be served by the current network unless some measures are taken to switch to an alternate network. Therefore, significant emphasis is being placed on the development of mobility management solutions that enable to handle the user mobility and portability across available heterogeneous networks for TCP based applications. Mobility management across heterogeneous networks involves the exchange of control signaling, resulting into handover delay, overhead and a possible degradation in throughput. This delay, overhead and throughput degradation time needs to be assessed to determine its impact on the performance of the TCP applications. In this work we focus on end-to-end mobility management solutions for TCP applications and provide a comparative analysis using mathematical modeling techniques of performance evaluation, already available in literature. We evaluate the performance of end-to-end mobility management solutions on the basis of handover delay, throughput degradation time and protocol overhead. The results gathered in this work manifest that the protocols that incorporate mechanism for simultaneous communication through multiple interfaces like pTCP, SLM, Tsukamoto’s proposal and LCT-based protocol, permit seamless vertical
handover, consequently providing throughput stability. Similarly, protocols that use some mechanism to avoid throughput degradation time after vertical handover, like VA-TCP, performs better in overlapping as well as non-overlapping regions.
Acknowledgment

All Praises be to “ALLAH” Almighty who enabled us to complete this task successfully and our utmost respect to His last Prophet (P.B.U.H.).

I would like to express my cordial gratitude to my reverend supervisor Prof. Dr. Amir Qayyum whose persistent guidance, encouragement and support has always been a source of motivation. I am extremely grateful to his scholastic and sympathetic attitude.

I am also thankful to the members of Center of Research in Networks & Telecom (CoReNeT), especially Mr. Muhammad Yousaf, my co-supervisor in this research activity, for his generous assistance, constructive criticism, timely advice and enlightened supervision in the accomplishment of this manuscript.

I will not be out of place to express my profound admiration for my parents, sisters and brothers for their continued prayers and support.

Peer Azmat Shah
Table of Contents

ABSTRACT ............................................................................................................................... III
ACKNOWLEDGMENT ................................................................................................................ V
LIST OF FIGURES .................................................................................................................... VII
LIST OF TABLES ..................................................................................................................... VIII
LIST OF ACRONYMS ............................................................................................................... IX
CHAPTER 1 .............................................................................................................................. 1
INTRODUCTION .................................................................................................................... 1
  1.1 MOBILITY MANAGEMENT ............................................................................................... 1
     1.1.1 Functional Requirements of Mobility Management .................................................... 1
     1.1.2 Mobility Management at Different Layers of Protocol Stack ....................................... 3
     1.1.3 Desirable Characteristics of an End-to-End Mobility Management Solution ............... 4
  1.2 PROBLEM STATEMENT .................................................................................................. 5
  1.3 THESIS GOALS ............................................................................................................... 5
  1.4 THESIS ORGANIZATION ............................................................................................... 5
CHAPTER 2 ............................................................................................................................. 6
SURVEY OF END-TO-END MOBILITY MANAGEMENT SOLUTIONS FOR TCP .............. 6
  2.1 CHARACTERISTICS OF A MOBILITY MANAGEMENT SOLUTION ............................... 6
     2.1.1 Performance Characteristics ....................................................................................... 6
     2.1.2 Implementation and Deployment Related Characteristics ........................................... 7
     2.1.3 End User Related Characteristics .............................................................................. 8
  2.2 EXISTING END-TO-END MOBILITY MANAGEMENT SOLUTIONS .............................. 9
     2.2.1 Layer 3.5 Solutions .................................................................................................... 10
     2.2.2 Transport Layer Solutions ......................................................................................... 11
     2.2.3 Layer 4.5 and Upper Layer Solutions ....................................................................... 15
CHAPTER 3 .............................................................................................................................. 20
HANOVER DELAY ANALYSIS ............................................................................................... 20
  3.1 VERTICAL HANDOVER SCENARIOS ............................................................................ 20
  3.2 MODELING END-TO-END HANDOVER DELAY ............................................................. 25
  3.3 HANDOVER DELAY ANALYSIS .................................................................................... 33
CHAPTER 4 .............................................................................................................................. 40
ANALYSIS OF THROUGHPUT DEGRADATION TIME DURING VERTICAL HANDOVER .... 40
  4.1 MODELING THROUGHPUT DEGRADATION TIME FOR VERTICAL HANDOVER ........ 40
  4.2 THROUGHPUT DEGRADATION TIME ANALYSIS ......................................................... 44
CHAPTER 5 .............................................................................................................................. 48
PROTOCOL OVERHEAD ANALYSIS .................................................................................... 48
  5.1 PROTOCOL OVERHEAD ............................................................................................... 48
  5.2 MODELING PROTOCOL OVERHEAD ........................................................................... 48
     5.2.1 Protocol Overhead of Different End-to-End Mobility Management Protocols .......... 49
  5.2 RESULTS & DISCUSSIONS REGARDING PROTOCOL OVERHEAD ............................ 51
CHAPTER 6 .............................................................................................................................. 55
CONCLUSION & FUTURE WORK ......................................................................................... 55
REFERENCES .......................................................................................................................... 57
List of Figures

FIGURE 3.1 VERTICAL HANDOVER SCENARIOS ................................................................. 20
FIGURE 3.2 TIMING DIAGRAM NON OVERLAPPING REGIONS, WHEN MN IS RECEIVING DATA .............................................................................................................. 22
FIGURE 3.3 TIMING DIAGRAM NON OVERLAPPING REGIONS, WHEN MN IS SENDING DATA .............................................................................................................. 23
FIGURE 3.4 TIMING DIAGRAM SCENARIO 2B, WHERE MN HAS ALREADY PERFORMED L2 CONNECTIVITY .................................................................................. 24
FIGURE 3.5 TIMING DIAGRAM FOR SCENARIO 2B, WHERE MN HAS ALREADY ACQUIRED IP ADDRESS ............................................................................................. 24
FIGURE 3.6 HAN DOVER DELAY IN NON OVERLAP Ping REGIONS ........................................ 27
FIGURE 3.7 HAN DOVER DELAY, WHEN MN MOV ES INTO OVERLAPPING REGION AND PROTOCOL DOESN'T ALLOW SIMULTANEOUS COMMUNICATION .................. 28
FIGURE 3.8 HANDOVER DELAY AS FUNCTION OF PACKET LOSS PROBABILITY IN SCENARIO 1, NON OVERLAPPING REGION .......................................................... 33
FIGURE 3.9 HANDOVER DELAY AS FUNCTION OF PACKET LOSS PROBABILITY IN SCENARIO 2A, WHERE MN ENTERS INTO OVERLAPPING REGION .................................. 34
FIGURE 3.10 HANDOVER DELAY AS FUNCTION OF PACKET LOSS PROBABILITY IN SCENARIO 2B, WHERE MN LEAVES THE OVERLAPPING REGION .................................. 35
FIGURE 3.11 HANDOVER DELAY AS FUNCTION OF BANDWIDTH IN SCENARIO 1, NON OVERLAPPING REGION ............................................................................ 36
FIGURE 3.12 HANDOVER DELAY AS FUNCTION OF BANDWIDTH IN SCENARIO 2A, WHERE MN ENTERS INTO OVERLAPPING REGION .................................................. 37
FIGURE 3.13 HANDOVER DELAY AS FUNCTION OF BANDWIDTH IN SCENARIO 2B, WHERE MN LEAVES THE OVERLAPPING REGION .................................................. 38
FIGURE 4.1 THROUGHPUT DEGRADATION TIME ACROSS NON OVERLAPPING REGIONS .................................................................................................................. 45
FIGURE 4.2 THROUGHPUT DEGRADATION TIME, WHEN MN MOVES INTO OVERLAPPING REGION ........................................................................................................ 46
FIGURE 4.3 THROUGHPUT DEGRADATION TIME, WHEN MN MOVES OUT OF OVERLAPPING REGION .............................................................................................. 46
FIGURE 5.1 PROTOCOL OVERHEAD OF DIFFERENT MOBILITY MANAGEMENT PROTOCOLS .............................................................................................................. 52
FIGURE 5.2 PROTOCOL OVERHEAD OF MOBILITY SOLUTIONS THAT SEND ADDITIONAL INFORMATION WITH EACH DATA PACKET ...................................................... 53
FIGURE 5.3 MN'S MOVEMENT IN DIFFERENT ACCESS NETWORKS ..................................... 53
FIGURE 5.4 PROTOCOL OVERHEAD AS A FUNCTION OF NUMBER OF HANDOVERS ........ 54
List of Tables

TABLE 2.1 ANALYSIS MATRIX OF EXISTING END-TO-END MOBILITY MANAGEMENT PROTOCOLS ........................................................................................................................................... 18

TABLE 3.1 HANDOVER DELAY COMPARISON OF DIFFERENT MOBILITY MANAGEMENT PROTOCOLS ........................................................................................................................................ 39

TABLE 4.1 THROUGHPUT DEGRADATION TIME COMPARISON OF DIFFERENT MOBILITY MANAGEMENT PROTOCOLS DURING VERTICAL HANDOVER .............................................. 47

TABLE 5.1 PROTOCOL OVERHEAD COMPARISON OF DIFFERENT END-TO-END MOBILITY MANAGEMENT PROTOCOLS ........................................................................................................ 50
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>Mobile Node</td>
</tr>
<tr>
<td>RN</td>
<td>Remote Node</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>HIP</td>
<td>Host Identity Protocol</td>
</tr>
<tr>
<td>TCP-R</td>
<td>TCP-Redirection</td>
</tr>
<tr>
<td>SL-TCP</td>
<td>Socketless-TCP</td>
</tr>
<tr>
<td>LCT</td>
<td>Local Connection Translation</td>
</tr>
<tr>
<td>VA-TCP</td>
<td>Vertical handoff Aware-TCP</td>
</tr>
<tr>
<td>SLM</td>
<td>Session Layer mobility Management</td>
</tr>
<tr>
<td>RVS</td>
<td>Rendezvous Server</td>
</tr>
<tr>
<td>mSCTP</td>
<td>mobile Stream Control Transmission Protocol</td>
</tr>
<tr>
<td>SIP</td>
<td>Session Initiation Protocol</td>
</tr>
<tr>
<td>NAT</td>
<td>Network Address Translation</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

Mobile Internet applications have become popular and are being widely used. With the evolution of 4G wireless heterogeneous networks, all time access and seamless mobility across different networks, like WLAN, WiMAX, UMTS and WWAN etc., is desirable. On the other hand, Internet users want to use the best access network according to the network characteristics or their own preferences. These heterogeneous wireless networks may use a common IP core network. More than 90% of Internet traffic is of Transmission Control Protocol (TCP) and TCP connection's association is 4-tuple, Source IP: Destination IP: Source Port: Destination Port. Whenever a Mobile Node (MN) moves from one access network to another, this 4-tuple is disturbed and the already established communication with Remote Node (RN) remains no longer active.

1.1 Mobility Management

In wireless networks, mobility can be defined as changing the point of attachment without losing ability to communicate with other nodes in the network. Whenever a MN changes its point of attachment, its Internet Protocol (IP) address changes. As TCP connection is associated with IP address and port number, thus whenever one of the communicating node changes its IP address, all on-going TCP connections are no more active. Hence data cannot be sent/received over these TCP connections. In this situation, handover from one access network to another is required for the
continuation of communication. Handover is categorized into two types: horizontal and vertical.

**Horizontal handover:**

When a MN moves between different base stations / access points within the same link layer technology then horizontal handover is made. It is usually due to the geographical movement of a MN in a particular access network.

**Vertical handover:**

A vertical handover means that MN moves from one access network to another access network of a different type/technology. It is usually due to the topological movement of MN in different access networks.

### 1.1.1 Functional Requirements of Mobility Management

A common goal (Eddy, 2004) among all the approaches for mobility management is that mobility should not break the existing connections between two communicating nodes and MN should not become unreachable for future connections. To achieve this goal, two basic functionalities are performed by the mobility management protocol.

**Handover Management:**

Keeping ongoing communication between two communicating nodes alive, while MN moves and changes its point of attachment to the Internet, is called handover management. As MN can't send/receive data after changing the access network, thus in order to continue the communication, handover management is required.

**Location Management:**

Another important function needed to support mobility is the reliable and timely notification of MN's current location to a location server so that other nodes may
communicate with it for future connections. Location management involves identifying the current location of MN and also keeping track of location changes as it moves on. As Domain Name System (DNS) is used for resolving hostname to IP address in the Internet, therefore DNS dynamic updates are usually used by MN for location updation.

1.1.2 Mobility Management at Different Layers of Protocol Stack

Mobility can be handled at different layers of the traditional TCP/IP protocol stack (Eddy, 2004). It can be handled at Link layer, Network layer e.g. Mobile IP (Perkins, 1998) etc., layer 3.5 e.g. Host Identity Protocol (HIP) (Moskowitz & Nikander, 2006) (So & Wang, 2006), Transport layer e.g. SCTP (Koh & Kim, 2005), variants of TCP etc. and at Session layer e.g. SIP based mobility management (Banerjee, Acharya & Das, 2006), SLM (Landfeldt et al., 1999) etc. At each of these layers, mobility management schemes have corresponding strengths and weaknesses.

Link layer mobility management is used for horizontal handover within the same link layer technology, causing horizontal handover. As we are focusing on mobility of MN in heterogeneous environments, thus we are not considering it. At Network layer, mobility management has several shortcomings (Zeadally & Siddiqui, 2007), for example intermediate entities (Home Agent, Foreign Agent, Proxy etc.) are introduced between MN and RN. These approaches (Mobile IPv4) do not require any changes to be made on the fixed side but require changes in the existing network infrastructure. This is the main problem with network layer approaches, as service providers are not willing to change the network infrastructure due to additional deployment cost. Session layer mobility management approaches are quite favorable but they are used only for real time data, like multimedia applications, so cannot be
considered as the best for Internet traffic, which mainly (around 90%) consists of TCP traffic. Transport layer mobility management techniques are considered strongest, due to end-to-end nature, despite the fact that they require changes to the existing protocols like TCP. These transport layer techniques require no changes in the network infrastructure and mobility is handled by two end nodes. As no additional deployment cost is involved in end-to-end solutions thus service providers are facilitated to implement these solutions.

1.1.3 Desirable Characteristics of an End-to-End Mobility Management Solution

Mobility management scheme, operating at any layer of the protocol stack, should have the following characteristics that attempt to keep the Internet infrastructure unchanged by allowing the end nodes to take care of mobility.

- There should be no change in the network infrastructure
- No need to maintain states in the network like tunnel state
- Low handover latency should be involved
- Direct communication should be done between MN and RN, avoiding the packet redirection
- Application should remain unaware of the handover
- Authentication should be simple but highly secure and no third parties should be involved
- Mobility management scheme should not affect the TCP’s reliability and in-order delivery features

These characteristics can be achieved if a solution is based on end-to-end philosophy.
1.2 Problem Statement

The scope of this research is to analyze existing end-to-end mobility management solutions for TCP. Main focus is particularly around the performance metrics such as handover delay, throughput degradation time, protocol overhead, security issues and implementation issues. This research activity will identify the characteristics of an end-to-end mobility management solution and the behavior of different protocols in different mobility scenarios using common performance metrics. This performance evaluation will help to identify the characteristics that make a mobility solution better or otherwise, in comparison to other solutions. If possible, some improvements to existing solutions may also be suggested to overcome the limitation of these protocols.

1.3 Thesis Goals

The major goal of this thesis is to analyze performance of some existing end-to-end mobility management solutions for TCP. As a consequence of this evaluation, different protocols have been analyzed, using mathematical modeling techniques of performance evaluation already available in literature, in order to identify the characteristics that make a mobility solution better.

1.4 Thesis Organization

Rest of the thesis is organized as follows: In chapter 2 a survey of existing end-to-end mobility management protocols is presented. In chapter 3, 4 and 5, performance evaluation of different mobility management protocols on the basis of handover delay, throughput degradation time and protocol overhead is made. Finally in chapter 6, conclusion is drawn on the basis of analysis done.
Chapter 2
Survey of End-to-End Mobility Management Solutions for TCP

This chapter identifies various characteristics of mobility management protocols and surveys the existing techniques in the domain of end-to-end mobility management.

2.1 Characteristics of a Mobility Management Solution

Besides two basic components: handover management and location management, as discussed in chapter 1, mobility management solutions are supposed to have some other characteristics as well. Below mentioned are some of these characteristics.

2.1.1 Performance Characteristics

While developing an Internet wide mobility solution, performance characteristics should be given special attention. These characteristics have an impact on the overall performance of communication between communicating nodes, during the handover process. Following performance characteristics are most relevant for mobility support.

*Seamless handover:*

A seamless handover allows MN to continue communication without any disruption. Thus seamless handover can also be defined as make-before-break strategy for connection establishment.
Multi-homing:

Today’s mobile devices are equipped with multiple network interfaces. Thus it is possible for a mobile device to have multiple IP addresses. Using multi-homing, an association between Mobile Node (MN) and Remote Node (RN) can be spanned across multiple IP addresses.

Bandwidth Aggregation:

When a multi-homed MN is in the coverage areas of different wireless access networks, it can use multiple interfaces simultaneously for sending or receiving data. Some mobility management schemes use multiple TCP connections, between MN and RN, for a single stream of data. In such a situation, a scheduler is required for efficient bandwidth aggregation of data on parallel connections.

Simultaneous movement awareness:

In some scenarios it is possible that two communicating nodes may move simultaneously. A mobility management scheme is said to be simultaneous movement aware if it supports simultaneous movement of two communicating nodes.

2.1.2 Implementation and Deployment Related Characteristics

In this category, characteristics regarding the network infrastructure and the current implementation of TCP are included.

Changes in the current implementation of TCP:

Some mobility management schemes require changes in the current implementation of TCP, means that they do not use the conventional TCP control and data signaling. These schemes may be termed as the TCP-variants for mobility management.
Requires changes at end nodes:

A mobility scheme may require changes in the communication protocol stack at either end or at the both ends.

Additional support from network:

A mobility management scheme may require additional support from the network for handover control and data signaling, for example use of gateways, proxies, mobility agent, etc.

Lower layers awareness:

A mobility scheme that takes information about signal strengths etc. from lower layers, for handover decision, is said to be lower layer aware. Cross layer optimization may be used to pass the lower layer information to upper layers, bypassing the intermediate layers.

2.1.3 End User Related Characteristics

In this category, those characteristics are included which are related to the end user like willful handover and application transparency.

Willful handover:

Handover should satisfy the requirements of an end user. In case of overlapping regions, i.e. strong signals of different access networks are received by MN, user may not want to perform handover. Thus mobility management protocol should not perform handover if user is not willing to do it.

Application transparency:

A mobility management scheme is said to be application transparent if user application remains unaware of the underlying handover process.
Security issues:

Any mobility solution should protect itself against misuse of mobility features and mechanism, e.g. generating a false handover message. Therefore, security is an important concern when providing Internet wide mobility support.

2.2 Existing End-to-End Mobility Management Solutions

To handle the end-to-end mobility, various mobility management protocols have been proposed, operating at different layers of TCP/IP protocol stack. These solutions eliminate the need for making changes in the network infrastructure and tunneling of data packets like Mobile IP.

TCP-R (Funato et al., 1997), TCP-Migrate (Snoeren & Balakrishnan, 2000), pTCP (Hsieh & Sivakumar, 2004), Socketless-TCP (SL-TCP) (Beda & Ventura, 2005), Tsukamoto’s proposal (Tsukamoto, Shigeru & Yuji, 2007) etc., are some transport layer end-to-end mobility management solutions that have been proposed in the literature for TCP based application flows. All these solutions handle the mobility by making some modifications in TCP implementation at both ends. Mobile Stream Control Transmission Protocol (mSCTP) (Koh, Chang & Lee, 2004) is also an end-to-end transport layer protocol to provide mobility support but it is not based on TCP traffic. Layer 3.5 solutions are also proposed like Host Identity Protocol (HIP) (Moskowitz and Nikander, 2006) (So & Wang, 2006) to handle the mobility by separating the dual role of IP address as a location identifier and an end point identifier. Layer 4.5 solutions like LCT-based (Guo, Guo, Zhang & Zhu, 2003) and VA-TCP (Lin & Chang, 2007) are also proposed. Session layer mobility management is also discussed in literature like Session Layer mobility Management (SLM)
(Landfeldt, Larsson, Ismailov & Seneviratne, 1999) and Session Initiation Protocol (SIP) based mobility management (Banerjee, Acharya & Das, 2006). SLM is designed for TCP and UDP applications however SIP based mobility management solutions are not for TCP traffic, rather they are for real time traffic, especially for multimedia applications.

2.2.1 Layer 3.5 Solutions

In this category, those protocols have been included that introduce additional layer between Network and Transport layers of TCP/IP protocol stack, to handle the mobility.

2.2.1.1 HIP based Mobility Management

In current Internet, IP address is used both as location identifier and an end point identifier. Host Identity Protocol (HIP) (Moskowitz & Nikander, 2006) (So & Wang, 2006) separates the dual role of IP address as end point identifier and location identifier. It introduces a Host Identifier (HI) used as end point identifier at transport layer and IP address is used as location identifier at network layer. This is termed as a new layer, Host Identity layer, which is seen as layer 3.5 in TCP/IP protocol stack.

In this approach, seamless handover can be done as multi-homing is allowed. However parallel transmission on two links is limited due to the fact that when there is data flowing on one link, control packets for handover flow on other link. As HIP header is sent with each data packet, an efficient bandwidth aggregation mechanism can be implemented easily by sending some sequencing information in each data packet. For simultaneous movement support, HIP introduces a Rendezvous Server (RVS) in the network. HIP requires changes in TCP implementation and both ends need to be HIP aware. In HIP-based solutions, no option for willful handover is
considered and application transparency is provided by keeping user unaware of the underlying handover process.

2.2.2 Transport Layer Solutions

As mentioned in section 2.1.2, in this category transport layer solutions are included that handle mobility by making modifications in TCP implementation and are referred as TCP variants for mobility management.

2.2.2.1 TCP-Redirection

For continuous operation on Mobile Node, TCP-Redirection (TCP-R) (Funato et al., 1997) revises the pair of IP addresses and port numbers in the existing connection, whenever there is a change in the MN’s IP address.

After getting a new IP address in the new access network, Mobile Node sends Redirect message to Remote Node (RN). On receiving the Redirect message, RN sends Authentication request to MN. MN replies with connection Authentication reply message. RN checks the authentication reply message and if found correct, then it revises the pair of addresses of existing TCP connection. In the meanwhile, MN also revises its pair of addresses. DNS is proposed for location management in this protocol.

This scheme can be implemented as a complete mobility solution but seamless handover is not possible due to the fact that TCP-R does not support multi-homing. Moreover, due to non multi-homing nature, bandwidth aggregation cannot be done. No additional mechanism is proposed for simultaneous movement support. This protocol requires changes in TCP implementation at both ends. No option for willful handover is present but application transparency is provided.
2.2.2.2 TCP-Migrate

In TCP-Migrate (Snoeren & Balakrishnan, 2000), for connection migration, authors have proposed a Migrate option for TCP that allows an existing TCP connection to be migrated by either host from an old IP address to a new IP address. This TCP connection migration is done by exchanging newly introduced TCP SYN with Migrate option. In migrate option, a token (identifier) is included that identifies the previously established connection. This token is negotiated during initial connection establishment using migrate-permitted option. Location management is done through DNS dynamic updates.

This scheme is a complete end-to-end mobility management solution but the handover involved is hard handover. Just like Freeze-TCP (Goff, Moronski, Phatak, & Gupta, 2000), the connection is stopped whenever MN looses connectivity in a network. No option for multi-homing and bandwidth aggregation is available. It does not require any changes in the network infrastructure but TCP at both ends needs to be changed. Willful handovers are not supported thus providing application transparency. This scheme lacks simultaneous movement support. An enhanced version of TCP-Migrate with simultaneous movement support is proposed in (Wu, Le & Zhang, 2007).

2.2.2.3 pTCP

pTCP (Hsieh & Sivakumar, 2002 and 2004) is a transport layer solution for multi-homed MN. It creates and maintains one TCP state for each active network interface. TCP-virtual (TCP-v) pipes are created for each active interface in a pTCP connection. Whenever a pTCP socket is opened by application for a connection, then a TCP-v pipe for that connection using active interface is created. Application on the MN
opens a pTCP socket with RN and pTCP creates the first TCP-v pipe. This pTCP with only one pipe behaves the same as traditional TCP connection. When MN moves to a new network and gets a new IP address then pTCP creates a new pipe but does not closes the previously established pipe.

pTCP provides seamless handover and reliable sequenced delivery of data to application using its multi-homing feature. For effective throughput, in case of overlapping regions, it can perform bandwidth aggregation on parallel interfaces. No additional mechanism is proposed for simultaneous movement support. This protocol requires both ends to be pTCP aware. No additional support is required from the network, like proxy, etc. It does not support willful handovers but application transparency is maintained.

2.2.2.4 SL-TCP

Socketless-TCP (SL-TCP) (Beda & Ventura, 2005) has adopted a unique number, Connection IDentifier (CID), at both ends separately, to maintain the connection continuity, in case MN moves across heterogeneous networks. This CID, which is a combination of local port and destination IP, does not depend on the physical attachment of MN in the network; hence it remains unique for the duration of the communication session on MN. CID depends upon the location of RN in the network; hence whenever MN moves across heterogeneous networks and obtains a new IP address then CID at RN is updated. SL-TCP performs its own connection initialization, connection management and connection migration. These functionalities are implemented within the transport layer, as a modification to TCP (Beda & Ventura, 2005).
When a MN moves and changes its point of attachment (IP address), then SL-TCP connection migration depends upon the role of the end node. In case of MN, it has the correct destination IP address of RN within its CID mapping, thus outgoing packets will be addressed correctly (Beda & Ventura, 2005). In case of RN, whenever a packet is received from network layer, SL-TCP retrieves the source IP address from network layer header, and extracts the CID from transport layer header. It then checks the CID to (local port, destination IP) mapping and compares the source address obtained from the network layer with the destination IP, if the address is different it updates the destination IP address in CID, otherwise no action is taken. Packets are then routed to respective application via the applications port number (Beda & Ventura, 2005).

This protocol does not provide seamless handover and multi-homing facility. Simultaneous movement is also not supported by SL-TCP. Changes are to be made in TCP implementation and protocol stack at both ends. No details are given for location management but DNS dynamic updates can be used. Authors have not discussed that how lower layer information will be passed to upper layers for handover decision. Willful handover is not handled thus provides application transparency.

2.2.2.5 Tsukamoto’s Proposal for Multiple TCP Connections

Tsukamoto proposed a new mobility management scheme that modifies the transport protocol for handling multiple connections (Tsukamoto, Shigeru & Yuji, 2007) during handover process. A cross layer approach is used that detects any change in the status of wireless networks and selects the optimal interface. A Handover Manager (HM) is implemented at transport layer that performs the vertical
handover. Link layer information is sent to the HM at transport layer, bypassing the intermediate layers using cross layer manager. Using this information, HM can detect a link-up or link-down state. During communication, if HM detects a link-up state for a new wireless network, then it starts a new TCP connection via that interface and starts parallel transmission using two TCP connections. For each connection, end-to-end performance measures, such as Bandwidth, Bandwidth Delay product (BDP) or bottleneck bandwidth are calculated and compared, using first few packets in the start-up period (slow start). On the basis of this comparison, handover takes place.

In this scheme, no details of the control and data flow are given. Seamless handovers can be done using the functionality of multi-homing. Bandwidth aggregation is possible but authors have not given any details about this. No option is available for simultaneous movement handling. This protocol requires changes in TCP implementation at both ends but does not require any additional support from network. Cross layer optimization is used to pass lower layer information to higher layers. No provision of willful handover is there, providing application transparency.

2.2.3 Layer 4.5 and Upper Layer Solutions

In this category those solutions are included that introduce abstraction layer between transport and application layers.

2.2.3.1 LCT-based Handoff Protocol

To maintain connection’s continuity Local Connection Translation (LCT) based handoff protocol (Guo, Guo, Zhang & Zhu, 2003) introduces two components, Connection Manager (CM) and Virtual Connectivity (VC). Responsibility of CM includes the detection of link layer conditions, like signal strength, and network conditions like available bandwidth and end-to-end delay etc. VC is responsible for
maintaining connection continuity in case of mobility. VC has an internal component LCT, which maintains mapping relationship between the original connection information and the current connection information for each active connection, thus making the mobility transparent to applications. Location management is done through DNS. This protocol introduces, at application layer, Subscription/Notification (S/N) service to support mobility, in case Network Address Translator (NAT) and simultaneous movement are involved. To support these two functionalities, an additional entity, S/N server, is required in the network, implementing the subscription/notification functionalities. A S/N client is also required at each end node exchanging information with the S/N server using a S/N protocol.

VC handles the mobility by sending a Connection Update (CU) message to RN, when MN's IP address changes. Connection Update ACK (CUA) is an acknowledgment message for a received CU when RN trusts that the MN is really at its claimed current address carried in the CU. If RN does not trust the CU message, then it sends a Connection Update Challenge (CUC), which is acknowledge by MN using Connection update Challenge Response (CCR) message.

LCT-based handoff protocol is a complete mobility management solution. Multi-homing support is available hence seamless handover can be done in case of overlapping regions. No support for bandwidth aggregation is provided by this protocol. LCT-based protocol requires changes in the protocol stack by introducing new layer at both ends. To handle simultaneous movement, additional support is required from the network in the form of Rendezvous Server (RVS). Willful handovers are not handled by this protocol but application transparency is provided.
2.2.3.2 VA-TCP

Vertical handoff Aware-TCP (VA-TCP) (Lin & Chang, 2007) is an end-to-end mobility management scheme that uses the abstraction layer introduced by LCT-based handoff protocol to maintain the application transparency. For connection continuity, this protocol adds some new messages to support vertical handover and also maintains a LCT table, purposed by LCT-based protocol, for maintaining a relationship between original connection information and new connection information. When a MN roams in a new network, then VA-TCP dynamically estimates the connection parameters like bandwidth, delay and the bandwidth-delay product from the information provided by the lower layers, so that it can fit into the new environment. MN sends a notification message to RN telling that it has changed its IP address. RN updates the connection parameters. In the meanwhile MN also updates the connection parameters.

In VA-TCP multi-homing is available hence seamless handover is done when MN is present in overlapping regions. This protocol requires changes in the protocol stack by introducing an abstraction layer and in TCP implementation at both ends. Lower layer information is required for handover initiation. Handover is not done according to user requirements but is done according to the new network conditions. Application transparency is maintained by using the abstraction layer.

2.2.3.3 Session Layer Mobility management (SLM)

Session Layer mobility Management (SLM) (Landfeldt, Larsson, Ismailov & Seneviratne, 1999) is a framework that supports mobility management by introducing Session layer between Transport and Application layers. This layer operates above TCP and switches data streams between multiple connections. The end-to-end network path is divided into three separate paths. Two paths are between the
applications on the two hosts and socket connectors and the third path between the socket connections of two hosts. Location management is done through User Location Server (ULS).

In this scheme, multi-homing is available and bandwidth aggregation can be done, but no details are given about the aggregation mechanism. Seamless handover is performed in case of overlapping regions. Simultaneous movement problem is not handled by this technique. SLM does not require any changes in TCP implementation and uses the conventional TCP for communication. Both ends need to be SLM aware and User Location Server (ULS) is used as an alternate to DNS for location management. A new layer is inserted in TCP/IP protocol stack and no option is available for willful handover. This scheme provides application transparency.

Table 2.1 shows a qualitative analysis of different end-to-end mobility management solutions. Each column corresponds to an end-to-end mobility management solution.
management protocol and each row represents a characteristic of mobility management protocol that are discussed in section 2.1. Application transparency is an important feature that is provided by all the protocols but none of them have option for willful handover. All the protocols discussed in this chapter require some changes in TCP implementation for mobility support. SLM, HIP-based and LCT-based protocols require additional support from network, but are still included in the list of end-to-end protocols. Reason is that the additional network entities are not required for handover management, but they are required for location management to support simultaneous movement feature.

In this chapter, we have discussed different end-to-end mobility management solutions for TCP, and a qualitative analysis is made. In the upcoming chapters, we will discuss the performance analysis of these protocols on the basis of handover delay, throughput degradation time and protocol overhead. The reason for the selection of these metrics is that, these metrics affect the overall performance of a mobility management protocol when vertical handover is done by MN across heterogeneous access networks.
Chapter 3
Handover Delay Analysis

In this chapter, handover delay of end-to-end mobility management protocols in different vertical handover scenarios is discussed. Already proposed modeling techniques of performance evaluation are used and the handover delay of mobility management protocols is analyzed using common parameters.

3.1 Vertical Handover Scenarios

Consider a multi-homed Mobile Node (MN) with two network interfaces. Two different vertical handover scenarios are possible here: movement of MN in non-overlapping regions and movement of MN in overlapping regions, as shown in Figure 3.1.
Scenario 1: Movement of MN in non-overlapping regions

When MN moves out of network's coverage area and enters into the coverage area of some other access network, connectivity of MN in old access network is no more active. In this scenario, MN is connected to single access network at a time, as shown as S1 in Figure 3.1. Here break-before-make handover is done. MN has to perform the scanning for new network then association with the Access Point (AP) or Base Station (BS) and finally handover signaling will be done with Remote Node (RN).

Scenario 2: Movement of MN in overlapping regions

In this scenario, two different cases are possible: MN enters into the overlapping region or leaves the overlapping region.

Case a: MN enters into overlapping region

MN moves into the coverage area of a new access network while remaining in the coverage area of old access network. Here MN has arrived in the overlapping region and is accessible through two different access networks, as shown as S2 in Figure 3.1.

Case b: MN leaves the overlapping region

MN moves out of the overlapping region and loses connectivity to the network through which it was connected before. This scenario is shown as S3 in Figure 3.1. Here MN can perform a make-before-break handover if the mobility management protocol has the intelligence to detect a link going down event.

To analyze the handover delay in above mentioned scenarios we have used the reference network as shown in Figure 3.1. Let $T_d$ is the one way delay for data packets between MN and RN in new access network, $T_{ip}$ is the time required to acquire new IP address, $T_{L2}$ is the time required for layer2 connectivity and $T_{sig}$ is the time
required for the signaling of handover control messages between two communicating nodes.

Figures 3.2 and 3.3 show timing diagram of handover delay for scenario1 when MN is receiving data and MN is sending data respectively. Here MN moves between non overlapping regions of two different access networks. It loses its connectivity in the old network thus becomes unable to send/receive data until it moves into the coverage area of new access network and performs handover signaling.

![Figure 3.2 Timing diagram non overlapping regions, when MN is receiving data](image-url)
In scenario 2a, MN enters into the overlapping region of two access networks. In this situation the handover delay depends upon the nature of mobility management protocol. Some mobility solutions stops sending/receiving data through old interface after MN moves into coverage area of new access network and perform the same procedure as discussed above for scenario1; thus handover delay will be same as shown in Figure 3.2 and Figure 3.3. If the mobility management solution allows simultaneous communication through two different interfaces then MN will be receiving data through old network while it is performing layer2 connectivity and then exchanging handover control messages (connection migration / new connection establishment) with RN through new interface. Here MN will not suffer from any handover delay because MN is still receiving data through old interface while it is performing handover signaling through new interface.

Figure 3.3 Timing diagram non overlapping regions, when MN is sending data
In scenario 2b, MN was in overlapping region before moving out thus we assume that it has already established layer2 connectivity with AP/BS of new access network. Handover delay will be worst when MN has not acquired the IP address in new network and best when MN has already acquired the IP address. Figure 3.4 and 3.5 depicts the timing diagrams for handover delay in this scenario.

Figure 3.4 Timing diagram scenario 2b, where MN has already performed L2 connectivity

Figure 3.5 Timing diagram for scenario 2b, where MN has already acquired IP address
In all of the above discussed scenarios, handover signaling between the communicating nodes is considered as three-way handshake. However different end-to-end mobility management protocols exchange different number of messages for handover signaling thus their handover delay will be different.

3.2 Modeling End-to-End Handover Delay

Total handover delay of any end-to-end mobility management solution is the time interval between receiving the last packet in old access network and the first packet in new access network. It is the sum of all individual delays involved in handover. Usually handover includes the following delay components:

a. Layer 2 connectivity delay \( (T_{L2}) \)
b. IP address acquisition delay in new network \( (T_{IP}) \)
c. Connection setup/migration delay \( (T_{sig}) \)
d. One way delay for data packet in new network \( (T_d) \)

a. Layer 2 connectivity delay:

Link layer connectivity is the time required to establish layer2 connectivity \( (T_{L2}) \) between the mobile node and Access Point (AP)/Base Station (BS) of new access network. It includes the time spent in scanning of available AP/BS and then authenticating and associating with one of them. Beacon messages are broadcast by the AP that includes an Extended Service Set Identifier (ESSID) of new access network. MN chooses an AP depending upon the handover policy used by the mobility management protocol.
b. IP address acquisition delay in new network:

After performing layer 2 connectivity, an IP address for new UP interface is acquired in the new network. This time is said to be IP address acquisition delay \( T_{IP} \) in new access network.

c. Connection setup/migration delay:

After making decision for handover and getting new IP address, mobility management protocols either migrate the old connection or establish a new TCP connection, depending upon the mobility management solution, in a secure or non secure manner. This process is referred to as connection setup/migration delay or handover control messages signaling delay \( T_{sig} \) during handover.

d. One way delay for data packet in new network:

It is the one way delay between MN and RN in the new access network. After handover control messages signaling, RN sends first data packet to the MN’s new location which requires one way delay, \( T_d \) in new network.

Summing up all the individual delay components, total handover delay will be:

\[
T_{\text{handover}} = T_{L2} + T_{IP} + T_{sig} + T_d
\]  

(3-1)

For our performance analysis, \( T_{sig} \) is the major parameter that will affect the total handover delay of different end-to-end mobility management protocols. \( T_{IP} \) and \( T_{L2} \) usually remain same for a particular access network.

(Kim & Koh, 2008) has discussed a mathematical model for the evaluation of handover delay of Mobile IPv6 (MIPv6) and mSCTP. We are using the same model for end-to-end protocols. According to this model, signaling delay of any mobility management protocol \( x \) can be modeled as a function of number of signaling
messages exchanged and the end-to-end delay between communicating nodes (Kim & Koh, 2008).

\[ T_{\text{sig}_x} = N(D_{MN-RN}) + M(D_{RN-MN}) \]  \hspace{1cm} (3-2)

Where

\[ T_{\text{sig}_x} = \text{Handover signaling delay for protocol } x \]

\[ N = \text{number of handover signaling messages sent from MN to RN} \]

\[ M = \text{number of handover signaling messages sent from RN to MN} \]

\[ D_{x-y} = \text{one way end-to-end delay for handover signaling messages between node } x \text{ and } y \]

Handover delay in non overlapping regions, discussed in section 3.1, includes all the four delay components discussed above. The reason is that MN looses connectivity in the old access network and enters into the network coverage area of a new access network, where it has to face all the handover delay components. Details are shown in Figure 3.6.

\[ T_{\text{handover}_x} = T_{L2} + T_{IP} + T_{\text{sig}_x} + T_d \]  \hspace{1cm} (3-3)

Figure 3.6 Handover delay in non overlapping regions
Handover delay when MN enters into coverage area of overlapping regions depends upon the mobility management protocol, whether it allows simultaneous communication through two interfaces or not. If mobility management protocol allows simultaneous communication through two network interfaces then handover delay will be zero (MN is receiving data through old network interface while it is performing handover signaling through new network interface), otherwise it will be the sum of handover signaling delay and $T_d$, as shown in Figure 3.7.

$$T_{handover} = \begin{cases} 0 & : \text{protocol allows simultaneous communication} \\ T_{sig} + T_d & : \text{protocol does not allow simultaneous communication} \end{cases}$$ (3-4)

Figure 3.7 Handover delay, when MN moves into overlapping region and protocol doesn't allow simultaneous communication

In case MN moves out of overlapping regions, then handover delay will depend upon the presence/absence of Link Going Down (LGD) intelligence (Lampropoulos, Salkintzis & Passas, 2008). If mobility management protocol has the intelligence to predict a LGD, then MN can initiate handover signaling through alternate interface before the old link goes down resulting zero handover delay, otherwise handover
delay will be equal to the sum of handover signaling delay and one way delay for data packet between MN and RN in new network.

\[
T_{\text{handover} - x} = \begin{cases} 
0 & \text{; protocol has LGD intelligence} \\
T_{\text{sig} - x} + T_d & \text{; protocol has no LGD intelligence}
\end{cases} \tag{3-5}
\]

Majority of end-to-end protocols for mobility management use connection re-establishment or connection migration, between MN and RN in new network, to process the handover. (Cardwell, Savage \& Anderson, 2000) modeled the TCP connection setup delay as a function of end-to-end packet loss probabilities in forward and reverse directions.

To calculate the end-to-end packet loss probabilities, we use the same reference network as shown in Figure 3.1. MN is attached to wireless Access Point (AP)/ Base Station (BS) and AP/BS is connected to RN using wired medium in the Internet. The path between MN and RN consists of two parts: wireless link between MN and BS/AP and the wired link between BS/AP and the RN.

The end-to-end packet loss probability between MN and RN as discussed by (Mohanty \& Akyildiz, 2007):

\[
p = 1 - (1 - P_{wl}) (1 - P_{wr}) \tag{3-6}
\]

Where \(P_{wl}\) is the packet loss probability in wireless part and \(P_{wr}\) is the packet loss probability in wired part.

Packet loss probability in wireless link can be represented as a function of link layer frame error rate and the number of link layer frames per packet (Mohanty \& Akyildiz, 2007).

\[
p_{wl} = 1 - (1 - FER)^k \tag{3-7}
\]
Where FER is the link layer Frame Error Rate (FER) and \( k \) is the number of link layer frames per packet. \( k \) can be calculated by dividing the packet length with link layer frame length. Packet loss probability in the wired link is very small. Thus the end-to-end packet loss probability between MN and RN from equations (3-6) and (3-7) (Mohanty & Akyildiz, 2007):

\[
p = 1 - (1 - \text{FER})^k (1 - p_{\text{wr}})
\]

(3-8)

Analytical model proposed by (Mohanty & Akyildiz, 2007), discussed the connection setup/migration delay as function of end-to-end packet loss probability, number of messages exchanged, their number of retries and the retransmission timeout. According to this analytical model, signaling delay of protocols that exchange three messages for handover (1.5 RTT), like TCP-Migrate, TCP-Redirect (TCP-R), Tsukamoto’s proposal and parallel-TCP (pTCP) will be (Mohanty & Akyildiz, 2007):

\[
L_{\text{sig}_{3h}} = 1.5RTT + \sum_{a=0}^{i-1} 2^a \text{RTO} + \sum_{b=0}^{j-1} 2^b \text{RTO} + \sum_{c=0}^{k-1} 2^c \text{RTO}
\]

or

\[
L_{\text{sig}_{3h}} = 1.5RTT + (2^i + 2^j + 2^k - 3) \text{RTO}
\]

(3-9)

Where \( i-1, j-1 \) and \( k-1 \) are the number of unsuccessful tries for handover control messages and protocol doubles Retransmission Time Out (RTO) after each RTO expiration.

Analytical model proposed by (Mohanty & Akyildiz, 2007) discussed the handover signaling delay for TCP-Migrate only. We are using the same model,
equation (3-9), to derive handover signaling delay for other end-to-end mobility management protocols.

The signaling delay of protocols like Local Connection Translation (LCT) based handoff protocol that exchange two messages for handover signaling (1 RTT) will be:

\[ L_{\text{sig}_{-x}} = RTT + \sum_{a=0}^{i-1} 2^a RTO + \sum_{b=0}^{j-1} 2^b RTO \]

or

\[ L_{\text{sig}_{-x}} = RTT + (2^i + 2^j - 2)RTO \]  (3-10)

Similarly, the signaling delay of protocols like SocketLess-TCP (SL-TCP) and Vertical handoff Aware-TCP (VA-TCP) that exchange single message (0.5 RTT) for handover signaling is:

\[ L_{\text{sig}_{-x}} = 0.5RTT + \sum_{a=0}^{i-1} 2^a RTO \]

or

\[ L_{\text{sig}_{-x}} = 0.5RTT + (2^i - 1)RTO \]  (3-11)

Let \( p_1, p_2 \) and \( p_3 \) are the end-to-end packet loss probabilities of three handover control messages, which are calculated using equation (3-8), then probability of handover signaling completion for different protocols, discussed in equations (3-9), (3-10) and (3-11), will be (Mohanty & Akyildiz, 2007):

\[ P_{\text{sig}_{-x}} = p_1^i (1 - p_1) p_2^j (1 - p_2) p_3^k (1 - p_3) \]  (3-12)

\[ P_{\text{sig}_{-x}} = p_1^i (1 - p_1) p_2^j (1 - p_2) \]  (3-13)

\[ P_{\text{sig}_{-x}} = p_1^i (1 - p_1) \]  (3-14)
Where $P_{\text{sig}_x}$ in equations (3-12) to (3-14) is the probability of handover signaling completion for protocol $x$ for three different handover signaling types discussed above in equations (3-9) to (3-11). The average value of handover signaling delay (Mohanty & Akyildiz, 2007):

$$E[T_{\text{sig}_x}] = \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{k=0}^{N} P_{\text{sig}_x} L_{\text{sig}_x}$$

(3-15)

Where $N$ is the maximum allowed number of retries for handover signaling messages. (Cardwell et al., 2000) has discussed that most TCP implementations abort connection establishment attempt after 4-6 failures.

Analytical model discussed in (Munasinghe & Jamalipour, 2007), also attempts to model the signaling delay during handover. Main focus of this technique is to find the handover delay of Mobile IP but it can be used for end-to-end mobility management solutions. It has discussed signaling delay as a function of packet size, bandwidths and delays of wired and wireless parts, average number of hops and the processing delay at each hop (Munasinghe & Jamalipour, 2007).

$$T_{\text{sig}} = n \left[ \left( \frac{S}{B_{wl}} + L_{wl} \right) + \left( \frac{S}{B_{wr}} + L_{wr} \right) H_{a-b} + L_{\text{proc}} (H_{a-b} + 1) \right]$$

(3-16)

Where $S$ is the average size of handover signaling messages exchanged between MN and RN, $B_{wl}$ and $B_{wr}$ are the bandwidths of wireless and wired links, $L_{wl}$ and $L_{wr}$ are the latencies of wireless and wired link respectively, $H_{a-b}$ is the average number of hops between two communicating nodes, $L_{\text{proc}}$ is the processing delay at each node and $n$ is the number of handover control messages exchanged during handover between MN and RN.
3.3 Handover Delay Analysis

For our performance comparison, we assume that Layer 2 connectivity delay and IP address acquisition delay remain same for all end-to-end protocols in a particular access network. The only thing that affects the handover delay is the handover signaling delay that depends upon mobility management protocol.

To compare the handover delay performance of end-to-end mobility management solutions for TCP connection, we assume following values for different parameters (Mohanty & Akyildiz, 2007): $T_{IP}=20\text{msec}$, $T_{L2}=10\text{msec}$ and $T_d=50\text{msec}$. Packet loss probability in the wired network is considered as $P_{wr}=1e-6$, $RTT=100\text{msec}$ and $RTO=200\text{msec}$.

Figure 3.8 shows the comparison of handover delay as a function of packet loss probability of different end-to-end mobility management solutions in scenario 1, when MN is receiving data. These results are obtained using equations (3-6) to (3-15) and equation (3-1). Results show that the protocols can be categorized in three groups.

Figure 3.8 Handover delay as function of packet loss probability in scenario 1, non overlapping region
keeping in mind the observed handover delay. This categorization is made on the basis of number of handover signaling messages exchanged. SL-TCP and VA-TCP performs much better than the other protocols. For example, for a loss probability of 0.15, the delay for LCT-based protocol is twice than that for the SL-TCP and VA-TCP whereas for the TCP-Migrate and related protocols, the delay is about 3.6 times that of the SL-TCP and VA-TCP. Further, the difference in delay performance of different protocols increases with increasing packet loss probability. It can be seen that the SL-TCP and VA-TCP are the least affected with increase in packet loss probability whereas the TCP-Migrate, pTCP, HIP, Tsukamoto’s proposal and TCP-R are the most affected. The reason is that for SL-TCP and VA-TCP the number of handover control messages is smaller as compared to other protocols.

Figure 3.9 shows the handover delay in scenario 2a, when MN enters into overlapping region of two access networks. These results are obtained from equation (3-6) to (3-15) and equation (3-1). Behavior of TCP-Migrate, TCP-R and SL-TCP is
same as was in Figure 3.8. The reason is that these protocols do not support multi-homing and simultaneous communication through parallel network interfaces, which causes handover delay. VA-TCP, SLM, pTCP, LCT-based handoff protocol and HIP-based solutions have zero handover delay. These protocols allow MN to exchange data through old network interface while performing handover signaling through new interface thus zero handover delay is observed in case MN moves into overlapping region.

Figure 3.10 shows the handover delay in scenario 2b, when MN moves out of overlapping region. These results are obtained from equations (3-6) to (3-15) and equation (3-1). Here handover delay of VA-TCP has become smaller than SL-TCP, as compared to Figure 3.8, where both have the same handover delay. The reason for this decrease is that, VA-TCP allows multi-homing, thus MN has already performed layer2 connectivity and IP address acquisition in the new network. Similarly TCP-
-Migrate and TCP-R have higher delays as compared to SLM, pTCP and HIP-based solutions, due to non multi-homed nature.

To compare the handover delay as a function of bandwidth, we used following values for different parameters (Munasinghe et al., 2008): $S=60$ bytes, $L_{proc}=1 \times 10^{-3}$ msec, $H_{a-b}=10$, $L_{wl}=0.1$ msec against different values of $B_{wb}$, $L_{wr}=1 \times 10^{-3}$ msec and $B_{wr}=1$ Mbps. When we plotted the graph for $T_{sig}$ against different values of $B_{wb}$ we found that signaling delay decreases exponentially up to an increase in bandwidth of 500 Kbps. After which this decrease becomes linear by increasing the bandwidth but upto a certain limit, 1.5Mbps, and after which it remains same irrespective of the increase in bandwidth as shown in Figures 3.11, 3.12 and 3.13.

Figure 3.11 shows the handover delay as a function of bandwidth in scenarios where MN moves across non overlapping regions of two access networks. Results are obtained from equation (3-1) and (3-16). Behavior of protocols is same as was in Figure 3.8. It can be seen that SL-TCP and VA-TCP perform better as compared to

![Figure 3.11 Handover delay as function of bandwidth in scenario 1, non overlapping region](image)

36
other protocols. These two protocols, as discussed earlier, exchange less number of handover messages resulting minimum handover delay.

Figure 3.12 shows the handover delay as a function of bandwidth in scenario 2a when, MN while connected to an access network enters into the overlapping region. Results are obtained from equation (3-1) and (3-16). Here handover delay of a group of protocols is found to be zero. These protocols (VA-TCP, SLM, pTCP, LCT-based, HIP-based, Tsukamoto's proposal) allow MN to exchange data through old network interface while performing handover signaling through new interface. Thus MN will be receiving data through old network interface while mobility management protocol is performing handover process, resulting zero handover delay.
of SL-TCP (from 401 ms to 328 ms). The reason for this decrease is that, VA-TCP allows multi-homing, thus layer2 connectivity delay and IP address acquisition delay will not be involved in the handover delay. Similarly TCP-Migrate and TCP-R have higher handover delays as compared to SLM, pTCP, Tsukamoto's proposal and HIP-based solutions, due to non multi-homed nature.

![Figure 3.13 Handover delay as function of bandwidth in scenario 2b, where MN leaves the overlapping region](image)

Figure 3.13 Handover delay as function of bandwidth in scenario 2b, where MN leaves the overlapping region
Table 3.1 shows the handover delay comparison of mobility management protocols for different scenarios. These values represent the same results that were discussed in previous graphs. SL-TCP's and VA-TCP's performance is better as compared to other protocols in non-overlapping regions but protocols having option for simultaneous communication, like pTCP, Tsukamoto's proposal, SLM etc., perform better in overlapping regions.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>MN moves across non overlapping regions</th>
<th>MN enters the overlapping region</th>
<th>MN leaves the overlapping region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>As a function of packet loss probability</td>
<td>As a function of bandwidth</td>
<td>As a function of packet loss probability</td>
</tr>
<tr>
<td>HIP-based</td>
<td>187-637</td>
<td>943-724</td>
<td>0</td>
</tr>
<tr>
<td>SL-TCP</td>
<td>92-142</td>
<td>401-328</td>
<td>92-142</td>
</tr>
<tr>
<td>Tsukamoto's proposal</td>
<td>187-637</td>
<td>943-724</td>
<td>0</td>
</tr>
<tr>
<td>pTCP</td>
<td>187-637</td>
<td>943-724</td>
<td>0</td>
</tr>
<tr>
<td>VA-TCP</td>
<td>92-141</td>
<td>401-328</td>
<td>0</td>
</tr>
<tr>
<td>LCT-based</td>
<td>127-327</td>
<td>672-526</td>
<td>0</td>
</tr>
<tr>
<td>SLM</td>
<td>187-637</td>
<td>943-724</td>
<td>0</td>
</tr>
</tbody>
</table>

In this chapter, we have discussed the handover delay of different end-to-end mobility management protocols. The results manifest that the protocols that incorporate mechanism for simultaneous communication through multiple interfaces permit seamless vertical handover, consequently providing less handover delay.
Chapter 4
Analysis of Throughput Degradation Time during Vertical Handover

After moving out of the coverage area of an access network, Mobile Node (MN) becomes unable to exchange data, thus a zero throughput is observed. When MN enters into a new network, end-to-end mobility management protocols migrate the old TCP connection or establish a new one. Normally a new TCP connection starts from slow start and gradually moves to a steady state. In this situation, MN suffers from throughput degradation. In this chapter, throughput degradation time is discussed in different vertical handover scenarios, as discussed in chapter 3, and mobility management protocols are evaluated for throughput degradation time during vertical handover with the help of mathematical modeling techniques.

4.1 Modeling Throughput Degradation Time for Vertical Handover

Throughput degradation time during vertical handover is the sum of handover delay and time spent by TCP connection to reach steady state. If the handover delay is zero (as discussed in chapter 3), then throughput degradation time will be the time spent by TCP connection to reach steady state.

\[ T_{TD} = T_{HO} + T_{ss} \]  \hspace{1cm} (4-1)

Where
$T_{TD}$ = Throughput degradation time

$T_{HO}$ = Time for handover

$T_{ss}$ = Time to reach at steady state

To model the throughput degradation time, we calculate the time spent by TCP connection in slow start phase i.e. time to reach steady state. (Mohanty & Akyildiz, 2007) has modeled the time to reach steady state as a function of number of slow start rounds and the RTT value i.e.

$$T_{ss} = [\log_2 (1 + CW_{ss}) - 1] \times RTT$$

(4-2)

Where $\log_2 (1 + CW_{ss}) - 1$ is the number of slow start rounds, that can be calculated from equation (4-3), and RTT is the average round trip time.

For simplicity, we assume that there is no packet loss during slow start and TCP doubles its congestion window every $RTT$ and after time $T_{ss}$, TCP connection reaches at steady state. Congestion window at steady state, $CW_{ss}$ is (Mohanty & Akyildiz, 2007):

$$CW_{ss} = 1 + 2 + 2^2 + 2^3 + \ldots + 2^i = 2^{i+1} - 1$$

(4-3)

Here $i$ is the number of rounds taken by TCP to reach at steady state.

To calculate the congestion window size and time spent in slow start, (Cardwell et al., 2000) considered $\gamma$ to denote the rate of exponential growth of congestion window during slow start. According to this model, amount of data sent to reach at steady state, $data_{ss}$, in $i$ number of rounds can be modeled using geometric series as (Cardwell et al., 2000):

$$data_{ss} = w_1 + w_1 \gamma + w_1 \gamma^2 + \ldots + w_1 \gamma^{i-1}$$

or
data
ss = \( w_1 \cdot \frac{\gamma^i - 1}{\gamma - 1} \)  \hspace{1cm} (4-4)

\( w_1 = \text{initial congestion window} \)

Where \( w_1 \) is the number of segments in initial congestion window when TCP sender starts transmission, normally \( w_1 = 1 \) (Stevens, 1997).

Number of rounds \( i \), that TCP has spent to reach at steady state, can be derived from equation (4-4) as (Cardwell et al., 2000):

\[
i = \log_{\gamma} \left( \frac{\text{data}_{ss} (\gamma - 1)}{w_1} + 1 \right)
\]

(4-5)

The expected time spent in slow start can be expressed as (Cardwell et al., 2000):

\[
E[T_{ss}] = RTT \cdot i
\]

(4-6)

Here \( E[d_{ss}] = \text{avg}[\text{data}_{ss}] \)

In chapter 3, we have discussed different vertical handover scenarios. In those scenarios, throughput degradation time of different protocols is discussed in the following.

Throughput degradation time in scenario 1:

In case of MN’s movement in non-overlapping regions, throughput degradation time will be the sum of handover delay and time spent in slow start, for protocols that do not have any throughput degradation avoidance mechanism. It is due to the fact that MN loses its connectivity in the old access network and after some time, enters into coverage area of another access network where it first establishes a new TCP connection and then starts sending/receiving data in slow start phase. For protocols
that use throughput degradation avoidance mechanism (like VA-TCP), this time will be equal to the time spent for handover. Let \( T_{TD,x} \) is the throughput degradation time in vertical handover for protocol \( x \), then:

\[
T_{TD_{-x}} = \begin{cases} 
T_{HO} + T_{ss} ; & \text{no throughput degradation avoidance mechanism} \\
T_{HO} ; & \text{having throughput degradation avoidance mechanism} 
\end{cases} \tag{4-7}
\]

**Throughput degradation time in scenario 2a:**

In case MN moves into overlapping region of two access networks, throughput degradation time depends upon the mobility management protocol.

*Case 1.* If mobility management protocol allows to exchange data through old connection until new connection is established, then throughput degradation time will be the time spent by new TCP connection to reach steady state, for protocols that don't use any throughput degradation avoidance mechanism otherwise it will be zero. Let \( T_{TD,x} \) is the throughput degradation time after vertical handover for protocol \( x \), then:

\[
T_{TD_{-x}} = \begin{cases} 
T_{ss} ; & \text{no throughput degradation avoidance mechanism} \\
0 ; & \text{having throughput degradation avoidance mechanism} 
\end{cases} \tag{4-8}
\]

*Case 2.* If the mobility management solution does not allow parallel communication through two different interfaces, then after detecting new access network, MN decides for handover and will stop sending/receiving data through old connection. After establishing/migrating connection in the new access network, TCP at MN will start in slow start phase. In this case throughput degradation time will be same as discussed in equation (4-7).
Throughput degradation time in scenario 2b:

In case MN moves out of overlapping region, throughput degradation time depends upon the mobility management protocol.

Case 1. If the mobility management protocol has the LGD (Link Going Down) intelligence (Lampropoulos et al., 2008), to detect in advance that interface through which it is connected will go down, then it can establish/migrate the connection through new interface before losing connectivity of old access network. Here throughput degradation time will be same as discussed in equation (4-8).

Case 2. If the mobility management solution can't detect a LGD, then after losing the connectivity in old access network, MN will establish/migrate the connection through new access network. Here throughput degradation time will be same as discusses in scenario1, equation (4-7).

4.2 Throughput Degradation Time Analysis

To compare the throughput degradation time during vertical handover using models discussed in section 4.1, we considered following values for different parameters used: \( w_i = 1 \text{MSS} \) (Stevens, 1997), \( d_{ss} = 1024 \) segments (Cardwell et al., 2000). By using these values, and for simplicity assuming that there is no packet loss during data transmission (Mohanty et al., 2007), we calculated the number of slow start rounds \( i \), \( E[W_{ss}] \), \( E[T_{ss}] \), \( CW_{ss} \) and \( T_{ss} \). For both of the two models (equation 4-2 and 4-6) we found the same values for throughput degradation time.

Figure 4.1 shows the throughput degradation time for scenario 1. Here it includes the time to reach at steady state and time for handover except VA-TCP. VA-TCP has the lowest throughput degradation time. It is because, when roaming into a new network, VA-TCP-enabled MN dynamically estimates the bandwidth, delay and the
bandwidth-delay product. On the basis of estimated bandwidth-delay product, VA-TCP sets slow start threshold, congestion window, RTO, and RTT appropriately to fit into the new network conditions. Hence VA-TCP does not start from slow start phase, to avoid throughput degradation time.

Figure 4.2 shows the throughput degradation time of different protocols in scenario 2a. Here MN moves into the overlapping region and some mobility management protocols (like pTCP) allow to send/receive data through old connection. VA-TCP, SLM and pTCP have zero throughput degradation time. It is because, pTCP and SLM allows simultaneous communication through parallel interfaces, while VA-TCP adjusts the congestion window according to the new network conditions. HIP-based handover protocol allows simultaneous communication through two interfaces but restricted to the fact that when data is traveling on old network interface, handover control signaling messages are being exchanged through new network interface. After handover TCP starts from slow start causing throughput degradation.
Figure 4.2 Throughput degradation time, when MN moves into overlapping region

Figure 4.3 shows the throughput degradation time in scenario 2b for different protocols, when MN moves out of overlapping region. As no protocol discussed so far in this research work has the intelligence to detect a LGD (Link Going Down), hence handover delay will be added in the throughput degradation time thus eliminating the possibility of zero throughput degradation time in scenario 2b.

Figure 4.3 Throughput degradation time, when MN moves out of overlapping region
Table 4.1 Throughput degradation time comparison of different mobility management protocols during vertical handover

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Throughput degradation time (msec)</th>
<th>MN's movement across non-overlapping region</th>
<th>MN moves into overlapping region</th>
<th>MN moves out of overlapping region</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP-based</td>
<td></td>
<td>1550-2000</td>
<td>1362</td>
<td>1519-1969</td>
</tr>
<tr>
<td>SL-TCP</td>
<td></td>
<td>1454-1505</td>
<td>1454-1505</td>
<td>1454-1505</td>
</tr>
<tr>
<td>Tsukamoto’s proposal</td>
<td></td>
<td>1550-2000</td>
<td>1362</td>
<td>1519-1969</td>
</tr>
<tr>
<td>pTCP</td>
<td></td>
<td>1550-2000</td>
<td>0</td>
<td>1519-1969</td>
</tr>
<tr>
<td>VA-TCP</td>
<td></td>
<td>92-142</td>
<td>0</td>
<td>62-112</td>
</tr>
<tr>
<td>LCT-based</td>
<td></td>
<td>1490-1690</td>
<td>1362</td>
<td>1460-1660</td>
</tr>
<tr>
<td>SLM</td>
<td></td>
<td>1550-2000</td>
<td>0</td>
<td>1519-1969</td>
</tr>
</tbody>
</table>

Table 4.1 shows the throughput degradation time of different end-to-end mobility management protocols. Results show that protocols that allow simultaneous communications after vertical handover (like pTCP, Tsukamoto's proposal and SLM), in overlapping regions, perform better with minimum or zero throughput degradation time as compared to protocols that do not allow simultaneous communication through parallel interfaces. Also, protocols that use some mechanism to avoid throughput degradation time after vertical handover, like VA-TCP, performs better in overlapping as well as non-overlapping regions.

We have discussed so far the handover delay and throughput degradation time after vertical handover for different end-to-end mobility management protocols in different handover scenarios. Now we move on to the next step which is the protocol overhead analysis, which will compare protocols on the basis of overhead traffic being created by mobility management protocol.
Chapter 5
Protocol Overhead Analysis

5.1 Protocol Overhead

Protocol overhead is the traffic created by mobility management protocol in the network. As protocol overhead we have considered the amount of information being sent by the mobility management protocol with each data packet and in handover control messages. Amount of information being sent with each data packet and in each handover control message, depends upon the nature of mobility management protocol.

5.2 Modeling Protocol Overhead

The protocol overhead for protocol $x$, $PO_x$, of different end-to-end mobility management protocols is the sum of overhead created by protocol for handover signaling and the overhead of the additional bytes sent with each data packet, i.e.

$$PO_x = Overhead_{HO} + Overhead_{AD}$$  \hspace{1cm} (5-1)

Where

$Overhead_{HO}$ = handover signaling overhead

$Overhead_{AD}$ = overhead of the additional bytes sent with each data packet by the mobility management protocol

Handover signaling overhead can be computed as:

$$Overhead_{HO} = a.(S_{MN-RN}) + b.(S_{RN-MN}) + c.(S_{HP})$$  \hspace{1cm} (5-2)
Where

\( a \) = number of messages sent from MN to RN for connection migration

\( S_{MN-RN} \) = size of message sent from MN to RN for connection migration

\( b \) = number of messages sent from RN to MN for connection migration

\( S_{RN-MN} \) = size of message sent from RN to MN for connection migration

\( c \) = number of handover packets

\( S_{HP} \) = size of handover packet sent from MN to RN

Overhead of additional bytes sent with each data packets for \( n \) data packets can be computed as:

\[
Overhead_{AD} = \sum_{i=1}^{n} S_{Add\_bytes}
\]  

(5-3)

Where \( S_{add\_bytes} \) is the additional bytes sent with each data packet by the end-to-end mobility management protocol and \( n \) is the number of data packets.

If \( N \) number of handovers are performed by MN, then the total overhead for \( N \) handovers, \( Overhead_{N\_HO} \), will be:

\[
Overhead_{N\_HO} = N . Overhead_{HO}
\]  

(5-4)

### 5.2.1 Protocol Overhead of Different End-to-End Mobility Management Protocols

Table 5.1 shows a comparison of different end-to-end mobility management protocols regarding protocol overhead that is computed using equation (5-2). Here, handover signaling message packet size includes the 60 bytes of TCP, IP and link layer headers also, except for HIP that does not contain TCP header. An analysis of protocol overhead of different protocols is done in the next section and a description of each protocol regarding protocol overhead is given here:
Table 5.1 Protocol overhead comparison of different end-to-end mobility management protocols

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Handover signaling messages size</th>
<th>Handover overhead</th>
<th>Number of additional bytes sent with each data packet</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTCP</td>
<td>60,60,60,72 B</td>
<td>252 B</td>
<td>12 B</td>
</tr>
<tr>
<td>TCP-Migrate</td>
<td>80,80,80 B</td>
<td>240 B</td>
<td>0 B</td>
</tr>
<tr>
<td>SLM</td>
<td>60,60,60,64B</td>
<td>244 B</td>
<td>4 B</td>
</tr>
<tr>
<td>TCP-R</td>
<td>70,66,70 B</td>
<td>206</td>
<td>0 B</td>
</tr>
<tr>
<td>HIP based</td>
<td>84,84,84 B</td>
<td>252 B</td>
<td>40 B</td>
</tr>
<tr>
<td>LCT-based</td>
<td>Not defined</td>
<td>Not defined</td>
<td>Cid (size not defined)</td>
</tr>
<tr>
<td>SL-TCP</td>
<td>66B</td>
<td>66 B</td>
<td>06B</td>
</tr>
<tr>
<td>VA-TCP</td>
<td>Not defined</td>
<td>Not defined</td>
<td>Not defined</td>
</tr>
<tr>
<td>Tsukamoto’s proposal</td>
<td>Not defined</td>
<td>Not defined</td>
<td>Not defined</td>
</tr>
</tbody>
</table>

In pTCP (Hsieh et al. 2004), after initiating a new TCP connection on TCP-v pipe, Mobile Node (MN) sends a pTCP packet with pTCP header of 12 bytes.

pTCP header is additional to the TCP-v pipe (TCP) header and is also sent with each data packet in addition to traditional TCP header. In TCP-Migrate (Snoeren & Balakrishnan, 2000), handover control messages include the migrate option of 20B and no additional information is sent with each data packet. In SLM (Landfeldt et al., 1999), after connection establishment in new network, MN sends the Session ID to Remote Node (RN). Details are not given that what will be the size of Session ID. As most protocols use a Session ID of 4 bytes, we assume that SLM also have a Session ID of 4 bytes. As parallel transmission is done over two or more interfaces for a single stream of data, therefore, some sequencing information should also be sent with each data packet. Details are not given about this sequencing information.

TCP-R (Funato et al., 1997) has introduced different options for TCP having different sizes, e.g. handover related message Redirection Request packet has an option field of 10B, Authentication Request packet has an option field of 6B, and Authentication Reply has an option field of 10B. As TCP-R does not support multi-
homing and bandwidth aggregation, thus no additional information is sent with each TCP data packet.

HIP (So et al., 2006) (Moskowitz et al., 2006) has a base header of 40B, which is sent with each data packet, and some extension headers containing HIP parameters with variable sizes, that are sent in special circumstances. Handover control message is sent from HIP layer (layer 3.5) that does not include the TCP header. In LCT-based handoff protocol (Guo et al., 2003), no details about the size of handover signaling message is given but a format of signaling message is given that represents following fields: Type, Cid, current IP, current port, sequence number, signature. With each data packet, an additional field Cid is sent. Size of Cid is not defined.

In SL-TCP (Beda et al., 2005), only a single message is sent from MN to RN, containing the CID, for handover. CID contains the local port number (2 bytes) and destination IP address (4 bytes) consuming 6 bytes. This CID is also sent with each SL-TCP data packet. In VA-TCP (Lin & Chang, 2007), a notification message is sent to RN about the new location of MN. No details have been given about the packet format of notification message. Also no details are given that whether any information will be sent with each data packet or not. In Tsukamoto's proposal (Tsukamoto et al. 2004 and 2007), no details are given about the handover signaling packet format and about the information sent with data packets.

5.2 Results & Discussions Regarding Protocol Overhead

Figure 5.1 shows the protocol overhead of different end-to-end mobility management protocols after handover initiation using equation (5-1) for n=1 to 10. Protocols that send additional information with each data packet have higher protocol overhead as compared to protocols that do not send any additional information with
data packets. Protocol overhead of *HIP-based* mobility management protocol is highest of all because an additional *HIP* base header of 40 bytes is sent with each data packet. Using equation (5-3) for n=1000, we found that *HIP* creates 1MB of overhead traffic for sending 25,000 packets over the network. Protocol overhead for *TCP-Migrate* and *TCP-R*, remains constant after second packet. It is due to the fact that these protocols do not send any information with data packets, keeping the protocol overhead constant at 160 and 144 bytes for *TCP-Migrate* and *TCP-R* respectively. For *SL-TCP*, protocol overhead starts from 66 bytes and is increased by 6 bytes on every packet sent, by sending an additional field, *CID*, with each data packet. This increase does not create as much overhead as created by *HIP*, *pTCP* and *SLM*.

![Figure 5.1 Protocol overhead of different mobility management protocols](image)

Figure 5.1 shows the overhead comparison of protocols that send additional information with each data packet. This overhead is computed for n=1000 using equation (5-1). It was found that protocol overhead of *HIP-based*, *pTCP*, *SLM* and *SL-TCP* for sending 1000 packets is 40252, 12252, 4244 and 6066 bytes respectively.
Thus HIP generates an overhead traffic of around 40KB, pTCP around 12KB, SLM around 4KB and SL-TCP generates an overhead traffic around 6KB, for sending 1000 packets over the network.

![Protocol Overhead of Mobility Solutions](image)

Figure 5.2 Protocol overhead of mobility solutions that send additional information with each data packet

To analyze the protocol overhead of each protocol as a function of number of handovers, we assume that, while downloading a file, MN moves from one access network to another according to the mobility trajectory as shown in Figure 5.3. We partition the period of mobility into 6 phases. At the end of each phase vertical handover is performed.

![MN's Movement in Different Access Networks](image)

Figure 5.3 MN's movement in different access networks
Figure 5.4 shows the cumulative protocol overhead for each vertical handover performed by mobility management protocol using equation (5-4). pTCP and SLM have the highest overhead as compared to other protocols. The reason is that, both of these two protocols send an additional packet for handover containing previous connection information to RN after establishing new connection. Also the handover packet size of these two protocols is large as compared to others. SL-TCP has the lowest protocol overhead, because it exchanges minimal information for the signaling of handover.

![Figure 5.4 Protocol overhead as a function of number of handovers](image)

In this chapter, we have discussed the protocol overhead of different end-to-end mobility management solutions. Results show that protocols that exchange less number of handover signaling messages have low protocol overhead. Also, protocols that send additional information with each data packet, creates more overhead as compared to protocols that do not send any additional information with data packets.
Chapter 6
Conclusion & Future Work

In this work, we analyzed the performance of different end-to-end mobility management protocols in terms of three performance metrics: handover delay, throughput degradation time and protocol overhead. We used analytical models specifically developed for evaluation based on these metrics in wireless heterogeneous networks for end-to-end mobility management protocols. We have considered different vertical handover scenarios in which mobile node moves from one access network to another, subject to varying network connectivity options.

Results show that in a scenario, where MN moves between two non-overlapping network regions, handover delay performance of VA-TCP and SL-TCP is observed to be 2-4 times better as compared to other protocols. This is mainly due to less number of control messages exchanged for handover by these protocols. The throughput degradation time in this scenario is lowest for VA-TCP. The reason is that VA-TCP uses a throughput degradation avoidance mechanism in which TCP connection after handover does not start from slow start and congestion window is adjusted according to the new network's conditions. Hence in scenario of non overlapped coverage access, VA-TCP's throughput degradation time is same as the handover delay but other protocols have the throughput degradation time as the sum of handover delay and the time to reach at steady state. In the second scenario where MN moves between regions with overlapping network coverage, handover delay is found to be
zero for a group of protocols including pTCP, VA-TCP, Tsukamoto’s proposal and SLM. These protocols use simultaneous communication through two different interfaces which renders delay to zero. Similarly, throughput degradation time for these protocols is also zero in case of overlapping regions. This is because mobile node can send data through old network interface while it performs handover signaling through new network interface. In the third scenario, where MN moves out of overlapping region, handover delay of none of the protocols is found to be zero, as discussed by the analytical models. This is because, none of the protocols have link going down prediction intelligence that can be used to predict in advance that a link is going down. If a link going down is detected by the mobility management protocol then it can initiate a new TCP connection in advance using the alternate available interface, thus reducing the handover delay and throughput degradation time.

Similarly, protocol overhead of mobility management protocols that do not send any additional information with data packets, like TCP-Migrate and TCP-R, is low as compared to the protocols that send additional information with each data packet like pTCP, SLM and HIP-based handoff protocol.

The results gathered in this work manifest that the protocols that incorporate mechanism for simultaneous communication through multiple interfaces permit seamless vertical handover, consequently providing throughput stability. As an extension of this work we intend to develop an end-to-end mobility management solution for TCP applications that will try to overcome the limitations (higher handover delay and throughput degradation time) of different end-to-end mobility management protocols identified in this research work.
REFERENCES


